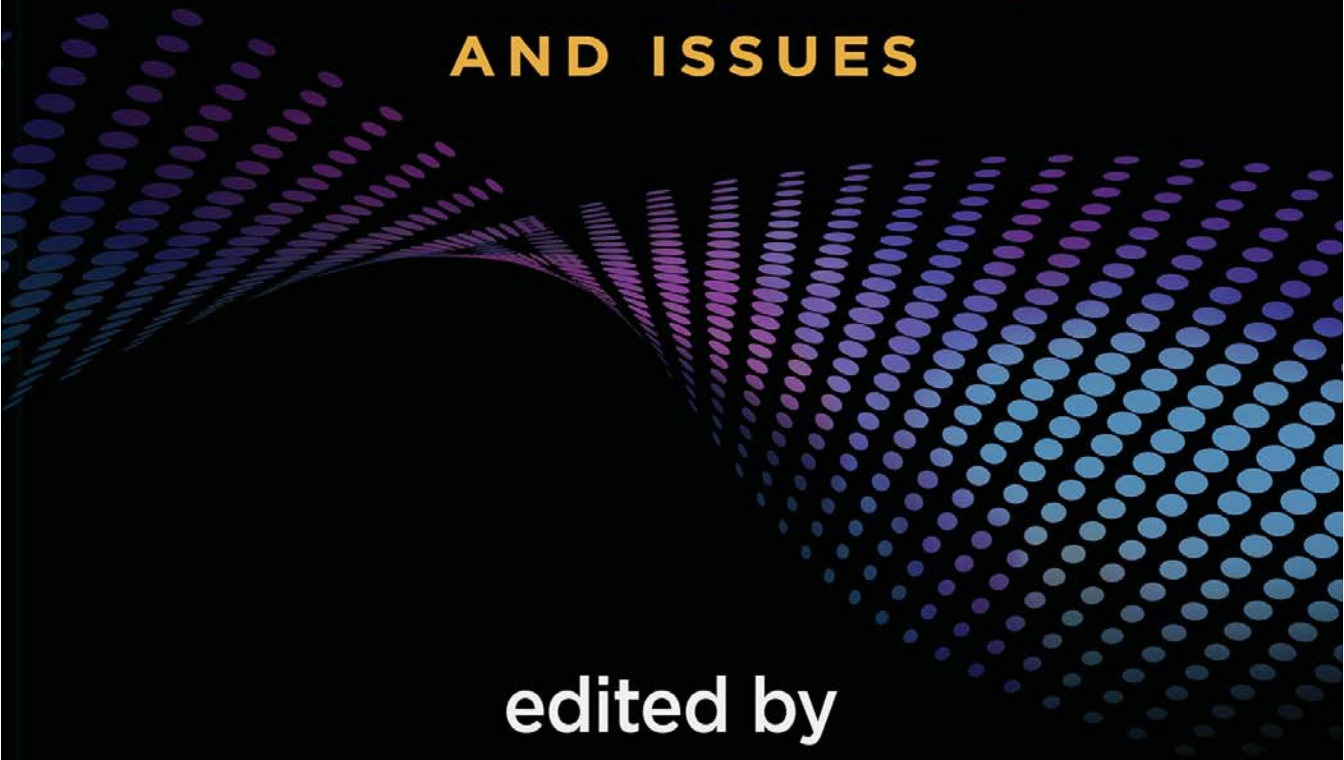
FOURTH EDITION

CONTEMPORARY INTELLECTUAL ASSESSMENT

THEORIES, TESTS, AND ISSUES



edited by

Dawn P. Flanagan

Erin M. McDonough

CONTEMPORARY INTELLECTUAL ASSESSMENT

CONTEMPORARY INTELLECTUAL ASSESSMENT

THEORIES, TESTS, AND ISSUES

FOURTH EDITION

edited by
Dawn P. Flanagan
Erin M. McDonough

Foreword by Alan S. Kaufman

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
370 Seventh Avenue, Suite 1200, New York, NY 10001
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Names: Flanagan, Dawn P., editor. | McDonough, Erin M., editor.
Title: Contemporary intellectual assessment : theories, tests, and issues / edited by Dawn P. Flanagan, Erin M. McDonough ; foreword by Alan S. Kaufman.
Description: Fourth Edition. | Fourth edition. | New York : The Guilford Press, 2018. | Revised edition of Contemporary intellectual assessment, c2012. | Includes bibliographical references and index.
Identifiers: LCCN 2017055258 | ISBN 9781462535781 (hardback : alk. paper)
Subjects: LCSH: Intelligence tests. | Intelligence tests—History. | BISAC: PSYCHOLOGY / Assessment, Testing & Measurement. | EDUCATION / Testing & Measurement. | MEDICAL / Psychiatry / Child & Adolescent. | PSYCHOLOGY / Psychotherapy / Child & Adolescent. | EDUCATION / Educational Psychology.
Classification: LCC BF431 .C66 2018 | DDC 153.9/3—dc23
LC record available at *https://lccn.loc.gov/2017055258*

About the Editors

Dawn P. Flanagan, PhD, is Professor of Psychology at St. John's University in Queens, New York, and Affiliate Clinical Professor at Yale Child Study Center, Yale University School of Medicine. She serves as an expert witness, learning disability consultant, and test/measurement consultant and trainer for organizations nationally and internationally. Dr. Flanagan is a Fellow of the American Psychological Association (APA) and a Diplomate of the American Board of Psychological Specialties. She received the inaugural Contributions to Practice Award from Division 16 (School Psychology) of the APA. Dr. Flanagan has published extensively on cognitive assessment, specific learning disabilities, and psychometric theories of the structure of cognitive abilities. She is also an author of the Cross-Battery Assessment Software System (X-BASS).

Erin M. McDonough, PhD, is Clinical Assistant Professor in the School Psychology Program of the Graduate School of Applied and Professional Psychology at Rutgers, The State University of New Jersey, and serves as Director of the Rutgers School Psychology Internship Consortium. Dr. McDonough is also the founder and Director of Psychological Diagnostic Evaluations of New York, where she sees clients and supervises psychologists and interns. Dr. McDonough publishes on issues related to psychological assessment of students with learning disabilities, and lectures and conducts workshops in this area at the state, regional, and national levels.

Contributors

Phillip L. Ackerman, PhD, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia

Vincent C. Alfonso, PhD, School of Education, Gonzaga University, Spokane, Washington

Erin K. Avirett, PhD, PLLC, private practice, Amarillo, Texas

Bruce A. Bracken, PhD, School of Education, College of William and Mary, Williamsburg, Virginia

Kristina C. Breaux, PhD, Pearson Clinical Assessment, San Antonio, Texas

Jie-Qi Chen, PhD, Erikson Institute Graduate School in Child Development, Chicago, Illinois

Michael Costa, MS, Department of Psychology, St. John's University, Jamaica, New York

Rik Carl D'Amato, PhD, Department of School Psychology, Chicago School of Professional Psychology, Chicago, Illinois

Mark Daniel, PhD, Learning Assessment Products, Bloomington, Minnesota

John L. Davis, PhD, Department of Educational Psychology, University of Utah, Salt Lake City, Utah

Scott L. Decker, PhD, Department of Psychology, University of South Carolina, Columbia, South Carolina

Felicia A. Dixon, PhD, Department of Educational Psychology, Ball State University, Selma, Indiana

Lisa Whipple Drozdick, PhD, Pearson Clinical Assessment, San Antonio, Texas

Ron Dumont, EdD, NCSP, School of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey

Agnieszka M. Dynda, PsyD, Department of Psychology, St. John's University, Jamaica, New York

Jamie Eiseman, MSEd, NCSP, School of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey

Colin D. Elliott, PhD (deceased), Department of Counseling, Clinical, and School Psychology, University of California, Santa Barbara, Santa Barbara, California

Ryan L. Farmer, PhD, BCBA, Department of Psychology, Western Kentucky University, Bowling Green, Kentucky

Joseph Ferraracci, MA, Department of Psychology, University of South Carolina, Columbia, South Carolina

Catherine A. Fiorello, PhD, NCSP, ABPP, Department of Psychological Studies in Education, Temple University, Philadelphia, Pennsylvania

Dawn P. Flanagan, PhD, Department of Psychology, St. John's University, Jamaica, New York

Randy G. Floyd, PhD, Department of Psychology, University of Memphis, Memphis, Tennessee

Howard Gardner, PhD, Graduate School of Education, Harvard University, Cambridge, Massachusetts

Richard J. Haier, PhD, Department of Pediatrics, School of Medicine, University of California, Irvine, Irvine, California

Rex E. Jung, PhD, Department of Neurosurgery, University of New Mexico, Albuquerque, New Mexico

Randy W. Kamphaus, PhD, Department of Special Education and Clinical Sciences, College of Education, University of Oregon, Eugene, Oregon

Alan S. Kaufman, PhD, Child Study Center, Yale University, New Haven, Connecticut

James C. Kaufman, PhD, Neag School of Education, University of Connecticut, Storrs, Connecticut

Nadeen L. Kaufman, EdD, Child Study Center, Yale University, New Haven, Connecticut

Timothy Z. Keith, PhD, Department of Educational Psychology, University of Texas at Austin, Austin, Texas

David A. Kilpatrick, PhD, Psychology Department, State University of New York College at Cortland, Cortland, New York

Sangwon Kim, PhD, NCSP, Department of Psychology, Humboldt State University, Arcata, California

Meghan A. Leahy, MS, Department of School Psychology, St. John's University, Jamaica, New York

Elizabeth O. Lichtenberger, PhD, private practice, Carlsbad, California

Denise E. Maricle, PhD, NCSP, Department of Psychology and Philosophy, Texas Woman's University, Denton, Texas

Nancy Mather, PhD, Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona

Robb N. Matthews, PhD, INTEGRIS Medical Group, Edmond, Oklahoma

R. Steve McCallum, PhD, Department of Educational Psychology and Counseling, University of Tennessee, Knoxville, Tennessee

George McCloskey, PhD, Psychology Department, Philadelphia College of Osteopathic Medicine, Philadelphia, Pennsylvania

Erin M. McDonough, PhD, Graduate School of Applied and Professional Psychology, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

Ryan J. McGill, PhD, Department of School Psychology, College of William and Mary, Williamsburg, Virginia

Kevin S. McGrew, PhD, Institute for Applied Psychometrics, St. Joseph, Minnesota

David E. McIntosh, PhD, Department of Special Education, Ball State University, Muncie, Indiana

Daniel C. Miller, PhD, Woodcock Institute for the Advancement of Neurocognitive Research and Applied Practice, Denton, Texas

Alyssa Montgomery, PhD, School of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey

Jaclyn Danielle Morrison, MA, Pearson Clinical Assessment, San Antonio, Texas

Jack A. Naglieri, PhD, Curry School of Education, University of Virginia, Charlottesville, Virginia; Department of Psychology, George Mason University, Fairfax, Virginia

Salvador Hector Ochoa, PhD, College of Education, University of New Mexico, Albuquerque, New Mexico

Samuel O. Ortiz, PhD, Department of Psychology, St. John's University, Jamaica, New York

Tulio M. Otero, PhD, Chicago School of Professional Psychology, Chicago, Illinois

Katherine Palma, BA, Department of School Psychology, St. John's University, Jamaica, New York

Nicole Piazza, BA, Department of Psychology, St. John's University, Jamaica, New York

Eric E. Pierson, PhD, Department of Educational Psychology, Ball State University, Muncie, Indiana

Elizabeth M. Power, EdD, Department of School Psychology, The College of Saint Rose, Albany, New York

Aurelio Prifitera, PhD, private consultant, San Antonio, Texas

Susan Engi Raiford, PhD, Pearson Clinical Assessment, San Antonio, Texas

Tara C. Raines, PhD, NCSP, Department of Teaching and Learning Sciences, Morgridge College of Education, University of Denver, Denver, Colorado

Cecil R. Reynolds, PhD, Department of Educational Psychology, Texas A&M University, College Station, Texas

Matthew R. Reynolds, PhD, Department of Educational Psychology, University of Kansas, Lawrence, Kansas

Alycia M. Roberts, PhD, Child and Family Psychological Services, Norwood, Massachusetts

Ellen W. Rowe, PhD, Department of Psychology, George Mason University, Fairfax, Virginia

Deanna Rumohr, EdS, Psychology Department, Philadelphia College of Osteopathic Medicine, Philadelphia, Pennsylvania

Joseph D. Salerno, PsyD, Sovereign Health Group, Marina del Rey, California

W. Joel Schneider, PhD, Department of Psychological Studies in Education, Temple University, Philadelphia, Pennsylvania

Fredrick A. Schrank, PhD, Department of Educational and Counseling Psychology, McGill University, Montreal, Quebec, Canada

Jennie Kaufman Singer, PhD, Division of Criminal Justice, California State University, Sacramento, Sacramento, California

Jaime Slonim, EdS, Office of Student Disabilities Services, University of Pennsylvania, Philadelphia, Pennsylvania

Marlene Sotelo-Dynega, PsyD, ABSNP, NCSP, School Psychology Program, St. John's University, Jamaica, New York

Robert J. Sternberg, PhD, Department of Human Development, Cornell University, Ithaca, New York

Julia Englund Strait, PhD, Department of School Psychology and Health Services Psychology, University of Houston–Clear Lake, Houston, Texas

Megan C. Sy, PsyD, Johns Hopkins All Children's Hospital, St. Petersburg, Florida

Erica Torres, PsyD, School of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey

Dustin Wahlstrom, PhD, Pearson Clinical Assessment, San Antonio, Texas

John D. Wasserman, PhD, independent practice, Burke, Virginia

Lawrence G. Weiss, PhD, Pearson Clinical Assessment, San Antonio, Texas

Barbara J. Wendling, MA, educational consultant, Dallas, Texas

John O. Willis, EdD, Regional Services and Education Center, Amherst, New Hampshire

Anne Pierce Winsor, PhD, private practice, Athens, Georgia

Richard W. Woodcock, EdD, Department of Psychology and Philosophy, Texas Woman's University, Denton, Texas

Kirby L. Wycoff, PsyD, NCSP, Department of Education, Worcester State University, Worcester, Massachusetts

Jianjun Zhu, PhD, Pearson Clinical Assessment, San Antonio, Texas

Foreword

Alan S. Kaufman

Dawn Flanagan and her colleagues have been on the cutting edge of assessment with *Contemporary Intellectual Assessment: Theories, Tests, and Issues* since the first edition was published in 1997. This first edition set the tone with a trend-setting array of amazing chapters by a "who's who" of authors, including John Horn and John Carroll, providing in-depth coverage of theory and practice. The book introduced Cattell–Horn–Carroll (CHC) theory to a generation of psychologists, integrated theory with practice against a backdrop of the history of assessment and test interpretation, and became so popular that it helped *define* the cutting edge of assessment.

Psychologists could barely wait for the second edition of *Contemporary Intellectual Assessment* to be published in 2005, given the ongoing dramatic changes in the field of assessment. Between 2001 and 2004, cognitive tests received major "facelifts" with the advent of two new Wechsler scales (the Wechsler Preschool and Primary Scale of Intelligence—Third Edition [WPPSI-III] and the Wechsler Intelligence Scale for Children—Fourth Edition [WISC-IV]), as well as an array of theory-based tests (the Woodcock–Johnson III [WJ III], the Kaufman Assessment Battery for Children [KABC-II], and the Stanford–Binet Intelligence Scales, Fifth Edition [SB5]). The field needed to have the state of the art redefined. Theory had never before so influenced practice; the psychometric-based CHC theory and Luria-based neuropsychological theories moved from the ivory tower to testing rooms; and the second edition of *Contemporary Intellectual Assessment* helped practitioners ease into the transition.

The exceptional third edition, published in 2012, tried to unify a field that had been fractured by the Individuals with Disabilities Education Act (IDEA) and response to intervention (RTI). And this new fourth edition, edited by Dawn Flanagan and Erin McDonough, captures the essence of where the field of assessment has ventured and where it is today. But I must admit that the field is *not* where I thought it would be now, and it is not where test publishers (at least Pearson, the publisher of the Wechsler and Kaufman batteries) thought it would be. A half-dozen years ago, really closer to a dozen years ago, we thought that the assessment of intelligence and achievement would follow the rest of the world into the digital age. We thought that somewhere in basements, groups of 19-year-old nerds were applying the latest technology to assemble the next generation of IQ tests that would take the field by storm. We (my wife, Nadeen, and I) and Pearson were deep into the development of the KABC—Digital (KABC-D), hoping to beat the high-tech teens to the punch. The WISC-V, to the best of my knowledge, included *three* forms of the test, at least for a short while during its development: two that were published (clinical test kit and Q-interactive versions) and one that I believe was left on the cutting room floor (an all-digital version). We were advised by different professionals at Pearson that "paper-and-pencil" tests like the KABC-II had a shelf life of 5–7 years; that almost all examiners would switch to Pearson's Q-interactive administration in that time; and that the future was digital. Hence the push for the KABC-D and the all-digital WISC-V.

Well, it didn't happen. "Paper-and-pencil" tests have continued to thrive (how I hate that term when applied to clinical test kit versions of tests such as the Cognitive Assessment System—Second Edition [CAS2] or WISC-V!). The exodus from clinical versions to Q-interactive versions did not occur—at least, not as rapidly as anticipated—and it has met resistance from many school systems. It still may be the wave of the future, but it does not define the cutting edge of the present. The KABC-D, embraced by a vibrant, dynamic, innovative, intervention-oriented test development process for half a decade, is now "on hold." An all-digital version of the WISC-V has not yet been published (although the Wechsler Adult Intelligence Scale [WAIS 5] of the future may include such a version). Traditional clinical tests, like the amazing array of cognitive batteries published in 2014—the CAS2, WJ IV, and WISC-V—continue to dominate our field and are seemingly setting the tone for the future.

How did we get it so wrong? For one thing, we underestimated scientist-practitioners—most notably the school psychologists and neuropsychologists who had been leading the field into the future on the foundation of a sophisticated amalgamation of refined theory, state-of-the-art methodology, and empirically based educational interventions. They were neither ready nor willing to reduce the role of the examiner as clinician and keen observer, even if the burdens of administration, scoring, and timing of test items could easily be passed on to a machine. They have also been unwilling to give up concrete test materials, which afford the opportunity for firsthand observations of a child's or adult's problem-solving strategies (even though images on a computer screen can undoubtedly yield reliable and valid scores, just as group-administered IQ tests have done for a century). And they worry about confidentiality—an ethical concern that looms even larger now than before, in light of the impact of possible hacking on even the most seemingly secure areas.

These are all reasonable and thoughtful concerns. The field of assessment may ultimately yield to the pressures of digital technology, but not yet, and probably not soon. Computer-based administration and scoring such as Q-interactive methods will undoubtedly grow in popularity, but I no longer believe that the 19-year-olds will take over our field with digital technology, the way that Binet once ousted Galton or Wechsler once surpassed everyone.

The ultimate, perhaps inevitable, shift from clinical test kit to digital assessment will only happen when a computerized test is published that addresses the scientist-practitioner's important concerns and is clearly seen as a better alternative—one that takes full advantage of digital capabilities without squelching the role of the clinician.

This fourth edition of *Contemporary Intellectual Assessment*, once again, is on the cutting edge of the field of assessment. Since the third edition was published in 2012, the field has continued to shift. The advances are more subtle than in previous decades, but nonetheless tests have become more integrated, more complex, and more focused on both neuropsychological and educational interventions. The fourth edition has kept pace with these changes, which have included the following:

- *An integration of psychometric (CHC) and neuropsychological processing traditions.* This integration is evident in chapters by Joel Schneider and Kevin McGrew (Chapter 3) and by Daniel Miller and Denise Maricle (Chapter 33). This interface is taken a step further by Richard Woodcock and colleagues (Chapter 32), who present a "functional" classification system for CHC abilities—a system they propose as being more informative and useful for practitioners in school psychology, neuropsychology, and special education. And this interweaving of psychometric and neuropsychological processing approaches pervades the entire volume. Examples include the discussions of assessing special populations—gifted children, children with various disabilities, and English language learners (Part IV of the book).

- *An expansion—not a reduction—of the number of viable theories available to inform diagnosis and interventions.* All editions of *Contemporary Intellectual Assessment* have rigorously covered important theories of intelligence, even those that have *not* been translated to clinical tests of intelligence—most notably Robert Sternberg's triarchic theory of successful intelligence (Chapter 5) and Howard Gardner's multiple-intelligences theory (Chen & Gardner, Chapter 4). The fourth edition has expanded the theory section to include two innovative approaches—Richard Haier and Rex Jung's parieto-frontal integrative theory (PFIT) (Chapter 7) and Phillip Ackerman's intelligence-as-process, personality, interests, and intelligence-as-knowledge (PPIK) framework for adult intellectual development (Chapter 8). The PPIK theory, which is primarily focused on Raymond Cattell's ideas, also incorporates neurological concepts advocated by D. O. Hebb (always my favorite learning theorist); this new chapter ensures an increased focus on *adult* development and intelligence. Both the PFIT and PPIK theories highlight the rapidly growing influence of neuropsychology on the current and future assessment scene.

- *Enhanced psychometric and theoretical sophistication evident in assessment research that has accompanied the development and interpretation of all major cognitive, achievement, and neuropsychological tests.* Part III of this newest *Contemporary Intellectual Assessment* has separate chapters on the latest editions of all the major cognitive batteries—all of the Wechslers (including the WISC-V Integrated), as well as the Reynolds Intellectual Assessment Scales, Second Edition (RIAS-2), Universal Nonverbal Intelligence Test—Second Edition (UNIT2), CAS2, NEPSY-II, WJ IV, KABC-II Normative Update, and Differential Ability Scales–II (DAS-II). Major achievement batteries are included as well; see especially Chapter 29 by Jaclyn Morrison, Jennie Singer, and Susan Raiford, which juxtaposes the Kaufman Test of Educational Achievement, Third Edition (KTEA-3) with the WISC-V and WISC-V Integrated to link assessment results to interventions. Contemporary psychometric sophistication peaks in Chapter 31, as Timothy Keith and Matthew Reynolds address a heated and controversial debate in the literature regarding the efficacy of bifactor models versus hierarchical confirmatory factor analysis (CFA) models as the best explanation of the structure of intelligence.

- *A more in-depth neuropsychological understanding of reading disabilities and specific learning disabilities (SLD).* State-of-the-art approaches dominate the chapters by Marlene Sotelo-Dynega on the use of neuropsychological assessment in the identification of reading disorders (Chapter 34); by David Kilpatrick on the role of orthographic processing in reading (Chapter 35); by Dawn Flanagan and colleagues on the use of patterns of strengths and weaknesses (PSW) in the identification of SLD (Chapter 22); and by Erin McDonough and colleagues on the diagnosis of SLD via DSM-5 (Chapter 37).

- *Greater emphasis on the translation of theory, clinical insights, and test scores into hands-on educational interventions.* This focus on the direct translation of test results to the classroom permeates nearly all chapters in the book, but is especially salient in Part V, as luminaries in the field such as Dawn Flanagan, Vincent Alfonso, Samuel Ortiz, Nancy Mather, Barbara Wendling, Catherine Fiorello, and Susan Raiford share their insights and innovations.

The publication of the fourth edition of *Contemporary Intellectual Assessment* marks a half-century since I entered the field of assessment. In 1967, I took a course in the clinical psychology department at Columbia University that covered the Wechsler tests (WISC and WAIS), the 1960 Stanford–Binet, the Rorschach, and the Thematic Apperception Test (TAT). I was taught that IQ tests were personality tests, and that we shouldn't pay too much attention to the numbers. It was OK to deviate from standardized administration, because, after all, we are clinicians. Most subtests were interpreted from a neo-Freudian perspective, in which a low score on Information was more likely to denote repression than a lack of knowledge. Similarly, high Comprehension coupled with low Information signified a hysteric reaction; missing a few easy Comprehension items was likely to reflect schizophrenia or psychotic depression; higher Digits Backward than Digits Forward meant obstinacy or negativism; and relatively high Picture Completion suggested a paranoid trend. There was no statistical significance to worry about. How high meant "high"? How low was "low"? It just didn't matter. A scaled score of 11 on Digit Symbol (now Coding) coupled with an 8 or 9 on Digit Span was usually enough for a clinician to infer that the person was controlling strong and pressing anxiety by excessive activity.

I asked my lab instructor many questions: Why was a child or adult not told to give a second reason on WISC or WAIS Comprehension, when the scoring system gave 2 points only if two different ideas were expressed? Why were we told to keep the stopwatch hidden from view, and why were clients not told that they should work quickly when they could earn as many as 3 bonus points for quick perfect performance on a single nonverbal item? What was the validity evidence for interpreting specific verbal responses to Vocabulary and Comprehension items as evidence of social incompetence, psychopathology, dependence, or grandiosity? She laughed at my ignorance and impertinence and embarrassed me in front of the other lab students.

That was 1967. In 1968, I selected psychometrics as my main area of study, and Robert L. Thorndike agreed to be my chair. That same year, I started working at The Psychological Corporation, where my boss was Alexander Wesman (who coined the term *intelligent testing* in a 1968 *American Psychologist* article). By 1970, I had begun my 4-year apprenticeship with David Wechsler. Revising the WISC and developing the WISC-R became my life; learning everything I could from my brilliant mentor became my passion; applying that knowledge became my life's work. I was given the chance to fix the flaws in Wechsler's scales, to the extent that the Master was accepting of change. Later in the 1970s, Nadeen and I were developing the K-ABC and training brilliant doctoral students at the University of Georgia who would change the face of cognitive, behavioral, and neuropsychological assessment—Bruce Bracken, Jack Cummings, Patti Harrison, Randy Kamphaus, Steve McCallum, Jack Naglieri, and Cecil Reynolds. I realized that I had totally freed myself from the shackles of my early training: the notion that IQ tests were no more and no less than measures of personality.

In the 1970s, I joined other psychologists, most notably Joe Matarazzo and Jerry Sattler, in writing books that attempted to bridge the gap between psychometrics and school/clinical psychology. When Nadeen and I published the K-ABC in 1983 (with the spectacular help of the Georgia doctoral students mentioned above), we started the movement toward theory-based tests and the integration of psychometric and neuropsychological processing traditions. But that was just a small beginning. The 1980s saw new approaches to theory-based assessment, highlighted by the Horn-inspired WJ-R in 1989. Dawn Flanagan and her colleagues, many of them coauthors of chapters in this volume, used those benchmarks as merely the starting point of what was to come—namely, the explosion of sophisticated theoretical, clinical, and psychometric constructs, and the interdisciplinary expansion of the fields of test construction and clinical assessment (in the present volume, see Kamphaus et al., Chapter 2, for the history of intelligence test interpretation).

And here we are a half-century later, standing on the cutting edge of our field as we ingest the innovative ideas of brilliant professionals who have come together to write the 39 chapters that constitute the 2018 edition of *Contemporary Intellectual Assessment.* Not a weak link in the bunch, not a wasted word or graphic. Not a single chapter devoted to computerized assessment.

Preface

The history of intelligence testing has been well documented, from the early period of mental measurement to present-day conceptions of the structure of intelligence and valid assessment methods. The foundations of psychometric theory and practice were established in the late 1800s and set the stage for the ensuing measurement of human cognitive abilities. The technology of intelligence testing was apparent in the early 1900s, when Binet and Simon developed a test that adequately distinguished children with intellectual disabilities from children with normal intellectual capabilities, and was well entrenched when the Wechsler–Bellevue was published in the late 1930s. In subsequent decades, significant refinements and advances in intelligence-testing technology have been made, and the concept of individual differences has remained a constant focus of scientific inquiry.

Although several definitions and theories have been offered in recent decades, the nature of intelligence, cognition, and competence continues to be elusive. Perhaps the most popular definition was that offered by Wechsler in 1958. According to Wechsler, *intelligence* is "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (p. 7). It is on this conception of intelligence that the original Wechsler tests were built. Because the Wechsler batteries were the dominant intelligence tests in the field of psychology for decades and were found to measure global intelligence validly, they assumed "number one" status and remain in that position today. As such, Wechsler's definition of intelligence continues to guide and influence the present-day practice of intelligence testing.

In light of theoretical and empirical advances, however, it is clear that earlier editions of the Wechsler tests were not based on the most dependable or current evidence of science, and that overreliance on these instruments served to widen the gap between intelligence testing and cognitive science. From the 1980s through the 2000s, new intelligence tests have been developed to be more consistent with contemporary research and theoretical models of the structure of cognitive abilities, and the Wechsler scales (most notably the Wechsler Intelligence Scale for Children—Fifth Edition [WISC-V]) are now keeping pace with advances in theory and practice. In addition, changes in services, programs, legislation, and policy in psychology and education have had many implications for practical uses of cognitive assessments and for the client populations with whom they are applied. More recently, technological advances including tablet-based test administration have increased the ease and convenience of cognitive assessment. Since the publication of the first edition of *Contemporary Intellectual Assessment: Theories, Tests, and Issues* in 1997, there has been tremendous growth in research about cognitive constructs, and this research—for example, recent research in neuropsychology and executive functions—has in turn influenced contemporary purposes of cognitive assessment in psychology and education. These include the development and delivery of multi-tiered services for children experiencing learning problems, as well as increasing uses of cognitive and neuropsychological assessment for people who have specific learning disabilities, autism spectrum disorder, attention-deficit/hyperactivity disorder, and other challenges.

The information presented in this text on modern intelligence theory and assessment technology suggests that clinicians should be familiar with the many approaches to assessing intelligence that are now available. For the field of intellectual assessment to continue to advance, clinicians should use instruments that operationalize empirically supported theories of intelligence and should employ assessment techniques that are designed to measure the broad array of cognitive abilities represented in current theory and research. It is only through a broader measurement of intelligence—grounded in well-validated theories of the nature of human cognitive abilities—that professionals can gain a better understanding of the relationship between intelligence and important outcome criteria (e.g., school achievement, occupational success). They can also continue to narrow the gap between the professional practice of intelligence and cognitive ability testing on the one hand, and theoretical, empirical, and practical advances in psychology and education on the other.

PURPOSE AND OBJECTIVES

The purpose of the fourth edition of this book is to provide a comprehensive conceptual and practical overview of current theories of intelligence, individual measures of cognitive ability, and uses of intellectual assessments. This text summarizes the latest research in the field of intellectual assessment and includes comprehensive treatment of critical issues that should be considered when the use of intelligence tests is warranted. The three primary objectives of this book are as follows: (1) to present in-depth descriptions of prominent theories of intelligence, tests of cognitive abilities, and neuropsychological instruments, and issues related to the use of these tests; (2) to provide important information about the validity of contemporary intelligence, cognitive, and neuropsychological tests and their use in diagnosis and intervention planning; and (3) to demonstrate the utility of a well-validated theoretical and research foundation for developing cognitive tests and interpretive approaches, and for guiding research and practice. The ultimate goal of this book is to provide professionals with the knowledge necessary to use the latest cognitive instruments effectively.

Practitioners, university faculty, researchers, undergraduate and graduate students, and other professionals in psychology and education will find this book interesting and useful. It would be appropriate as a primary text in any graduate (or advanced undergraduate) course or seminar on cognitive psychology, clinical or psychoeducational assessment, or measurement and psychometric theory.

ORGANIZATION AND THEMES

This book consists of 39 chapters, organized into six parts.

Part I, "The Origins of Intellectual Assessment," traces the historical roots of test conceptualization, development, and interpretation up to the present day. The updated chapters provide readers with an understanding of how current practices evolved, as well as a basis for improving contemporary approaches to test interpretation. Chapters provide a necessary foundation for understanding and elucidating the contemporary and emerging theories, tests, and issues in the field of intellectual assessment that are presented in subsequent sections of this volume.

Part II, "Contemporary Theoretical Perspectives," updates several models presented in the previous editions of this text (e.g., Cattell–Horn–Carroll [CHC] theory; planning, attention, simultaneous, and successive [PASS] theory; the triarchic theory of successful intelligence) and presents two alternative theoretical conceptualizations of human intelligence: the parieto-frontal integration theory (PFIT) and the intelligence-as-process, personality, interests, and intelligence-as-knowledge (PPIK) framework for adult intellectual development. These theories are described in terms of (1) how they reflect recent advances in psychometrics, neuropsychology, and cognitive psychology, as well as neuroimaging; (2) what empirical evidence supports them; and (3) how they have been operationalized and applied. A comprehensive description of each theory is provided, focusing specifically on its historical origins, as well as the rationale and impetus for its development and modification. The theories represent viable foundations from which to develop and interpret cognitive measures—measures that may lead to greater insights into the nature, structure, and neurobiological substrates of cognitive functioning.

Part III, "Contemporary Intelligence, Cognitive, and Neuropsychological Batteries and Associated Achievement Tests," includes comprehensive chapters on the most widely used and most current individual intelligence batteries and their utility in understanding the cognitive capabilities of individuals from early school age through adulthood. This new edition of *Contemporary Intellectual Assessment: Theories, Tests, and Issues* updates information found in the third edition: It provides current research about the WISC-V and the other most recent revisions of the Wechsler intellectual, memory, and achievement scales; the Kaufman Assessment Battery for Children—Second Edition, and Normative Update and the Kaufman Test of Educational Achievement, Third Edition; the Woodcock–Johnson IV Tests of Cognitive Abilities, Oral Language, and Achievement; the Differential Ability Scales—Second Edition; the Universal Nonverbal Intelligence Test—Second Edition; the Cognitive Assessment System—Second Edition; and the Reynolds Intellectual Assessment Scales, Second Edition. In general, the authors provide descriptions of their assessment instruments and discuss the instruments' theoretical and research underpinnings, organization and format, and psychometric characteristics. The authors also summarize the latest research and practical uses and provide recommendations for interpreting the abilities measured by their instruments.

Part IV, "Relevance of Tests of Intelligence, Cognitive Abilities, and Neuropsychological Processes in Understanding Individual Differences," includes chapters about the use of cognitive tests to assess individuals for giftedness, intellectual disability, specific learning disabilities, and traumatic brain injury, as well as individuals from culturally and linguistically diverse populations.

Part V, "Linking Assessment Data to Intervention," includes chapters about using test data to inform interventions aimed at ameliorating the manifestations of cognitive weaknesses.

Part VI, "Contemporary and Emerging Issues in Intellectual, Cognitive, and Neuropsychological Assessment," includes updated chapters related to the validity of intelligence batteries and presents an updated practical and simplified nomenclature for CHC abilities. In addition, this section's chapters discuss the relationship between cognitive assessment and reading disorders, as well as the role of cognitive testing in the assessment of executive functions and specific learning disorder as defined in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5). The integration of neuropsychological assessment into school-based practice is discussed in two chapters. Suggestions and recommendations regarding the appropriate use of intelligence, cognitive ability, and neuropsychological tests, as well as future research directions, are provided throughout this section of the book.

DAWN P. FLANAGAN
ERIN M. MCDONOUGH

Acknowledgments

We are deeply grateful for the extraordinary contributions of our chapter authors. It has been a rewarding experience to work with such a dedicated group of people, many of whom are nationally and internationally recognized authorities in their respective areas. We thank Katherine Palma and Alexa Pate for their editorial assistance in the preparation of this book. Their careful attention to detail improved the quality of this work. We thank Natalie Graham for her support of another edition of this book, Marie Sprayberry for her outstanding copy-editing skills, and Laura Specht Patchkofsky for her skillful management of the production process. The staff at The Guilford Press are gratefully acknowledged for their expertise and their pleasant and cooperative working style. The professionalism and expertise of our contributors and the staff at Guilford made this project a most enjoyable and productive endeavor.

Contents

I. The Origins of Intellectual Assessment

II. Contemporary Theoretical Perspectives

III. Contemporary Intelligence, Cognitive, and Neuropsychological Batteries and Associated Achievement Tests

IV. Relevance of Tests of Intelligence, Cognitive Abilities, and Neuropsycholgical Processes in Understanding Individual Differences

V. Linking Assessment Data to Intervention

VI. Contemporary and Emerging Issues in Intellectual, Cognitive, and Neuropsychological Assessment

PART I

The Origins of Intellectual Assessment

CHAPTER 1

A History of Intelligence Assessment

The Unfinished Tapestry

John D. Wasserman

> When our intelligence scales have become more accurate and the laws governing IQ changes have been more definitively established it will then be possible to say that there is nothing about an individual as important as his IQ, except possibly his morals; that the greatest educational problem is to determine the kind of education best suited to each IQ level; that the first concern of a nation should be the average IQ of its citizens, and the eugenic and dysgenic influences which are capable of raising or lowering that level; that the great test problem of democracy is how to adjust itself to the large IQ differences which can be demonstrated to exist among the members of any race or nationality group.
>
> —LEWIS M. TERMAN (1922b)

This bold statement by the author of the first Stanford–Binet intelligence scale captures much of both the promise and the controversy that have historically surrounded, and that *still* surround, the assessment of intelligence. Intelligence tests and their applications have been associated with some of the very best and very worst human behaviors. On the one hand, intelligence assessment can provide a meaningful basis for understanding the strengths and weaknesses of misunderstood children, adolescents, or adults—thereby providing data that can be used to design and implement interventions to help people reach their potential more effectively. On the other hand, intelligence assessment can be used to segregate and label people—treating their future as a fixed outcome, an unchangeable fate. The history of forced sterilizations of individuals with intellectual disabilities in the United States and many other countries is a tragic example of how intelligence tests may be misused, exceeded only by the systematic extermination of intellectually disabled individuals in Nazi Germany (e.g., Friedlander, 1995). The topic of intelligence and its assessment deservedly elicits many strong feelings.

Intelligence is arguably the most researched topic in the history of psychology, and the concept of *general intelligence* has been described as "one of the most central phenomena in all of behavioral science, with broad explanatory powers" (Jensen, 1998, p. xii). Still, social, legal, and political forces have in some instances excluded intelligence test results from important types of educational and personnel decision-making processes. Tangible advances in assessment practices have been slow and episodic. Following Alfred Binet's initial successes, the beginning of the 20th century saw an accelerated pace of small- and large-scale applied intelligence testing, but many anticipated educational and occupational benefits were never realized. Buros (1977) considered 1927 as the "banner year" in which "the testing movement reached maturity" (p. 9). The middle of the century saw only incremental gains in testing, such as electronic scoring, analysis, and reporting of test results, but with comparatively "little progress" (Buros, 1977,

p. 10) and more than a little "stagnation" (Carroll, 1978, p. 93). A landmark quantitative review of factor-analytic investigations near the end of the 20th century (i.e., Carroll, 1993) stimulated a new school of thinking about intelligence assessment, but the story remains unfinished. In the United States, federal educational reforms and civil rights legislation have had pronounced effects upon the use of intelligence tests in education. It is possible to see the history of intelligence assessment as an unfinished tapestry depicting the rich saga of a developing discipline, with recurrent characters interwoven through different narratives, as well as more than a few loose and unresolved thematic threads.

In this chapter, the origins of intelligence assessment are recounted, with an emphasis on milestone events and seminal individuals. Thematic strands present from the early days are traced, including some that were resolved and some that remain unresolved. An effort has been made whenever possible to provide samples of primary source material. Finally, although we all tend to view history through the lens of our own experiences, it is helpful to appreciate the sociocultural context, institutional traditions, and professional *Zeitgeist* associated with historical events, as well as the experiences and personal motivations that may have driven the ideas and behaviors of historical figures.

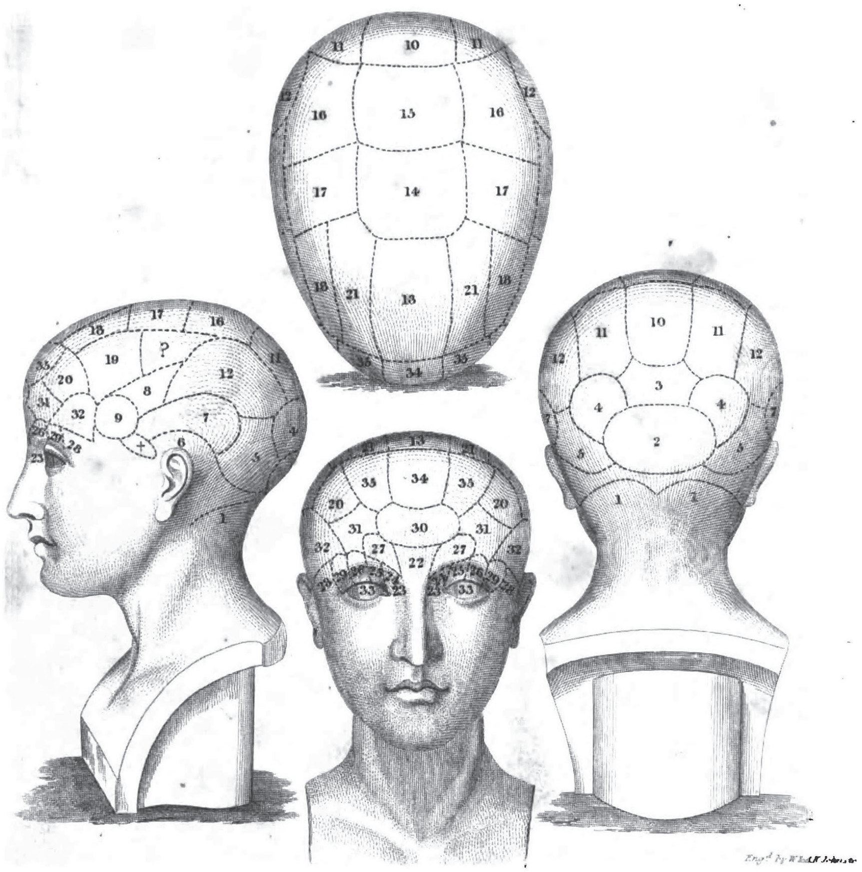
PSEUDOSCIENTIFIC ANTECEDENTS: PHRENOLOGY IN THE 19TH CENTURY

The first science purporting to be a "true science of mind" that could measure mental qualities and functions was *cranioscopy,* introduced at the beginning of the 19th century by Franz Joseph Gall, and later renamed *phrenology* by Gall's associate, Johann Gaspar Spurzheim. Gall (1758–1828) was a Viennese physician and neuroanatomist, and Spurzheim (1776–1832) was a physician and colleague who would ultimately be responsible for the widespread dissemination of phrenology. But it would be a Scotsman, George Combe (1788–1858)—who developed and published a two-volume system of phrenology in 1824, as well as launching a phrenology journal with his brother—who would prove most instrumental in the popularization of phrenology. Combe's system appears in Figure 1.1. He also wrote the immensely successful book *The Constitution of Man,* which advanced the idea that all the laws of nature were in harmony with one another, and that people could best fulfill God's will and obtain the greatest happiness by discovering these laws and obeying them. The book went through eight editions and sold approximately 350,000 copies between 1828 and 1900.

The basic tenets of phrenology can be summarized easily. In a letter to a Viennese official, Gall (1798/1857) asserted that the brain was the organ of the mind, that the mind could be reduced to a number of faculties, that the faculties were innate, that the faculties were located in distinct and particular organs of the brain, that the surface of the skull was determined by the external form of the brain, and that phrenologists could judge the development of individual faculties merely by examining the form of the skull. A well-developed faculty was thought to have a large cerebral organ that corresponded to a cranial protuberance. Gall originally described and localized 27 distinct faculties; Spurzheim (1815) increased the list to 32 faculties; Combe (1853) further expanded the list to 35; and others expanded the list to 43 (e.g., Payne, 1920).

Gall and Spurzheim traveled through Europe promoting phrenology, which Gall advocated as a science and Spurzheim as a way to reform education, religion, and penology. It quickly became popular in the United Kingdom, and Spurzheim came to the United States in 1832 to promote phrenology to a scientific community that was already quite familiar with it. By the time Combe conducted his 1839 American phrenology lecture tour, audiences averaged over 500 across each of the 16 lectures (Walsh, 1976). A satirical depiction of a phrenological examination from about the same time appears in Figure 1.2.

Gall and Spurzheim are today credited with recognizing the significance of gray matter as the source of nerve fibers; most importantly, they are credited with introducing the neuroscientific concept of functional localization in the cerebral cortex (Simpson, 2005). Dallenbach (1915) provides evidence that they should be credited with the terms *mental functions* and *faculties*. British philosopher and critic G. H. Lewes (1867) went a step further, asserting that Gall laid the groundwork for psychology as a science rather than philosophy: "Gall rescued the problem of mental functions from Metaphysics and made it one of Biology" (p. 407). Even so, there is a long history of disparaging efforts to localize mental functions in specific regions in the brain by calling them a new "phrenology" (e.g., Franz, 1912; Fuster, 2008, p. 346).



Names of the Phrenological Organs

REFERRING TO THE FIGURES INDICATING THEIR RELATIVE POSITIONS.

AFFECTIVE						INTELLECTUAL					
I. PROPENSITIES			II. SENTIMENTS			I. PERCEPTIVE			II. REFLECTIVE		
1	Amativeness	Page 116	10	Self-esteem	231	22	Individuality	380	34	Comparison	460
2	Philoprogenitiveness	121	11	Love of approbation	243	23	Form	385	35	Causality	474
3	Concentrativeness	134	12	Cautiousness	252	24	Size	389			
4	Adhesiveness	151	13	Benevolence	261	25	Weight	393			
5	Combativeness	157	14	Veneration	274	26	Colouring	399			
6	Destructiveness	165	15	Firmness	285	27	Locality	414			
+	Alimentiveness	184	16	Conscientiousness	288	28	Number	420			
7	Secretiveness	190	17	Hope	304	29	Order	424			
8	Acquisitiveness	203	18	Wonder	309	30	Eventuality	425			
9	Constructiveness	217	19	Ideality	322	31	Time	434			
			?	Unascertained	330	32	Tune	430			
			20	Wit or Mirthfulness	340	33	Language	446			
			21	Imitation	353						

FIGURE 1.1. George Combe, the best-known phrenologist of the 19th century, divided the brain into intellectual faculties and feelings. The plate of the phrenological bust faces the title page in Combe (1830). In the public domain.

FIGURE 1.2. Illustration from a fictional story of a member of a phrenology society who decides to use phrenology to identify a possible thief in his household. The drawing shows a servant who was paid five shillings to shave his head so that the phrenological organs could be traced in ink, not a standard part of phrenology practice. From Prendergast (1844, p. 17). In the public domain.

PHILOSOPHICAL AND SCIENTIFIC ANTECEDENTS

The most prominent British philosopher of his era, Herbert Spencer (1820–1903) sought to synthesize universal natural laws (especially *evolution*) across the disciplines of biology, psychology, sociology, and ethics. Spencer coined the phrase "survival of the fittest" (p. 444) in *The Principles of Biology* (1864) after reading Charles Darwin (1859), although he was reluctant to accept Darwin's evolutionary mechanism of natural selection. In *The Principles of Psychology* (1855), Spencer described how the behavior of the individual organism adapts through interaction with the environment, and defined *intelligence* as a "continuous adjustment" of "inner to outer relations" (p. 486). Spencer's ideas persist in a number of ways to this day. Intelligence, as we shall see, is still widely considered to represent a capacity associated with adaptation to one's environment. In a critical review of Spencer's synthesis, John Dewey (1904) was struck by the luck that Spencer and Darwin published almost simultaneously, thereby making their very different concepts of evolution indistinguishable to the public.

Beyond Spencer's philosophical influence, the foundation for psychology as a science, as well as for the scholarly study of intelligence, was laid by naturalist Charles Darwin (1809–1882), who is most remembered for his theory of evolution by natural selection. In his writings, Darwin frequently referred to adaptive behavior in animals and humans as "intelligent"; more importantly, he argued that the same forces that act on animal evolution also apply to human mental abilities: "There is no fundamental difference between man and the higher mammals in their mental faculties" (Darwin, 1871, p. 35). In *The Descent of Man*, Darwin (1871) went even further in applying his evolutionary theory to human mental characteristics—probably after reading the work of his half-cousin Francis Galton, the Victorian polymath, whose drive for scientific measurement of human capabilities would start the race to develop measures of intelligence in motion.

It is difficult to overstate the impact of Darwin's theory of evolution on psychology. By considering human behavior in an evolutionary context, Darwin treated the study of psychology as no less a science than biology and other natural sciences. His influence was substantial and may be seen, for example, in Joseph Jastrow's (1901) American Psychological Association (APA) presidential address to start the 20th century. Jastrow described psychology as both a laboratory science and an applied science, setting the study of intelligence in a somewhat Spencerian evolutionary context:

> Intelligence must first be realized as an advantage-gaining factor in the evolutionary struggle; that struggle is not merely, and indeed in all the stages that here come into consideration, not mainly a conflict of tooth and nail, a contest of strength of claw and fleetness of foot, but a war of wits, an encounter of skill and cunning, a measure of strategy and foresight. (p. 9)

Francis Galton and the Anthropometric Laboratory

If you lived in London in the mid-1880s or 1890s, you could pay three- or fourpence for you or your children to undergo a variety of physical measurements and tests, with the option to register results for future reference and follow-up. The measurements were available from Francis Galton's Anthropometric Laboratory, first located at the International Health Exhibition (see Figure 1.3), then at the University of Cambridge, the South Kensington Museum, and finally at the Clarendon Museum at Oxford. *Anthropometry* referred to the "measurement of man," and Galton's laboratory was, according to Diamond (1977), "a device to tease the public into providing the data he needed for his research" (p. 52). As a lifelong advocate for objective scientific measurement, Galton (1822–1911; see Figure 1.4) was a pioneer in the use of test batteries and questionnaires for data collection, the concept of control groups in research, and statistical methods (as developer of the techniques of regression and correlation).

Galton introduced his system of anthropometric measurements in *Inquiries into Human Faculty and Its Development* (1883), where he wrote, "It is needless for me to speak here about the differences in intellectual power between different men and different races, or about the convertibility of genius as shown by different members of the same gifted family achieving eminence in varied ways" (pp. 82–83). He conceptualized his measurements as constituting indicators of physical efficiency to complement performance on formal academic written literary examinations, which he thought were the best available measures of intelligence (e.g., Galton, 1884, 1891).

FIGURE 1.3. Francis Galton's first Anthropometric Laboratory was featured at the International Health Exhibition held in London in 1884–1885. Nearly 10,000 people paid threepence each to be examined and receive a copy of their measurements. From Pearson (1924, Plate L). Reprinted by permission of Cambridge University Press.

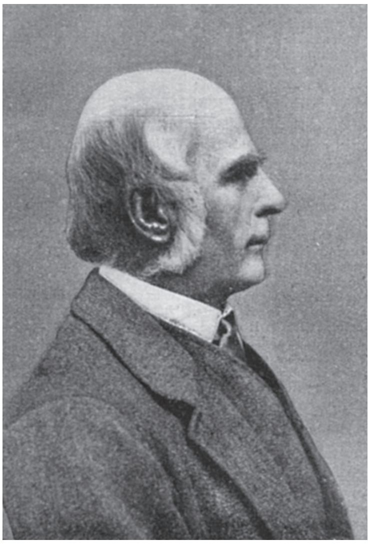


FIGURE 1.4. Francis Galton in 1888 at the age of 66, when the Anthropometric Laboratory remained active. From the copperplate prepared for *Biometrika.* In the public domain.

The examination took less than 1 hour. Although the makeup of the battery changed slightly over time, each session began with the completion of a card recording age, birthplace, marital status (married, unmarried, or widowed), residence (urban, suburban, or country), and occupation. The examinee's name, birth date, and initials were collected in the laboratory's later years, with the full name indexed in a separate list. The examiner then recorded the color of the examinee's eyes and hair, followed by tests and measurements of sensory acuity, stature, strength, and lung capacity:

- Eyesight keenness, color sense, and judgment in estimating length and squareness
- Hearing keenness and highest audible note
- Height standing, without shoes
- Height sitting, from seat of chair
- Span of arms (between opposite fingertips, with arms fully extended)
- Weight, in usual indoor clothing
- Breathing capacity (volume of air exhaled after a deep breath)
- Strength of pull (as an archer draws a bow)
- Strength of grasp (squeeze with the strongest hand)
- Swiftness of blow with fist (against a flat bar with pad at one end)

Specialized instruments (some invented by Galton) were employed, such as the spirometer, which required exhaling into a tube to measure the number of cubic inches of water displaced in a tank. Galton (1890b) interpreted breathing (lung) capacity as an indicator of energy level:

> The possession of a considerable amount of breathing capacity and of muscular strength is an important element of success in an active life, and the rank that a youth holds among his fellows in these respects is a valuable guide to the selection of the occupation for which he is naturally fitted, whether it should be an active or a sedentary one. (p. 238)

Galton constructed normative distributions for each measurement, including mean values and *percentile ranks* (i.e., the percentage of cases falling below the obtained score) in specified age ranges, differentiated by gender. Some measures, like breathing capacity and strength of grip, were assessed in relation to stature. It was possible to look at a normative chart and instantly know your approximate percentile rank. After collecting data on nearly 10,000 examinees at the International Health Exhibition, Galton's laboratory at South Kensington collected data on an additional 3,678 examinees (Galton, 1892), so adult norms were based on fairly large samples.

Galton never directly asserted that his tests measured *intelligence.* Instead, he observed that sensory measures are relevant in determining the breadth of experience upon which intelligence can operate:

> The only information that reaches us concerning outward events appears to pass through the avenue of our senses; and the more perceptive our senses are of difference, the larger is the field upon which our judgment and intelligence can act. (1907, p. 19)

In 1890, Galton acknowledged that only research could reveal the most important areas of human functioning to measure, through careful examination of test results and correlations with external criteria:

> One of the most important objects of measurement is hardly if at all alluded to here and should be emphasized. It is to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. We thus may learn which of the measures are the most instructive. (1890a, p. 380)

The uncertainty, of course, was where to sink the "shafts"—or, in other words, which abilities to measure.

With the methods initiated by Galton, favorable perspectives about the scientifically based mental measurement of individual differences began to crystallize in the 1890s, and many independent research efforts were launched in the United States and Europe. Charles E. Spearman (1904, pp. 206–219) counted over 30 international investigators studying mental tests, and this was probably an underestimate. The quest for mental tests is generally agreed to have started in Great Britain with Galton's initiatives, but Spearman's discovery of a *general intellectual factor,* described in a later section, would almost immediately begin to guide theory development. The earliest U.S. efforts in mental testing came through large-scale studies from James McKeen Cattell (Cattell & Farrand, 1896) at Columbia University; Franz Boas, then at Clark University (see Bolton, 1892); J. Allen Gilbert (1894) at Yale University; and Joseph Jastrow (1893) at the University of Wisconsin–Madison. In France, Alfred Binet and his colleagues (principally Victor Henri and then Théodore Simon) were the pioneers. Germany's early contributors included Hermann Ebbinghaus and Emil Kraepelin, especially his student Axel Oehrn (1896).

It is debatable whether efforts to develop a working intelligence test ever became a scientific race, like the competitive quest to identify the molecular structure of deoxyribonucleic acid (DNA) or the pursuit of space travel technology to become the first nation to land a person on the moon. Certainly, there was constant comparison between test development efforts in different nations. For example, Sharp (1899) reviewed the competing perspectives of "M. Binet and the French psychologists," "Prof. Kraepelin and the German psychologists," and the American psychologists (p. 334), pitting the assertions of each research group against one another. In journals like *L'Année Psychologique,* Binet and his colleagues could be found reviewing work by all competing laboratories, even commenting on Sharp's paper. After Spearman (1904) described his statistical method of quantifying "general" intelligence, competition between research groups may have become even more pronounced because a more focused end goal had been specified (i.e., a test of "general" intelligence per se, rather than random tests of associated mental processes). As shown in Figure 1.5, the practice of intelligence assessment in the earliest years of the 20th century essentially consisted

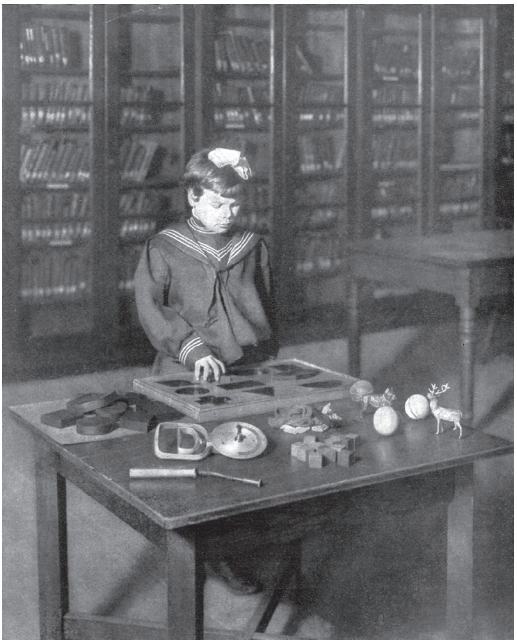


FIGURE 1.5. A photograph depicting the array of tasks used to measure intelligence at Lightner Witmer's Psychological Clinic at the University of Pennsylvania in about 1908. On the table are a Galton whistle for testing the upper limit of sound perception; a dynamometer for testing hand strength; colored yarns and blocks for testing counting skills and color perception; toys to test common knowledge, play, instinctive reactions, and coordination; and the formboard for identifying nonverbal problem solving and detecting feeble-mindedness. From Carter (1909, p. 166). In the public domain.

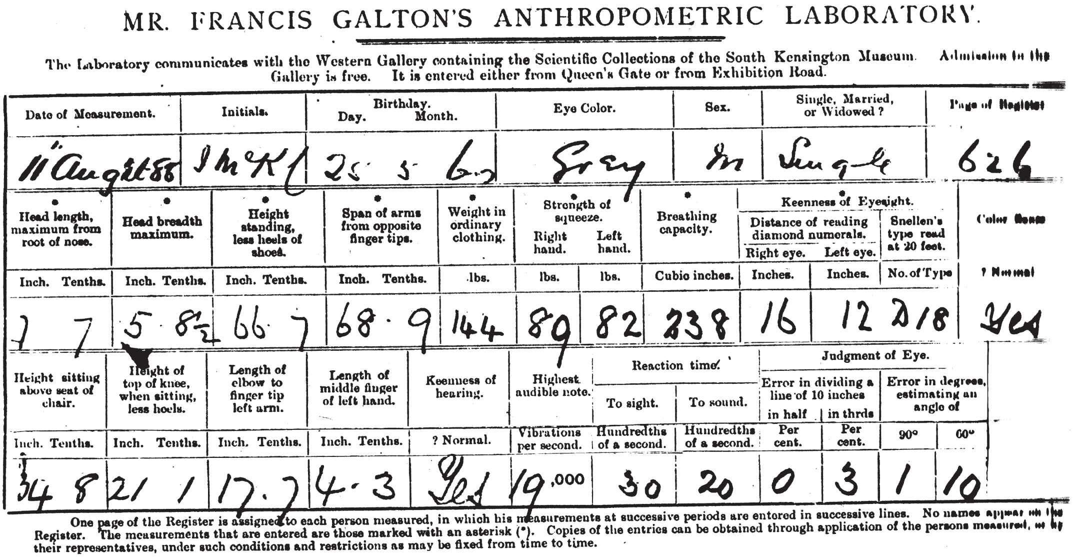
of an array of sensory and motor measures, with a formboard to measure higher mental processes.

James McKeen Cattell and the End of Anthropometrics

If there were royalty in the field of psychology, James McKeen Cattell (1860–1944) might qualify. He was the son of a professor at (and later the president of) Lafayette College in Easton, Pennsylvania, where he graduated as valedictorian in 1880. After studying for 2 years in Germany, he won a fellowship at Johns Hopkins University, where he began researching the timing of various mental processes in G. Stanley Hall's "physiologico-psychological laboratory" (Sokal, 1981, p. 64). He left to study with Wilhelm Wundt, the father of experimental psychology, at the University of Leipzig, in Germany, where he worked from 1883 to 1886 before receiving his doctorate. His article "The Time It Takes to See and Name Objects" (Cattell, 1886) summarized two of his studies on basic reading processes, which are now considered to be the first research studies to support a whole-word, sight-reading approach to reading instruction (Venezky, 2002, p. 6). Rejecting Wundt's reliance on experimenter introspection, Cattell conducted reaction time experiments with some of his own instruments, growing interested in the measurement of individual differences. According to his biographer, Michael M. Sokal, Cattell "refocused psychological research away from experimenters' self-observation of their mental activity and toward subjects' behavior in a laboratory setting precisely defined by experimenters" (Sokal, 2006, p. 25). In just a few years, Cattell would become the leading American experimental psychologist of his time.

In 1887, Cattell took a position at the University of Cambridge, where he came to know and work closely with Francis Galton. Cattell's data card from his personal anthropometric measurements appears in Figure 1.6. Cattell helped Galton set up the Anthropometric Laboratory at South Kensington. Cattell would remain devoted to Galton for the rest of his life, acknowledging in his late 60s that Galton was "the greatest man whom I have known" (Cattell, 1930, p. 116). For 2 years, Cattell split his time between work in Galton's laboratory, lecturing and establishing a laboratory at Cambridge University, and lecturing also at Bryn Mawr College and the University of Pennsylvania in the United States. In 1888, Cattell became a professor of psychology at the University of Pennsylvania (the first such professorship established anywhere, he claimed). In 1891, Cattell relocated to Columbia University, where he became the administrative head—beginning Columbia's experimental psychology laboratory and mentoring doctoral students like Edward L. Thorndike, Robert S. Woodworth, and Harry L. Hollingworth, who would themselves become faculty and leading figures in psychology. Over 40 students would take their doctorates with Cattell, seven of them becoming presidents of the APA. Cattell himself served as president of the APA in 1895.

With respect to intelligence testing, Cattell is a seminal historical figure due to his tireless advocacy for psychology as a science, his own test development efforts, and his advocacy for psychometrics and testing, as well as his emphasis on statistical analyses of individual differences, all of



MR. FRANCIS GALTON'S ANTHROPOMETRIC LABORATORY.

The Laboratory communicates with the Western Gallery containing the Scientific Collections of the South Kensington Museum. Admission to the Gallery is free. It is entered either from Queen's Gate or from Exhibition Road.

Date of Measurement.	Initials.	Birthday. Day. Month.	Eye Color.	Sex.	Single, Married, or Widowed?	Page of Register
11 Augt 88	J McK	25 5 60	Grey	m	Single	626

• Head length, maximum from root of nose.		• Head breadth maximum.		• Height standing, less heels of shoes.		• Span of arms from opposite finger tips.		• Weight in ordinary clothing.	• Strength of squeeze.		• Breathing capacity.	• Keenness of Eyesight.			Color Sense
									Right hand.	Left hand.		Distance of reading diamond numerals.		Snellen's type read at 20 feet.	
												Right eye.	Left eye.		
Inch.	Tenths.	Inch.	Tenths.	Inch.	Tenths.	Inch.	Tenths.	lbs.	lbs.	lbs.	Cubic inches.	Inches.	Inches.	No. of Type	? Normal
7	7	5	8½	66	7	68	9	144	89	82	238	16	12	D18	Yes

Height sitting above seat of chair.		Height of top of knee, when sitting, less heels.		Length of elbow to finger tip left arm.		Length of middle finger of left hand.		Keenness of hearing.	Highest audible note.	Reaction time.		Judgment of Eye.			
										To sight.	To sound.	Error in dividing a line of 10 inches		Error in degrees, estimating an angle of	
												in half	in thrds		
Inch.	Tenths.	Inch.	Tenths.	Inch.	Tenths.	Inch.	Tenths.	? Normal.	Vibrations per second.	Hundredths of a second.	Hundredths of a second.	Per cent.	Per cent.	90°	60°
34	8	21	1	17	7	4	3	Yes	19,000	30	20	0	3	1	10

One page of the Register is assigned to each person measured, in which his measurements at successive periods are entered in successive lines. No names appear on the Register. The measurements that are entered are those marked with an asterisk (*). Copies of the entries can be obtained through application of the persons measured, or by their representatives, under such conditions and restrictions as may be fixed from time to time.

FIGURE 1.6. Measurement data card recorded in 1888 at Galton's Anthropometric Laboratory for J. McKeen Cattell, who was deeply influenced by Francis Galton. Papers of James McKeen Cattell, 1835–1948, Library of Congress, Washington, D.C. In the public domain.

which established a fertile environment for test development at Columbia University and in American psychology in general. In the British journal *Mind,* Cattell (1890) used the term *mental tests* for the first time:

> Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement. A step in this direction could be made by applying a series of mental tests and measurements to a large number of individuals. The results would be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting, and perhaps, useful in regard to training, mode of life or indication of disease. (p. 373)

Cattell made his principal research initiative at Columbia an investigation to determine whether a battery of Galtonian anthropometric tests and sensory, motor, and higher cognitive tasks could measure intelligence. Beginning in 1894, the Cattell–Columbia Tests (as Cattell referred to them in 1924) were given to freshmen at Columbia's School of Arts and School of Mines. With student consent, the tests were to be repeated at the end of the sophomore and senior years. In the course of an hour, 26 measurements were made in the laboratory, and 44 observations were recorded. Later, each student sent in answers to 50 questions with regard to background, health, physical condition, habits (including coffee, smoking, alcohol use, and exercise), and interests. Cattell also had access to student academic records and athletic accomplishments.

Tests and measurements conducted in the laboratory included some of Galton's sensory measures; some of Cattell's reaction time measures; and some newer measures, including letter cancellation, rapid color naming, memory for digits, logical memory, self-reported retrieval of mental images, and a word association test. The battery was something of a hybrid of anthropometric, lower-order, and higher-order measures. Cattell had always relied on the experimental approach as producing descriptive results that would speak for themselves; he did not offer a priori hypotheses or even articulate his concept of intelligence. Cattell's commitment to quantitative measurement and statistical analysis of experimental results was unshakeable, and as late as 1924, Cattell, pictured in Figure 1.7, still expressed a belief that his test battery might correlate with long-term student accomplishments. He would not have the chance to find out, as two studies would put a conclusive end

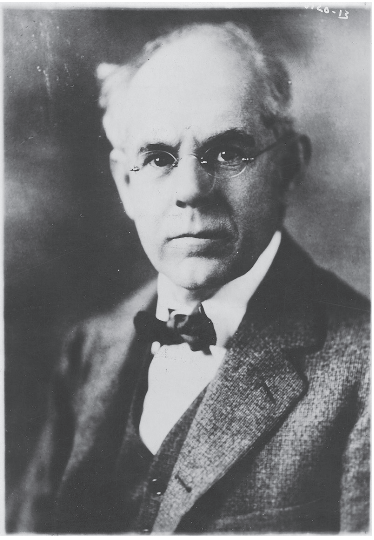


FIGURE 1.7. James McKeen Cattell at the age of 63 in December 1923. Long after the failure of his anthropometric testing program and after his 1917 dismissal from Columbia University, Cattell founded The Psychological Corporation and continued to edit several scientific journals. From the chapter author's personal collection.

to Cattell's approach to intelligence testing and his experimental research efforts.

First, a dissertation completed by Stella Sharp (1899) in Edward B. Titchener's laboratory at Cornell University sought to examine the variability of complex mental processes and the relations between complex mental processes, with the intention of demonstrating the practicality of testing complex processes rather than the simpler mental processes endorsed by Cattell and Galton. She assessed seven advanced philosophy students at the university with the test battery formulated by Binet and Henri (1895), including measures of memory, mental images, imagination, attention, observation/comprehension, suggestibility, and aesthetic tastes. Her results listed the scores of individual participants and described these results in terms of rank order and variability. Sharp concluded:

> We concur with Mm. Binet and Henri in believing that individual psychical differences should be sought for in the complex rather than in the elementary processes of mind, and that the test method is the most workable one that has yet been proposed for investigating these processes. (p. 390)

She further concluded that the Binet–Henri measures required modification but were practical and yielded considerable variation in scores. She offered only qualitative observations about relations between tests of different mental processes, however. Although she did not collect data on other assessment approaches, she was critical of the anthropometric tests as unproven and lacking an explanatory theory.

The second blow to Cattell's testing program, and its *coup de grâce*, came from a Columbia University psychology graduate student, Clark Wissler (1901). Wissler examined the correlations between the Cattell–Columbia Tests and student grades for 300 undergraduates at Columbia and Barnard Colleges. He reported that while isolated correlations were large (e.g., height and weight, $r = .66$; Latin and Greek grades, $r = .75$), the laboratory mental tests had negligible correlations with each other and with college class grades. The failure to correlate with academic grades was considered fatal to Cattell's testing program because academic performance had long been considered an independent criterion measure of intelligence. In the words of Cattell's biographer, Wissler's analysis would definitively "discredit anthropometric testing" (Sokal, 2006, p. 29).

It remains to note that over a century after Galton's and Cattell's testing programs were discredited, the relations of elementary cognitive processes (reaction time and sensory discrimination) to mental abilities and intelligence are now being revisited. Jensen (2006) has effectively summarized the literature relating reaction time to intelligence, while Deary and his colleagues (Deary, 1994; Deary, Bell, Bell, Campbell, & Fazal, 2004) have documented findings with sensory discrimination and intelligence. There is uniform agreement as to the serious methodological flaws in the Sharp and Wissler studies, including small sample size, restriction of range, and unreliability of measures (e.g., Buckhalt, 1991; Deary, 1994; Jensen, 2006).

THE ORIGINS OF CONTEMPORARY INTELLIGENCE TESTING

I have described the pseudoscience of phrenology and the visionary science of Galton, who inspired the search for effective ways to measure intelligence and who pioneered many statistical methods that would be critical for norm-referenced assessment. I have also recounted the tale of the psychologist that Galton so profoundly influenced, J. McKeen Cattell, who threw down a gauntlet of sorts when proposing that *mental tests* should constitute part of establishing psychology as a science that can measure individual differences. The unfortunate fate of Cattell's test battery has been told. Even after his assessment work was discredited, however, Cattell remained a highly connected institutional scientist and a pioneer in the development of scientific psychology in American universities.

Ironically, the problem of developing a working intelligence test would be solved by an outsider, a man with few friends, who worked without pay and who had no institutional connections of any benefit. He did have his own journal, however, where he reviewed the work of his contemporaries. His name was Alfred Binet.

Alfred Binet: The Innovative Outsider

Alfred Binet (1857–1911) is generally acknowledged as the father of intelligence tests, having developed the first working measure of intelligence (see Figure 1.8). Binet's remarkable history has been most definitively documented by biographer Theta Wolf (1973). He was educated as a lawyer, but chose not to practice; some historical accounts also report that Binet studied medicine until his father, a physician, traumatized him by showing him a cadaver. As an only child of a wealthy family, he could afford to pursue a career with little remuneration, and he developed a consuming interest in the study of psychology. He was a voracious reader across several languages who educated himself as a psychologist, spending considerable time studying in the Bibliothèque Nationale [French National Library]. He wrote his first article at age 23 and completed a doctorate in the natural sciences at age 37. According to long-time colleague Théodore Simon, for most of Binet's career "psychology was his sole occupation" (quoted by Wolf, 1973, p. 9).

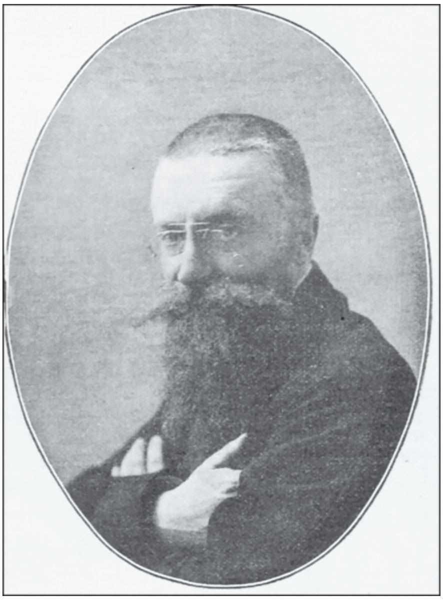


FIGURE 1.8. Alfred Binet in 1910. Reprinted courtesy of the Professor Serge Nicolas Private Collection.

Although he is remembered for his intelligence test (from which he does not appear to have profited financially), Alfred Binet was a remarkably productive and versatile researcher, authoring nearly 300 works during his career; he is now credited with pioneering experimental investigations in areas of abnormal, cognitive, developmental, educational, forensic, personality, and social psychology (e.g., Siegler, 1992; Wolf, 1973). Regrettably, most of his work has never been translated into English, although nearly all of it has been brought back into print in the last two decades. Personally, he has been described as a loner, "a reserved man with few friends" (Tuddenham, 1974, p. 1071), and as a domineering individual who antagonized many of his coworkers (cf. Henri Piéron, according to an interview with Wolf, 1961, p. 246). In 1901, Binet wrote a friend, "I educated myself all alone, without any teachers; I have arrived at my present scientific situation by the sole force of my fists; *no one,* you understand, no one, has ever helped me" (quoted by Wolf, 1973, p. 23). Lacking patronage, he was denied academic positions in France (Nicolas & Ferrand, 2002), and his efforts for educational reform and mental measurement in the military were resisted by a rigid French establishment (e.g., Carson, 2007; Zazzo, 1993). Several scholars have sought to explain why so much of his work was forgotten after his death (e.g., Fancher, 1998; Schneider, 1992; Siegler, 1992; Wolf, 1973), and the answer seems to lie in his disconnection from the professional and academic community in France: He did not present at conferences, he did not leave students to continue his work, and he preferred to work alone or with a collaborator. Only at the 2005 centennial of the first Binet–Simon scale did he begin to garner national recognition in his native France for his remarkable contributions.

From 1890 through his death, Binet published more than 200 articles, many in the journal *L'Année Psychologique,* which he cofounded and edited. In 1891, he became an unpaid staff member at the Laboratory of Physiological Psychology at the Sorbonne. Three years later, he became director of the laboratory, a position he held until his death. Between 1894 and 1898, Binet and Victor Henri sought new methods that would "substitute for vague notions of man in general, of the archetypal man, precise observations of individuals considered in all the complexity and variety of their aptitudes" (Binet & Henri, 1894, p. 167; translated and cited by Carson, 2007, p. 132). In 1899, Binet was approached by a young intern physician and *aliéniste* (psychiatrist), Théodore Simon, who had access to clinical populations (Wolf, 1961). Simon completed his doctoral thesis under Binet's supervision, and their subsequent collaborations included Binet's most important work in intelligence assessment.

The creative work that culminated in the intelligence scales began in 1890, when Binet published three papers describing experimental studies with his two young daughters, Madeleine and Alice (given the pseudonyms Marguerite and Armande), whom he had carefully observed and tested with a variety of cognitive and personality tasks. In describing their attentional styles, he wrote that Madeleine was "silent, cool, concentrated, while Alice was a laugher, gay, thoughtless, giddy, and turbulent" (translated by Wolf, 1966, p. 234). Madeleine had greater "stability" and had better voluntary control of her attention; she could more effectively focus on assigned work and memorize neutral, uninteresting material; and she tended to respond with shorter and more constant reaction times. Alice presented more "variability"; she was more imaginative and emotional; and material to be learned had to be of interest to her, or she would have difficulty. She could not memorize long literary passages verbatim as her sister could, but she could accurately remember a series of ideas provided just once (see Wolf, 1973, p. 132). Binet continued to test his daughters through midadolescence with a battery of cognitive and personality tests, including measures of attention, language, reasoning, and memory (many repeated multiple times in alternative forms over several years), always accompanied by careful qualitative observation and interview inquiries. He reported the results in 1903 in *L'étude Expérimentale de l'Intelligence [The Experimental Study of Intelligence]*.

Comparison of his children's performances with each other and with those of adults led Binet to conclude that complex, multidimensional tasks were more sensitive to developmental changes than narrow, unidimensional tasks. He further concluded that a mental developmental progression from childhood through adulthood should be reflected in task performance:

> In case one should succeed in measuring intelligence—that is to say, reasoning, judgment, memory, the ability to make abstractions—which appears not absolutely impossible to me, the figure that would represent the average intellectual development of an adult would present an entirely different relation to that of the figure representing the intellectual development of a child. (1890, p. 74; translated by Wolf, 1966, p. 235)

In 1895, Binet and Henri outlined the project for the development of an intelligence test, specifying 10 discrete mental faculties that would be measured: memory, imagery, imagination, attention, comprehension, suggestibility, aesthetic sentiment, moral sentiment, muscular strength/willpower, and motor ability/hand–eye coordination. Higher-order, complex processes were considered to show greater variability among individuals and to constitute better measures of intelligence than simpler sensory and motor processes:

> The higher and more complex a process is, the more it varies in individuals; sensations vary from one individual to another, but less so than memory; memory of sensations varies less than memories of ideas, etc. The result is, that if one wishes to study the differences existing between two individuals, it is necessary to begin with the most intellectual and complex processes, and it is only secondarily necessary to consider the simple and elementary processes. (Binet & Henri, 1895, p. 417; translated by Sharp, 1899, p. 335)

In a passage that made direct reference to the work of Galton and Cattell, Binet and Henri (1895) rebutted the claim that greater experimental precision was possible in the measurement of simpler mental processes:

> If one looks at the series of experiments made—the *mental tests* as the English say—one is astonished by the considerable place reserved to the sensations and the simple processes, and by the little attention lent to the superior processes. . . . The objection will be made that the elementary processes can be determined with much more precision than the superior processes. This is certain, but people differ in these elementary ones much more feebly than in the complex ones; there is no need, therefore, for as precise a method for determining the latter as for the former. . . . Anyway, it is only by applying one's self to this point that one can approach the study of

> individual differences. (pp. 426, 429; translated by Siegler, 1992, p. 181)

The formal mandate that led to the development of the intelligence test came in October 1904, when Joseph Chaumié, the Minister of Public Instruction, established a commission chaired by Léon Bourgeois and charged it with studying how France's 1882 mandatory public education laws could be applied to abnormal [*anormaux*] children, including students who were blind, deaf-mute, and backward [*arriérés*] (Carson, 2007; Wolf, 1969). The ministry was persuaded to take this initiative by public pressure, including a resolution from the 1903 Third National Congress of Public and Private Welfare held at Bordeaux, where critics noted France's failure to comply with its own special education laws. A resolution from an educational advocacy group, La Société Libre pour l'Étude Psychologique de l'Enfant [Free Society for the Psychological Study of the Child], was also a reason for creation of the Bourgeois Commission. Binet was a leader of La Société Libre and an author of the resolution, and he became a member of the commission, which began its work by circulating questionnaires to teachers and principals throughout France. The commission met on numerous occasions in 1904 and 1905, issuing its report in 1906.

Binet saw in the commission's mandate an opportunity to complete his efforts toward a norm-referenced standard for diagnosis and educational decision making. Building on the earlier work with Henri (who had departed), Binet and Simon developed and tested a series of cognitive tests. Their collaborations worked in tandem: One of them would talk with and question the examinee, while the other wrote the replies and noted salient behaviors. The assessments had "the air of a game" for children, with encouragement being constantly provided (1905/1916a, p. 141). The work culminated in the 1905 publication of the Binet–Simon Intelligence Scale (Binet & Simon, 1905/1916c), consisting of 30 items that could be given in about 20 minutes; it was normed on some 50 children from ages 3 through 11 years; and one of its chief advances may have been to combine a wide range of cognitive tasks to obtain a global estimate of intelligence (e.g., DuBois, 1970). Binet and Simon (1905/1916c) sequenced tasks in a cognitive-developmental order from easy to hard and from simpler to more complex, while sampling a wide range of tasks tapping various abilities. In general, they sought tasks that tapped the higher-order ability of *judgment*—especially procedures that had demonstrated the capacity to differentiate groups on the basis of intelligence. For example, individuals considered *idiots* generally could not move beyond the sixth of the 30 tasks; individuals considered *imbeciles* rarely went beyond the 15th task (Binet & Simon, 1905/1916a).

The Bourgeois Commission issued its report early in 1906, based primarily on a subcommittee report drafted by Binet. Recommendations were that the *anormaux* be educated through *classes spéciales* annexed to ordinary primary schools and, in certain situations, through separate institutions. A five-part classification of exceptional students was proposed, identifying students who were blind, deaf, medically abnormal, intellectually backward, and emotionally unstable. The commission recommended that students who did not benefit from education, teaching, or discipline should receive a "medico-pedagogical examination" before being removed from primary schools, and that such children, if educable, should be placed in special classes. The examination was to be overseen by an examination committee consisting of an inspector of primary schools, a physician, and a director of the separate special school. The commission did not offer any specific content for the examination, recommending that the Minister of Public Instruction appoint a competent person to draw up a scientific guide for the school examination committee (Carson, 2007). Undoubtedly Binet hoped to draw up the scientific guide, and Binet and Simon's book *Les Enfants Anormaux* (1907/1914) was probably intended to serve as the guide; it even contained a preface by Léon Bourgeois, the head of the commission.

Unfortunately, Binet's efforts were almost completely rebuffed by the French establishment. When the French legislature enacted the law of April 15, 1909, on the education of the *anormaux*, it stated that the commission determining eligibility for special education should be composed of a physician, school inspector, and director or teacher at an *école perfectionnement*. It highlighted the medical examination and made no mention of any role for psychologists or use of special methods (i.e., intelligence tests) for assessing students (Carson, 2007). Binet's efforts had little visible impact on practice in his native France.

In the 1908 revision, the Binet–Simon Scale took its definitive revolutionary form, the "graded scale of intelligence" [*L'échelle métrique de l' intelligence*], which was easier to use and interpret (Binet & Simon, 1908/1916b). It featured 56 tests arranged by difficulty so that tests were placed at levels, or grades, corresponding to approximately a 75% pass rate for children of a given age, based on normative performances of about 200 children between the ages of 3 and 15. The 1908 scale permitted a student's mental level [*niveau mental*] to be estimated through what later became interpreted in the United States as a mental age level. The mental level was determined by the highest age at which a child passed four or five tests (the basal year), with an additional year credited for each of the five tests passed beyond the basal. By the completion of the 1911 edition (Binet, 1911/1916), the scale was extended from age 3 through adulthood, with 11 levels and five items administered at each level. Table 1.1 lists content from the final 1911 scale. The Binet–Simon Scale never yielded an intelligence quotient (IQ), but Binet endorsed the convention of identifying intellectual disability [*arriérés*] for a mental level delay of "two years when the child is under [age] nine, and three years when he is past his ninth birthday" (Binet & Simon, 1907/1914, p. 42). Long after Binet's death, Simon indicated that the use of a summary IQ score was a betrayal [*trahison*] of the scale's objective (cited by Wolf, 1973, p. 203).

In the spring of 1908, Henry H. Goddard, director of the psychological research laboratory at the New Jersey Training School for Feeble-Minded Girls and Boys (later known as the Vineland Training School), traveled to Europe. He visited doctors and teachers working in 19 different institutions and 93 special classes. Ironically, he did not even look up Binet in Paris, having been told by Pierre Janet that "Binet's Lab. is largely a myth . . . Not much being done—says Janet," according to his journal (cited by Zenderland, 1998, pp. 92–93). In Brussels, he met Ovide Decroly, a Belgian teacher, physician, and psychologist, engaged in a tryout of the 1905 Binet–Simon Scale. Decroly provided him with a copy of the test, and upon his return home, Goddard began to use the test on the children at the training school. In the words of Goddard's biographer Leila Zenderland (1998), Goddard immediately understood the significance of the Binet–Simon Scale:

> Two years of frustrating institutional experience had prepared him to see what Janet, Cattell, and even [G. Stanley] Hall, the most prescient of contemporary psychological entrepreneurs, had missed. Contained within Binet's articles, Goddard quickly realized, was an entirely new psychological approach toward diagnosing and classifying feeble minds. (p. 93)

In a short time, Goddard would become the United States' leading advocate for Binet's approach to assessment and diagnosing intellectually disabled individuals. He described his evaluation of the ease, simplicity, and the utility of the 1908 scale as "a surprise and a gratification" (1916, p. 5), and he promoted the test widely. The Binet–Simon Scale was both praised and criticized widely in professional journals; for example, several consecutive issues of the *Journal of Educational Psychology* in April, May, and June 1916 were dedicated to "Mentality Tests: A Symposium," a wide-ranging exchange of experiences with the Binet–Simon Scale (and other tests) by 16 leading psychologists. Goddard arranged for Elizabeth S. Kite, his laboratory's field worker and contributor to the famous Kallikak study, to complete the definitive translations into English of Binet and Simon's writings on their intelligence scale. By 1916, the Vineland laboratory had distributed 22,000 copies of a pamphlet describing administration of the Binet–Simon Scale and 88,000 record forms, as well as publishing a two-volume translation of the Binet–Simon articles (Goddard, 1916). By 1939, there were some 77 available adaptations and translations of the Binet–Simon Scale (Hildreth, 1939), including the most used psychological test of all, the Stanford–Binet. According to Théodore Simon, Binet gave Lewis M. Terman at Stanford University the rights to publish an American revision of the Binet–Simon Scale "for a token of one dollar" (cited by Wolf, 1973, p. 35). Terman's work would change the landscape for mental testing in the United States.

The Binet–Simon Intelligence Scale represented a major paradigm shift for the young field of psychology. It tapped intelligence through assessment of complex mental abilities, as opposed to the narrow sensory and motor measures dominating the Galton–Cattell batteries. It was standardized, with explicit procedures for administration and objective scoring guidelines. It was norm-referenced, permitting an individual's performance to be compared with that of his or her age peers. It was reliable, yielding consistent scores from one occasion to another. It was developmentally sensitive, recognizing that mental abilities in children develop in a meaningful progression and that the abilities of children differ substantially from that of adults. It was efficient and engaging, administered in an adaptive format in which content changed frequently. It offered clinical assessment, aimed at diagnosing intellectual disabilities, identifying cognitively advanced students, and describing

TABLE 1.1. Contents of the Binet–Simon (Binet, 1911/1916) Intelligence Scale [*L'Échelle Métrique de l' Intelligence*]

Three years

Show eyes, nose, mouth
Name objects in a picture
Repeat 2 figures
Repeat a sentence of 6 syllables
Give last name

Four years

Give sex
Name key, knife, penny
Repeat 3 figures
Compare 2 lines

Five years

Compare 2 boxes of different weights
Copy a square
Repeat a sentence of 10 syllables
Count 4 sous
Put together two pieces in a "game of patience"

Six years

Distinguish morning and evening
Define by use
Copy diamond
Count 13 pennies
Compare 2 pictures esthetically

Seven years

Right hand, left ear
Describe a picture
Execute 3 commissions
Count 3 single and 3 double sous
Name 4 colors

Eight years

Compare 2 objects from memory
Count from 20 to 0
Indicate omission in pictures
Give the date
Repeat 5 digits

Nine years

Give change out of 20 sous
Definitions superior to use
Recognize the value of 9 pieces of money
Name the months
Comprehend easy questions

Ten years

Place 5 weights in order
Copy a design from memory
Criticize absurd statements
Comprehend difficult questions
Place 3 words in 2 sentences

Twelve years

Resist the suggestion of lines
Place 3 words in 1 sentence
Give more than 60 words in 3 minutes
Define 3 abstract words
Comprehend a disarranged sentence

Fifteen years

Repeat 7 figures
Find 3 rhymes
Repeat a sentence of 26 syllables
Interpret a picture
Solve a problem composed of several facts

Adults

Comprehend a cut in a folded paper
Reversed triangle
Answer the question about the President
Distinguish abstract words
Give the sense of the quotation from Hervieu

Note. The final 1911 Binet–Simon Scale extended from 3 years into adulthood. In this edition, an individual's mental level [*niveau mental*] was estimated by identifying the highest age at which all the tests were passed (the basal year), to which is added one-fifth of a year for every test passed. The Binet–Simon Scale never yielded an intelligence quotient (IQ), but Binet endorsed the convention of identifying intellectual disability for a mental-level delay of 2 years when a child is under age 9, and 3 years past his or her 9th birthday. From Binet (1911/1916). In the public domain.

the characteristics of both "normal" and "abnormal" individuals. Finally and most importantly, it seemed to work fairly well, providing an empirical foundation for the nascent study of intelligence and cognitive abilities.

Lewis M. Terman: Defender of the Discipline

> I hate the impudence of a claim that in fifty minutes you can judge and classify a human being's predestined fitness in life. I hate the pretentiousness of that claim. I hate the abuse of scientific method which it involves. I hate the sense of superiority which it creates, and the sense of inferiority which it imposes.
>
> —WALTER LIPPMANN (1923)

When journalist Walter Lippmann launched the first high-profile public attack on intelligence testing in a series of articles in *The New Republic* (Lippmann, 1922a, 1922b, 1922c, 1922d, 1922e, 1922f, 1923), it was Lewis M. Terman (1922a) who responded and defended the new discipline. He was the natural choice—developer of the Stanford University revision of the Binet–Simon Intelligence Scale (later called the Stanford–Binet Intelligence Scale); member of the National Research Council team that created the Army mental tests in 1917 and 1918; coauthor of the National Intelligence Tests and Terman Group Test of Mental Ability, released in 1920; principal investigator on the longitudinal Genetic Studies of Genius, initiated in 1921–1922; and coauthor of the Stanford Achievement Test, which would be released in 1923. For decades, Terman would be the living American most strongly associated with intelligence testing and its value for educational decision making.

The 12th of 14 children from a rural Indiana farming family, Lewis M. Terman (1877–1956) was a brilliant, hard-working, and determined student from an early age; he accelerated from first grade to third grade and memorized most of his textbooks. Graduating early from eighth grade (the conclusion of education in typical Midwest farming communities of that era), he began teacher's college at the age of 15, attending when he could and taking breaks to earn enough money to return. Terman pursued training in education, as teaching was the "only avenue of escape for the youth who aspired to anything beyond farm life" (1932, p. 300); eventually he would teach for one year in a one-room schoolhouse. By the age of 21, he had earned three baccalaureate degrees from Central Normal College in Danville, Indiana, and he became a principal of a small high school. He then pursued a master's degree in psychology at Indiana University, followed by a doctorate at Clark University. In 1905, recurrent tubercular hemorrhages in his lungs (eventually the cause of his death) forced Terman to relocate his family to Southern California, where he worked again as a high school principal and then as a professor of pedagogy at Los Angeles State Normal School (later UCLA) before accepting a position in 1909 at Stanford University, where he remained for the duration of his career. Figure 1.9 shows Terman at about the time he started his career at Stanford University.

Terman is described by two biographers, Henry L. Minton (1988) and May V. Seagoe (1975), as having been a highly gifted man and voracious learner, who was tirelessly persistent, intense, and sensitive. As a rigorous and careful researcher, he became a pioneer in mental testing by creating the best of many adaptations of the Binet–Simon Scale. He also harbored a progressive vision of large-scale testing to identify the individual differences and needs of schoolchildren, as well as to identify intellectually gifted children (Chapman, 1988). Like Cattell, Terman had been seen by a

FIGURE 1.9. Lewis M. Terman in 1910, the year he arrived at Stanford University. Terman was the leading advocate for intelligence testing in the first half of the 20th century. Reprinted by courtesy of the Department of Special Collections and University Archives, Stanford University Libraries.

phrenologist as a child; he was deeply impressed by the experience and remembered that the phrenologist "predicted great things of me" (Terman, 1932, p. 303). Having spent 6 months each year during his adolescence toiling at farmwork from 5:00 A.M. through about 7:00 or 8:00 P.M., Terman considered his intellectual abilities to have been inherited; he remembered his lengthy stints at farmwork as periods without mental stimulation, contributing to his conviction that environment was substantially less important than heredity in explaining intelligence.

Terman's master's thesis on leadership, his doctoral dissertation on genius, and his longitudinal study of gifted children beginning in 1921–1922 all contributed to his status as founder of the "gifted child" movement. Terman's thesis, published as a journal article in 1904, used experimental methodology (from Binet's suggestibility studies), teacher ratings, and questionnaires to examine leadership in male and female schoolchildren from grades 2 through 8. It is a qualitatively rich study that identifies different types of leaders and subtly links leadership with perceived intelligence. Terman's dissertation, completed in 1905 and published as a journal article in 1906, was entitled "Genius and Stupidity: A Study of Some of the Intellectual Processes of Seven 'Bright' and Seven 'Stupid' Boys." For his dissertation, Terman administered a variety of higher-order mental tests to seven boys identified by teachers as the "brightest" and seven boys identified as the "dullest," based on a holistic review (i.e., not merely based on classwork) of willing boys. All of the boys were 10–13 years of age. Terman tested the boys for about 20–40 hours in each of eight areas: creative imagination, logical processes, mathematical ability, mastery of language, interpretation of fables, ease of learning to play chess, powers of memory, and motor abilities. Some tests were culled from the literature, including measures from Binet and Henri, Ebbinghaus, and others; other tests were tasks developed by Terman that would reappear in the Stanford–Binet. Terman found that the bright boys were superior to the dull boys in all but the motor tests, with creative imagination tests showing modest differences between bright and dull boys. Most tests administered tended to agree with one another—a finding that Terman interpreted as supporting the presence of Spearman's general factor. Bright children preferred to read, while dull children preferred to play games; there was little difference between the two groups in terms of persistence.

In 1910, Terman began his revision of the Binet–Simon Scale, a technical *tour de force* that would be published in 1916. Terman began initial studies by administering the 1908 Binet–Simon Scale to some 400 schoolchildren, as well as examining all available published studies of age-level placement for the Binet tests. It soon became evident that some tests were misplaced, with tests at the lower age levels too easy and those at the upper age levels too hard. He also wanted to add tests to reach six at each age level, eventually augmenting the Binet–Simon with 36 new tasks and clarifying administration and scoring criteria. Terman, his students, and his colleagues tested some 700 additional children in pilot studies. Some of Terman's new tasks were noteworthy, including a 100-word vocabulary test yielding full credit for correct definitions, half credit for partially correct definitions, and no credit for incorrect responses; and (arguably) the first executive function measure, the Ball and Field Test of Practical Judgment (see Littman, 2004, for an account of its origins). Terman and Childs (1912a, 1912b, 1912c, 1912d) published a "tentative revision and extension" of the Binet–Simon Scale, but further revision was necessary, given the 1911 extension of the Binet–Simon Scale through adulthood. As Seagoe (1975) reports, Terman's "unfamiliarity with statistics" and dislike of the "drudgery of computation" (p. 47) caused him to rely heavily on Arthur S. Otis, and for later editions on Truman L. Kelley and Quinn McNemar, for statistical analyses and data management. Dahlstrom (1985) noted the critical contribution of Otis's statistical knowledge and skills to the 1916 Stanford–Binet. Otis would later make important contributions to the development of the Army mental tests.

Terman's final standardization sample for the 1916 Stanford–Binet included 905 participants between the ages of 5 and 14 years, all within 2 months of a birthday and drawn from public schools in California and Nevada. No foreign-born or minority children were included. Special population studies included 200 "defective" and "superior" children. The adult sample consisted of 150 adolescent delinquents, 150 unemployed men, 50 high school students, and 30 businessmen across California and Oregon. The overall sample was predominantly white, urban, and middle-class, with an average adult mental age of 15–17 years. The final 1916 Stanford–Binet consisted of 90 items—six at each age level from ages 3 to 10; eight items at age 12; six items at age 14; and six items at each of two adult levels (average adult, superior adult). Sixteen alternative tests were available for tests that were inappropriate or otherwise spoiled. Of the final 90 items, 60% were drawn from the Binet–Simon and 40% from Terman and other sources. Terman adapted William Stern's (1914) "mental quotient" to generate the IQ (mental age divided by chronological age, with the product multiplied by 100 to remove decimals). Although Terman was critical of Spearman's work, he explicitly stated that the Stanford–Binet measured general intelligence, in effect making the single IQ score a functional estimate of Spearman's *g* and treating intelligence as a unitary construct:

> The scale does not pretend to measure the entire mentality of the subject, but only *general intelligence*. There is no pretence of testing the emotions or the will beyond the extent to which these naturally display themselves in the tests of intelligence. (1916, p. 48; original emphasis)

Terman retained Binet's adaptive testing format, which permitted flexibility in determining at which level to start the test, and different item types were intermixed to make the testing experience a fast-moving experience with tasks changing frequently.

Terman's Stanford–Binet was a resounding success, becoming the most frequently used psychological test (and intelligence test) in the United States for decades (Louttit & Browne, 1947). The Stanford–Binet would be renormed and expanded to create two parallel forms (Form L for Lewis, and Form M for coauthor Maud A. Merrill) spanning the ages 2 years through Superior Adult III in a remarkable 1937 revision (Terman & Merrill, 1937). The best items from the two forms would be retained in a single form for two updates (Terman & Merrill, 1960, 1973). From sales of test record forms, R. L. Thorndike (1975) estimated that the Stanford–Binet was administered to an average of about 150,000 persons a year from 1916 to 1937, to about 500,000 persons a year from 1937 to 1960, and to about 800,000 a year from 1960 to 1972. The fourth edition would make radical changes, including conversion to a point scale format and assessment of discrete factors of ability according to extended Gf-Gc theory (Thorndike, Hagen, & Sattler, 1986), but the fifth edition would endeavor to restore some of the features that distinguished the Stanford–Binet from its start (Roid, 2003).

Terman was also responsible, more than any other psychologist, for the rapid growth of intelligence and achievement tests in schools. The "Oakland experiment" of 1917–1918 was one of the first systematic attempts to use intelligence/ability tests as a basis for grouping students—a movement that is well documented in Chapman's *Schools as Sorters* (1988). Beginning in 1917, one of Terman's students, Virgil E. Dickson, became director of research for the Oakland Public Schools and organized the testing of 6,500 schoolchildren with the Stanford–Binet, the Otis Absolute Point Scale, and other tests in all of Oakland's 45 elementary schools. From his findings, Dickson concluded that many students fail because their ability levels make mastery of the ordinary curriculum impossible; furthermore, he asserted, the "mentally superior" are in need of accelerated curricula. Dickson called for segregation of students into special classes based on their ability levels. Figure 1.10 depicts the introduction of intelligence tests in the schools. Receiving enthusiastic endorsements from administrators and teachers, Dickson (1919) concluded:

> Standard tests, both psychological and pedagogical—group and individual—should be of great assistance in classification of pupils according to ability and capacity to do the work. They should inspire better teaching and better educational guidance through a more intimate knowledge of the individual child. (p. 225)

In 1923, Dickson published *Mental Tests and the Classroom Teacher,* the first in a series of "measurement and adjustment" books to be edited by Terman and published through the World Book Company. In 5 years, Terman would oversee nine additional titles, each focusing on problems of student testing and "adjustments to meet the problems of instruction and school administration arising out of individual differences" (see Chapman, 1988, p. 104, for a description of Terman's blueprint for the series).

Large-Scale Assessments and the Army Mental Tests

In retrospect, it was a remarkable accomplishment: In the span of only 18 months during World War I, a small team of psychologists developed, tried out, and then directed the administration of the first intelligence measures designed for large-scale adult testing. By the time the Armistice was signed in November 1918, an estimated 1,726,966 Army enlisted men and officers had been tested with the new group tests. More than 83,500 en-

FIGURE 1.10. After the success of the "Oakland experiment" of 1917–1918, Terman and other psychologists advocated successfully for the use of intelligence tests to group students according to their ability levels. Educators recognized the value of measuring "individual differences" but were wary of the proliferating tests (e.g., Hines, 1922). From Heaton (1922). In the public domain.

listed men were also given individual examinations. Although the military was not particularly appreciative of the testing program, psychologists used the perceived success of the Army mental tests to sell the general public on the value of mental testing; large-scale assessment thus found its way into American education system, where it remains prominent today. Accounts of World War I Army mental testing are available in official narratives from the psychologist directing the process (e.g., Yerkes, 1919, 1921; Yoakum & Yerkes, 1920), as well as a number of independent scholars (e.g., Camfield, 1969; Carson, 1993; Kevles, 1968; Napoli, 1981; Samelson, 1977; Sokal, 1987; Spring, 1972; von Mayrhauser, 1986, 1987, 1989). I draw on these sources and others for the following history.

The story of the Army mental tests begins with the United States' lack of preparation for the war. World War I, which started in 1914, was fought mainly in Europe between the Allied Powers (the Russian and British Empires, France, and later Italy and the United States) and the Central Powers (the Austro-Hungarian, German, and Ottoman Empires and Bulgaria). An isolationist United States, under the leadership of President Woodrow Wilson, sought neutrality in what was perceived as a European conflict, leaving the U.S. military unprepared to enter the war. As of April 1917, the strength of the U.S. Army was below 200,000 men, the smallest number since the Civil War (e.g., Yockelson, 1998).

Wilson finally asked Congress for a declaration of war against Germany on April 2, 1917 (Wilson, 1917). Congress declared war 4 days later. President Wilson signed the Selective Service Act into law on May 18, 1917; within a few months, 10 million men had registered for the draft, with almost 2.8 million men actually being drafted by the U.S. Army (Baker, 1918). Under General John J. Pershing, troops of the American Expeditionary Forces (AEF) began arriving in Europe in June 1917.

The draft, however, had no established procedures to identify and exclude men who were unfit for service. There was also no way to identify large numbers of potential officers, since the existing officers had been selected and trained through the U.S. Military Academy at West Point and fell far short of needs. Secretary of War Newton Baker (1918, p. 15) wrote that "one of the most serious problems confronting the War Department in April 1917, was the procurement of sufficient officers to fill the requirements of the divisions that were to be formed for overseas duty." Moreover, there was no systematic way to assign men to specialized military jobs similar to those they had held in civilian life (e.g., assigning a practicing accountant to military requisitions tracking or record keeping). The massive draft provided an opportunity for the young scientific discipline of psychology to demonstrate the value of its still-new technologies—the intelligence test and personnel selection procedures—to efficiently screen large numbers of enlisted men.

Yerkes and the Army Mental Tests

The involvement of psychologists in the war effort formally began on April 6, 1917, at the annual meeting of Edward B. Titchener's Society of Experimental Psychologists at Harvard University. When war was officially declared by the U.S. Congress on that day, Robert M. Yerkes, the president of the APA, asked the assembled psychologists how they could assist the government in time of war. A committee was proposed under Yerkes's chairmanship, "to gather information concerning the possible relations of psychology to military problems" (Yerkes, 1921, p. 7). Almost 2 weeks later, on April 21, the executive council of the APA met in the Hotel Walton in Philadelphia. In preparation, Yerkes had been busy behind the scenes, touring Canadian hospitals, interviewing military doctors, and soliciting support from APA council members and members of the National Research Council. According to historian Richard T. von Mayrhauser (1987), Yerkes would use the military crisis to assert "near-dictatorial power within the profession" of psychology (p. 135).

The meeting at the Hotel Walton was misguided from the start because it involved a discussion among academic psychologists about what the military needed, rather than a request to the military as to how psychology might serve military needs. Moreover, a heavy-handed Yerkes sought to impose his narrow vision of mental testing on psychology, while suppressing input from another council member, Walter Dill Scott, who had more applied experience in personnel selection than anyone else at the meeting. With simultaneous authorization from the APA council and the National Research Council Psychology Committee, Yerkes appointed a dozen war-related psychology committees and chairs, dealing with areas such as aviation, recreation, propaganda, vision, acoustics, shellshock, emotional stability, and deception. Yerkes appointed himself chair of the "Committee on the Psychological Examining of Recruits," which was charged with preparation and standardization of testing methods and the demonstration of their effectiveness. Yerkes's initial testing plan—10-minute individual mental testing of at least 20% of "exceptional or unsatisfactory" recruits (von Mayrhauser, 1987, p. 141) by psychologists working under the supervision of military physicians—was in part a recapitulation of his own experiences working half-time directing research in the Psychopathic Department at Boston State Hospital under the supervision of Harvard psychiatrist Elmer Ernest Southard. At the same hospital, Yerkes had developed his own point scale adaptation of the Binet–Simon (Yerkes, Bridges, & Hardwick, 1915), which he probably hoped would be prominent in any testing program.

APA council member Walter V. Bingham later described his (and colleague Walter Dill Scott's) revulsion at events in Yerkes's meeting: "Meeting of the council in the smoke-filled room of a Philadelphia hotel. Midnight. Scott's utter disgust with the shortsighted self-interest revealed. His insurrection not previously told" (cited by von Mayrhauser, 1987, p. 139). Elsewhere in Bingham's papers appears the following disclosure:

> As the meeting proceeded it became clear to Scott and Bingham that Yerkes and the others were interested primarily in going into the army in order to acquire new psychological knowledge. They seemed to be more concerned with what the army could do for them than with what they could do for the army. Angrily, Scott and Bingham walked out in a huff. (cited by von Mayrhauser, 1987, p. 139)

With this divisive start, Yerkes alienated Scott, who had experience and skills he sorely needed. There was much at stake, as George Ellery Hale, famed astronomer and organizer of the National Research Council, warned Yerkes in May 1917:

> In the case of psychology, it is obvious that the first thing to do is to prove conclusively that the psychologists can perform service of unquestioned value to the government. . . . It is of fundamental importance that no tests be adopted which are not absolutely conclusive because if they were, the science of psychology would suffer an injury from which it would not recover for many years. (G. E. Hale to R. M. Yerkes, 1917; cited by Camfield, 1992, p. 107)

Yerkes's Committee and Arthur Otis

The Committee on the Psychological Examining of Recruits, made up of Robert M. Yerkes, Walter V. Bingham, Henry H. Goddard, Thomas H.

Haines, Lewis M. Terman, F. Lyman Wells, and Guy M. Whipple, met at the Vineland Training School in New Jersey from May 28 to June 9 to develop the Army mental tests. After reaching agreement that the tests had the goals of eliminating "unfit" recruits and identifying those with "exceptionally superior ability" (who might become officers), discussion turned to the merits of brief individually administered tests versus group-administered tests. Deciding that efforts should be made to test all recruits, the committee concluded that brief individual tests were problematic in terms of reliability and uniformity of method and interpretation, opting instead for group administration (Yerkes, 1921, p. 299). At this point, Lewis Terman presented the group-administered tests developed by his Stanford graduate student Arthur S. Otis. According to Yerkes (1921, p. 299), 4 of the 10 tests in the original Army scale for group testing were accepted with little change from the Otis scale, and certain other tests were shaped in part by the content and format of the Otis series.

Committee members identified a dozen criteria to use for selection of additional tests: suitability for group use; interest and appeal; economy of administration time; score range and variability; scoring objectivity; scoring ease and rapidity; minimal writing requirements; resistance to coaching; resistance to malingering; resistance to cheating; independence from educational influences; and convergent validity with independent measures of intelligence. Each test was to consist of 10–40 items, with a time limit not to exceed 3 minutes. Moreover, oral directions needed to be simple, and written instructions easy to read. All tests needed to be accompanied by two or three completed sample items to ensure that examinees understood task requirements.

Psychologists around the country were recruited to write additional items to create 10 parallel equivalent forms of the Army mental tests. The tests underwent a series of pilot studies with 400 examinees drawn from different settings. After revisions were made, a larger trial with the 10 forms was conducted on 4,000 recruits in Army and Navy settings during July and August 1917. The final test occurred in the fall of 1917, when 80,000 men in four national Army cantonments were tested, along with 7,000 college, high school, and elementary school students to check the Army results. All processing of record forms and statistical analyses were conducted by a small group working out of Columbia University, directed by Edward L. Thorndike with assistance from Arthur Otis and Louis L. Thurstone. Thorndike and his statistical analysis group endorsed the psychometric properties of the group tests, although clearly not all forms were equivalent, and some had to be dropped in the end.

Examination Beta was developed after Alpha, when it became evident that a different approach was needed for valid assessment of recruits who were either illiterate or limited in their English proficiency. It included ideas from Otis, Terman, and others and was tested at several training camps and at the Vineland Training School. After some 15 tests were reduced to 8 tests, the Beta was completed in April 1918. It was designed to correlate well with Examination Alpha, to differentiate average from very low levels of ability, and to be easily understood and administered, yielding few zero scores.

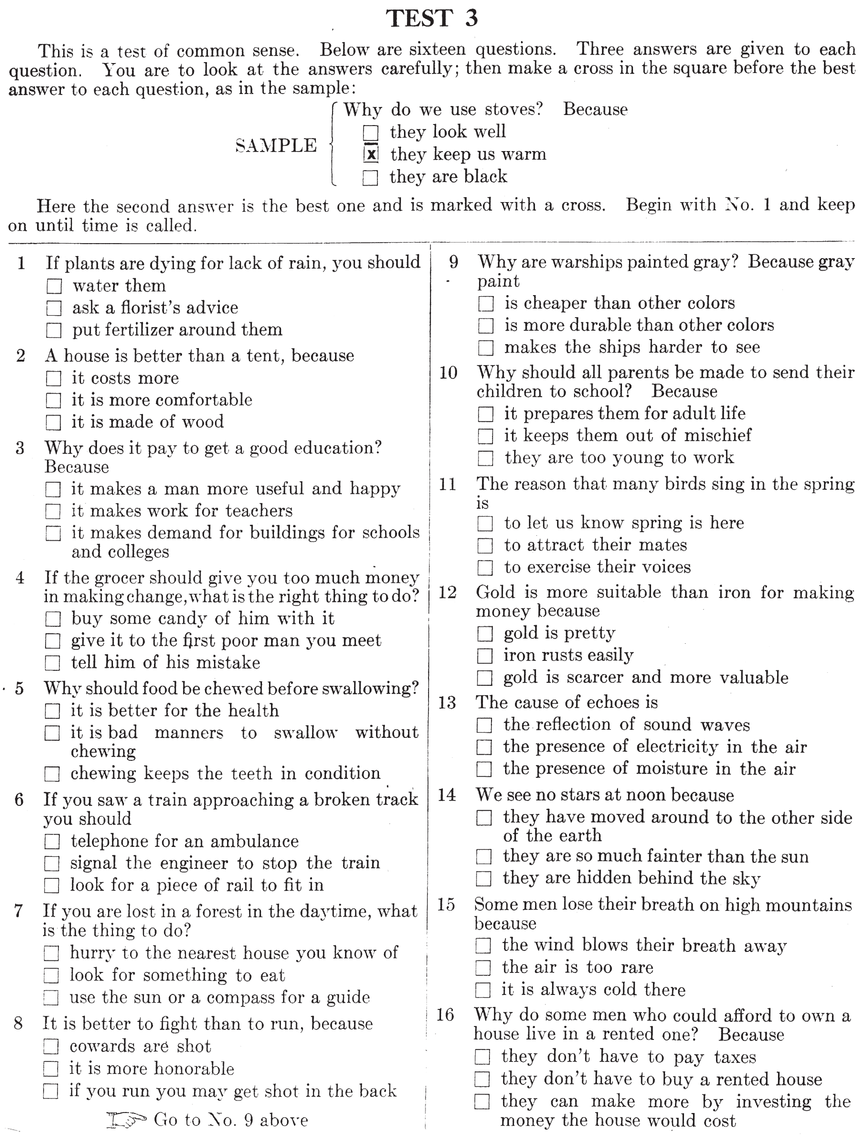
In December 1917, the Surgeon General recommended to the Secretary of War the continuance and extension of psychological examining to the entire Army. In January 1918, the Secretary of War authorized creation of a division of psychology in the Sanitary Corps out of the Surgeon General's office and expansion of the psychological examining program. A school for military psychology was organized with the Medical Officers Training Camp in Fort Oglethorpe, Georgia. While the school was active in 1918, approximately 100 officers of the Sanitary Corps and 300 enlisted men were given special training in military psychology. By the end of the war, psychological examining occurred at 35 army training camps and several army hospitals (Yerkes, 1919, 1920).

Examinations Alpha and Beta

The Army Alpha was intended for fluent and literate English-language speakers. Alpha was typically administered to men who could read newspapers and write letters home in English, with at least a fourth-grade education and five years of residency in the United States (Yerkes, 1921, p. 76). The Army Beta was a largely nonverbal scale intended for examinees with inadequate English-language proficiency or illiteracy (Yoakum & Yerkes, 1920). Beta was also given to low scorers on the Alpha. Men who had difficulty reading or writing in English were to be given both Alpha *and* Beta. E. G. Boring (1961) described the informal process of separating recruits into those suitable for Alpha or Beta: "You went down the line saying 'You read

American newspaper? No read American newspaper?'—separating them in that crude manner into those who could read English and take the Alpha examination and those who must rely for instructions on the pantomime of the Beta examination" (p. 30).

Examination Alpha consisted of eight tests, required approximately 40–50 minutes to administer, and could be given to groups as large as 500. A sample test from Alpha appears in Figure 1.11. Examinees were provided with the test form and a pencil. Responses were scored with stencils based on examinee responses (which usually involved writing numbers, underlining, crossing out, or checking a selected answer). After illiterate and non-English-speaking examinees were removed, and all recruits were seated with pencils and test forms, the examiner said:



TEST 3

This is a test of common sense. Below are sixteen questions. Three answers are given to each question. You are to look at the answers carefully; then make a cross in the square before the best answer to each question, as in the sample:

SAMPLE
Why do we use stoves? Because
☐ they look well
☒ they keep us warm
☐ they are black

Here the second answer is the best one and is marked with a cross. Begin with No. 1 and keep on until time is called.

1 If plants are dying for lack of rain, you should
☐ water them
☐ ask a florist's advice
☐ put fertilizer around them

2 A house is better than a tent, because
☐ it costs more
☐ it is more comfortable
☐ it is made of wood

3 Why does it pay to get a good education? Because
☐ it makes a man more useful and happy
☐ it makes work for teachers
☐ it makes demand for buildings for schools and colleges

4 If the grocer should give you too much money in making change, what is the right thing to do?
☐ buy some candy of him with it
☐ give it to the first poor man you meet
☐ tell him of his mistake

5 Why should food be chewed before swallowing?
☐ it is better for the health
☐ it is bad manners to swallow without chewing
☐ chewing keeps the teeth in condition

6 If you saw a train approaching a broken track you should
☐ telephone for an ambulance
☐ signal the engineer to stop the train
☐ look for a piece of rail to fit in

7 If you are lost in a forest in the daytime, what is the thing to do?
☐ hurry to the nearest house you know of
☐ look for something to eat
☐ use the sun or a compass for a guide

8 It is better to fight than to run, because
☐ cowards are shot
☐ it is more honorable
☐ if you run you may get shot in the back

☞ Go to No. 9 above

9 Why are warships painted gray? Because gray paint
☐ is cheaper than other colors
☐ is more durable than other colors
☐ makes the ships harder to see

10 Why should all parents be made to send their children to school? Because
☐ it prepares them for adult life
☐ it keeps them out of mischief
☐ they are too young to work

11 The reason that many birds sing in the spring is
☐ to let us know spring is here
☐ to attract their mates
☐ to exercise their voices

12 Gold is more suitable than iron for making money because
☐ gold is pretty
☐ iron rusts easily
☐ gold is scarcer and more valuable

13 The cause of echoes is
☐ the reflection of sound waves
☐ the presence of electricity in the air
☐ the presence of moisture in the air

14 We see no stars at noon because
☐ they have moved around to the other side of the earth
☐ they are so much fainter than the sun
☐ they are hidden behind the sky

15 Some men lose their breath on high mountains because
☐ the wind blows their breath away
☐ the air is too rare
☐ it is always cold there

16 Why do some men who could afford to own a house live in a rented one? Because
☐ they don't have to pay taxes
☐ they don't have to buy a rented house
☐ they can make more by investing the money the house would cost

FIGURE 1.11. The Army Examination Alpha Practical Judgment Test. Soldiers were allowed 1½ minutes for this test. From Yerkes (1921). In the public domain.

Attention! The purpose of this examination is to see how well you can remember, think, and carry out what you are told to do. We are not looking for crazy people. The aim is to help find out what you are best fitted to do in the Army. The grade you make in this examination will be put on your qualification card and will also go to your company commander. Some of the things you are told to do will be very easy. Some you may find hard. You are not expected to make a perfect grade, but do the very best you can. (Yoakum & Yerkes, 1920, p. 53; emphasis added)

Beta was typically administered with task performance modeled through pantomimed demonstrations and some brief verbal directions (e.g., "Fix it!" while pointing to the incomplete pictures

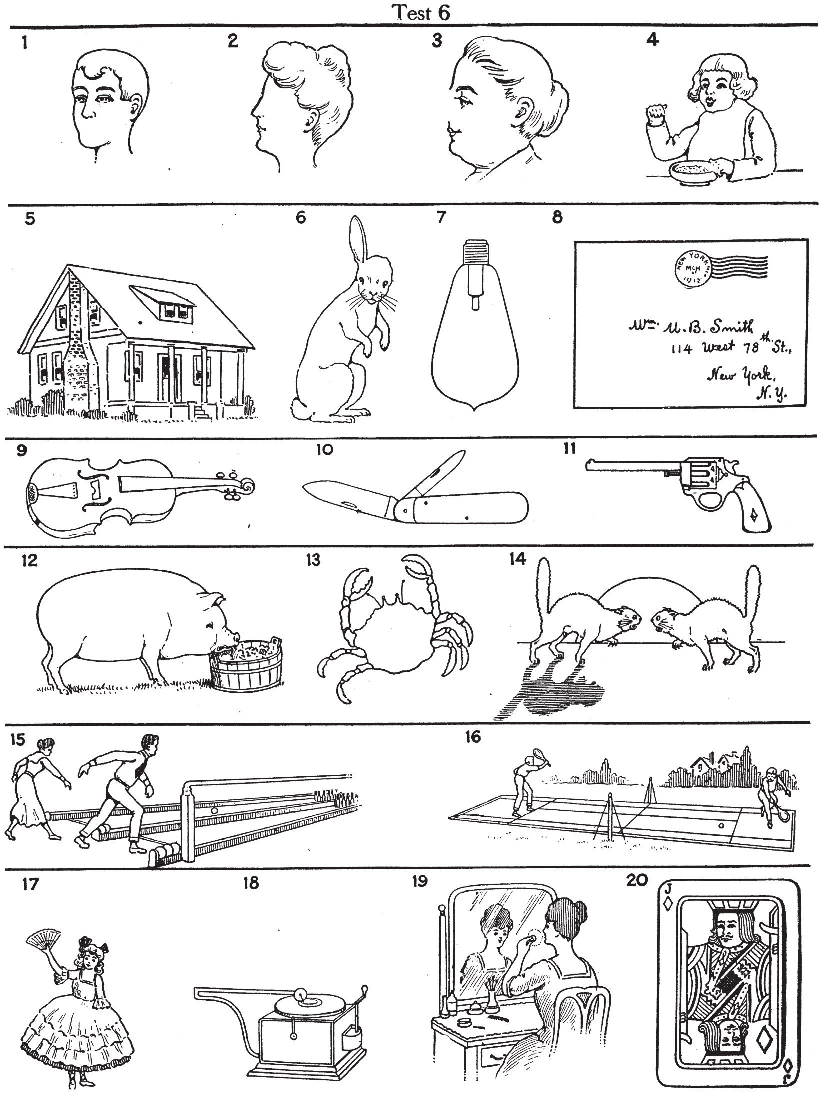


FIGURE 1.12. The Army Examination Beta Picture Completion Test. Instructions: "This is Test 6 here. Look. A lot of pictures . . . Now watch." Examiner points to separate sample at front of room and says to Demonstrator, "Fix it." After pausing, the Demonstrator draws in the missing part. Examiner says, "That's right." The demonstration is repeated with another sample item. Then Examiner points to remaining drawings and says, "Fix them all." Demonstrator completes the remaining problems. When the samples are finished, Examiner says to all examinees, "All right. Go ahead. Hurry up!" At the end of 3 minutes, Examiner says, "Stop!" From Yerkes (1921). In the public domain.

on Pictorial Completion; Yoakum & Yerkes, 1920, p. 87). A sample test from Beta appears in Figure 1.12. Administered to groups as large as 60, it was typically completed in about 50–60 minutes and required a blackboard with chalk, eraser, curtain, and chart (on a roller to show 27 feet of pictorial instructions in panels). The examiner gave brief instructions, while a demonstrator pantomimed how to complete tasks correctly on the blackboard panels corresponding to the test response form. There were seven final tests in Beta.

Reports of intelligence ratings derived from test scores were typically made within 24 hours and entered on service records and qualification cards that were delivered to commanding officers and personnel officers. Individual examinations with the Yerkes–Bridges Point Scale, the Stanford–Binet Intelligence Scale, or the Army Performance Scale were usually reserved as checks on questionable or problematic group examination results. The test scores yielded grade ratings from A to E, with the following descriptions drawn from Yoakum and Yerkes (1920):

- *A (Very Superior).* An A grade was earned by only 4–5% of drafted men. These men were considered to have high officer characteristics when they were also endowed with leadership and other necessary qualities. They were shown to have the ability to make a superior record in college or university.
- *B (Superior).* A B grade was obtained by 8–10% of examinees. This group typically contained many commissioned officers, as well as a large number of noncommissioned officers. A man with B-level intelligence was capable of making an average record in college.
- C+ *(High Average).* The C+ group included 15–18% of all soldiers and contained a large number of recruits with noncommissioned officer potential and occasionally commissioned officer potential, when leadership and power to command were rated as being high.
- C *(Average).* The C group included about 25% of soldiers who made excellent privates, with a certain amount of noncommissioned officer potential.
- C– *(Low Average).* The C– group included about 20% of soldiers; these men usually made good privates and were satisfactory in routine work, although they were below average in intelligence.
- *D (Inferior).* Men in the D group were likely to be fair soldiers, but they were usually slow in learning and rarely went above the rank of private. They were considered short on initiative and required more than the usual amount of supervision.
- *D– and E (Very Inferior).* The last group was divided into two classes: D–, consisting of men who were very inferior in intelligence but who were considered fit for regular service, and E, consisting of men whose mental inferiority justified a recommendation for development battalion, special service organization, rejection, or discharge. The majority of men receiving these two grades had a mental age below 10 years. Those in the D– group were thought only rarely able to go beyond the third or fourth grade in primary school, however long they attended.

To his chagrin, Yerkes's division of psychology was appointed to the Sanitary Corps instead of the Medical Corps (where he had hoped psychologists would be classified), but he was still a member of the Surgeon General's staff. Yerkes encountered near-continual resistance to the testing program from the military establishment, and the Army mental examiners often had inadequate testing facilities or faced a deeply entrenched military establishment that did not see the value in intelligence tests. In response to queries about testing from Army officers, Yerkes gave the psychological examiners standard responses to provide as needed—specifying the potential value of the Alpha and Beta in military decision making, but also emphasizing that test scores alone should not constitute the sole basis for making military service decisions:

> The rating a man earns furnishes a fairly reliable index of his ability to learn, to think quickly and accurately, to analyze a situation, to maintain a state of mental alertness, and to comprehend and follow instructions. The score is little influenced by schooling. Some of the highest records have been made by men who had not completed the eighth grade. . . . The mental tests are not intended to replace other methods of judging a man's value to the service. It would be a mistake to assume that they tell us infallibly what kind of soldier a man will make. They merely help to do this by measuring one important element in a soldier's equipment, namely, intelligence. They do not measure loyalty, bravery, power to command, or the emotional traits that make a man "carry on." (Yoakum & Yerkes, 1920, pp. 22–24)

According to Yerkes (1918a, 1918b), the Army testing program was tasked with four military objectives: (1) aiding in the identification and elimination of "mentally defective" men who were unfit for service; (2) identifying men of exceptional intelligence for special responsibilities or possible officer training; (3) balancing military units in terms of intelligence; and (4) assisting personnel officers in the camps with the classification of men. The tests appear to have functioned well in identifying recruits of very high intelligence and very low intelligence, although research findings showed a disproportionate number of minority, foreign-born, and illiterate recruits as having very low intelligence, in spite of efforts to correct for the language and literacy demands of the Alpha with the Beta (Yerkes, 1921). There is little evidence that the Alpha and Beta were effectively used to balance the intellectual composition of military units. Although Army battalions ideally should be comparable and interchangeable in terms of effectiveness, individual battalion commanders no doubt wanted the best available recruits and held onto the recruit who received A and B grades. Yoakum and Yerkes (1920) described the challenge:

> In making assignments from the Depot Brigade to permanent organizations it is important to give each unit its proportion of superior, average, and inferior men. If this is left to chance there will inevitably be "weak links" in the army chain. Exception to this rule should be made in favor of certain arms of the service which require more than the ordinary number of mentally superior men; for example, Signal Corps, Machine Gun, Field Artillery and Engineers. These organizations ordinarily have about twice the usual proportion of "A" and "B" men and very much less than the usual proportion of "D" and "D–" men. (p. 25)

With respect to the final objective, of assisting with personnel decisions, the Army mental tests provided a single piece of information—intelligence level—that was of considerable value. It would be Walter Dill Scott's Army Committee on Classification of Personnel that provided a context for the Army mental test scores in making military personnel decisions.

Scott's System of Personnel Selection

Walter Dill Scott (1869–1955) was a pioneering industrial psychologist, applying principles of experimental methodology to practical business problems (Ferguson, 1962, 1963a; Lynch, 1968; Strong, 1955). An interest in identifying and selecting successful salesmen led Scott (1916) to develop a multimethod quantitative personnel selection approach consisting of historical information from former employers (i.e., a model letter soliciting information and ratings, which was included in a *personal history record*); performance on tests of intellectual ability devised by Scott; performance on tests of technical ability (written calculation and clerical transcription) scored for accuracy, speed, and legibility; and multiple ratings based on a series of "interviews" with trained raters (in which the examinee was to introduce himself and try to sell merchandise to a series of interviewers posing as merchants). In 1916, Walter V. Bingham, the head of the division of applied psychology at Carnegie Institute of Technology, offered Scott the opportunity to direct the newly formed Bureau of Salesmanship Research and to become the first professor of applied psychology in the United States (Ferguson, 1964a). In a remarkable partnership between the Carnegie Bureau and 30 large national businesses, Scott had the opportunity to test his personnel selection methods with the hiring of 30,000 new salesmen each year, and comparison of personnel decisions against actual sales performances. It was a highly successful arrangement, possibly unprecedented in the history of psychology, and Scott's work was well regarded by the business community.

The history of Scott's personnel selection system in the military may be found in several resources, including official accounts from the Army (Committee on Classification of Personnel in the Army, 1919a, 1919b) and contemporary accounts from von Mayrhauser (1987, 1989); the most in-depth accounts are available from Ferguson (1963b, 1963c, 1964b, 1964c). When war was declared in 1917, Scott realized that his existing methods could readily be applied to personnel selection in the military. At the Hotel Walton meeting on April 21, Scott objected to Yerkes's positions on the war as an opportunity to advance the prominence of psychology. Scott and Bingham were the only psychologists at the meeting with experience in personnel selection, and they knew that Scott's system already had demonstrated effectiveness. In Scott's system, the mental test results had value, but Scott and Bingham were certain that interview ratings would be more important in the selection of officers. Moreover, Scott did not want to subordinate psychologists to psychiatrists, but instead thought they should report to a high official such as the Secretary of War. Offended by Yerkes's self-serving agenda, Scott and Bingham walked out.

Scott decided to launch his own initiative, independent of Yerkes. Scott revised his existing salesman rating scale, completing A Rating Scale for Selecting Captains by May 4, 1917. He shared with it several psychologists and asked Edward L. Thorndike to write a letter of support to Frederick P. Keppel, who had been a dean at Columbia and was now Third Assistant Secretary of War. Keppel invited Scott to Washington, D.C., where Scott presented his scale, did some testing with it, made some improvements, and overcame institutional resistance (including having his scale ripped "to tatters" by officers in Plattsburg who had been invited to suggest improvements [Committee on Classification of Personnel in the Army, 1919a, p. 50]). When he finally met directly with Secretary of War Newton D. Baker, Scott suggested that a group of psychologists and experienced employment managers be appointed to advise the Army on personnel selection, volunteering to assemble such a group. On August 5, 1917, Scott received approval of a plan to include scientific staff, a group of civilian experts for research and planning, and a board of military representatives to bring problems to the Committee on Classification of Personnel in the Army and help implement its recommendations. Within 6 weeks, the committee created and implemented a classification and assignment system for the Army where none had existed before. It was the largest program of personnel selection ever attempted up to that time. Scott was the committee's director, Bingham was its executive secretary, and they answered directly to the Adjutant General of the Army. They began with a single office that grew to 11 rooms in the War Building (then the central hub of military decision making, housing the offices of the Secretary of War, Chief of Staff, and Adjutant General).

Scott's personnel system for the army included a Soldier's (or Officer's) Qualification Card, grades on the Army mental tests and proficiency on specialized trade tests, and the Officers' Rating Scale in various forms for noncommissioned and commissioned officers. The Qualification Card relied on interviews to obtain occupational history, education, leadership experience, and military history. Test scores on the Army mental tests ranged from A through E and were provided by Yerkes's examiners. For recruits claiming experience in specific trades of value to the military, Scott's committee oversaw development of special trade tests that measured specific proficiencies, generating a range of scores from "Expert" through "Novice." Finally, the Officers' Rating Scale became the main tool used for the selection and promotion of officers, with all officers receiving quarterly ratings by the end of the war. This scale involved ratings in five areas: physical qualities, intelligence, leadership, personal qualities, and general value to the service.

If the Army mental tests were intended to measure general intelligence, the trade tests measured specific ability and knowledge related to the performance of several hundred specific occupations needed by the military. Vocational training was impractical, and men were frequently found to have misrepresented their civilian jobs and skills on the Soldier's Qualification Card. In order to identify personnel requirements for specific jobs, occupational titles were compiled and detailed personnel specifications were prepared by Scott's team. With the criteria of covering all trades rapidly and objectively by examiners who did not have to be knowledgeable about each individual trade, a series of oral, picture, and/or performance trade tests were administered and scored so that the number of questions correctly answered predicted status as a novice, apprentice, journeyman, or expert. There were 84 trade tests for jobs as varied as butchers, electricians, pipefitters, and most other specialties needed by the military. For example, the trade test officer issued driver's licenses for all drivers of touring cars, motorcycles, and trucks (Committee on Classification of Personnel in the Army, 1919a, p. 135). Examples of trade tests appear in the committee's personnel manual (1919b), and after the war compilations of trade tests were published in Chapman (1921) and Toops (1921).

From the military's perspective, it is clear that Scott's personnel selection procedures were much more valued than Yerkes's testing program. At the end of the war, Yerkes's Division of Military Psychology was summarily and completely shut down. Scott's Committee on Classification of Personnel in the Army was transferred to the General Staff and merged with the newly created Central Personnel Branch, in effect institutionalizing Scott's personnel selection procedures within the Army (Yerkes, 1919). The War Department awarded Scott the Distinguished Service Medal when he left the service in 1919, and asked *him* to convey its appreciation to Major Yerkes. Scott became the highest-ranking psychologist in the Army, having been commissioned as a colonel in the Adjutant General's Department in November 1918.

Undoubtedly, multiple factors explained the military's different responses to Scott and to Yerkes. Scott adapted his system to military needs, while Yerkes sought to impose academic know-how on an unreceptive army. Scott partnered with military personnel, while Yerkes's examiners were seen as unwelcome, externally imposed "pests" and "mental meddlers" by camp commanders (cited by Kevles, 1968, p. 574). No less than three independent investigations were launched by Army personnel suspicious of Yerkes and his men (Zeidner & Drucker, 1988, p. 11). Scott worked initially in an advisory capacity, while Yerkes continually sought authority. Scott had considerable personal skills in persuasion and salesmanship (Strong, 1955), whereas Yerkes was a strong planner but a poor manager (Dewsbury, 1996). Scott's system had substantial and understandable face validity for military performance, while Yerke's Examinations Alpha and Beta did not have obvious relevance for soldiering. From the perspective of the history of intelligence testing, however, a broader argument should be considered: Yerkes's committee created the tests and his examiners generated scores, but Scott's committee provided a systematic context (including recruits' history and specific skills) within which the test scores made sense and could be used to make practical decisions by teams of military personnel not schooled in psychology.

World War II Assessment Procedures

In World War II, the plan developed by Scott was streamlined and implemented again, this time with Walter V. Bingham in charge of the personnel system and mental tests (Bingham, 1942, 1944). Bingham served as chair of the Committee on Classification of Military Personnel, the committee having been appointed in 1940 by the National Research Council at the request of the Adjutant General, months before passage of the Selective Service and Training Act (Bingham, 1944). In contrast to the unwelcoming reception Yerkes had received in World War I, Bingham and the infrastructure he established were valued by the Army (Zeidner & Drucker, 1988). The Army Alpha was replaced by the Army General Classification Test, a shorter version of the Alpha; initial versions were administered in spiral omnibus form in about 40 minutes, and there were four parallel forms (Bittner, 1947; Staff, Personnel Research Section, 1945). Conceptualized as a test of "general learning ability," it consisted of vocabulary items (intended to tap verbal comprehension), arithmetic problems (thought to tap quantitative reasoning), and block-counting problems (intended to measure spatial thinking), all endeavoring to deemphasize speed somewhat. Grades of A to E were replaced with five levels of learning readiness, I to V. Terms like *mental age* and *IQ* were largely eliminated from group tests.

As in World War I, specialized Non-Language Tests were developed and standardized to test illiterate and non-English-speaking recruits (Sisson, 1948). In 1944, the Wechsler Mental Ability Scale, also known as the Army Wechsler, was replaced by the Army Individual Test, which included three verbal subtests (Story Memory, Similarities–Differences, and Digit Span) and three nonverbal subtests (Shoulder Patches, Trail Making, and Cube Assembly) (Staff, Personnel Research Section, 1944). Rapaport (1945) praised the Army Individual Test, noting that it was "admirably well-constructed" (p. 107) as a measure of general mental ability, but he also raised cautions about its diagnostic limitations. Numerous specialized trade tests and aptitude tests (e.g., mechanical aptitude, clerical aptitude) were developed as well. Most importantly, the Personnel Research Section of the Adjutant General's Office that Bingham established quickly earned the military's trust, leading to the creation of the Army Research Institute (which still exists). One of Bingham's charges was to put the "brakes on projects . . . of great scientific interest" if they did not help the Army "toward early victory" (Zeidner & Drucker, 1988, p. 24). It was a lesson in military priorities that Yerkes, whose agenda included advancing psychology as a science, may not have learned in World War I.

David Wechsler: The Practical Clinician

The practice of contemporary applied intelligence assessment in the second half of the 20th century may arguably be said to have been most strongly and directly influenced by the measurement instruments developed by David Wechsler (1896–1981). Beginning in the 1960s, the Wechsler intelligence scales supplanted the Stanford–Binet as the leading intelligence tests (Lubin, Wallis, & Paine, 1971). Surveys of psychological test usage decades after his death show that Wechsler's intelligence tests continue to dominate intelligence assessment among school psychologists, clinical psychologists, and neuropsychologists (Camara, Nathan, & Puente, 2000; Wilson & Reschly, 1996). Early in the 21st century, the Wechsler scales for adults, children, and preschoolers are still taught at much higher frequencies than any other intelligence tests in North American clinical and school psychology training programs (Cody & Prieto, 2000; Ready & Veague, 2014).

In many ways, David Wechsler was an unexpected success—coming to the United States as a child amid a flood of Eastern European immigrants, losing both parents by the age of 10, compiling a relatively ordinary academic record in high school and college (while graduating early), registering as a conscientious objector to the 1917 World War I draft (a risky decision at the time, when "slackers" were universally condemned), and not having become a naturalized citizen by the time of the war. Even so, these risk factors may have been somewhat ameliorated by the guidance of an accomplished older brother (pioneering neurologist Israel S. Wechsler), who became his caretaker and role model; by the opportunity to provide military service as an Army mental test examiner, thereby quickly learning about assessment and making key professional contacts; and by receiving his graduate education and professional psychology training at an opportune time and place in the development of what eventually would become "clinical" psychology.

Wechsler's Early Life and Education

David Wechsler was the youngest of three boys and four girls born in Romania to Moses Wechsler, a merchant, and Leah (Pascal) Wechsler, a shopkeeper (see, e.g., Matarazzo, 1972). At the time the Wechsler family emigrated in 1902, poor harvests in 1899 and 1900 had produced famine and an economic downturn in Romania, worsening the scapegoating of Jews and resulting in severe applications of existing anti-Jewish decrees (e.g., Kissman, 1948). The family's new life on the Lower East Side of New York City was marked by tragedy. Within 5 years of their arrival, both Moses and Leah Wechsler passed away from malignancies ("Deaths reported Aug. 23," 1903; Wechsler, 1903; Wexler, 1906). The effects of these losses upon the family, particularly David as the youngest, are likely to have been profound. By 1910, David's older brother Israel S. Wechsler, then a physician in general practice, appears to have taken over as the head of the family.

Wechsler was educated in the public schools on the Lower East Side (see Wechsler, 1925, p. 181). After high school graduation, Wechsler attended the College of the City of New York (now known as City College) from 1913 to 1916, graduating with an AB degree at the age of 20 ("206 get degrees at City College," 1916). Following his graduation, Wechsler enrolled in graduate studies in psychology at Columbia University, where he would complete his master's degree in 1917 and his doctorate in 1925. His decision to continue his education beyond college had family precedent; Israel had graduated from New York University and Bellevue Medical College in 1907 at the age of 21. Israel would take a position in neurology at Mount Sinai Hospital in 1916 and begin teaching neurology at the outpatient clinic of the Columbia University College of Physicians and Surgeons in 1917 (Stein, 2004; see also "Israel Wechsler, Neurologist, Dies," 1962). Israel initially intended to become a psychiatrist and was self-taught in psychoanalysis, but personally identified as a neurologist because, as he later explained, "If the brain is the organ of thought and disturbance of its functions expresses itself in disorders which are called neuroses and psychoses, it seemed reasonable and necessary to know neurology" (Wechsler, 1957, pp. 1113–1114). Israel Wechsler was involved in the early-20th-century struggles between psychiatry and neurology, when each medical discipline was vying for control of the care of those with mental illness. David instead pursued psychology, but he would follow his brother's lead in becoming a practicing clinician, a hospital-based academic, and the author of professional textbooks.

Columbia was one of the few major universities that provided graduate experimental psychology training with a willingness to address applied problems, termed *experimental abnormal psychology* by Woodworth (1942, p. 11)—an educational orientation that would eventually evolve into clinical psychology (Routh, 2000). The Columbia graduate psychology department in this era was made up primarily of commuter students "who emerged from the subway for their classes and research and departed immediately into the subway afterwards" (Thorne, 1976, p. 164). Columbia University was the academic home of faculty J. McKeen Cattell (until his dismissal in October 1917), Robert S. Woodworth, and Edward L. Thorndike, three of the most influential psychologists of the early 20th century. "Cattell, Woodworth, and Thorndike were the trio at Columbia," said F. L. Wells, who had worked as an assistant to Cattell and Woodworth, adding, "Cattell might inspire awe, Thorndike admiration, and Woodworth affection. Affection toward Cattell and Thorndike was not possible" (quoted in Burnham, 2003, p. 34).

For his master's thesis, completed (and published as a journal article) in 1917, Wechsler (1917a) patched together a clinical memory battery from existing published and unpublished tests, closely following the memory framework suggested by Whipple (1915). Wechsler spent 2½ months conducting in-depth assessment of six patients with Korsakoff psychosis at the Manhattan State Hospital on Ward's Island. He saw each patient as many as 20 times, and he also had the opportunity to observe psychiatric assessment on the wards. Wechsler's master's thesis represented his first known attempt to build a test battery. He established a pattern he was later to follow with intelligence tests, memory tests, and a failed personality test: that of appropriating practical and clinically useful procedures from other authors, making slight improvements and modifications, and synthesizing them into a battery of his own.

World War I Service

After the U.S. Congress declared war on Germany in April 1917, Wechsler (1917b) completed his required registration for the draft, listing himself as a "Consciencious [*sic*] Objector" and as an alien who was a citizen of Romania, who was disabled by "Near Sightness [*sic*]" and "Physical Unfitness." Wechsler's draft registration thus used multiple methods to avoid being drafted—claiming status as a conscientious objector, claiming exemption from military service by reason of alien (noncitizen) status, and claiming physical deficiencies that would disqualify him for military service. We do not know what motivated David Wechsler to try to avoid military service at age 21, but the public press treated conscientious objectors with contempt, and some were arrested and imprisoned. Even Army mental examiners considered conscientious objector status to be a form of psychopathology (May 1920). Wechsler's status as a noncitizen native of Romania, some 15 years after his arrival in the United States, also put him at risk. As an alien, he could not be drafted, but he could be deported. The U.S. Congress tried to close this draft loophole in response to the perceived "alien slacker" problem, but treaty obligations circumvented final passage ("Alien Slackers May Not Escape Service," 1917; "Pass Alien Slacker Bill," 1918). Within military training camps, however, some officers considered all aliens who had not become naturalized citizens as suspect.

Becoming an Army mental test examiner represented a way by which Wechsler could avoid seeing combat, and it was probably through back-channel communications from his professor Robert S. Woodworth to Robert M. Yerkes that Wechsler was identified as a prospective tester. In May 1918, Yerkes requested in writing that Wechsler and 13 others who had "qualifications for psychological service" be sent authorization for military induction and be assigned for course instruction in military psychology at Camp Greenleaf, Chickamauga Park, Georgia. Shown in Figure 1.13 at the time of his military service, Wechsler reported to the School for Military Psychology, where he was taught the Army Alpha, Army Beta, Stanford–Binet, Yerkes Point Scale, and other tests. Trainees also received instruction in military practices, including military law, field service, honors and courtesies, equipment, and gas attack defense instructions and drills. E. G. Boring, who reported to Camp Greenleaf as a captain in February 1918, described what may have also been Wechsler's experience:

> We lived in barracks, piled out for reveillé, stood inspection, drilled and were drilled, studied testing procedures, and were ordered to many irrelevant lectures. As soon as I discovered that everyone else resembled me in never accomplishing the impossible, my neuroses left me, and I had a grand time, with new health created by new exercise and many good friendships formed with colleagues under these intimate conditions of living. (1961, p. 30)

FIGURE 1.13. David Wechsler at the age of 23, from his 1918 passport application. Wechsler used a program designed to educate World War I veterans in Europe to pursue educational opportunities in France and London, including time with Charles E. Spearman and Karl Pearson. From Wechsler (1918b). National Archives and Records Administration, Washington, D.C. In the public domain.

In May 1918, Congress enacted legislation that allowed aliens serving in the U.S. armed forces to file a petition for naturalization without having made a declaration of intent or proving 5 years' residence (e.g., Scott, 1918). Under this new law, Wechsler became a naturalized citizen in June 1918, with Captain John E. Anderson and Lt. Carl A. Murchison, two psychologists who would have noteworthy careers, serving as his witnesses (Wechsler, 1918b). Wechsler completed his training at Camp Greenleaf in July, was promoted to the rank of corporal, and was assigned to Camp Logan in Houston, Texas, in early August 1918. There he would give individual psychological assessments to recruits who had failed the Alpha and/or the Beta, largely because of limited English proficiency or illiteracy. Conditions at Camp Logan were poor, with inadequate space and support, but the Army examiners administered over 300 individual assessments (Yerkes, 1921, p. 80).

It was during his time as an Army examiner that many of Wechsler's core ideas about assessment were born, especially his idea to construct an intelligence scale combining verbal and nonverbal tests, paralleling the Army Alpha and Army Beta/performance exams (Wechsler, 1981). Most of the assessment procedures appropriated by Wechsler for his intelligence scales appear in Yerkes (1921). Matarazzo (1981) relates that Wechsler realized the value of individual assessment when group tests yielded misleading results, as many of his examinees functioned adequately in civilian life in spite of their low group test scores. Wechsler also reportedly learned the value of nonverbal assessment and the limitations of the Stanford–Binet with adults. He even (Wechsler, 1932) described an approach to profile analysis of Army Alpha subtests—a clear antecedent to the intraindividual (ipsative) profile analyses still used in interpreting the Wechsler intelligence scales.

With the signing of the armistice, Wechsler participated in AEF University, a program created by order of General John J. Pershing and other military leaders to serve the 2 million idle (and bored) American servicemen who remained stationed in Europe, waiting to be shipped home (Cornebise, 1997; "Education for American Soldiers in France," 1919). Although Wechsler had never served overseas, he arranged to spend time in France (December 1918–March 1919) and then in London (March 1919–July 1919) as part of this program. Some 2,000 soldiers attended the Sorbonne, while about 2,000 soldier-students attended British universities, with 725 going to University College London ("U.S. Maintains Great Schools on Foreign Soil," 1919). At University College London, Wechsler had the opportunity to work for 3 months with Charles E. Spearman and to meet Karl Pearson, becoming familiar with Spearman's work on the general intelligence factor and Pearson's correlation statistic, as well as to note their professional rivalry (Wechsler, Doppelt, & Lennon, 1975). Wechsler was honorably discharged from the military in July 1919. Given his efforts to avoid military service in 1917, it might be considered ironic that the skills he acquired and contacts he made during his military service would shape his career in assessment and test development.

From 1919 to 1921, Wechsler studied and conducted research at the University of Montpelier and principally at the Sorbonne, under the supervision of Henri Pieron and Louis Lapicque (Rock, 1956, p. 675; Wechsler, 1925, p. 8). Wechsler used the research to complete his doctorate at Columbia, under the guidance of Robert Woodworth (Wechsler, 1925, p. 8). The opportunity to study at the Sorbonne came through Wechsler's application for an American Field Service fellowship from the Society for American Fellowships in French Universities (Wechsler, 1918b). In its first year (1919–1920), there were eight fellows, one of whom was Wechsler.

Bellevue Psychiatric Hospital and Other Clinical Experiences

After completing his fellowship at the Sorbonne, Wechsler traveled through France, Switzerland, and Italy before reluctantly returning to the United States (Wechsler, 1921). Once he was settled in New York, he began practicing psychology, mostly conducting assessments, in a variety of clinical and industrial settings. His ambivalence about returning, as disclosed to Edwards (1974), was reflected in his 1922 paper on the psychopathology of indecision.

Wechsler spent the summer of 1922 working with F. L. Wells at the Psychopathic Hospital in Boston, followed by 2 years as a psychologist with the New York Bureau of Children's Guidance. The Bureau of Children's Guidance was a psychiatric clinic, operating under the aegis of the New York School of Social Work and reflecting the values of the popular child guidance movement. Directed by Bernard Glueck, the bureau served troubled children referred by school principals or selected teachers for problems in the areas of scholarship, attendance, behavior, or general welfare. It was staffed by social workers, psychiatrists, and psychologists. The bureau emphasized problems with delinquency, with the objective of "a keener understanding of the child as an individual, and assistance to the school in working out needed readjustments, whether they be physical, social or educational" ("Crime Clinics Growing," 1922).

From 1925 to 1927, Wechsler worked with J. McKeen Cattell as acting secretary and research associate of The Psychological Corporation (Wasserman & Maccubbin, 2002). Created by Cattell, The Psychological Corporation did not directly employ any psychologists at the time; instead, consulting psychologists worked in nonsalaried, commission-based arrangements, undertaking projects for businesses and dividing the payment between themselves and the corporation. A 29-year-old David Wechsler, having completed his dissertation, had difficulty finding a job and contacted his old professor, Cattell, who hired him; according to Wechsler, Cattell told him, "You can get the *pro tem* acting secretary here. You have to get your own business and whatever business you get, the company will get half of your remunerations" (Wechsler et al., 1975). Wechsler undertook two known projects at The Psychological Corporation: the development of an automobile driving simulator and psychometric tests for taxicab drivers (Wechsler, 1926) and a tabloid newspaper study with a *New York World* reporter to test the intelligence of Ziegfeld chorus girls with the Army Alpha.

In 1932, following the tragic death of two Bellevue Hospital staff psychologists in a boating accident ("*Sea Fox* Wreckage," 1931), Wechsler was hired as a psychologist by the Psychiatric Division of Bellevue Hospital, New York. Bellevue was the oldest public hospital in the United States, but its psychopathic wing was scheduled for replacement by the Bellevue Psychiatric Hospital, described at its groundbreaking as the "chief battle-ground in the war against diseases of the mind" ("Old Bellevue and New," 1930). When the new unit finally opened in 1933, its capacity was planned at 600 patients to "give wide scope and facility for modern methods of investigating and treating mental disorders" ("A Bellevue Unit Formally Opened," 1933). By 1941, Wechsler had become chief psychologist and a clinical faculty member at the New York University College of Medicine, supervising more than 15 clinical psychologists, five interns, and two research psychologists on grants (Weider, 2006). Wechsler would retire from Bellevue in 1967, after having pioneered the role of the psychologist in a psychiatric hospital (Wechsler, 1944), and his clinical experiences would help him remain oriented to the use of psychological testing as it relates to practical patient care.

Concept of Intelligence

In his earliest scholarly statement on intelligence in his brother's neurology book, Wechsler (1927) ventured a definition: "All definitions of intelligence refer essentially to ability to learn and adapt oneself to new conditions; that is, not knowledge and practical success, but ability to acquire knowledge and ability to cope with experience in a successful way" (p. 105). It is Wechsler's (1939) definition, which built on his previous efforts and borrowed elements from his predecessors, that remains best known among definitions of intelligence:

> Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence. But intelligence is not identical with the mere sum of these abilities, however inclusive. (p. 3)

The long-standing popularity of this definition is probably due to the enduring popularity of the Wechsler intelligence scales with which it is associated. The definition reflects Wechsler's generally cautious writing style; it was exceptionally rare that he made any bold statement in writing that might alienate any colleagues. The phrase "aggregate or global capacity" appears to encompass Spearman's general factor, *g*—but Wechsler included an accommodation for the group factors, which, "though not entirely independent, are qualitatively differentiable." According to Wechsler (Wechsler et al., 1975), this definition also subsumes Binet's emphasis on adaptation. The phrase "to deal effectively with his environment" recapitulates Binet's (1911/1916) observation that "Intelligence marks itself by the best possible adaptation of the individual to his environment" (p. 301), as well as the use of adaptation in the definition of intelligence by others. In one of his final publications, Binet (1910) also took the position that intelligence is a dynamic synthesis, more than the different "pieces of the machine" that comprise it; this may have influenced Wechsler's statement that intelligence is more than the sum of its constituent abilities.

Creation and Development of the Wechsler Intelligence Scales

Of course, it is for his intelligence tests that David Wechsler is best remembered. Wechsler's gifts in the area of test development lay in his ability to synthesize the work of others—that is, to recognize clinically useful measurement procedures and to streamline and package them so as to be maximally useful for the practicing psychologist (Wasserman & Kaufman, 2015). His test work was unoriginal, and his intelligence tests consist entirely of tests (sometimes incrementally improved) that were originally devised by other psychologists. Several researchers have sought to trace the origins of the specific Wechsler intelligence subtests (e.g., Boake, 2002; Frank, 1983), a historically important endeavor, but it is notable that from the start Wechsler (1939) openly disclosed the sources he drew upon. As Boake (2002) suggested, it is most unfortunate that the names of the original innovators who created the Wechsler subtest procedures have been forgotten, omitted from mention in contemporary test manuals.

The Bellevue Intelligence Scale was originally subsidized by a Works Progress Administration grant during the Great Depression (Wechsler, 1981; Wechsler et al., 1975). Wechsler (1939, p. 137) reported that the test took 7 years to develop, and it first underwent trials in 1937 and 1938 at the Bellevue Psychiatric Hospital, the Court of General Sessions of New York City, and the Queens General Hospital. The need for a new adult test stemmed largely from the inadequacy of the Stanford–Binet, particularly its poor normative sample for adults, and the poor fit of the Army mental tests for clinical decision making. As the chief psychologist in a large public hospital, Wechsler had the opportunity to appreciate the needs and applications for an adult intelligence test. After careful review, Wechsler essentially cherry-picked his subtests from the most clinically useful and psychometrically adequate tests of his era; he thus provided practitioners with an easy transition to make from using many separate, independently normed tests with a variety of instructions and scoring rules to a single battery of co-normed tests, with streamlined administration and fairly uniform scoring rules. Wechsler acknowledged, "Our aim was not to produce a set of brand new tests but to select, from whatever source available, such a combination of them as would best meet the requirements of an effective adult scale" (1939, p. 78). Most of the standardization sample of 1,586 participants was collected in the city and state of New York; the sample was stratified by age, sex, education, and occupation, but was limited to English-speaking white examinees.

The Bellevue consisted of 10 subtests, with the Vocabulary subtest serving as an alternate. With the exception of a single speeded subtest (Digit Symbol), items on each subtest were sequenced in approximate order of difficulty, from easiest to hardest. Performance on the first five subtests contributed to the Verbal IQ, and performance on the second five subtests contributed to the Performance IQ. Full Scale IQ scores ranged from 28 to 195. Subtest raw scores were converted to a mean of 10 and standard deviation of 3, while IQ scores approximated a mean of 100 and standard deviation of 15. Wechsler's subtests dichotomized the composition of his test battery into Verbal and Performance/nonverbal, just as the Army mental tests had distinguished between the Alpha and the Beta/performance tests. This dichotomy remained of value for the same reasons it was helpful with Army mental testing: It permitted valid assessment of individuals whose intelligence was likely to be underestimated by verbal intelligence tests alone (i.e., those who were poorly educated, from non-English-language origins, or otherwise disadvantaged by language-dependent tests). Moreover, Wechsler considered distinctive Verbal and Performance intelligence tasks to sample behaviors in multiple areas of interest, generating important diagnostic information rather than representing different forms of intelligence (Wechsler, 1939). He considered the Verbal and Performance tests to be equally adequate measures of general intelligence, but he emphasized the importance of appraising people "in as many different modalities as possible" (Wechsler et al., 1975, p. 55).

The 1939 test battery (and all subsequent Wechsler intelligence scales) also offered a deviation IQ, the index of intelligence based on statistical distance from the normative mean in standardized units, as Arthur Otis (1917) had proposed. Wechsler deserves credit for popularizing the deviation IQ, although the Otis Self-Administering Tests and the Otis Group Intelligence Scale had already used similar deviation-based composite scores in the 1920s. Inexplicably, Terman and Merrill made the mistake of retaining a ratio IQ (i.e., mental age/chronological age) on the 1937 Stanford–Binet, even though the method had long been recognized as producing distorted IQ estimates for adolescents and adults (e.g., Otis, 1917). Terman and Merrill (1937, pp. 27–28) justified their decision on the dubious ground that it would have been too difficult to reeducate teachers and other test users familiar with the ratio IQ.

Wechsler first introduced the Bellevue Intelligence Scale at a meeting at the New York Academy of Medicine in 1937, and the first edition of *The Measurement of Adult Intelligence*—which would include the manual for the test soon known as the Wechsler–Bellevue Form I—was published in 1939. Early after its publication, Wechsler was approached by George K. Bennett, director of the Tests Division of The Psychological Corporation, who was impressed by the test and asked to produce the test materials (Edwards, 1974). Critics generally praised the "organization of well-known tests into a composite scale" with "considerable diagnostic as well as measurement value" (Lorge, 1943, p. 167), but Wechsler was faulted on technical errors (Anastasi, 1942; Cureton, 1941; McNemar, 1945) and theoretical shortcomings (e.g., Anastasi, 1942; Cronbach, 1949). Figure 1.14 shows Wechsler in the 1940s, after his test had become a success.

FIGURE 1.14. David Wechsler was chief psychologist at New York's Bellevue Psychiatric Hospital when he published his Bellevue Intelligence Scale (later known as the Wechsler–Bellevue), which quickly became the intelligence test of choice for adults. Reprinted by courtesy of the late Arthur Weider.

Among practicing psychologists and researchers working with adults, the Wechsler–Bellevue was a resounding success. In his review of research on the Wechsler–Bellevue in its first 5 years, Rabin (1945) concluded:

> The Wechsler–Bellevue Scales have stimulated considerable psychometric research and have supplanted some time-honored diagnostic tools. The reliability and validity of Wechsler's scales, as a whole and in part, have been proved in several studies. The consensus of opinion is that the test correlates highly with some of the best measures of intellect and that it tends to differentiate better than other measures between the dull and feebleminded. (p. 419)

In an update 6 years later, Rabin and Guertin (1951) noted the "vast popularity and wide usage of the test" (p. 239) and a "veritable flood" of research (p. 211), making the Wechsler–Bellevue "a commonly used measuring rod for comparison and validation, if not actual calibration of newer and more recent techniques" (p. 239).

From 1941 to 1945, Wechsler served as an expert civilian consultant to the Adjutant General's Office, preparing the Wechsler Mental Ability Scale, Form B (Wechsler, 1942, cited by Altus, 1945), also known as the Army Wechsler, and the Wechsler Self-Administering Test. These tests appear to have been of limited use for the military, in large part because they were too difficult for many Army recruits. The Wechsler Mental Ability Scale, Form B is of interest because it consisted of seven Verbal and nine Performance subtests, including Mazes and Series Completion (Altus, 1945), signaling possible additions to the battery. Wechsler also taught in the Army Psychological Training Program (Seidenfeld, 1942).

In the years and decades after the war, Wechsler developed the Wechsler–Bellevue Form II (Wechsler, 1946), the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967). Although David Wechsler died in 1981, most of these tests have gone through multiple editions, with staff test development specialists and external expert advisors substituting for a living author in recent decades. In 1975, Wechsler expressed support for measuring intelligence in individuals older than age 65 "without exposing the older person to tests involving speed, perception, and so forth." He proposed to call this test the Wechsler Intelligence Scale for the Elderly, or the WISE (Wechsler et al., 1975; also D. O. Herman, personal communication, November 9, 1993). Wechsler never proposed or wrote about achievement tests or nonverbal tests like those that currently carry his name.

In creating his intelligence scales, Wechsler combined popular and clinically useful existing tests into a streamlined, well-organized, and psychometrically innovative battery. Although his tests have become established as industry standards over many decades, Chattin and Bracken (1989) surveyed practicing school psychologists and reported that efficiency and practicality remain the central reasons why the Wechsler intelligence scales remain popular.

LOOSE THREADS: RESOLVED AND UNRESOLVED ISSUES IN INTELLIGENCE TESTING

Students of history are likely to find intelligence and its assessment a fascinating and frustrating subject—full of remarkable characters and events like those I have described—but also with many problems that surface over and over again. Because intelligence testing is a young science, it should be no surprise that so many strands in its story remain loose and unresolved, and there is sufficient diversity in thought among psychologists that even the most scientifically proven ideas will have dissenters. At the same time, it does not seem scientifically unreasonable to expect at some point a consensus-based definition of *intelligence,* agreement on the existence of a general factor of intelligence, and establishment of a uniform framework for understanding the structure of human cognitive abilities (all of which are discussed below). The historical association of intelligence testing with eugenics, however, is an ideological problem that may be harder to resolve; it may forever taint the tests with the appearance of social inequity and racism, in spite of many efforts to enhance the fairness of intelligence tests. In this section, I describe a few of many loose thematic threads that have contributed to breaks in the fabric of applied intelligence testing from its early days.

Before I begin describing long-standing unresolved issues in intelligence, it may be helpful first to note areas that appear to be resolved. In response to the public controversy associated with Herrnstein and Murray's book *The Bell Curve* (1994), Linda S. Gottfredson of the University of Delaware contacted an editor at the *Wall Street Journal,* who agreed to publish a statement signed by experts about mainstream scientific thinking on intelligence. Gottfredson drafted the statement, had it reviewed by several authorities, and solicited signatures of agreement from experts across psychology and other disciplines. The resulting statement with 25 conclusions, "Mainstream Science on Intelligence," was published in late 1994 with 52 signatories (Gottfredson, 1994); it was later reprinted with supplemental information as an editorial in the journal *Intelligence* (Gottfredson, 1997). In another response to Herrnstein and Murray's book, the APA Board of Scientific Affairs created a task force to issue an authoritative scientific statement about intelligence and its assessment, titled "Intelligence: Knowns and Unknowns" (Neisser et al., 1996). These two statements represent relatively rare scientific consensus statements about intelligence in the history of psychology. Ironically, there are many areas in which they appear to disagree.

The Definition of *Intelligence*

An initial step in any scholarly endeavor is to define one's terms, but the term *intelligence* still has no consensus-based definition. Efforts to arrive at a consensus date back about a century, as do criticisms that "psychologists have never agreed on a definition" (Lippmann, 1922c, p. 213). In a frequently quoted but much reviled definition, E. G. Boring (1923) wrote:

> Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test. This is a narrow definition, but it is the only point of departure for a rigorous discussion of the tests . . . no harm need result if we but remember that measurable intelligence is simply what the tests of intelligence test, until further scientific observation allows us to extend the definition. (p 35)

The failure to arrive at a consensus on defining *intelligence* after a century of research constitutes one of the most surprising loose threads in the history of psychology. Terman (1916) demurred, essentially arguing that we can work with the construct of intelligence without arriving at a definition:

> To demand, as critics of the Binet method have sometimes done, that one who would measure intelligence should first present a complete definition of it, is quite unreasonable. As Stern points out, electrical currents were measured long before their nature was well understood. Similar illustrations could be drawn from the processes involved in chemistry, physiology, and other sciences. In the case of intelligence it may be truthfully said that no adequate definition can possibly be framed which is not based primarily on the symptoms empirically brought to light by the test method. (p. 44)

As demonstrated in the statements above, Boring and Terman expected that research would eventually lead to a definition of *intelligence*. How much longer must we wait?

As we have reported, the association of intelligence with evolutionary *adaptation* dates back to Spencer (1855), who described intelligence as "an adjustment of inner to outer relations" (p. 486). This definition may be understood as suggesting that intelligence confers a capacity to adapt to environmental change, but principles of neo-Darwinian evolution hold that natural selection favors adaptations that enhance survival and reproductive fitness. In order to validate a definition of intelligence featuring adaptation, then, the logical and empirical question is whether intelligence confers any advantages in terms of longer lifespans, fecundity, or other aspects of reproductive fitness. Studies relating intelligence to evolutionary fitness (e.g., family size, number of children) date back to the 1930s, and clearly a meta-analysis is needed to make sense of the many contradictory findings. Gottfredson (2007) recently reported evidence that higher intelligence may improve overall survival rate, and that lower intelligence may be associated with a disproportionately elevated risk of accidental death. Together with colleagues, she has also reported findings of a fitness factor that is related to intelligence (Arden, Gottfredson, Miller, & Pierce, 2009).

Several formal meetings or print symposia have sought a definition of *intelligence,* and the clear-cut conclusion from these efforts is that the experts do not agree on a definition. A list of proposed definitions for the term appears in Table 1.2. The earliest symposium I can identify, entitled "Instinct and Intelligence" (e.g., Myers, 1910), was held in London in July 1910, at a joint meeting of the Aristotelian and British Psychological Societies and the Mind Association, with resulting papers appearing in the *British Journal of Psychology*. The best-known print symposium is "Intelligence and Its Measurement: A Symposium," appearing in the *Journal of Educational Psychology* (Peterson, 1921; Pintner, 1921; Thorndike, 1921). The symposium asked 17 leading investigators explicitly what they conceived intelligence to be. Another symposium, "The Nature of General Intelligence and Ability," was conducted at the Seventh International Congress of Psychology, held at Oxford University in 1923 (e.g., Langfeld, 1924). In a follow-up to the 1921 *Journal of Educational Psychology* symposium, Sternberg and Detterman (1986) asked 25 authorities to write essays conveying what they believed intelligence to be. Sternberg and Berg tabulated facets of the definitions provided: In descending order, the most frequent attributes in definitions of intelligence were higher-level cognitive functions (50%), that which is valued by culture (29%), executive processes (25%), elementary processes (perception, sensation, and/or attention; 21%), knowledge (21%), and overt behavioral manifestations of intelligence (such as effective or successful responses; 21%). By comparison, the most frequent attributes in definitions from the 1921 symposium were higher-level cognitive functions (57%), adaptation (29%), ability to learn (29%), physiological mechanisms (29%), elementary processes (21%), and overt behavioral manifestations of intelligence (21%). Even efforts to seek definitions of intelligence among laypeople have found that definitions vary; moreover, people can be self-serving and seem to offer definitions that also capture some quality readily found in themselves (e.g., Gay, 1948).

Never one to embrace diverse perspectives, Charles E. Spearman (1927) disparaged "repeated recourse to symposia" (p. 8) and surveys of expert opinion in efforts to define intelligence:

> Chaos itself can go no further! The disagreement between different testers—indeed, even the doctrine and the practice of the selfsame tester—has reached its apogee. If they still tolerate each other's proceedings, this is only rendered possible by the ostrich-like policy of not looking facts in the face. In truth, "intelligence" has become a mere vocal sound, a word with so many meanings that it finally has none. (p. 14)

Jensen (1998) echoed Spearman's sentiment, recommending that psychologists "drop the ill-fated word from our scientific vocabulary, or use it only in quotes, to remind ourselves that it is not only scientifically unsatisfactory but wholly unnecessary" (p. 49).

TABLE 1.2. Selected Definitions of Intelligence (Arranged Chronologically)

Herbert Spencer (1855): “Instinct, Reason, Perception, Conception, Memory, Imagination, Feeling, Will, &c., &c., can be nothing more than either conventional groupings of the correspondences; or subordinate divisions among the various operations which are instrumental in effecting the correspondences. However widely contrasted they may seem, these various forms of intelligence cannot be anything else than either particular modes in which the adjustment of inner to outer relations is achieved; or particular parts of the process of adjustment” (p. 486).

Alexander Bain (1868): “The functions of Intellect, Intelligence, or Thought, are known by such names as Memory, Judgment, Abstraction, Reason, Imagination” (p. 82).

Hermann Ebbinghaus (1908): “Intelligence means organization of ideas, manifold interconnection of all those ideas which ought to enter into a unitary group because of the natural relations of the objective facts represented by them. The discovery of a physical law in a multitude of phenomena apparently unrelated, the interpretation of an historical event of which only a few details are directly known, are examples of intelligence thought which takes into consideration innumerable experiences neglected by the less intelligent mind. Neither memory alone nor attention alone is the foundation of intelligence, but a union of memory and attention” (pp. 150–151).

Charles S. Myers (1910): “As the organism becomes endowed with an increasingly larger number of mutually incompatible modes of reaction, the intelligent aspect apparently comes more and more to the fore while the instinctive aspect apparently recedes *pari passu* into the background” (p. 214).

C. Lloyd Morgan (1910): “I regard the presence of implicit expectation (in the lower forms) or explicit anticipation (in the higher forms) as distinguishing marks or criteria of intelligence. In other words for the intelligent organism the present experience at any given moment comprises more or less ‘meaning’ in terms of previously-gotten experience” (p. 220).

H. Wildon Carr (1910): “Intelligence is the power of using categories, it is knowledge of the relations of things. It is a knowledge that gives us the representation of a world of objects externally related to one another, a world of objects in space, or measurable actions and reactions. . . . Intelligence is an outward view of things, never reaching the actual reality it seeks to know” (pp. 232–233).

Alfred Binet and Théodore Simon (Binet, 1911/1916): “Intelligence serves in the discovery of truth. But the conception is still too narrow; and we return to our favorite theory; the intelligence marks itself by the best possible adaptation of the individual to his environment” (pp. 300–301).

William Stern (1914): “Intelligence is a general capacity of an individual consciously to adjust his thinking to new requirements: it is general mental adaptability to new problems and conditions of life” (p. 3).

M. E. Haggerty (1921). “In my thinking the word intelligence does not denote a single mental process capable of exact analytic definition. It is a practical concept of connoting a group of complex mental processes traditionally defined in systematic psychologies as sensation, perception, association, memory, imagination, discrimination, judgment and reasoning “(p. 212).

V. A. C. Henmon (1921): “Intelligence . . . involves two factors—the capacity for knowledge and knowledge possessed” (p. 195).

Joseph Peterson (1921): “Intelligence seems to be a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior. It is a mechanism for adjustment and control, and is operated by internal as well as by external stimuli. The degree of a person’s intelligence increases with his range of receptivity to stimuli and the consistency of his organization of responses to them” (p. 198).

Rudolf Pintner (1921): “I have always thought of intelligence as the ability of the individual to adapt himself adequately to relatively new situations in life. It seems to include the capacity for getting along well in all sorts of situations. This implies ease and rapidity in making adjustments and, hence, ease in breaking old habits and in forming new ones” (p. 139).

Lewis M. Terman (1921): “The essential difference, therefore, is in the capacity to form concepts to relate in diverse ways, and to grasp their significance: *An individual is intelligent in proportion as he is able to carry on abstract thinking*” (p. 128; emphasis in original).

Edward L. Thorndike (1921): “Realizing that definitions and distinctions are pragmatic, we may then define intellect in general as *the power of good responses from the point of view of truth or fact*, and may separate it according as the situation is taken in gross or abstractly and also according as it is experienced directly or thought of” (p. 124; emphasis in original).

L. L. Thurstone (1921): “Intelligence as judged in everyday life contains at least three psychologically differentiable components: a) the capacity to inhibit an instinctive adjustment, b) the capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trial and error, c) the volitional capacity to realize the modified instinctive adjustment into overt behavior to the advantage of the individual as a social animal” (pp. 201–202).

Herbert Woodrow (1921): “Intelligence . . . is the capacity to acquire capacity” (p. 208).

(continued)

TABLE 1.2. *(continued)*

E. G. Boring (1923): "Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test" (p. 35).

Édouard Claparède (1924): "[Intelligence is] the ability to solve new problems" (quoted by Langfeld, 1924, p. 149).

Godfrey H. Thomson (1924): "[Intelligence is] the ability to meet new situations with old responses and to discard those responses which prove unsuccessful" (quoted by Langfeld, 1924, p. 149).

David Wechsler (1939): "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (p. 3).

Anne Anastasi (1986): "Intelligence is not an entity within the organism but a quality of behavior. Intelligent behavior is essentially adaptive, insofar as it represents effective ways of meeting the demands of a changing environment" (pp. 19–20).

Jonathan Baron (1986): "I define intelligence as the set of whatever abilities make people successful at achieving their rationally chosen goals, whatever those goals might be, and whatever environment they are in. . . . To say that a person has a certain level of ability is to say that he or she can meet a certain standard of speed, accuracy, or appropriateness in a component process defined by the theory in question" (p. 29).

J. W. Berry (1986): "At the present time intelligence is a construct which refers to the end product of individual development in the cognitive-psychological domain (as distinct from the affective and conative domains); this includes sensory and perceptual functioning but excludes motor, motivational, emotional, and social functioning . . . it is also adaptive for the individual, permitting people to operate in their particular cultural and ecological contexts" (p. 35).

J. P. Das (1986): "Intelligence, as the sum total of all cognitive processes, entails planning, coding of information and attention arousal. Of these, the cognitive processes required for planning have a relatively higher status in intelligence. Planning is a broad term which includes among other things, the generation of plans and strategies, selection from among available plans, and the execution of those plans. . . . Coding refers to two modes of processing information, simultaneous and successive. . . . The remaining process (attention arousal) is a function basic to all other higher cognitive activities" (pp. 55–56).

Douglas K. Detterman (1986): "In my opinion, intelligence can best be defined as a finite set of independent abilities operating as a complex system" (p 57).

John Horn (1986): "'What do I conceive intelligence to be?' This is rather like asking me: 'What do I conceive invisible green spiders to be?' For current knowledge suggests to me that intelligence is not a unitary entity of any kind. Attempts to describe it are bound to be futile" (p. 91).

Earl Hunt (1986): "'Intelligence' is solely a shorthand term for the variation in competence on cognitive tasks that is statistically associated with personal variables. . . . Intelligence is used as a collective term for 'demonstrated individual differences in mental competence'" (p. 102).

James W. Pellegrino (1986): "The term intelligence denotes the general concept that individuals' responses to situations vary in quality and value as judged by their culture" (p. 113).

Sandra Scarr (1986): "To be an effective, intelligent human being requires a broader form of personal adaptation and life strategy, one that has been described in 'invulnerable' children and adults: They are copers, movers, and shapers of their own environments" (p. 120).

Richard E. Snow (1986): "[Intelligence can be defined in several ways:] . . . [1] the incorporation of concisely organized prior knowledge into purposive thinking—for short, call it *knowledge-based thinking.* . . . [2] *apprehension* captures the second aspect of my definition—it refers to Spearman's (1923, 1927) principle that persons (including psychologists) not only feel, strive, and know, but also *know* that they feel, strive, and know, and can anticipate further feeling, striving, and knowing; they monitor and reflect upon their own experience, knowledge, and mental functioning in the past, present, and future tenses. . . . [3] *adaptive purposeful striving.* It includes the notion that one can adopt or shift strategies in performance to use what strengths one has in order to compensate for one's weaknesses. . . . [4] agile, analytic reasoning of the sort that enables significant features and dimensions of problems, circumstances, and goals to be decontextualized, abstracted, and interrelated rationally . . . *fluid-analytic reasoning.* . . . [5] *mental playfulness* . . . able to find or create interesting problems to solve and interesting goals toward which to strive. This involves both tolerance of ambiguity and pursuit of novelty. . . . [6] *idiosyncratic learning* . . . Persons differ from one another in the way they assemble their learning and problem-solving performance, though they may achieve the same score. Persons differ *within* themselves in how they solve parts of a problem, or different problems in a series" (pp. 133–134; emphasis in original).

Robert J. Sternberg (1986): "Intelligence is mental self-government. . . . The essence of intelligence is that it provides a means to govern ourselves so that our thoughts and actions are organized, coherent, and responsive to both our internally driven needs and to the needs of the environment" (p. 141).

The argument has also been made that a structural/statistical understanding of intelligence may serve as an adequate substitute for a verbal/descriptive definition. Gottfredson and Saklofske (2009) suggest that definitional issues of intelligence are "now moot because the various empirical referents to which the term is commonly applied can be distinguished empirically and related within a common conceptual structure [i.e., the Cattell–Horn–Carroll model of human cognitive abilities]" (p. 188).

To *g* or Not to *g*?

Another long-standing unresolved thread in the history of intelligence testing has to do with the general factor of intelligence, psychometric *g*. General intelligence was affirmed in the 1994 "Mainstream Science on Intelligence" statement (Gottfredson, 1997), but the 1996 "Intelligence: Knowns and Unknowns" statement hedged on *g*, stating that "while the *g*-based factor hierarchy is the most widely accepted current view of the structure of abilities, some theorists regard it as misleading" (Neisser et al., 1996, p. 81). Here I describe some history for *g*.

In 1904, Charles E. Spearman (1863–1945) published a groundbreaking paper reporting the discovery of a factor of "general intelligence," derived from positive intercorrelations between individual scores on tests of sensory discrimination, musical talent, academic performance, and common sense. Although the correlation coefficient statistic was still relatively new, Spearman realized that previous studies (e.g., those by Gilbert and by Wissler) had failed to account for measurement error—that is, reduced score reliability, which invariably reduces the magnitude of correlations. He devised a method to correct the correlation coefficient for attenuation, reporting subsequently that his correlational analyses showed "all branches of intellectual activity have in common one fundamental function (or group of functions)" (p. 284), which he later described using concepts from physics such as "the amount of a general mental energy" (Spearman, 1927, p. 137). The *g* factor, or psychometric *g*, was a mathematically derived general factor, stemming from the shared variance that saturates batteries of cognitive/intelligence tests. Jensen (1998) has summarized the literature showing that correlates of *g* include scholastic performance, reaction time, success in training programs, job performance in a wide range of occupations, occupational status, earned income, and creativity, among others.

Critics of general intelligence appeared quickly. Edward L. Thorndike, who challenged Spearman's work for decades, reported no support for *g* on a set of measures similar to those originally used by Spearman, finding a weak correlation between sensory discrimination and general intelligence, and stating that "one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatever* common to all mental functions, or to any half of them" (Thorndike, Lay, & Dean, 1909, p. 368; original emphasis).

Until Spearman's death, Thorndike; a Scotsman, Godfrey Thomson; and two Americans, Truman L. Kelley and Louis L. Thurstone, participated in an ongoing scholarly debate with him on the existence and nature of *g*, as well as other aspects of the structure of intelligence. Spearman devoted the rest of his career to elaboration and defense of his theory, authoring *The Nature of "Intelligence" and the Principles of Cognition* (Spearman, 1923), *The Abilities of Man: Their Nature and Measurement* (Spearman, 1927), and *Human Ability: A Continuation of "The Abilities of Man"* (Spearman & Wynn Jones, 1950). A good account of this debate may be found in R. M. Thorndike and Lohman (1990). Newly discovered exchanges among Thorndike, Thomson, and Spearman in the 1930s serve to highlight Spearman's dogmatism (Deary, Lawn, & Bartholomew, 2008).

The leading intelligence test developers generally accepted the existence of a psychometric *g* factor. After initial reticence, Alfred Binet eventually embraced a general factor; in *Les Idées Modernes sur les Enfants,* Binet (1909/1975) wrote that "the mind is unitary, despite the multiplicity of its faculties . . . it possesses one essential function to which all the others are subordinated" (p. 117). In the 1916 Stanford–Binet, Lewis M. Terman accepted the concept of general intelligence and conceded that the IQ score provided a good estimate of *g*:

> It is true that more than one mental function is brought into play by the test. The same may be said of every other test in the Binet scale and for that matter of any test that could be devised. It is impossible to isolate any function for separate testing. In fact, the functions called memory, attention, perception, judgment, etc., never operate in isolation. There are no separate and special "faculties" corresponding to such terms, which are merely convenient names for characterizing mental processes of various types. In any test it is "general ability" which is operative, perhaps now *chiefly* in remembering, at another time

> *chiefly* in sensory discrimination, again in reasoning, etc. (p. 194; original emphasis)

David Wechsler, who had been deeply impressed with Spearman during his few months at University College London in 1919, wrote that Spearman's theory and its proofs constitute "one of the great discoveries of psychology" (1939, p. 6). He further noted that "the only thing we can ask of an intelligence scale is that it measures sufficient portions of intelligence to enable us to use it as a fairly reliable index of the individual's global capacity" (p. 11).

What is the current status of *g*? When Reeve and Charles (2008) surveyed 36 experts in intelligence, they found a consensus that *g* is an important, nontrivial determinant (or at least predictor) of important real-world outcomes, and that there is no substitute for *g* even if performance is determined by more than *g* alone. With the leading authors of intelligence tests accepting psychometric *g*, and with authorities in intelligence research consensually accepting its importance, the thread of general intelligence would appear to be well secured in our metaphorical tapestry of the history intelligence.

Yet the concept of general intelligence continues to be challenged, most often on theoretical grounds but also on statistical grounds. Stephen J. Gould (1996) forcefully challenged *g*, associating it with many of the historically negative (and shameful) applications of intelligence testing. Several intelligence theorists, including Raymond B. Cattell, J. P. Das, Howard Gardner, and Robert J. Sternberg, have also rejected the concept of general intelligence. The most cogent challenges to *g* have come from John L. Horn (Horn & Noll, 1994, 1997), who pointed out fallacies of extracting *g* from the *positive manifold* (i.e., the finding that almost all tests that reliably measure a cognitive ability correlate positively with all other such tests).

The Structure of Intelligence

The struggle to construct a complex model of intelligence probably began with the phrenologists, who specified individual faculties (each corresponding to an "organ" of the brain) that together constituted intelligence. For example, Combe (1830) described faculties of perception (e.g., form, size, weight, eventuality, language) and faculties of reflection (e.g., comparison, causality) that altogether constituted intellectual faculties; he also described a separate set of affective faculties. With the discovery of *g* by Spearman (1904), the notion of a unitary intelligence gained traction, but by the end of the 1930s, psychologists and educators were again embracing the complexity of the mind (e.g., Ackerman, 1995). Current hierarchical models of intelligence feature broad ability factors, which have grown steadily in number: from the two factors enumerated by Cattell (1941) and Vernon (1950) to the eight specified by Carroll (1993) to about 10 factors specified by Carroll (2003) to about 15 or 16 broad factors in 2010 (e.g., McGrew, 2009; Newton & McGrew, 2010). The question that appears to be unresolved in this thread is this: Just how many group factors constitute the structure of intelligence?

For much of the 20th century and into the 21st, the complex structure of intelligence has been revealed through statistical methodologies that discover and define sources of test performance variance, usually through factor analyses. Factor analysis is a statistical technique capable of reducing many variables into a few underlying dimensions. The foundation for use of factor analysis in understanding the structure of cognition was laid with Spearman (1904). Spearman's theory encompassing general intelligence was originally called *two-factor theory* because it partitioned performance variance into a *general factor* shared across tasks, and *specific factors* that were unique to individual tasks. Following the contributions of Kelley, Thorndike, and Thurstone (among others), Spearman (1927) reluctantly came to acknowledge the existence of *group factors* formed by clusters of tests that yielded higher-than-expected intercorrelations by virtue of similarities in their content, format, or response requirements: "Any element whatever in the specific factor of an ability will be turned into a group factor, if this ability is included in the same set with some other ability which also contains this element" (p. 82). The extraction of a general factor and group factors (now called *broad ability factors*) contributed to the development of *hierarchical* structural analyses of intelligence. In hierarchical factor analyses, a general factor is first extracted; the residual variance is factored to extract any group factors; and the remaining variance is often said to be specific.

Although there have been well over 1,000 factor-analytic investigations in the literature of intelligence and cognitive abilities (see Carroll, 1993), many of which remain important in understanding the structure of cognitive abilities, space only permits coverage of a few prototypal models with distinctive characteristics.

Thurstone's Primary Mental Abilities

Louis L. Thurstone (1887–1955) developed the statistical technique of multiple factor analysis and is best remembered for his theory of primary mental abilities, a factor-analysis-derived model of multiple cognitive abilities that effectively challenged Spearman's single general factor of intelligence. Thurstone developed factor analysis techniques permitting the extraction of factors that are orthogonal to each other (i.e., separate, independent, and unrelated). From a battery of 56 paper-and-pencil tests administered in about 15 hours to each of 240 superior, college-level students, Thurstone (1938) extracted seven primary factors: spatial/visual, perception of visual detail, numerical, two verbal factors (logic and words), memory, and induction. From a study of over 700 students age 14, who were given 60 tests in 11 sessions lasting 1 hour each, Thurstone and Thurstone (1941) extracted six factors: verbal comprehension, word fluency, space, number, memorizing, and reasoning/induction. By 1945, Thurstone had settled on eight primary mental abilities, each denoted by a letter: Verbal Comprehension (V), Word Fluency (W), Number Facility (N), Memory (M), Visualizing or Space Thinking (S), Perceptual Speed (P), Induction (I), and Speed of Judgment (J). Although Thurstone (1947) eventually accepted the existence of a general factor, he considered the use of a single score such as the IQ to be inadequate, and urged the use of cognitive profiles describing strengths and weaknesses among the fundamental abilities (Thurstone, 1945).

Vernon's Hierarchical Model

In what has been called the first truly hierarchical model of intelligence, Philip E. Vernon (1905–1987) proposed that a higher-order *g* factor dominates two lower-order factors, v:ed (verbal:educational) and k:m (spatial:mechanical); in turn, v:ed and k:m subsume various minor group factors, which in turn dominate very narrow and specific factors. Based on his review of factor-analytic investigations through 1950, Vernon (1950, 1961) considered v:ed to dominate verbal, number, reasoning, attention, and fluency factors, while k:m dominates spatial ability, mechanical ability, psychomotor coordination, reaction time, drawing, handwork, and various technical abilities. He considered it a likely oversimplification to assume that there are just two factors at the level below *g*, although his simple dichotomy may be seen as having supported the verbal–performance dichotomy traditionally associated with the Wechsler intelligence scales.

Cattell, Horn, and Carroll's Model of Fluid and Crystallized Intelligence

Arguably the most important contemporary structural and hierarchical model of intelligence is based on extensions of the theory of fluid (Gf) and crystallized (Gc) intelligence first proposed by Raymond B. Cattell (1905–1998) in a 1941 APA convention presentation. Cattell, who completed his doctorate in 1929 at University College London with Spearman, joined E. L. Thorndike's research staff at Columbia University in 1937, where he worked closely with proponents of multifactor models of intelligence. He authored over 500 articles and 43 books during his career. In his 1941 APA presentation, Cattell asserted the existence of two separate general factors: g_f (fluid ability or fluid intelligence) and g_c (crystallized ability or crystallized intelligence). The convention was later adopted that these factors would be represented by uppercase G, whereas a single general factor would be represented by lowercase *g*.

Fluid ability was described by Cattell (1963, 1971) and Horn (1976) as a facility in reasoning, particularly where adaptation to new situations is required and crystallized learning assemblies are of little use. Ability is considered to be fluid when it takes different forms or utilizes different cognitive skill sets according to the demands of the problem requiring solution. For Cattell, fluid ability is the most essential general-capacity factor, setting an upper limit on the possible acquisition of knowledge and crystallized skills. In contrast, *crystallized* intelligence refers to accessible stores of knowledge and the ability to acquire further knowledge via familiar learning strategies. It is typically measured by recitation of factual information, word knowledge, quantitative skills, and language comprehension tasks because these include the domains of knowledge that are culturally valued and educationally relevant in the Western world (Cattell, 1941, 1963, 1971, 1987; Horn & Cattell, 1966).

Cattell's model of fluid and crystallized intelligence was energized by the contribution of John L. Horn (1928–2006). Not only was Horn's (1965) dissertation the first empirical study of the theory since 1941; it also showed that fluid and crystallized abilities have different developmental trajectories over the lifespan (McArdle, 2007). Cattell and Horn expanded the number of ability factors from two to five (adding visualization, retrieval capacity, and cognitive speed; Horn & Cattell, 1966). In the next 25 years or so, Horn had arrived at nine ability factors (Horn & Noll, 1994, 1997), while Cattell's list had grown to six ability factors (adding distant memory and retrieval) plus three smaller provincial factors (visual, auditory, and kinesthetic; Cattell, 1998). The growth of the number of factors in this model continues, and a 2001 symposium at the University of Sydney enumerated even more potential ability factors (Kyllonen, Roberts, & Stankov, 2008). As noted earlier, McGrew (2009; see also Newton & McGrew, 2010) now lists 15 or 16 broad ability factors.

In 1993, John B. Carroll (1916–2003) built upon the work of Cattell and Horn by proposing a hierarchical, multiple-stratum model of human cognitive abilities with the general intelligence factor, *g*, at the apex (or highest stratum); eight broad factors of intelligence at the second stratum; and at least 69 narrow factors at the first (or lowest) stratum. Carroll was the author of nearly 500 books and journal articles over the span of 60 years; he had been mentored early in his career by L. L. Thurstone, and some years later after Thurstone's death he became director of the Thurstone Psychometric Laboratory at the University of North Carolina, Chapel Hill (Jensen, 2004). For a dozen years after his retirement, Carroll (1983, 1993, 1994) accumulated over a thousand archival datasets related to human cognitive test performance; 461 of the datasets were ultimately judged adequate for his analyses. He then conducted iterative principal-factor analyses requiring convergence to a strict criterion, followed by varimax rotation of the principal-factor matrix, with the requirement that each extracted factor contain salient loadings on at least two variables. If necessary, promax or other rotational procedures were used. Factorization was then carried up to the highest viable order. The data were subjected to the Schmid–Leiman orthogonalized hierarchical-factor procedure, and factor interpretations were based on the resulting hierarchical-factor matrix. Carroll's results showed general intelligence (*g*) as appearing in the highest stratum; the second stratum, listed in descending strength of association with *g*, consisted of fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gsm), broad visual perception (Gv), broad auditory perception (Ga), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (reaction time decision speed); finally, very narrow and specific factors were placed in the lowest stratum. Although Carroll's three-stratum model is historically young, its early reception suggests that it has quickly become a landmark study. The following samples from reviews are fairly representative:

- "Further research may alter details of the map, although it is unlikely that any research for some years to come will lead to a dramatic alteration in Carroll's taxonomy." (Brody, 1994, p. 65)
- "It is simply the finest work of research and scholarship I have read and is destined to be the classic study and reference work of human abilities for decades to come." (Burns, 1994, p. 35)
- "[It is] a truly monumental work." (Jensen, 2004, p. 3)
- "Carroll's work represents what may well be the most extensive, indeed, exhaustive analysis of a data case that has ever been attempted in the field of intelligence. The theory deserves to be taken seriously." (Sternberg, 1994, p. 65)

A note of caution for applied practitioners, however, comes from Carroll himself: He indicated that his survey of cognitive abilities "paid very little attention to the importance, validity, or ultimate usefulness of the ability factors that have been identified" (1993, p. 693). Carroll's three-stratum theory has been integrated with extended Gf-Gc theory to form the Cattell–Horn–Carroll (CHC) framework, a name to which Horn and Carroll both agreed a few years after Cattell's death (Newton & McGrew, 2010). The CHC framework already appears to have exerted a strong influence upon the development of contemporary intelligence tests (e.g., Keith & Reynolds, 2010). Shortly before his death, Carroll (2003) expanded his model to include 10 second-stratum factors, indicating that even this definitive model may be expanded.

Intelligence and Eugenics

We have seen more than once that the public welfare may call upon the best citizens for their lives. It would be strange if it could not call upon those who already sap the strength of the State for these lesser sacrifices, often not felt to be such by those concerned, in order to prevent our being swamped with incompetence. It is better for all the world, if instead of waiting to execute degenerate offspring for crime, or to let them starve for their imbecility, society can prevent those who are manifestly unfit from continuing their kind. The principle that sustains compulsory vaccination is broad enough to cover cutting the Fallopian tubes. Three generations of imbeciles are enough.

—OLIVER WENDELL HOLMES (*Buck v. Bell,* 1927)

So wrote Justice Oliver Wendell Holmes, Jr., for the majority opinion of the U.S. Supreme Court in 1927, in the case of Carrie Buck versus James Hendren Bell, Superintendent of the Virginia State Colony for Epileptics and Feeble Minded. Carrie Buck was an 18-year-old woman with the mental age equivalent of 9 when the superintendent of the Virginia State Colony petitioned to have her sterilized. She was reported to be the daughter of a feeble-minded mother in the same institution and the mother of a feeble-minded child—hence Holmes's statement that "Three generations of imbeciles are enough." By an 8-to-1 margin, the court upheld the 1924 Virginia statute, the Eugenical Sterilization Act of 1924, authorizing the compulsory sterilization of "mental defectives," including individuals who were "feeble-minded" (i.e., intellectually disabled). On October 19, 1927, Carrie Buck was sterilized. Although the Supreme Court ruling has never been challenged or reversed, the Commonwealth of Virginia repealed the 1924 sterilization law in 1974. Historian Paul A. Lombardo (2008) recently reexamined this case, finding that there was insufficient evidence ever to assert cognitive impairment in Buck or her daughter, based on their school records.

For our purposes, it may be enough to cite compulsory sterilization laws for those with intellectual disabilities as a historical illustration of how intelligence test results may be (mis)used. In the broader context are questions about the wisdom of making legal, political, and public policy decisions on the basis of intelligence test research. Scholars in intelligence are at risk when they stray too far from psychological science into the realm of social engineering.

Francis Galton coined the term *eugenics* in 1883, describing it as "the science of improving stock" and defining it as "all influences that tend in however remote a degree to give to the more suitable races or strains of blood a better chance of prevailing speedily over the less suitable than they otherwise would have had" (p. 25). In *Hereditary Genius* (1869), he had already presented evidence that superior abilities are found more often among eminent families (i.e., those of judges, statesmen, premiers, commanders, scientists, scholars, etc.), and he proposed to increase the proportion of individuals with superior genetic endowments and thereby benefit the national intelligence through selective early marriages. "A man's natural abilities are derived by inheritance," Galton wrote, "under exactly the same limitations as are the form and physical features of the whole organic world" (1869, p. 1). He related his vision of a eugenics-practicing society in an unpublished fictional tale entitled "Kantsaywhere" (Galton, 1930). In this utopia, an individual's hereditary worth was measured by anthropometric tests, genetic failures were placed in labor colonies, enforced celibacy was the rule, and childbirth for the "unfit" was a crime. Karl Pearson noted in a footnote to this tale (p. 416 in Galton, 1930) that Galton's fictional laboratory in Kantsaywhere bears an uncanny resemblance to his anthropometric laboratory at South Kensington, one of the places where intelligence testing began. A photograph of Galton toward the end of his life, with his friend, colleague, and biographer, Karl Pearson, appears in Figure 1.15. Pearson, the renowned statistician, was also a dedicated eugenicist.

Almost all of the early authorities in the field of intelligence either wrote favorably about eugenics or belonged to organizations advocating eugenics. Some of the authorities on record as favoring eugenics in one form or another include J. McKeen Cattell, Raymond B. Cattell, Henry H. Goddard, Lewis M. Terman, Edward L. Thorndike, and Robert M. Yerkes. Until the horrors of Nazi genocide were exposed, including euthanasia of individuals with intellectual disabilities or mental disorders, eugenics was commonly seen as a contribution of

FIGURE 1.15. Francis Galton at age 87 with his biographer, Karl Pearson. Both were dedicated eugenicists. Photo from Pearson (1930, Plate 36). Reprinted by permission of The Pearson Papers, UCL Library Services, Special Collections.

science to human (and national) improvement. Lewis M. Terman, author of the Stanford–Binet, took a particularly active role in advocating for eugenics. For example, in a report to the California state legislature, Terman (1917) saw those with intellectual disabilities as having only negative impacts on society:

> Feeble-mindedness has always existed; but only recently have we begun to recognize how serious a menace it is to the social, economic, and moral welfare of the state. Extensive and careful investigations, in large numbers and in diverse parts of the United States, have furnished indisputable evidence that it is responsible for at least one-fourth of the commitments to state penitentiaries and reform schools, for the majority of cases of chronic and semi-chronic pauperism, and for much of our alcoholism, prostitution, and venereal diseases. (p. 45)

Terman's solutions were to segregate "feeble-minded" students in special classes so as not to "interfere with instruction" or "be a source of moral contagion" for other students (p. 51). He did not overtly recommend sterilization, but he implied that some action was necessary to prevent reproduction: "Three-fourths of the cases of feeble-mindedness are due to a single cause, heredity; and the one hopeful method of curtailing the increasing spawn of degeneracy is to provide additional care for our higher-grade defectives during the reproductive period" (p. 52).

Two of the most important 20th-century figures in applied intelligence testing, however, Alfred Binet and David Wechsler, are on record as having rejected perspectives associated with eugenics. Binet argued that intelligence can be changed, and he even developed a program of "mental orthopedics" to make educational interventions:

> I have often observed, to my regret, that a widespread prejudice exists with regard to the educability of intelligence. . . . A few modern philosophers seem to lend their moral support to these deplorable verdicts when they assert that an individual's intelligence is a fixed quantity, a quantity which cannot be increased. We must protest and react against this brutal pessimism. We shall attempt to prove that it is without foundation. (1909/1975, pp. 105–106)

David Wechsler found a more oblique way to criticize the eugenicists—by associating them with totalitarianism. In a 1961 paper, shortly after defining the terms *eugenes* and *apartheid,* he wrote, "The belief in class distinctions, whether considered innate or acquired, is . . . an essential tenet of all groups who are afraid of being ousted or displaced, and in particular of totalitarian governments" (p. 421). As Wechsler was a member of an oppressed immigrant group (Eastern European Jews) threatened with genocide in Romania and Germany, his condemnation of eugenics should not be surprising.

How does eugenics constitute a loose thread in the history of intelligence? Although no mainstream authorities today advocate for eugenics of the type that led to tragedies in the past, scholarship in the biology and heredity of intelligence remains extremely controversial, with recent accounts including threats to the academic freedom and even loss of lifetime recognition of achievements in psychology for those who conduct research in associated areas, including heritability (e.g., APA, 1997; Gottfredson, 2010; Horn, 2001; Tucker, 2009). Moreover, it might be argued that the history of intelligence and eugenics has contributed to a public perception that intelligence is about elitism, racism, and exclusion. In spite of over a century of research, the study of intelligence remains controversial for its social applications and implications.

TOWARD THE FUTURE

It may be beyond the scope of a historical chapter to look to the future, especially because this entire volume represents diverse visions for the future of intellectual assessment. At the same time, there are enough recurring events in our field that it sometimes seems as if we do not learn from mistakes (or, worse, as if we do not know our own history). For example, the concept of a general ability factor proposed by Spearman in 1904 has achieved as much of a consensus as is possible in psychology (e.g., Reeve & Charles, 2008), but still proponents of new theories (and tests) of intelligence often try to deny *g*—although their work usually proves to be saturated with *g*. Examples include Howard Gardner's (1983) theory of *multiple intelligences* and Keith Stanovich's newly proposed *rationality quotient* (Stanovich, West, & Toplak, 2016), each of which has been shown to be quite *g*-loaded (Ritchie, 2017; Visser, Ashton, & Vernon, 2006).

The critics of intellectual assessment return time and time again to issues of perceived bias against racial, ethnic, and linguistically different minorities. From a historical perspective, concerns about the effects of socioeconomic conditions on intelligence test performance were first openly recognized and articulated by Alfred Binet (1911/1916):

> The little children of the upper classes understand better and speak better the language of others. We have also noted that when they begin to compose, their compositions contain expressions and words better chosen than those of poor children. This verbal superiority must certainly come from the family life; the children of the rich are in a superior environment from the point of view of language; they hear a more correct language and one that is more expressive. (p. 320)

Issues of fairness have persisted for a century, perhaps reaching a low point in Carl C. Brigham's *A Study of American Intelligence* (1923), based on data from the Army mental tests. Brigham (1930), who created the SAT for the College Board, later recanted, acknowledging serious methodological flaws in his study of racial differences. Even now, a very active literature addresses the fairness of intelligence tests, often emitting more heat than light. Nancy M. Robinson (2003) has wisely pointed out that poverty and social inequities are longstanding problems in our society, with intelligence tests reflecting (but not causing) societal barriers and inequitable outcomes.

Sadly, psychologists do not receive credit for historical lessons learned, beginning with the terrible consequences of eugenics. But these shameful aspects of our history are still used to condemn intelligence testing, nearly 70 years after beliefs in eugenics were discarded. Moreover, support of eugenics is considered by some as a reason to discard the massive bodies of research accumulated by early 20th-century intelligence scholars like Lewis M. Terman, lead author of the Stanford–Binet. Among historians, the fallacy of *presentism* refers to the act of imposing the views of the present onto the past, instead of making a serious attempt to understand how historical figures themselves understood the world in their time (e.g., Fischer, 1970). Denigration of important scientific figures for advocacy of eugenics is a form of presentism—because no proponents of eugenics in the 1920s and 1930s knew what we now know, that Galton's (1883) concept of eugenics would ultimately prove to be a terrible, tragic, and malignant idea.

The study of history can yield some guidance in clearing the way going forward. Sage advice may be found in Edward L. Thorndike's constant admonition from the early history of psychology, "Look to the evidence!" (quoted in Goodenough, 1950, p. 301). To borrow from Leta S. Hollingworth, a pioneer in the psychology of women and exceptional children, the best solution to ongoing myths and controversies is to trust the "literature of fact" ascertained experimentally over the "literature of opinion" expressed by experts (Hollingworth, 1916). The most productive future for intelligence must be grounded in evidence-based research, not sociocultural ideology. This statement may seem self-evident to most readers, but leading figures have expressed concerns about the potential for public policy to be ideology-driven rather than data-driven where intelligence is involved (e.g., Sternberg, 2005).

Evidence that is undergirded by psychological theory will be more explanatory and predictive than research that is absent a theory. Again, this may be self-evident, but what passes for theory is often a list of factor-derived abilities and not a coherent account of mental processing that goes beyond factor structure. Recognizing this, John B. Carroll (1983) identified a number of additional forms of validation that transcend factor analysis, including establishing the nature of a factor, its psychometric characteristics, its place in a hierarchy of abilities, its developmental characteristics, its genetic and environmental determinants, the presence of any demographically based group mean differences, its susceptibility to intervention or more transient influences such as drugs or fatigue, its relationship to noncognitive variables, its ecological relevance and validity, and its implications for psychological theory as a whole. I also believe that any theory of intelligence must be meaningfully anchored in the functioning of the human brain. E. L. Thorndike (1901) was among the first psychologists to specify that any theory of intelligence must offer "some sort of brain correlate for ideational life and reasoning" (p. 61). Few contemporary theories of intelligence meet Thorndike's criterion.

Although there are certainly benefits to be obtained from bottom-up, data-driven tests, some caution needs to be exerted with regard to intelligence tests developed *sans* theories. Wasserman and Kaufman (2016) recently reported that intelligence tests without substantive associated theories (e.g., the Wechsler scales, which borrow from many theories but subscribe to none) may be more resilient to scholarly criticism because they cannot be tested and falsified. They stay popular for longer periods of time because it is harder to conduct research to challenge them. Some scientific courage is required to create a new intelligence measure with a testable (and therefore disprovable) theory, if the absence of a theory promises that the test will have a longer shelf life.

As readers of this book know, the practice of intelligence assessment is an *applied* science, used for decision making in clinical, educational, and vocational settings. From the introduction of Binet and Simon's (1908/1916b, 1905/1916c; Binet, 1911/1916) scales, intelligence tests have been used to guide placements for exceptional students, although their remarkable diagnostic and predictive capacities have shown them to be of value for many other kinds of decisions. Test developers would be well advised to attend to the "ultimate usefulness" of scores from their measures, following Carroll's (1993, p. 693) caution. Intelligence test scores that are most meaningful relative to the type of decisions being made need to be given the greatest emphasis in reporting. Not every factor or score from an intelligence test is clinically or educationally meaningful, and research is continually needed to discern which scores are most valid for predictive and decision-making purposes.

It remains to note a few contemporary exemplar test developers who have boldly tried to build bridges among these many roads forward. Alan S. Kaufman—coauthor with his wife, Nadeen L. Kaufman, of numerous mental tests—published *Intelligent Testing with the WISC-R* in 1979, becoming the leading authority on applied interpretation and utility of the Wechsler scales. While some of the score interpretive methods he advocated have been challenged (e.g., Watkins, 2000), his approaches to the clinical use of intelligence tests have become industry standards. Just 4 years after *Intelligent Testing* was published, the Kaufmans released the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983), a theory-based battery anchored in Luria's model of brain functioning. This is one of the few 20th-century intelligence tests meeting Thorndike's (1901) criterion. In the words of his students Randy W. Kamphaus and Cecil R. Reynolds (2009), Kaufman's contribution that has had the greatest long-term impact is his "joining of the two disciplines of measurement science and clinical assessment practice" (p. 148).

Another pioneer test developer bridging multiple paths forward is Richard W. Woodcock, an early advocate for what came to be known as CHC theory. Woodcock included both John L. Horn and John B. Carroll among his advisors in developing the Woodcock–Johnson—Revised Tests of Cognitive Ability (WJ-R; Woodcock & Johnson, 1989), recognizing the emergence of Gf-Gc theory. In 1990, Woodcock published the first cross-battery factor analysis, demonstrating that the inclusion of new marker subtests in factor analyses of existing cognitive test batteries may fundamentally alter a battery's factor structures in the direction of the CHC model. Woodcock was also a pioneer in the use of item response theory to build his tests, experimenting with multiple types of Rasch-derived scores in order to extract more useful applied information from test scores. Woodcock's development-referenced and criterion-referenced test scores (e.g., *W* scores, Relative Proficiency Index) represent efforts to move psychology beyond its overdependence on norm-referenced measurement (see Woodcock, 1999, for a complete discussion of these types of scores).

It is not obvious whether any contemporary thinkers will have lasting influence into the 21st century, and history seems to demonstrate that the United States is not an easy residence for intelligence testing. In *Democracy in America,* written nearly two centuries ago, French historian Alexis de Tocqueville (1835/1839) described the "great equality" (p. 42) that takes precedence over all other values and social conditions in American culture. The American people were uncomfortable with social or intellectual distinctions among people, he explained, and there was a fixed "middling standard" of expectations (p. 48) for learning and acquired knowledge because every citizen had access to the same educational opportunities. While God-given "gifts of intellect" (p. 49) might account for different capacities among individuals, he observed, American people see themselves and are seen by others "on a greater equality in point of fortune and intellect . . . than in any other country of the world, or, in any age of which history has preserved the remembrance" (p. 49).

REFERENCES

Ackerman, M. (1995). Mental testing and the expansion of educational opportunity. *History of Education Quarterly, 35,* 279–300.

Alien slackers may not escape service. (1917, April 22). *New York Times,* p. E3.

Altus, W. D. (1945). The differential validity and difficulty of subtests of the Wechsler Mental Ability Scale. *Psychological Bulletin, 42,* 238–249.

American Psychological Association (APA). (1997). Gold Medal Award for Life Achievement in Psychological Science: Raymond B. Cattell. *American Psychologist, 52,* 797–799.

Anastasi, A. (1942). Review: *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology, 55,* 608–609.

Anastasi, A. (1986). Intelligence as a quality of behavior. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 19–21). Norwood, NJ: Ablex.

Arden, R., Gottfredson, L. S., Miller, G., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence, 37,* 277–282.

Bain, A. (1868). *Mental science: A compendium of psychology, and the history of philosophy.* New York: Appleton.

Baker, N. D. (1918). *Annual report of the Secretary of War, 1918* (Vol. 1, pp. 5–65). Washington, DC: U.S. Government Printing Office.

Baron, J. (1986). Capacities, dispositions, and rational thinking. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 29–33). Norwood, NJ: Ablex.

A Bellevue Unit formally opened. (1933, November 3). *New York Times,* p. 20.

Berry, J. W. (1986). A cross-cultural view of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 35–38). Norwood, NJ: Ablex.

Binet, A. (1890). The perception of lengths and numbers in some small children. *Revue Philosophique, 30,* 68–81.

Binet, A. (1894). *Introduction à la psychologie expérimentale.* Paris: Baillière et Cie.

Binet, A. (1903). *L'étude expérimentale de l'intelligence.* Paris: Schleicher Frères.

Binet, A. (1910). Avant-propos: Le bilan de la psychologie en 1910. *L'Année Psychologique, 17,* v–xi.

Binet, A. (1916). New investigations upon the measure of the intellectual level among school children. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 274–329). Baltimore: Williams & Wilkins. (Original work published 1911)

Binet, A. (1975). *Modern ideas about children* (S. Heisler, Trans.). Menlo Park, CA: Suzanne Heisler. (Original work published 1909)

Binet, A., & Henri, V. (1894). Le développement de la mémoire visuelle chez les enfants. *Revue Générale des Sciences Pures et Appliquées, 5,* 162–169.

Binet, A., & Henri, V. (1895). La psychologie individuelle. *L'Année Psychologique, 2,* 411–465.

Binet, A., & Simon, T. (1914). *Mentally defective children* (W. B. Drummond, Trans.). New York: Longmans, Green. (Original work published 1907)

Binet, A., & Simon, T. (1916a). Application of the new methods to the diagnosis of the intellectual level among normal and subnormal children in institutions and in the primary schools. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 91–181). Baltimore: Williams & Wilkins. (Original work published 1905)

Binet, A., & Simon, T. (1916b). The development of intelligence in the child. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 182–273). Baltimore: Williams & Wilkins. (Original work published 1908)

Binet, A., & Simon, T. (1916c). New methods for the diagnosis of the intellectual level of subnormals. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 37–90). Baltimore: Williams & Wilkins. (Original work published 1905)

Bingham, W. V. (1942). The Army Personnel Classification System. *Annals of the American Academy of Political and Social Science, 220,* 18–28.

Bingham, W. V. (1944). Personnel classification testing in the Army. *Science (New Series), 100,* 275–280.

Bittner, R. H. (1947). The Army General Classification Test. In G. A. Kelly (Ed.), *New methods in applied psychology* (pp. 45–55). College Park: University of Maryland.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 383–405.

Bolton, T. L. (1892). The growth of memory in school children. *American Journal of Psychology, 4,* 362–380.

Boring, E. G. (1923, June 6). Intelligence as the tests test it. *The New Republic, 35,* 35–37.

Boring, E. G. (1961). *Psychologist at large: An autobiography and selected essays.* New York: Basic Books.

Brigham, C. C. (1923). *A study of American intelligence.* Princeton, NJ: Princeton University Press.

Brigham, C. C. (1930). Intelligence tests of immigrant groups. *Psychological Review, 37,* 158–165.

Brody, N. (1994). Cognitive abilities. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by J. B. Carroll. *Psychological Science, 5,* 63, 65–68.

Buck v. Bell, 274 U.S. 200 (1927).

Buckhalt, J. A. (1991). Reaction time and intelligence: Contemporary research supports Cattell's abandoned hypothesis. *School Psychology International, 12,* 355–360.

Burnham, J. C. (2003). Interviewing as a tool of the trade: A not-very-satisfactory bottom line. In D. B. Baker (Ed.), *Thick description and fine texture: Studies in the history of psychology* (pp. 19–37). Akron, OH: University of Akron Press.

Burns, R. B. (1994). Surveying the cognitive terrain. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by John B. Carroll. *Educational Researcher, 23,* 35–37.

Buros, O. K. (1977). Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Research,* 6(7), 9–15.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000).

Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31,* 141–154.

Camfield, T. M. (1969). *Psychologists at war: The history of American psychology and the First World War.* Unpublished doctoral dissertation, University of Texas at Austin. (AAT No. 7010766)

Camfield, T. M. (1992). The American Psychological Association and World War I: 1914 to 1919. In R. B. Evans, V. S. Sexton, & T. C. Cadwallader (Eds.), *The American Psychological Association: A historical perspective* (pp. 91–118). Washington, DC: American Psychological Association.

Carr, H. W. (1910). Instinct and intelligence. *British Journal of Psychology, 3,* 230–236.

Carroll, J. B. (1978). On the theory-practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), *Impact of research on education: Some case studies* (pp. 1–105). Washington, DC: National Academy of Education.

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 1–33). New York: Academic Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1994). Cognitive abilities: Constructing a theory from data. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 43–63). Norwood, NJ: Ablex.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Boston: Pergamon Press.

Carson, J. (1993). Army Alpha, Army brass, and the search for Army intelligence. *Isis, 84,* 278–309.

Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and American republics, 1750–1940.* Princeton, NJ: Princeton University Press.

Carter, M. H. (1909). The conservation of the defective child. *McClure's Magazine, 33*(2), 160–171.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15,* 373–381.

Cattell, J. M. (1924). The interpretation of intelligence tests. *Scientific Monthly, 18,* 508–516.

Cattell, J. M. (1930). Psychology in America. *Scientific Monthly, 30,* 114–126.

Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review, 3,* 618–648.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing [Abstract]. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Cattell, R. B. (Ed.). (1987). *Intelligence: Its structure, growth and action.* New York: Elsevier Science.

Cattell, R. B. (1998). Where is intelligence?: Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 29–38). Mahwah, NJ: Erlbaum.

Chapman, J. C. (1921). *Trade tests: The scientific measurement of trade proficiency.* New York: Holt.

Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890–1930.* New York: New York University Press.

Chattin, S. H., & Bracken, B. A. (1989). School psychologists' evaluation of the K-ABC, McCarthy scales, Stanford–Binet IV, and WISC-R. *Journal of Psychoeducational Assessment, 7,* 112–130.

Cody, M. S., & Prieto, L. R. (2000). Teaching intelligence testing in APA-accredited programs: A national survey. *Teaching of Psychology, 27,* 190–194.

Combe, G. (1830). *A system of phrenology* (3rd ed.). Edinburgh, UK: Anderson.

Combe, G. (1853). *A system of phrenology* (5th ed.). Edinburgh, UK: Maclachlan & Stewart.

Committee on Classification of Personnel in the Army. (1919a). *The personnel system of the United States Army: Vol. 1. History of the personnel system.* Washington, DC: U.S. Government Printing Office.

Committee on Classification of Personnel in the Army. (1919b). *The personnel system of the United States Army: Vol. 2. The personnel manual.* Washington, DC: U.S. Government Printing Office.

Cornebise, A. E. (1997). *Soldier-scholars: Higher education in the AEF, 1917–1919.* Philadelphia: American Philosophical Society.

Crime clinics growing. (1922, June 18). *New York Times,* p. 75.

Cronbach, L. J. (1949). *Essentials of psychological testing.* New York: Harper & Brothers.

Cureton, E. (1941). Reviewed work(s): *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology, 54,* 154.

Dahlstrom, W. G. (1985). The development of psychological testing. In G. A. Kimble & K. Schlesinger (Eds.), *Topics in the history of psychology* (Vol. 2, pp. 63–113). Hillsdale, NJ: Erlbaum.

Dallenbach, K. M. (1915). The history and derivation of the word "function" as a systematic term in psychology. *American Journal of Psychology, 26,* 473–484.

Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London: Murray.

Darwin, C. R. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London: Murray.

Das, J. P. (1986). On definition of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 55–56). Norwood, NJ: Ablex.

de Tocqueville, A. (1839). *Democracy in America* (3rd ed., Vol. 1; H. Reeve, Trans.). New York: George Adlard. (Original work published 1835)

Deary, I. J. (1994). Sensory discrimination and intelligence: Postmortem or resurrection? *American Journal of Psychology, 107,* 95–115.

Deary, I. J., Bell, P. J., Bell, A. J., Campbell, M. L., & Fazal, N. D. (2004). Sensory discrimination and intelligence: Testing Spearman's other hypothesis. *American Journal of Psychology, 117,* 1–18.

Deary, I. J., Lawn, M., & Bartholomew, D. J. (2008). A conversation between Charles Spearman, Godfrey Thomson, and Edward L. Thorndike: The International Examinations Inquiry Meetings 1931–1938. *History of Psychology, 11,* 122–142.

Deaths reported Aug. 23. (1903, August 24). *New York Times,* p. 7.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 57–61). Norwood, NJ: Ablex.

Dewey, J. (1904). The philosophical work of Herbert Spencer. *Philosophical Review, 13,* 159–175.

Dewsbury, D. A. (1996). Robert M. Yerkes: A psychobiologist with a plan. In G. A. Kimble, C. A. Boneau, & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 2, pp. 87–105). Washington, DC: American Psychological Association.

Diamond, S. (1977). Francis Galton and American psychology. *Annals of the New York Academy of Sciences, 291,* 47–55.

Dickson, V. E. (1919). Report of the Department of Research. In *Superintendent's annual report, Oakland Public Schools, 1917–1918* (pp. 174–247). Oakland, CA: Tribune.

Dickson, V. E. (1923). *Mental tests and the classroom teacher.* Yonkers-on-Hudson, NY: World Book.

DuBois, P. H. (1970). Varieties of psychological test homogeneity. *American Psychologist, 25,* 532–536.

Ebbinghaus, H. (1908). *Psychology: An elementary textbook* (M. Meyer, Trans.). Boston: Heath.

Education for American soldiers in France. (1919). *Scientific Monthly, 8,* 475–477.

Edwards, A. E. (Ed.). (1974). *Selected papers of David Wechsler.* New York: Academic Press.

Fancher, R. E. (1998). Alfred Binet, general psychologist. In G. A. Kimble & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 3, pp. 67–83). Washington, DC: American Psychological Association.

Ferguson, L. W. (1962). Walter Dill Scott: First industrial psychologist. *Heritage of Industrial Psychology, 1,* 1–10.

Ferguson, L. W. (1963a). Bureau of Salesmanship Research: Walter Dill Scott, Director. *Heritage of Industrial Psychology, 5,* 53–66.

Ferguson, L. W. (1963b). Psychology and the Army: Classification of personnel (1). *Heritage of Industrial Psychology, 10,* 141–167.

Ferguson, L. W. (1963c). Psychology and the Army: Introduction of the rating scale. *Heritage of Industrial Psychology, 9,* 125–139.

Ferguson, L. W. (1964a). Division of applied psychology: Carnegie Institute of Technology. *Heritage of Industrial Psychology, 3,* 25–34.

Ferguson, L. W. (1964b). Psychology and the Army: Classification of personnel (2). *Heritage of Industrial Psychology, 11,* 169–205.

Ferguson, L. W. (1964c). Psychology and the Army: Classification of personnel (3). *Heritage of Industrial Psychology, 12,* 209–246.

Fischer, D. H. (1970). *Historians' fallacies: Toward a logic of historical thought.* New York: Harper & Row.

Frank, G. (1983). *The Wechsler enterprise: An assessment of the development, structure, and use of the Wechsler tests of intelligence.* New York: Pergamon Press.

Franz, S. I. (1912). New phrenology. *Science (New Series), 35,* 321–328.

Friedlander, H. (1995). *The origins of Nazi genocide: From euthanasia to the final solution.* Chapel Hill: University of North Carolina Press.

Fuster, J. M. (2008). *The prefrontal cortex* (4th ed.). Amsterdam: Elsevier.

Gall, F. J. (1857). Letter to Joseph Fr[eiherr] von Retzer, upon the functions of the brain, in man and animals. In D. G. Goyder (Ed., & Trans.), *My battle for life: The autobiography of a phrenologist* (pp. 143–152). London: Simpkin, Marshall. (Original work published 1798)

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* London: Macmillan.

Galton, F. (1883). *Inquiries into human faculty and its development.* London: Macmillan.

Galton, F. (1884, June 5). Notes. "Rede" Lecture: The measurement of human faculty. *Nature, 30,* 129–131.

Galton, F. (1890a). Remarks [after J. M. Cattell's "Mental tests and measurements"]. *Mind, 15,* 380–381.

Galton, F. (1890b). Why do we measure mankind? *Lippincott's Monthly Magazine, 45,* 236–241.

Galton, F. (1891). Useful anthropometry. In *Proceedings of the American Association for the Advancement of Physical Education at its Sixth Annual Meeting, April 3 and 4, 1891* (pp. 51–58). Ithaca, NY: Andrus & Church.

Galton, F. (1892). Retrospect of work done at my anthropometric laboratory at South Kensington. *Journal of the Anthropological Institute, 21,* 32–35.

Galton, F. (1907). *Inquiries into human faculty and its development* (2nd ed.). London: Dent.

Galton, F. (1930). Kantsaywhere. In K. Pearson (Ed.), *The life, letters and labours of Francis Galton* (Vol. 3A, pp. 414–424). London: Cambridge University Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gay, C. J. (1948). The blind look at an elephant: 10 experts, 25 "people" define "intelligence." *The Clearing House, 22,* 263–266.

Gilbert, J. A. (1894). Researches on the mental and physical development of school children. In E. W. Scripture (Ed.), *Studies from the Yale Psychological Laboratory* (Vol. 2, pp. 40–100). New Haven, CT: Yale University.

Goddard, H. H. (1916). Editor's introduction. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (pp. 5–8). Baltimore: Williams & Wilkins.

Goodenough, F. L. (1950). Edward Lee Thorndike: 1874–1949. *American Journal of Psychology, 63*(2), 291–301.

Gottfredson, L. S. (1994, December 13). Mainstream science on intelligence. *Wall Street Journal,* p. A18.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13–23.

Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove, UK: Psychology Press.

Gottfredson, L. S. (2010). Lessons in academic freedom as lived experience. *Personality and Individual Differences, 49,* 272–280.

Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundation and issues in assessment. *Canadian Psychology, 50,* 183–195.

Gould, S. J. (1996). *The mismeasure of man* (rev. ed.). New York: Norton.

Haggerty, M. E. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 212–216.

Heaton, H. [Artist]. (1922). The pupil becomes an individual [Illustration]. *American School Board Journal, 64*(4), cover page.

Henmon, V. A. C. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 195–198.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hildreth, G. H. (1939). *A bibliography of mental tests and rating scales* (2nd ed.). New York: Psychological Corporation.

Hines, H. C. (1922). Measuring the intelligence of school pupils. *American School Board Journal, 64*(4), 35–37, 135.

Hollingworth, L. S. (1916). The vocational aptitudes of women. In H. L. Hollingworth (Ed.), *Vocational psychology: Its problems and methods* (pp. 222–244). New York: Appleton.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign. (AAT 6507113)

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. *Annual Review of Psychology, 27,* 437–485.

Horn, J. L. (1986). Some thoughts about intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 91–96). Norwood, NJ: Ablex.

Horn, J. L. (2001). Raymond Bernard Cattell (1905–1998). *American Psychologist, 56,* 71–72.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253–270.

Horn, J. L., & Noll, J. (1994). A system for understanding cognitive capabilities: A theory and the evidence on which it is based. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 151–203). Norwood, NJ: Ablex.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hunt, E. (1986). The Heffalump of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 101–107). Norwood, NJ: Ablex.

Israel Wechsler, neurologist, dies. (1962, December 7). *New York Times,* p. 39.

Jastrow, J. (1893). The section of psychology. In M. P. Hardy (Ed.), *Official catalogue—World's Columbian Exposition* (Part 7, pp. 50–60). Chicago: Conkey.

Jastrow, J. (1901). Some currents and undercurrents in psychology. *Psychological Review, 8,* 1–26.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R. (2004). Obituary: John Bissell Carroll. *Intelligence, 32,* 1–5.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier.

Kamphaus, R. W., & Reynolds, C. R. (2009). Kaufman's work in the penumbra between measurement science and clinical assessment. In J. C. Kaufman (Ed.), *Intelligent testing: Integrating psychological theory and clinical practice* (pp. 148–156). New York: Cambridge University Press.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children (K-ABC).* Circle Pines, MN: American Guidance Service.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650.

Kevles, D. J. (1968). Testing the Army's intelligence:

Psychologists and the military in World War I. *Journal of American History, 55,* 565–581.

Kissman, J. (1948). The immigration of Romanian Jews up to 1914. In N. Shlomo (Ed.), *YIVO annual of Jewish social science* (Vols. 2–3, pp. 160–179). New York: Yiddish Scientific Institute.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.). (2008). *Extending intelligence: Enhancement and new constructs.* New York: Erlbaum.

Langfeld, H. S. (1924). The Seventh International Congress of Psychology. *American Journal of Psychology, 35,* 148–153.

Lewes, G. H. (1867). *The history of philosophy: From Thales to Comte* (Vol. 2, 3rd ed.). London: Longmans, Green.

Lippmann, W. (1922a). The abuse of the tests. *The New Republic, 32,* 297–298.

Lippmann, W. (1922b). A future for the tests. *The New Republic, 33,* 9–10.

Lippmann, W. (1922c). The mental age of Americans. *The New Republic, 32,* 213–215.

Lippmann, W. (1922d). The mystery of the "A" men. *The New Republic, 32,* 246–248.

Lippmann, W. (1922e). The reliability of intelligence tests. *The New Republic, 32,* 275–277.

Lippmann, W. (1922f). Tests of hereditary intelligence. *The New Republic, 32,* 328–330.

Lippmann, W. (1923). The great confusion: A reply to Mr. Terman. *The New Republic, 33,* 145–146.

Littman, R. A. (2004). Mental tests and fossils. *Journal of the History of the Behavioral Sciences, 40,* 423–431.

Lombardo, P. A. (2008). *Three generations, no imbeciles: Eugenics, the Supreme Court, and Buck v. Bell.* Baltimore: Johns Hopkins University Press.

Lorge, I. (1943). Review of *The measurement of adult intelligence. Journal of Consulting Psychology, 7,* 167–168.

Louttit, C. M., & Browne, C. G. (1947). The use of psychometric instruments in psychological clinics. *Journal of Consulting Psychology, 11,* 49–54.

Lubin, B., Wallis, R. R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935–1969. *Professional Psychology, 2,* 70–74.

Lynch, E. C. (1968). Walter Dill Scott: Pioneer industrial psychologist. *Business History Review, 42,* 149–170.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford University Press.

Matarazzo, J. D. (1981). Obituary: David Wechsler (1896–1981). *American Psychologist, 36,* 1542–1543.

May, M. A. (1920). The psychological examination of conscientious objectors. *American Journal of Psychology, 31,* 152–165.

McArdle, J. J. (2007). Obituary: John Leonard Horn 1928–2006. *Intelligence, 35,* 517–518.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10.

McNemar, Q. (1945). Reviewed work(s): *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology, 58,* 420–422.

Minton, H. L. (1988). *Lewis M. Terman: Pioneer in psychological testing.* New York: New York University Press.

Morgan, C. L. (1910). Instinct and intelligence. *British Journal of Psychology, 3,* 219–229.

Myers, C. S. (1910). Instinct and intelligence. *British Journal of Psychology, 3,* 209–218.

Napoli, D. S. (1981). *Architects of adjustment; The history of the psychological profession in the United States.* Port Washington, NY: Kennikat Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell–Horn–Carroll-based assessment. *Psychology in the Schools, 47,* 621–634.

Nicolas, S., & Ferrand, L. (2002). Alfred Binet and higher education. *History of Psychology, 5,* 264–283.

Oehrn, A. (1896). Experimentelle Studien zur Individualpsychologie. *Psychologische Arbeiten, 1,* 92–151.

Old Bellevue and new. (1930, June 20). *New York Times,* p. 19.

Otis, A. S. (1917). A criticism of the Yerkes–Bridges Point Scale, with alternative suggestions. *Journal of Educational Psychology, 8,* 129–150.

Pass alien slacker bill. (1918, February 28). *New York Times,* p. 5.

Payne, A. F. (1920). The scientific selection of men. *Scientific Monthly, 11,* 544–547.

Pearson, K. (Ed.). (1924). *The life, letters and labours of Francis Galton: Vol. 2. Researchers of middle life.* London: Cambridge University Press.

Pearson, K. (Ed.). (1930). *The life, letters and labours of Francis Galton: Vol. 3A. Correlation, personal identification and eugenics.* London: Cambridge University Press.

Pellegrino, J. W. (1986). Intelligence: The interaction of culture and cognitive processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 113–116). Norwood, NJ: Ablex.

Peterson, J. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 198–201.

Pintner, R. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 139–143.

Prendergast, P. (1844). A "page" of phrenology. *The Illuminated Magazine, 11,* 17–20.

Rabin, A. I. (1945). The use of the Wechsler–Bellevue scales with normal and abnormal persons. *Psychological Bulletin, 42,* 410–422.

Rabin, A. I., & Guertin, W. H. (1951). Research with the Wechsler–Bellevue test: 1945–1950. *Psychological Bulletin, 48,* 211–248.

Rapaport, D. (1945). The new Army Individual Test of general mental ability. *Bulletin of the Menninger Clinic, 9,* 107–110.

Ready, R. E., & Veague, H. B. (2014). Training in psychological assessment: Current practices of clinical psychology programs. *Professional Psychology: Research and Practice, 45,* 278–282.

Reeve, C. L., & Charles, J. E. (2008). Survey of opinions on the primacy of *g* and social consequences of ability testing: A comparison of expert and non-expert views. *Intelligence, 36,* 681–688.

Ritchie, S. J. (2017). Review of *The rationality quotient: Toward a test of rational thinking. Intelligence, 61,* 46.

Robinson, N. M. (2003). Two wrongs do not make a right: Sacrificing the needs of gifted students does not solve society's unsolved problems. *Journal for the Education of the Gifted, 26*(4), 251–273.

Rock, G. (1956). *The history of the American Field Service, 1920–1955.* New York: American Field Service.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Routh, D. K. (2000). Clinical psychology training: A history of ideas and practices prior to 1946. *American Psychologist, 55,* 236–241.

Samelson, F. (1977). World War I intelligence testing and the development of psychology. *Journal of the History of the Behavioral Sciences, 13,* 274–282.

Scarr, S. (1986). Intelligence: Revisited. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 117–120). Norwood, NJ: Ablex.

Schneider, W. H. (1992). After Binet: French intelligence testing, 1900–1950. *Journal of the History of the Behavioral Sciences, 28,* 111–132.

Scott, J. B. (1918). The amendment of the Naturalization and Citizenship Acts with respect to military service. *American Journal of International Law, 12,* 613–619.

Scott, W. D. (1916). Selection of employees by means of quantitative determinations. *Annals of the American Academy of Political and Social Science, 65,* 182–193.

Sea Fox wreckage. (1931, May 28). *New York Times,* p. 1.

Seagoe, M. V. (1975). *Terman and the gifted.* Los Altos, CA: Kaufmann.

Seidenfeld, M. A. (1942). The Adjutant General's School and the training of psychological personnel for the Army. *Psychological Bulletin, 39,* 381–384.

Sharp, S. E. (1899). Individual psychology: A study in psychological method. *American Journal of Psychology, 10,* 329–391.

Siegler, R. S. (1992). The other Alfred Binet. *Developmental Psychology, 28,* 179–190.

Simpson, D. (2005). Phrenology and the neurosciences: Contributions of F. J. Gall and J. G. Spurzheim. *Australian and New Zealand Journal of Surgery, 75,* 475–482.

Sisson, E. D. (1948). The Personnel Research Program of the Adjutant General's Office of the United States Army. *Review of Educational Research, 18,* 575–614.

Smith, M. L. (2000). By way of Canada: U.S. records of Immigration across the U.S.–Canadian border, 1895–1954 (St. Albans Lists). *Prologue, 32,* 192–199.

Snow, R. E. (1986). On intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 133–139). Norwood, NJ: Ablex.

Sokal, M. M. (Ed.). (1981). *An education in psychology: James McKeen Cattell's journal and letters from Germany and England, 1880–1888.* Cambridge, MA: MIT Press.

Sokal, M. M. (Ed.). (1987). *Psychological testing and American society, 1890–1930.* New Brunswick, NJ: Rutgers University Press.

Sokal, M. M. (2006). James McKeen Cattell: Achievement and alienation. In D. A. Dewsbury, L. T. Benjamin, & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 6, pp. 18–35). Mahwah, NJ: Erlbaum.

Spanier, M., & Spanier, D. (1902, September 17). Canadian passenger manifest for the S. S. *Lake Megantic* (sailed from Liverpool 9 September, arrived 17). *Québec Ports, Immigrant Passenger Lists, 1865–1935 (RG 76): Québec 1865–1921 (includes Montreal). Microfilm Reel T-481 (1902 25 Aug to 1903 July 28).* Ottawa: Library and Archives Canada.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Spearman, C. (1923). *The nature of "intelligence" and principles of cognition.* London: Macmillan.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Spearman, C., & Wynn Jones, L. (1950). *Human ability: A continuation of "The Abilities of Man."* London: Macmillan.

Spencer, H. (1855). *The principles of psychology.* London: Longman, Brown, Green, & Longmans.

Spencer, H. (1864). *The principles of biology* (Vol. 1). London: Williams & Norgate.

Spring, J. H. (1972). Psychologists and the war: The meaning of intelligence in the Alpha and Beta tests. *History of Education Quarterly, 12,* 3–15.

Spurzheim, J. G. (1815). *The physiognomical system of Drs. Gall and Spurzheim; founded on an anatomical and physiological examination of the nervous system in general, and of the brain in particular; and indicating the dispositions and manifestations of the mind* (2nd ed.). London: Baldwin, Cradock, & Joy.

Staff, Personnel Research Section. (1942, March). *Analysis of Wechsler Self-Administering Test data* (PRS Report No. 286). Washington, DC: U.S. War Department, Adjutant General's Office.

Staff, Personnel Research Section. (1944). The new Army Individual Test of general mental ability. *Psychological Bulletin, 41,* 532–538.

Staff, Personnel Research Section. (1945). The Army General Classification Test. *Psychological Bulletin, 42,* 760–768.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking.* Cambridge, MA: MIT Press.

Stein, M. (2004). The establishment of the Department of Psychiatry in the Mount Sinai Hospital: A conflict between neurology and psychiatry. *Journal of the History of the Behavioral Sciences, 40,* 285–309.

Stern, W. (1914). *The psychological methods of testing intelligence* (Educational Psychology Monographs No. 13; G. M. Whipple, Trans.). Baltimore: Warwick & York. (Original work published 1912)

Sternberg, R. J. (1986). Intelligence is mental self-government. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 141–148). Norwood, NJ: Ablex.

Sternberg, R. J. (1994). 468 factor-analyzed data sets: What they tell us and don't tell us about human intelligence. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by J. B. Carroll. *Psychological Science, 5,* 63–65.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11*(2), 295–301.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of Intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 155–162). Norwood, NJ: Ablex.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence?: Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Strong, E. K. (1955). Walter Dill Scott: 1869–1955. *American Journal of Psychology, 68,* 682–683.

Terman, L. M. (1904). A preliminary study in the psychology and pedagogy of leadership. *Pedagogical Seminary, 11,* 413–451.

Terman, L. M. (1906). Genius and stupidity: A study of some of the intellectual processes of seven "bright" and seven "stupid" boys. *Pedagogical Seminary, 13,* 307–373.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet–Simon Intelligence Scale.* Boston: Houghton Mifflin.

Terman, L. M. (1917). Feeble minded children in the schools. In F. C. Nelles (Chair), *Report of 1915 legislature committee on mental deficiency and the proposed institution for the care of feeble-minded and epileptic persons* (pp. 45–52). Whittier, CA: Whittier State School, Department of Printing Instruction.

Terman, L. M. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 127–133.

Terman, L. M. (1922a). The great conspiracy, or the impulse imperious of intelligence testers, psychoanalyzed and exposed by Mr. Lippmann. *The New Republic, 33,* 116–120.

Terman, L. M. (1922b). Were we born that way? *The World's Work, 44,* 655–660.

Terman, L. M. (1932). Trails to psychology. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 2, pp. 297–331). Worcester, MA: Clark University Press.

Terman, L. M., & Childs, H. G. (1912a). A tentative revision and extension of the Binet–Simon measuring scale of intelligence. *Journal of Educational Psychology, 3,* 61–74.

Terman, L. M., & Childs, H. G. (1912b). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part II. Supplementary tests. 1. Generalization test: interpretation of fables. *Journal of Educational Psychology, 3,* 133–143.

Terman, L. M., & Childs, H. G. (1912c). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part II. Supplementary tests—continued. *Journal of Educational Psychology, 3,* 198–208.

Terman, L. M., & Childs, H. G. (1912d). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part III. Summary and criticisms. *Journal of Educational Psychology, 3,* 277–289.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new Revised Stanford–Binet tests of intelligence.* Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford–Binet Intelligence Scale: Manual for the third revision, Form L-M.* Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1973). *Stanford–Binet Intelligence Scale: Manual for the third revision, Form L-M (1972 table of norms by R. L. Thorndike).* Boston: Houghton Mifflin.

Thorndike, E. L. (1901). The evolution of the human intellect. *Popular Science Monthly, 60*(5), 58–65.

Thorndike, E. L. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 124–127.

Thorndike, E. L., Lay, W., & Dean, P. R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology, 20,* 364–369.

Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4,* 3–7.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Thorndike, R. M., & Lohman, D. F. (1990). *A century of ability testing.* Chicago: Riverside.

Thorne, F. C. (1976). Reflections on the golden age of Columbia psychology. *Journal of the History of the Behavioral Sciences, 12,* 159–165.

Thurstone, L. L. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 201–207.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Thurstone, L. L. (1945). Testing intelligence and aptitudes. *Hygeia, 23,* 32–36, 50, 52, 54.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of mind.* Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs, 2.*

Toops, H. A. (1921). *Trade tests in education.* New York: Teachers College, Columbia University.

Tucker, W. H. (2009). *The Cattell controversy: Race, science, and ideology.* Urbana: University of Illinois Press.

Tuddenham, R. D. (1974). Review: Fame and oblivion. *Science (New Series), 183,* 1071–1072.

206 get degrees at City College. (1916, June 23). *New York Times,* p. 7.

U.S. maintains great schools on foreign soil: Government trains soldiers for civilian life—thousands are enrolled. (1919, June 12). *Lima Daily News* [Lima, OH], p. 5.

Venezky, R. L. (2002). The history of reading research. In P. D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 3–38). Mahwah, NJ: Erlbaum.

Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). *g* and the measurement of multiple intelligences: A response to Gardner. *Intelligence, 34,* 507–510.

von Mayrhauser, R. T. (1986). *The triumph of utility: The forgotten clash of American psychologies in World War I.* Unpublished doctoral dissertation, University of Chicago. (AAT T-29951)

von Mayrhauser, R. T. (1987). The manager, the medic, and the mediator: The clash of professional psychological styles and the wartime origins of group mental testing. In M. M. Sokal (Ed.), *Psychological testing and American society, 1890–1930* (pp. 128–157). New Brunswick, NJ: Rutgers University Press.

von Mayrhauser, R. T. (1989). Making intelligence functional: Walter Dill Scott and applied psychological testing in World War I. *Journal of the History of the Behavioral Sciences, 25,* 60–72.

Walsh, A. A. (1976). The "new science of the mind" and the Philadelphia physicians in the early 1800s. *Transactions and Studies of the College of Physicians of Philadelphia, 43,* 397–413.

Wasserman, J. D., & Kaufman, A. S. (2015). David Wechsler (1896–1981). In R. Cautin & S. Lilienfeld (Eds.), *Encyclopedia of clinical psychology* (pp. 2942–2945). Hoboken, NJ: Wiley-Blackwell.

Wasserman, J. D., & Kaufman, A. S. (2016). A history of mental ability tests and theories. In W. B. Barr & L. Bieliauskas (Eds.), *The Oxford handbook of history of clinical neuropsychology [Oxford handbooks online].* New York: Oxford University Press. Available at *www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199765683.001.0001/oxfordhb-9780199765683.*

Wasserman, J. D., & Maccubbin, E. M. (2002, August). *David Wechsler at The Psychological Corporation, 1925–1927: Chorus girls and taxi drivers.* Paper presented at the annual meeting of the American Psychological Association, Chicago, IL.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*(4), 465–479.

Wechsler, D. (1917a). A study of retention in Korsakoff psychosis. *Psychiatric Bulletin, 2,* 403–451.

Wechsler, D. (1917b, June 5). *World War I Selective Service System Draft Registration Cards, 1917–1918* (Administration. M1509, Roll: 1765679, Draft Board: 96). Washington, DC: National Archives and Records.

Wechsler, D. (1918a, October 4). Passport Application No. 124984. In *Passport Applications, 1906–March 31, 1925.* Washington, DC: National Archives and Records Administration.

Wechsler, D. (1918b, June 7). *Petition for naturalization, David Wechsler. United States District Court of the Northern District of Georgia, Rome Division.* Morrow, GA: NARA Southeast Region.

Wechsler, D. (1921). Report of amendment of passport, August 18, 1921. In *Passport Applications, 1906–March 31, 1925.* Washington, DC: National Archives and Records Administration.

Wechsler, D. (1922). Quelques remarques sur la psychopathologie de l'indecision [Some remarks on the psychopathology of indecision]. *Journal de Psychologie, 19,* 47–54.

Wechsler, D. (1925). The measurement of emotional reactions. *Archives of Psychology, 76,* 1–181.

Wechsler, D. (1926). Tests for taxicab drivers. *Journal of Personnel Research, 5,* 24–30.

Wechsler, D. (1927). Psychometric tests. In I. S. Wechsler, *Textbook of clinical neurology* (pp. 104–116). Philadelphia: Saunders.

Wechsler, D. (1932). Analytic use of the Army Alpha examination. *Journal of Applied Psychology, 16,* 254–256.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1944). The psychologist in the psychiatric hospital. *Journal of Consulting Psychology, 8,* 281–285.

Wechsler, D. (1946). *Wechsler–Bellevue Intelligence Scale, Form II. Manual for administering and scoring the test.* New York: Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children (WISC).* New York: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale—Restandardized (WAIS).* New York: Psychological Corporation.

Wechsler, D. (1961). Conformity and the idea of being well born. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation* (pp. 412–423). New York: Harper & Row.

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI)*. New York: Psychological Corporation.

Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology, 6,* 82–85.

Wechsler, D., Doppelt, J. E., & Lennon, R. T. (1975). *A conversation with David Wechsler* (Unpublished transcript of interview). San Antonio, TX: Psychological Corporation.

Wechsler, I. (1957). A neurologist's point of view. *Medical Clinics of North America, 41,* 1111–1121.

Wechsler, M. (1903, August 23). *Certificate and record of death, Certificate No. 24787. City of New York, Department of Health, State of New York.* New York: New York City Department of Records and Information Services Municipal Archives.

Weider, A. (2006). *A Weider Weltanschauung: A wider perspective of psychology.* Bloomington, IN: Xlibris.

Wexler, L. (1906, October 31). *Certificate and record of death, Certificate No. 33534. City of New York, Department of Health, State of New York.* New York: New York City Department of Records and Information Services Municipal Archives.

Whipple, G. M. (1915). *Manual of mental and physical tests: Part 2. Complex processes* (2nd ed.). Baltimore: Warwick & York.

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25,* 9–23.

Wilson, W. (1917). Address of the President of the United States delivered at a joint session of the two houses of Congress, April 2, 1917. In *65th Congress, 1st Session, Senate* (Document No. 5, Serial No. 7264, pp. 3–8). Washington, DC: U.S. Government Printing Office.

Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Monographs, 3*(6), 1–62.

Wolf, T. (1961). An individual who made a difference. *American Psychologist, 16,* 245–248.

Wolf, T. (1966). Intuition and experiment: Alfred Binet's first efforts in child psychology. *Journal of the History of the Behavioral Sciences, 2,* 233–239.

Wolf, T. (1969). The emergence of Binet's conceptions and measurement of intelligence: II. A case history of the creative process. *Journal of the History of the Behavioral Sciences, 5,* 207–237.

Wolf, T. (1973). *Alfred Binet.* Chicago: University of Chicago Press.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.) *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Mahwah, NJ: Erlbaum.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.

Woodrow, H. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 207–210.

Woodworth, R. S. (1942). *The Columbia University psychological laboratory.* New York: Columbia University.

Yerkes, R. M. (1918a). Measuring the mental strength of an army. *Proceedings of the National Academy of Sciences USA, 4,* 295–297.

Yerkes, R. M. (1918b, May 3). Memorandum for Mr. Garrett. *Records of the Office of the Surgeon General (Army) (RG 112).* College Park, MD: National Archives and Records Administration.

Yerkes, R. M. (1919). Report of the Psychology Committee of the National Research Council. *Psychological Review, 26,* 83–149.

Yerkes, R. M. (1920). What psychology contributed to the war. In R. M. Yerkes (Ed.), *The new world of science: Its development during the war* (pp. 364–389). New York: Century.

Yerkes, R. M. (Ed.). (1921). *Memoirs of the National Academy of Sciences: Vol. 15. Psychological examining in the United States Army.* Washington, DC: U.S. Government Printing Office.

Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A point scale for measuring mental ability.* Baltimore: Warwick & York.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

Yockelson, M. (1998). They answered the call: Military service in the United States Army during World War I, 1917–1919. *Prologue: Quarterly of the National Archives and Records Administration, 30,* 228–234.

Zazzo, R. (1993). Alfred Binet (1857–1911). *Prospects: The Quarterly Review of Comparative Education, 23,* 101–112.

Zeidner, J., & Drucker, A. J. (1988). *Behavioral science in the army: A corporate history of the Army Research Institute.* Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing.* New York: Cambridge University Press.

CHAPTER 2

A History of Intelligence Test Interpretation

Randy W. Kamphaus
Anne Pierce Winsor
Ellen W. Rowe
Sangwon Kim

In this chapter, we focus exclusively on a historical account of dominant methods of intelligence test interpretation. Although the thoroughness of our research coverage is limited by space available, a historical overview offers important practical knowledge for the psychologist practicing today, and tomorrow. As E. G. Boring (1929) wisely observed long ago, "Without such [historical] knowledge he sees the present in distorted perspective, he mistakes old facts and old views for new, and he remains unable to evaluate the significance of new movements and methods" (p. vii). Boring's wisdom remains valuable today, even when we consider the date and context in which his work was written, as reflected by his exclusive use of the masculine pronoun. In recent decades, we have seen the emergence of "new" methods of intelligence test interpretation that are in fact merely reverberations of strategies that were considered either failed or questionable in a bygone era. Take, for example, the relatively widespread adoption of Howard Gardner's multiple-intelligences theory by schools around the United States and the world (see Chen & Gardner, Chapter 4, this volume). This potential "breakthrough" may yet be proven to be just that, but Gardner's theory is definitely a classic example of multiple-intelligence theory from the 1930s (Kamphaus, 2009), with the same attendant challenges of demonstrating valid inferences of test scores.

As would be expected, formal methods of intelligence test interpretation emerged soon after Binet and Simon's creation of the first successful intelligence scale (Kamphaus, 2001). These early interpretive methods, sometimes referred to colloquially as the "dipstick approach," attempted primarily to quantify a general level of intelligence—or, as Binet referred to it, a *developmental level*. He preferred this qualitative description to yielding a test score (Kamphaus, 2001). With the addition of subtest scores to clinical tests and the emergence of group tests measuring different abilities, *clinical profile analysis* replaced the "dipstick approach" as the dominant heuristic for intelligence test interpretation. All scores were deemed appropriate for interpretation, in the hope of providing more "insight" into an individual's cognitive strengths and weaknesses. *Psychometric profile analysis* soon followed. However, as measurement approaches to intelligence test interpretation developed, validity problems with the interpretation of the panoply of subtest scores and patterns surfaced. Today the gap between intelligence theory and test development has narrowed, and more validity evidence associated with score inferences is available; thus test interpretation is becoming both easier and more accurate. We trace all of these developments in this chapter and conclude with a look to the future.

QUANTIFICATION OF A GENERAL LEVEL: THE FIRST WAVE

The process of analyzing human abilities has intrigued scientists for centuries. Indeed, some method for analyzing individual's abilities has existed for over 2,000 years. The Chinese are credited with instituting civil service examinations and formulating a system to classify individuals according to their abilities (French & Hale, 1990). Early work in interpretation of intelligence tests focused extensively on classification of individuals into groups, often using terms no longer considered acceptable today. Early classification provided a way to organize individuals into specified groups based on scores obtained on intelligence tests—an organization that was dependent on the acceptance of intelligence tests by the public as well as by professionals. Today professionals in the fields of psychology and education benefit from the use of increasingly well-researched and objective instruments developed according to precise plans and test specifications. In order to put early intelligence test interpretation into context, the following section provides a brief description of some of the early work leading to the development of the tests themselves.

The Work of Early Investigators

At the beginning of the 20th century, practitioners in the fields of psychology and education were beginning to feel the compelling influence of Alfred Binet and his colleagues in France, most notably Théodore Simon. Binet's studies of the mental abilities of children for the purposes of school placement led to the first genuinely successful method for classifying persons with respect to their cognitive abilities (Goodenough, 1949). Binet and Simon's development of the first empirical and practical intelligence test for applied use in the classification of students represented a technological breakthrough in the field of intelligence assessment. The first version of the Binet–Simon Intelligence Scale (Binet & Simon, 1905) would lead to future scales and, according to Anastasi (1988), an overall increase in the use of intelligence tests for a variety of purposes.

Binet's efforts reflected his great interest in certain forms of cognitive activity. These included abilities related to thinking and reasoning, the development and application of strategies for complex problem solving, and the use and adaptation of abilities for success in novel experiences (Pintner, 1923). His work was rooted in an interest in measuring the complex cognitive processes of children and would eventually lead to a series of popular instruments, most recently represented in the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003).

At the same time, scientists such as James McKeen Cattell in the United States were conducting equally important work of a different kind. Cattell's investigations frequently focused on measures of perception and motor skills. Although different in scope and purpose from that of Binet and Simon, Cattell's work would ultimately have a profound effect on the popularization and use of intelligence tests (Pintner, 1923). Cattell's experimentation resulted in the appointment of a special committee whose members, with the assistance of the American Psychological Association, were charged with developing a series of mental ability tests for use in the classification and guidance of college students (Goodenough, 1949). The development of these tests placed great emphasis on the need for standardized administration and scoring procedures to improve the reliability of obtained scores.

Procedures for standardized test administration were introduced, with the idea that the measurements associated with an individual would be even more informative when compared to the measurements of another person in the same age group who was administered the same test under the same standard conditions (Pintner, 1923). Controlled and consistent conditions of test administration were considered prerequisite to advancing the goal of making interpretation of test scores more scientific (Anastasi, 1988). The earliest methods of test interpretation centered on the idea that individuals could be classified into groups according to their cognitive abilities, which would form the foundation for improved psychological diagnosis and educational classification. Of course these groups would need descriptive labels, which led in turn to the creation of classification systems as described in the next section.

Classification Schemes

The first well-documented efforts at intelligence test interpretation emphasized the assignment of a descriptive classification based on an overall intelligence test composite score. This practice seemed a reasonable first step, given that (1) the dominant scale of the day, the Stanford–Binet (Stanford Revisions and Extension of the Binet–Simon Scale [Terman, 1916] or the Revised Stanford–Binet [Terman & Merrill, 1937]), yielded only a single overall score; and (2) Spearman's (1927) general intelligence theory, the dominant theory of the day, was premised on the existence of a singular *mental energy*, or general intellectual ability.

According to Goodenough (1949), the identification of mental ability was regarded as a purely biological/physical/medical concern, given that intelligence was presumed to be genetically based and unmalleable. Wechsler (1944) made a similar statement, noting that the vocabulary of choice for a descriptive classification system included medical–legal terms such as *idiot, imbecile,* and *moron.* Levine and Marks (1928, p. 131) provided an example of a classification system incorporating terms of this ilk (see Table 2.1), which, fortunately and necessarily, have fallen into disuse and disfavor.

This classification system used descriptive terms that were evaluative and pejorative (especially when employed in the vernacular), leading to stigmatization of examinees. In addition, the many category levels contained bands of scores with different score ranges. The top and bottom three levels comprised bands of 24 score points each, while those in the middle, from *borderline* to *very bright,* comprised bands of 9 points each. Although the band comprising the *average* range was not far from our present conceptions of *average* (except for this example's upper limit), the use of numerous uneven levels was potentially confusing to the layperson.

TABLE 2.1. The Levine and Marks Intelligence Test Score Classification System

Level	Range in IQ
Idiots	0–24
Imbeciles	25–49
Morons	50–74
Borderline	75–84
Dull	85–94
Average	95–104
Bright	105–114
Very bright	115–124
Superior	125–149
Very superior	150–174
Precocious	175 or above

TABLE 2.2. Wechsler's Intelligence Classification According to IQ

Classification	IQ limits	% included
Defective	65 and below	2.2
Borderline	66–79	6.7
Dull normal	80–90	16.1
Average	91–110	50.0
Bright normal	111–119	16.1
Superior	120–127	6.7
Very superior	128 and over	2.2

Wechsler (1944) introduced another classification scheme that attempted to formulate categories according to a specific structural rationale. This system was based on a definition of intelligence levels related to statistical frequencies (i.e., percentages under the normal curve), in which each classification level was based on a range of intelligence scores lying specified distances from the mean (Wechsler, 1944). In an effort to move away from somewhat arbitrary qualities, his classification scheme incorporated estimates of the prevalence of certain intelligence levels in the United States at that time (see Table 2.2).

Components of Wechsler's system of bands of IQ limits has proved enduring. Some of the bands are close to those we use at the present time. Both the Levine and Marks (1928) and Wechsler (1944) schemes provide a glimpse at procedures used in early attempts at test interpretation. The potential for stigmatization has been lessened in the period since World War II; both scientists and practitioners have moved to a less evaluative vocabulary that incorporates parallel terminology around the mean, such as *above average* and *below average* (Kamphaus, 2001).

Considerations for Interpretation Using Classification Systems

We have made progress regarding the use of classification schemes in the evaluation of human cognitive abilities. The structure of classification systems appears to be more stable today than in the past. Previous practitioners often applied Terman's classification system, originally developed for interpretation of the Stanford–Binet, in their interpretation of many different tests that measured a variety of different abilities (Wechsler, 1944). Fortunately, many test batteries today provide their own classification schemes within the test manuals, providing an opportunity to choose among appropriate tests and classify the results accordingly. In addition, these classification systems are often based on deviations from a mean of 100, providing consistency across most intelligence tests and allowing easier comparison of an individual's performance on them.

Calculation of intelligence test scores, or IQs, became a common way of describing an individual's cognitive ability. However, test score calculation is only the first step in the interpretive process, which has been the case since the early days of testing (Goodenough, 1949). Although scores may fall neatly into classification categories, additional data should be considered when clinicians are discussing an individual's cognitive abilities. For example, individuals in the population who are identified as having below-average intellectual abilities do not necessarily manifest the same degree of disability, and in fact may demonstrate considerable variability in abilities (Goodenough, 1949). Wechsler's (1958) own view of classification schemes was that their use could remind clinicians that intelligence test scores are the results of comparisons to a normative population and not absolute quantities.

The early views of Goodenough and Wechsler have influenced intelligence test interpretation for decades. Clinicians continue to use classification schemes based on overall composite IQ scores for diagnosis and interpretation, and the concerns of Goodenough and Wechsler are still reflected in various texts on intelligence test interpretation (including this one). With the understanding that global IQ scores represent the most robust estimates of ability, they are frequently used in the diagnosis of intellectual disabilities, giftedness, learning disabilities, and other conditions. Still, we caution that global cutoff scores may not always be appropriate or adequate for the decisions typically made on the basis of intelligence test scores (Kaufman, 1990), since these scores constitute only one component of understanding an individual's range of developed cognitive abilities.

CLINICAL PROFILE ANALYSIS: THE SECOND WAVE

Rapaport, Gil, and Schafer's (1945–1946) seminal work has exerted a profound influence on intelligence test interpretation to the present day. These authors, recognizing an opportunity provided by the publication of the Wechsler–Bellevue Scale (Wechsler, 1939), advocated interpretation of the newly introduced subtest scores to achieve a more thorough understanding of an individual's cognitive skills; in addition, they extended intelligence test interpretation to include interpretation of individual test items and assignment of psychiatric diagnoses.

Profiles of Subtest Scores

Rapaport and colleagues (1945–1946) espoused an entirely new perspective for the interpretation of intelligence tests, focusing on the shape of subtest score profiles in addition to an overall general level of intellectual functioning. Whereas the pre–World War II psychologist was primarily dependent on the Binet scales and the determination of a general level of cognitive attainment, the post-Rapaport and colleagues psychologist became equally concerned with the shape of a person's profile of subtest scores. Specifically, patterns of high and low subtest scores could presumably reveal diagnostic and psychotherapeutic considerations:

> In our opinion, one can most fully exploit intelligence tests neither by stating merely that the patient was poor on some and good on other subtests, nor by trying to connect directly the impairments of certain subtest scores with certain clinical-nosological categories; but rather only by attempting to understand and describe the psychological functions whose impairment or change brings about the impairment of scores. . . . Every subtest score—especially the relationship of every subtest score to the other subtest scores—has a multitude of determinants. If we are able to establish the main psychological function underlying the achievement, then we can hope to construct a complex psychodynamic and structural picture out of the interrelationships of these achievements and impairments of functions . . . (Rapaport et al., 1945–1946, p. 106)

Rapaport and colleagues' (1945–1946) system had five major emphases, the first of which involved interpretation of item responses. The second emphasis involved comparing a subject's item responses within subtests. Differential responding to the same item type (e.g., information subtest items assessing U.S. vs. international knowledge) was thought to be of some diagnostic significance. The third emphasis suggested that meaningful interpretations could be based on within-subject comparisons of subtest scores. Rapaport and colleagues introduced the practice of deriving diagnostic information from comparisons between Verbal and Performance scales, the fourth interpretive emphasis. They suggested, for example, that a specific Verbal–Performance profile could be diagnostic of depression (p. 68). The fifth and final emphasis involved the comparison of intelligence test findings to other test findings. In this regard, they noted, "Thus, a badly impaired intelligence test achievement has a different diagnostic implication if the Rorschach test indicates a rich endowment or a poor endowment" (p. 68).

The work of Rapaport and colleagues (1945–1946) was a landmark due to its substantial scope. It provided diagnostic suggestions at each interpretive level for a variety of adult psychiatric populations. Furthermore, their work introduced an interpretive focus on intraindividual differences—a focus that at times took precedence over interindividual comparisons in clinical work with clients.

In addition to the breadth of their approach, the structure of Rapaport and colleagues' (1945–1946) approach gave clinicians a logical, step-by-step method for assessing impairment of function and for making specific diagnostic hypotheses. These authors directed clinicians to calculate a mean subtest score that could be used for identifying intraindividual strengths and weaknesses, and they gave desired difference score values for determining significant subtest fluctuations from the mean subtest score. The case of so-called "simple schizophrenia" (see Table 2.3) provides an example of the specificity of the diagnostic considerations that could be gleaned from a subtest profile.

Because of its thorough and clinically oriented approach, Rapaport and colleagues' (1945–1946) work provided a popular structure for training post–World War II clinical psychologists in the interpretation of intelligence test scores (i.e., the Wechsler–Bellevue Scale). This structure lingers today (Kamphaus, 2001).

TABLE 2.3. Diagnostic Considerations for the Case of "Simple Schizophrenia"

Subtest	Considerations
Vocabulary	Many misses on relatively easy items, especially if harder items are passed Relatively low weighted scores Parallel lowering of both the mean of the Verbal subtest scores (excluding Digit Span and Arithmetic) and the Vocabulary score
Information	Two or more misses on the easy items Relatively well-retained score 2 or more points above Vocabulary
Comprehension	Complete failure on any (especially more than one) of the seven easy items Weighted score 3 or more points below the Vocabulary score (or below the mean of the other Verbal subtests: Information, Similarities, and Vocabulary) Great positive Comprehension scatter (2 or more points superior to Vocabulary) is not to be expected
Similarities	Failure on easy items Weighted score 3 points below Vocabulary
Picture Arrangement	Tends to show a special impairment of Picture Arrangement in comparison to the other Performance subtests
Picture Completion	Weighted score of 7 or less
Object Assembly	Performance relatively strong
Block Design	No significant impairment from Vocabulary level Tends to be above the Performance mean
Digit Symbol	May show some impairment, but some "bland schizophrenics" may perform well

Verbal–Performance Differences and Subtest Profiles

Wechsler (1944), perhaps inadvertently, reinforced the practice of profile analysis by advocating a method of interpretation that also placed a premium on shape over a general level, with particular emphasis on subtest profiles and Verbal–Performance differences (*scatter*). His interpretive method is highlighted in a case example presented as a set of results for what he called "adolescent psychopaths" (see Table 2.4). It is noteworthy that Wechsler did not provide a Full Scale IQ (FSIQ) for this case example, focusing instead on shape rather than level. Wechsler offered the following interpretation of this "psychopathic" profile of scores:

> White, male, age 15, 8th grade. Continuous history of stealing, incorrigibility and running away. Several admissions to Bellevue Hospital, the last one after suicide attempt. While on wards persistently created disturbances, broke rules, fought with other boys and continuously tried to evade ordinary duties. Psychopathic patterning: Performance higher than Verbal, low Similarities, low Arithmetic, sum of Picture Arrangement plus Object Assembly greater than sum of scores on Blocks and Picture Completion. (p. 164)

This case exemplifies the second wave of intelligence test interpretation. This second wave was more sophisticated than the first, suggesting that intelligence test interpretation should involve more than mere designation of a general level of intelligence. However, methodological problems existed, eliciting one central question about these approaches: How do we know that these various subtest profiles accurately differentiate between clinical samples, and thus demonstrate diagnostic utility? The next wave sought to answer this salient question by applying measurement science to the process of intelligence test interpretation.

TABLE 2.4. Wechsler's Case Example for "Adolescent Psychopaths"

Subtest	Standard score
Comprehension	11
Arithmetic	6
Information	10
Digits	6
Similarities	5
Picture Arrangement	12
Picture Completion	10
Block Design	15
Object Assembly	16
Digit Symbol	12
Verbal IQ (VIQ)	90
Performance IQ (PIQ)	123

PSYCHOMETRIC PROFILE ANALYSIS: THE THIRD WAVE

The availability of computers and statistical software packages provided researchers of the 1960s and 1970s greater opportunity to assess the validity of various interpretive methods and the psychometric properties of popular scales. Two research traditions—*factor analysis* and *psychometric profile analysis*—have had a profound effect on intelligence test interpretation.

Factor Analysis

Cohen's (1959) seminal investigation addressed the second wave of intelligence test interpretation by questioning the empirical basis of the intuitively based "clinical" methods of profile analysis. He conducted one of the first comprehensive factor analyses of the standardization sample for the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), analyzing the results for 200 children from three age groups of the sample. Initially, five factors emerged: Factor A, labeled Verbal Comprehension I; Factor B, Perceptual Organization; Factor C, Freedom from Distractibility; Factor D, Verbal Comprehension II; and Factor E, quasi-specific. Cohen chose not to interpret the fourth and fifth factors, subsuming their loadings and subtests under the first three factors. Hence the common three-factor structure of the WISC was established as the de facto standard for conceptualizing the factor structure of the Wechsler scales. Eventually, Kaufman (1979) provided a systematic method for utilizing the three factor scores of the WISC-R (Wechsler, 1974) to interpret the scales as an alternative to interpreting the Verbal IQ (VIQ) and Performance IQ (PIQ), calling into question the common clinical practice of interpreting the Verbal and Performance scores as if they were measures of valid constructs. Cohen's labels for the first three factors were retained as names for the index scores through the third revision of the WISC (WISC-III; Wechsler, 1991). In addition, Cohen's study popularized the Freedom from Distractibility label for the controversial third factor (Kamphaus, 2001).

Cohen (1959) also introduced the consideration of subtest specificity prior to making subtest score interpretations. Investigation of the measurement properties of the subtests was crucial, as Cohen noted:

> A body of doctrine has come down in the clinical use of the Wechsler scales, which involves a rationale in which the specific intellective and psychodynamic trait-measurement functions are assigned to each of the subtests (e.g., Rapaport et al., 1945–1946). Implicit in this rationale lies the assumption that a substantial part of a test's variance is associated with these specific measurement functions. (p. 289)

According to Cohen (1959), *subtest specificity* refers to the computation of the amount of subtest variance that is reliable (not error) and specific to the subtest. Put another way, a subtest's reliability coefficient represents both reliable specific and shared variance. When shared variance is removed, a clinician may be surprised to discover that little reliable specific variance remains to support interpretation. Typically, the clinician may draw a diagnostic or other conclusion based on a subtest with a reliability estimate of .80, feeling confident of the interpretation. However, Cohen cautioned that this coefficient may be illusory because the clinician's interpretation assumes that the subtest is measuring an ability that is only measured by this subtest of the battery. The subtest specificity value for this same subtest may be rather poor if it shares considerable variance with other subtests. In fact, its subtest specificity value may be lower than its error variance (.20).

Cohen (1959) concluded that few of the WISC subtests could attribute one-third or more of their variance to subtest-specific variance—a finding that has been replicated for subsequent revisions of the WISC (Kamphaus, 2001; Kaufman, 1979). Cohen pointedly concluded that adherents to the "clinical" rationales would find no support in the factor-analytic studies of the Wechsler scales (p. 290). Moreover, he singled out many of the subtests for criticism; in the case of the Coding subtest, he concluded that Coding scores, when considered in isolation, were of limited utility (p. 295).

This important study set the stage for a major shift in intelligence test interpretation—that is, movement toward an emphasis on test interpretation supported by measurement science. Hallmarks of this approach are exemplified in Cohen's work, including the following:

1. Renewed emphasis on interpretation of the FSIQ (harkening back to the first wave), as a large second-order factor accounts for much of the variance of the Wechsler scales.
2. Reconfiguration of the Wechsler scales, proposing the three factor scores as alternatives or supplements to interpretation of the Verbal and Performance scales.
3. Deemphasis on individual subtest interpretation, due to limited subtest reliable specific variance (specificity).

Kaufman's Psychometric Approach

Further evidence of the influence of measurement science on intelligence test interpretation and the problems associated with profile analysis can be found in an influential book by Kaufman (1979), *Intelligent Testing with the WISC-R*. Kaufman provided a logically appealing and systematic method for WISC-R interpretation that was rooted in sound measurement theory. He created a hierarchy for WISC-R interpretation, which emphasized interpretive conclusions drawn from the most reliable and valid scores yielded by the WISC-R (see Table 2.5).

TABLE 2.5. Kaufman's Hierarchy for WISC-R Interpretation

Source of conclusion	Definition	Reliability	Validity
Composite scores	Wechsler IQs	Good	Good
Shared subtest scores	Two or more subtests combined to draw a conclusion	Good	Fair to poor
Single subtest scores	A single subtest score	Fair	Poor

Although such interpretive methods remained "clinical," in the sense that interpretation of a child's assessment results was still dependent on the child's unique profile of results (Anastasi, 1988), the reliance on measurement science for the interpretive process created new standards for assessment practice. Application of such methods required knowledge of the basic psychometric properties of an instrument, and consequently required greater psychometric expertise on the part of the clinician.

These measurement-based interpretive options contrasted sharply with the "clinical" method espoused by Rapaport and colleagues (1945–1946)—an approach that elevated subtest scores and item responses (presumably the most unreliable and invalid scores and indicators) to prominence in the interpretive process. The measurement science approach, however, was unable to conquer some lingering validity problems.

Diagnostic and Validity Problems

Publication of the Wechsler scales and their associated subtest scores created the opportunity for clinicians to analyze score profiles, as opposed to merely gauging an overall intellectual level from one composite score. Rapaport and colleagues (1945–1946) popularized this method, which they labeled *scatter analysis*:

> Scatter is the pattern or configuration formed by the distribution of the weighted subtest scores on an intelligence test . . . the definition of scatter as a configuration or pattern of all the subtest scores implies that the final meaning of the relationship of any two scores, or of any single score to the central tendency of all the scores, is derived from the total pattern. (p. 75)

However, Rapaport and colleagues began to identify problems with profile analysis of scatter early in their research efforts. In one instance, they expressed their frustration with the Wechsler scales as a tool for profile analysis, observing that "the standardization of the [Wechsler–Bellevue] left a great deal to be desired so that the average scattergrams of normal college students, Kansas highway patrolmen . . . and applicants to the Menninger School of Psychiatry . . . all deviated from a straight line in just about the same ways" (p. 161).

Bannatyne (1974) constructed one of the more widely used recategorizations of the WISC subtests into presumably more meaningful profiles (see Table 2.6). Matheson, Mueller, and Short (1984) studied the validity of Bannatyne's recategorization of the WISC-R, using a multiple-group factor analysis procedure with three age ranges of the WISC-R and data from the WISC-R standardization sample. They found that the four categories had high reliabilities, but problems with validity. For example, the Acquired Knowledge category had sufficiently high reliabilities, but it was not independent of the other three categories, particularly Conceptualization. As a result, Matheson and colleagues advised that the Acquired Knowledge category not be interpreted as a unique entity; instead, they concluded that the Acquired Knowledge and Conceptualization categories were best interpreted as one measure of verbal intelligence, which was more consistent with the factor-analytic research on the WISC-R and other intelligence test batteries.

Similarly, Kaufman (1979) expressed considerable misgivings, based on a review of research designed to show links between particular profiles of subtest scores and child diagnostic categories (although he too had provided detailed advice for conducting profile analysis). Kaufman noted that the profiles proved to be far less than diagnostic:

> The apparent trends in the profiles of individuals in a given exceptional category can sometimes provide one piece of evidence to be weighed in the diagnostic process. When there is ample support for a diagnosis from many diverse background, behavioral, and test-related (and in some cases medical) criteria, the emergence of a reasonably characteristic profile can

TABLE 2.6. Bannatyne's Recategorization of WISC-R Subtests

Spatial	Conceptualization	Sequencing	Acquired knowledge
Block Design	Vocabulary	Digit Span	Information
Object Assembly	Similarities	Coding	Arithmetic
Picture Completion	Comprehension	Arithmetic	Vocabulary
		Picture Arrangement	

> be treated as one ingredient in the overall stack of evidence. However, the lack of a characteristic profile should not be considered as disconfirming evidence. In addition, no characteristic profile, in and of itself, should ever be used as the primary basis of a diagnostic decision. We do not even know how many normal children display similar WISC-R profiles. Furthermore . . . the extreme similarity in the relative strengths and weaknesses of the typical profiles for mentally retarded, reading-disabled, and learning-disabled children renders differential diagnosis based primarily on WISC-R subtest patterns a veritable impossibility. (pp. 204–205)

Profile analysis was intended to identify intraindividual strengths and weaknesses—a process known as *ipsative interpretation*. In an ipsative interpretation, the individual client was used as his or her own normative standard, as opposed to making comparisons to the national normative sample. However, such seemingly intuitive practices as comparing individual subtest scores to the unique mean subtest score and comparing pairs of subtest scores are fraught with measurement problems. The clinical interpretation literature often fails to mention the poor reliability of a *difference score* (i.e., the difference between two subtest scores). Anastasi (1985) has reminded clinicians that the standard error of the difference between two scores is larger than the standard error of measurement of the two scores being compared. Thus interpretation of a 3- or 5-point difference between two subtest scores becomes less dependable for hypothesis generation or making conclusions about an individual's cognitive abilities. Another often-cited problem with ipsative interpretation is that the correlations among subtests are positive and often high, suggesting that individual subtests provide little differential information about a child's cognitive skills (Anastasi, 1985). Furthermore, McDermott, Fantuzzo, Glutting, Watkins, and Baggaley (1992), studying the internal and external validity of subtest strengths and weaknesses, found these measures to be wholly inferior to basic norm-referenced information.

Thus the long-standing practice of using profile analysis to draw conclusions about intraindividual strengths and weaknesses did not fare well in numerous empirical tests of its application. The lack of validity support for profile analysis remains unresolved (Kamphaus, 2009). Measurement problems remained, many of which were endemic to the type of measure used (e.g., variations on the Wechsler tradition). These validity challenges indicated the need for the fourth wave, wherein theory and measurement science became intermingled with practice considerations to enhance the validity of test score interpretation.

APPLYING MODERN FACTOR-ANALYTIC MODELS TO TEST DEVELOPMENT: THE FOURTH WAVE

Kaufman (1979) identified a disjuncture between the development of intelligence tests on the one hand and theories and research undergirding the intelligence construct on the other. A colleague of David Wechsler, he knew that Wechsler's measures were built to assess general intelligence, and included two subscales (Verbal and Performance) based not on theory or research findings supporting the existence of such constructs, but on the requirements of everyday assessment practice in linguistically diverse New York City. Wechsler's practical breakthrough allowed psychologists to use the Performance scale to assess the general intelligence of individuals with limited English fluency and comprehension.

Kaufman simultaneously noted that intelligence tests' lack of alignment with factor-analytic research compromised the validity of test score interpretations. He proposed reorganizing subtests into clusters that conformed to widely accepted theories of intelligence, thus allowing clinicians to produce more meaningful conclusions. This seminal insight has encouraged test developers over the past three decades to ensure that the scores offered are grounded in the known factor structure of intelligence. The fourth wave has addressed intelligence test validity through the development of contemporary instruments better grounded in theory, through research findings, and through integration of test results with multiple sources of information—hypothesis validation, as well as testing of rival hypotheses (Kamphaus, 2001).

Test Design for Interpretation

The history of intelligence test interpretation has been characterized by a disjuncture between the design of the tests and inferences made from those tests. A test, after all, should be designed a priori with a strong theoretical foundation, and supported by considerable validity evidence in order to measure a particular construct or set of constructs (and *only* those constructs). Prior to the 1990s, the interpretive process was conducted by clinicians who sometimes applied relatively subjective clinical acumen in the absence of empirically supported theoretical bases to interpret scores for their consumers. For more valid and reliable interpretation of intelligence tests, instrument improvement would now need to focus on constructing tests designed to measure a delimited and well-defined set of intelligence-related constructs.

During the second half of the 20th century, several theories of the structure of intelligence were introduced, promoting a shift to seeking theoretical support for the content of intelligence tests. Among the most significant theories have been Carroll's three-stratum theory of cognitive abilities, the Horn–Cattell fluid–crystallized (Gf-Gc) theory, the Luria–Das model of information processing, Gardner's multiple intelligences, and Sternberg's triarchic theory of intelligence (see Chapters 3–6 of the present volume for reviews).

Two popular theoretical models of intelligence have had the primary distinction of fostering this shift. First, the factor-analytic work of Raymond Cattell and John Horn (Horn & Cattell, 1966) describes an expanded theory founded on Cattell's (1943) constructs of *fluid intelligence* (Gf) and *crystallized intelligence* (Gc). Cattell described fluid intelligence as representing reasoning and the ability to solve novel problems, whereas crystallized intelligence was thought to constitute abilities influenced by acculturation, schooling, and language development. This fluid–crystallized distinction was supported by Horn (1988), who delineated additional contributing abilities such as visual–spatial ability, short-term memory, processing speed, and long-term retrieval.

Subsequent to this research was John Carroll's (1993) integration of findings from more than 460 factor-analytic investigations that led to the development of his three-stratum theory of intelligence. The three strata are organized by generality. Stratum III, the apex of the framework, consists of one construct only—general intelligence or *g*, the general factor that has been identified in numerous investigations as accounting for the major portion of variance assessed by intelligence test batteries. Stratum II contains eight broad cognitive abilities contributing to the general factor *g*, and is very similar to Gf-Gc abilities as described by Horn. Carroll's model proposes numerous narrow (specific) factors subsumed in stratum I. The two models are sometimes used together and are referred to in concert as the *Cattell–Horn–Carroll* (CHC) model of intelligence (see Schneider & McGrew, Chapter 3, this volume).

Merging Theory and Test Design

Most modern intelligence tests are based in part or whole on a few widely accepted theories of intelligence—theories built upon and consistent with decades of factor-analytic studies of intelligence test batteries (Kamphaus, 2001). The commonality of theoretical development is demonstrated in the following brief descriptions of several widely used tests. All are examples of a greater emphasis on theory-based test design. The intelligence tests are described in great detail in individual chapters of this book.

Among contemporary intelligence tests, the Woodcock–Johnson IV (WJ IV; Woodcock & Mather, 2014) is closely aligned with the Cattell–Horn (Cattell, 1943; Horn, 1988; Horn & Cattell, 1966) and Carroll (1993) theories of intelligence. According to the WJ III technical manual (McGrew & Woodcock, 2001), Cattell and Horn's Gf-Gc theory was the theoretical foundation for the Woodcock–Johnson Psycho-Educational Battery—Revised (WJ-R; Woodcock & Johnson, 1989). Four years after publication of the WJ-R, Carroll's text was published; professionals interested in theories of intelligence began to think in terms of a combination or extension of theories, the CHC theory of cognitive abilities (McGrew & Woodcock, 2001). CHC theory, in turn, served as the blueprint for the WJ III (Woodcock, McGrew, & Mather, 2001). The WJ III developers designed their instrument to broadly measure seven of the eight stratum II factors from CHC theory, providing the following cognitive cluster scores: Comprehension–Knowledge (crystallized intelligence), Long-Term Retrieval, Visual–Spatial Thinking, Auditory Processing, Fluid Reasoning (fluid intelligence), Processing Speed, and Short-Term Memory. Moreover, individual subtests are intended to measure several narrow abilities from stratum I. Finally, the General Intellectual Ability score serves as a measure of overall *g*, representing stratum III.

Similarly, the SB5 (Roid, 2003) is based on the CHC model of intelligence. The SB5 can be considered a five-factor model, in that it includes five of the broad stratum II factors having the highest loadings on *g*: Fluid Reasoning (fluid intelligence), Knowledge (crystallized knowledge), Quantitative Reasoning (quantitative knowledge), Visual–Spatial Processing (visual processing), and Working Memory (short-term memory). Among these factors, Visual–Spatial Processing is new to this revision—an attempt to enrich the nonverbal measures of the SB5, aiding in the identification of children with spatial talents and deficits. Moreover, the SB5 is constructed to provide a strong nonverbal IQ by creating nonverbal measures for all five factors.

The Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014) continues to improve the alignment of its scales with the factor-analytic work of Carroll and Horn. Among the primary index scales, the differentiation of Visual Spatial from Fluid Reasoning is just one example. The primary index scales now include:

- Verbal Comprehension Index (VCI)
- Visual Spatial Index (VSI)
- Working Memory Index (WMI)
- Fluid Reasoning Index (FRI)
- Processing Speed Index (PSI)

The factor-analytic work of Carroll (1993) also informed the creation of the Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003) and RIAS-2 (Reynolds & Kamphaus, 2015) by demonstrating that many of the latent traits assessed by intelligence tests are test-battery-independent. The RIAS focuses on the assessment of stratum III and stratum II abilities from Carroll's three-stratum theory. The RIAS-2 is designed to assess five important aspects of intelligence: general intelligence (stratum Ill), verbal intelligence (stratum II, crystallized abilities), nonverbal intelligence (stratum II, visualization/spatial abilities), and memory (stratum II, working memory, short-term memory, or learning) and processing speed. These five constructs are assessed by combinations of eight RIAS-2 subtests.

The effects of basing intelligence tests on the confluence of theory and research findings are at least threefold. First, test-specific training is of less value: A psychologist who knows the commonly accepted factor-analytic findings can interpret most modern intelligence tests with confidence. Second, priorities in pre- and inservice psychologist training can shift to developing knowledge of foundational factor-analytic findings that inform modern test construction and interpretation. Third, as intelligence tests seek to measure similar core constructs, they increasingly resemble commodities (Kamphaus, 2009). In other words, a psychologist's decision to use a particular test may be based less on differences in validity (because the validity of score inferences is similar for different tests measuring the same construct) as on differences in personal preferences for administration time, scoring and reporting options, packaging, price, and other practical considerations. The days of clearly inferior or superior tests are largely gone as publishing houses converge their efforts on assessing core accepted constructs as assessed by tests that adhere to the standards of the profession (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

Interpretation of Results

With tests based on strong factor-analytic evidence in place, the job of the assessing psychologist is made easier, but there is still ample room for making flawed test score interpretations. Like all tests, intelligence test results can only be interpreted meaningfully in the context of other assessment results, clinical observations, background/demographic information, and other sources of quantitative and qualitative information. Second, all interpretations made should be supported by research evidence and theory. Presumably, these two premises should mitigate against uniform interpretations that do not possess validity for a particular case (i.e., standard interpretations that are applied to case data but are at odds with information unique to an individual), as well as against interpretations that are refuted by research findings (i.e., interpretations that are based on a psychologist's unique observations or training but are nevertheless contradicted by research findings).

Failure to integrate intelligence test results with other case data (e.g., an examinee's history) is the primary culprit in flawed test interpretation that has been documented by many researchers and practitioners over numerous decades. Matarazzo (1990) offered the following example from a neuropsychological evaluation in which the clinician failed to interpret test results in the context of an individual's background information. This example has the additional advantage of demonstrating the problem of making test score interpretations that are not based on sound validity evidence (e.g., using selected intelligence test subtest results to assess preinjury general intelligence).

> There is little that is more humbling to a practitioner who uses the highest one or two Wechsler subtest scores as the only index of a patient's "premorbid" level of intellectual functioning and who therefore interprets concurrently obtained lower subtest scores as indexes of clear "impairment" and who is then shown by the opposing attorney elementary and

> high school transcripts that contain several global IQ scores, each of which were at the same low IQ levels as are suggested by currently obtained lowest Wechsler subtest scaled scores. (p. 1003)

There are many practical ways to protect against inaccurate test score interpretations. An example provided by Kamphaus (2001) suggested that the intelligence test user establish a standard for integrating intelligence test results with other contextual information by requiring. two pieces of corroborating evidence for each test interpretation made. Such a standard simply provides a helpful rubric that requires the examiner to carefully consider other findings and information before offering conclusions. A clinician, for example, may calculate a WISC-V FSIQ score of 84 (below average) for a young girl and conclude that she possesses below-average intelligence. Even this seemingly obvious conclusion should be corroborated by two external sources of information. If the majority of the child's achievement scores fall into this range, and her teacher reports that the child seems to be progressing more slowly than the majority of the children in her class, the conclusion of below-average intelligence has been corroborated by two sources of information external to the WISC-V. On the other hand, if this child has previously been diagnosed with an anxiety disorder, and if both her academic achievement scores and her progress as reported by her teacher are average, the veracity of the WISC-V scores may be in question. If she also appears highly anxious and agitated during the assessment session, the obtained scores may be even more questionable.

The requirement of research (i.e., validity) support for test-based interpretation is virtually mandatory in light of the publication of the *Standards for Educational and Psychological Testing* (AERA et al., 2014) and the increased expectations of consumers for assessment accuracy (Kamphaus, 2001). Clinical "impressions" of examiners, although salient, are no longer adequate for supporting interpretations of a child's intelligence scores (Matarazzo, 1990). Consider again the example above in which the young child obtains a WISC-V FSIQ score of 84. Let us assume that she has been independently found to have persistent problems with school achievement. Given the data showing the positive relationship between intelligence and achievement scores, the results seem consistent with the research literature and lend support to the interpretation of below-average developed cognitive abilities. Should it become necessary to support the conclusion of below-average intelligence, the clinician could give testimony citing studies supporting the correlational relationship between intelligence and achievement test scores (Matarazzo, 1990).

TESTING RIVAL HYPOTHESES

Some research suggests that clinicians routinely tend to overestimate the accuracy of their conclusions. There is virtually no evidence to suggest that clinicians underestimate the amount of confidence they have in their conclusions (Dawes, 1995). Therefore, intelligence test users should check the accuracy of their inferences by challenging them with alternative inferences.

A clinician may conclude, for example, that a client has a personal strength in verbal intelligence relative to nonverbal. An alternative hypothesis is that this inference is merely due to chance. The clinician may then use test manual discrepancy score tables to determine whether the difference between the two standard scores is likely to be reliable (i.e., statistically significant) and therefore not attributable to chance. Even if a difference is reliable, however, it may not be a "clinically meaningful" difference if it is a common occurrence in the population. Most intelligence test manuals also allow the user to test the additional hypothesis that the verbal–nonverbal score inference is reliable, but too small to be of clinical value for diagnosis or intervention, by determining the frequency of the score difference in the population. If a difference is also rare in the population, the original hypothesis (that the verbal–nonverbal difference reflects a real difference in the individual's cognitive abilities) provides a better explanation than the alternative rival hypothesis (that the verbal–nonverbal difference is not of importance) for understanding the examinee's cognitive performances.

Knowledge of theory, grounded in research, is important above and beyond isolated research findings, as theory allows the clinician to do a better job of conceptualizing an individual's scores. Clear conceptualization of a child's cognitive status, for example, allows the clinician to better explain the child's test results to parents, teachers, colleagues, and other consumers of the test findings. Parents will often want to know the etiology of the child's scores. They will question, "Is it my fault for not reading to her?" or "Did he inherit this problem? My father had the same problems in school." Without adequate theoretical knowledge, clinicians will find themselves unprepared to give reasonable answers to such questions.

CONCLUSIONS

In this chapter, we have presented several overarching historical approaches to the interpretation of intelligence test scores. For heuristic purposes, these approaches are portrayed as though they were entirely separate in their inception, development, and limitations. In the reality of clinical practice, however, much overlap exists.

Aspects of each of these approaches continue to influence modern test interpretation. For example, since Spearman's (1927) publication of findings in support of a central ability underlying performance on multiple tasks, clinicians typically have interpreted a single general intelligence score. Most intelligence tests yield a general ability score, and research continues to provide evidence for the role of a general ability or *g* factor (McDermott & Glutting, 1997). In Carroll's (1993) hierarchical theory, *g* remains at the apex of the model. Therefore, the ongoing practice of interpreting this factor is warranted. At the same time, clinicians continue to interpret an individual's profile of scores. For the most part, the days of making psychiatric diagnoses or predictions of psychiatric symptoms on the basis of intelligence test scores, as Rapaport and his colleagues (1945–1946) suggested, are past; however, profiles are still discussed—that is, in terms of ability profiles related to achievement or educational outcomes. Profile interpretations, however, have generally been superseded by factor-analytic evidence of the existence of stratum II and stratum III factors. Clinicians are on more solid research ground when they make "profile" interpretations based not on ad hoc or conceptual groupings of subtest scores, but rather on the "groupings" made by consistent factor-analytic evidence.

Furthermore, as was the case in what we describe as the third wave, results from psychometric analyses still inform and guide interpretations. Now, however, they are also integrated into broad descriptions and theories of intelligence. Carroll's theory is the result of factor-analytic research, and writers have labeled many of the dominant theories of intelligence as *psychometric* in their approach (Neisser et al., 1996). Thus we see progress in the area of intellectual assessment and interpretation as an evolution, rather than a series of disjointed starts and stops. This evolution has culminated in the integration of empirical research, theory development, and test design, resulting in more accurate and meaningful test interpretation.

What Will Be the Fifth Wave of Intelligence Test Interpretation?

What seems safe to predict is that ongoing educational reform and public policy mandates will continue to shape intellectual assessment and their associated interpretations. The influence of educational test and public policy were present when the first formal intelligence tests were introduced over a century ago, and their influence has not abated.

We hypothesize that the next wave will focus on the publication of new tests with stronger evidence of content validity; if the ultimate purpose of intelligence testing is to sample behavior representing a construct and then to draw inferences about that construct, the process of interpretation is limited by the clarity of the construct(s) being measured. It may also be time to apply a broader concept of *content validity* to intelligence test interpretation (e.g., Flanagan & McGrew, 1997). Cronbach (1971) suggested such an expansion of the term more than four decades ago, observing:

> Whether the operations that finally constitute the test correspond to the specified universe is the question of content validity. It is so common in education to identify "content" with the subject matter of the curriculum that the broader application of the word here must be stressed. (p. 452)

The nature–nurture debate in intelligence testing will continue to fade, thanks to findings indicating that cognitive ability is essentially "developed" through experience (Hart & Risley, 1995). Although the use of the term *IQ* has lessened, it is still popular among the general public. It has been replaced by terms such as *cognitive ability* and others, which have indirectly muted debates about the "excessive meanings" that have accompanied the IQ score, particularly its "native ability" interpretation among some segments of the public. It may now be time, in light of recent evidence, to adopt Anastasi's (1988) preferred term of long ago: *developed cognitive ability*.

We will also be likely to interpret intelligence tests in a more evidence-based manner. The evidence for the overall intelligence test score yielded by measures of developed cognitive ability is among the most impressive in the social sciences.

Lubinski (2004) summarizes some enduring findings:

> Measures of *g* covary .70–.80 with academic achievement measures, .70 with military training assignments, .20–.60 with work performance (correlations are moderated by job complexity), .30–.40 with income, and –.20 with unlawfulness. General intelligence covaries .40 with SES of origin and .50–.70 with achieved SES. As well, assortative mating correlations approach .50. These correlations indicate that *g* is among the most important individual differences dimensions for structuring the determinants of Freud's two-component characterization of life, *lieben* and *arbeiten,* working and loving (or resource acquisition and mating). (p. 100)

As intelligence tests incorporate current research-based theories of intelligence into their design, psychological interpretations will become more valid. This trend will be modified as changes occur in intelligence-testing technology, fostered by breakthrough theories (e.g., neurological) or empirical findings. Although it is difficult to draw inferences about the vast and somewhat undefined "universe" of cognitive functioning, it is also *de rigueur.* Psychologists routinely make such interpretations about the complex universe of human behavior. The emergence of tests that better measure well-defined constructs will allow psychologists to provide better services to their clients than were possible even a decade ago.

ACKNOWLEDGMENTS

We would like to express our gratitude to Martha D. Petoskey and Anna W. Morgan for their contributions to the first edition of this chapter.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

Anastasi, A. (1985). Interpreting results from multiscore batteries. *Journal of Counseling and Development, 64,* 84–86.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC scale scores. *Journal of Learning Disabilities, 7,* 272–274.

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [A new method for the diagnosis of the intellectual level of abnormal persons]. *L'Année Psychologique, 11,* 191–244.

Boring, E. G. (1929). *A history of experimental psychology.* New York: Century.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology, 23,* 285–299.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–506). Washington, DC: American Council on Education.

Dawes, R. M. (1995). Standards of practice. In S. C. Hayes, V. M. Vollette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 31–43). Reno, NV: Context Press.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314–325). New York: Guilford Press.

French, J. L., & Hale, R. L. (1990). A history of the development of psychological and educational testing. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 3–28). New York: Guilford Press.

Goodenough, F. L. (1949). *Mental testing: Its history, principles, and applications.* New York: Rinehart.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Brookes.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (2nd ed., pp. 645–865). New York: Academic Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253–270.

Kamphaus, R. W. (2001). *Clinical assessment of children's intelligence.* Needham Heights, MA: Allyn & Bacon.

Kamphaus, R. W. (2009). Assessment of intelligence and achievement. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 230–246). Hoboken, NJ: Wiley.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley-Interscience.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Needham Heights, MA: Allyn & Bacon.

Levine, A. J., & Marks, L. (1928). *Testing intelligence and achievement.* New York: Macmillan.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' Objectively Determined and Measured." *Journal of Personality and Social Psychology, 86*(1), 96–199.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing?: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45*(9), 999–1017.

Matheson, D. W., Mueller, H. H., & Short, R. H. (1984). The validity of Bannatyne's acquired knowledge category as a separate construct. *Journal of Psychoeducational Assessment, 2,* 279–291.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25,* 504–526.

McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—or, more illusion of meaning? *School Psychology Review, 26*(2), 163–176.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III technical manual.* Itasca, IL: Riverside.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Pintner, R. (1923). *Intelligence testing.* New York: Holt, Rinehart & Winston.

Rapaport, D., Gil, M., & Schafer, R. (1945–1946). *Diagnostic psychological testing* (2 vols.). Chicago: Year Book Medical.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales.* Lutz, FL: Psychological Assessment Resources.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Terman, L. M. (1916). *The measurement of intelligence: An explanation and a complete guide for the use of the Stanford revision and extensions of the Binet–Simon Intelligence Scale.* Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new Revised Stanford–Binet Tests of Intelligence.* Boston: Houghton Mifflin.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery— Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Mather, N. (2014). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside.

PART II

Contemporary Theoretical Perspectives

CHAPTER 3

The Cattell–Horn–Carroll Theory of Cognitive Abilities

W. Joel Schneider
Kevin S. McGrew

The Cattell–Horn–Carroll theory of cognitive abilities (CHC theory) is a comprehensive taxonomy of abilities embedded in multiple overlapping theories of cognition. Like all good taxonomies, its purpose is to orient the user toward distinctions that are meaningful and productive, and away from those that are irrelevant and pointless.

If CHC theory had a personality, it would be open-minded, ambitious, and polite. It does not explain everything about intelligence, but it wants to. It has a lot to say and is perhaps a bit long-winded, but it also listens to other theories. It has a "big-tent" mindset, tolerating ambiguities and disagreements wherever there are reasonable grounds for disagreement (e.g., the nature of general intelligence). It likes to cooperate with other theories when it can and enjoys a good debate when it cannot. Our hope for CHC theory is that it provides a common framework and nomenclature for intelligence researchers to communicate their findings without getting bogged down in endless debates about whose version of this or that construct is better.

A BASIC DESCRIPTION OF CHC THEORY

The basic idea of CHC theory is that intelligence is both multidimensional and functionally integrated. The dimensions of intelligence can be measured, studied, and understood in terms of their shared and separate antecedents, correlates, and causal effects. The source theorists of CHC theory, Raymond Cattell, John Horn, and John Carroll, worked primarily within the psychometric tradition of intelligence theory. That is, they measured individual differences and inferred truths about intelligence by studying the patterns of covariation among these differences. For example, they would try to make sense of the facts that all tests of cognitive ability are strongly correlated, but that there are many clusters of highly correlated tests that are weakly correlated with other clusters of highly correlated tests. The source theorists attempted to infer theoretical entities that would explain such patterns and at the same time be consistent with known facts from biological, cognitive, developmental, and personality psychology, as well as from other disciplines (e.g., genetics, neuroscience, and linguistics).

In CHC theory, the dimensions of ability have a hierarchical structure, meaning that some have a broader scope than others (see Figure 3.1). At the bottom of the hierarchy are *specific abilities,* which are tied to a specific task or test, such as the ability to repeat sentences back after hearing them once or the ability to discern whether two lines are parallel or skewed. Specific abilities are the only abilities that can be measured directly. Other types of

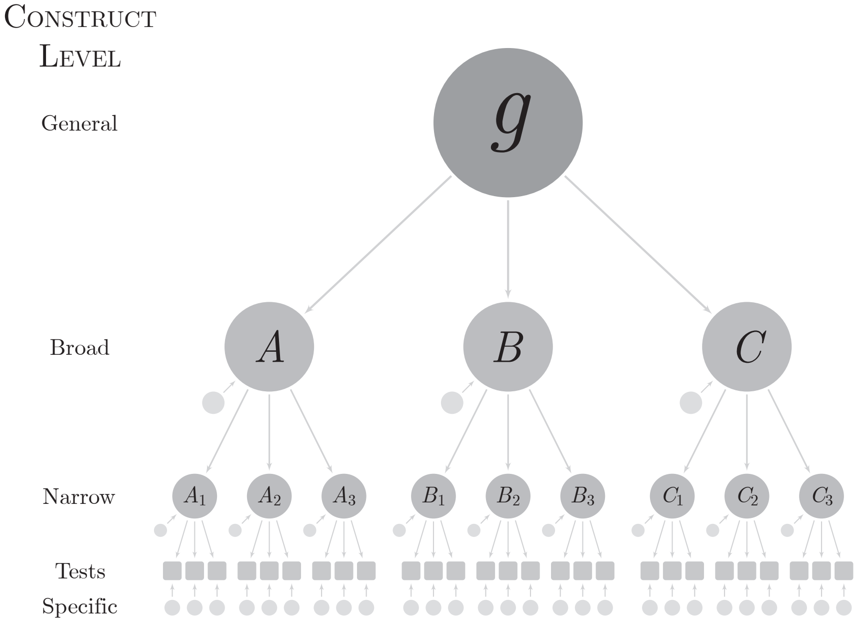


FIGURE 3.1. A tidy hierarchical structure.

abilities—narrow, broad, and general abilities—are theoretical entities inferred from the observed relations among specific abilities.

Narrow abilities are clusters of highly correlated specific abilities. For example, the ability to repeat back sentences is highly correlated with the ability to repeat back single words. These specific abilities are highly correlated with the ability to repeat back digits, letters, and nonsense words. What theoretical entity might explain this fact? Borrowing from cognitive psychology, we might hypothesize that this cluster of highly correlated abilities corresponds to individual differences in *auditory short-term memory storage capacity*.

If we also measure tests that require remembering different kinds of visual information over short periods, we might find that these also correlate highly with each other, prompting us to hypothesize the existence of a different narrow ability, *visual–spatial short-term memory storage capacity*. If we give tests requiring people not only to repeat back the information but also to manipulate it (e.g., repeat words backward, sort them alphabetically, group them by size), we might find that such tests form a separate cluster of highly correlated tests, and that the sensory modality (auditory vs. visual) does not matter much in distinguishing which tests correlate with other tests in this cluster. We might hypothesize that this cluster represents individual differences in a domain-general capacity to direct the focus of attention to process information, such as Baddeley and Hitch's (1974) *central executive* in their model of *working memory*.

Broad abilities are clusters of narrow abilities that are mutually more correlated with each other than with abilities in other broad-ability clusters. For example, auditory short-term storage, visual short-term storage, and attention control form a cluster of narrow abilities that are mutually more correlated with each other than with other kinds of abilities. We can call this collection of narrow abilities *working memory capacity* (Unsworth & Engle, 2007). By convention, all broad abilities in CHC theory have an abbreviation that starts with a capital G, followed by lowercase letters. In this case, the broad ability of working memory capacity would be abbreviated as *Gwm*. The letter G stands for *general*, though not all broad abilities are truly general. Cattell (1971) tried to fix his nomenclature by using the letters *g* and *p* to distinguish between domain-general capacities and "provincial" capacities, but this innovation never caught on.

Carroll (1993) found strong evidence for 8–10 broad clusters of narrow abilities. There may be as many as 20 of them. We are being coy about the exact number not only because of incomplete evidence, but because some ability clusters are difficult to categorize. The terms *broad* and *narrow* are not hard categories, but useful ways of describing degrees of breadth. As our understanding of intelligence has become more nuanced, we have added as many intermediate categories between broad and narrow as needed until the domain is adequately described. In this regard, CHC theory is becoming increasingly like Vernon's (1950) hierarchical group factor theory of intelligence of ability, and by extension, its successor, Johnson and Bouchard's (2005) verbal–perceptual–rotation model.

A full description of CHC theory's broad- and narrow-ability factors appears at the end of the chapter. As an advance organizer, we present the CHC broad abilities grouped conceptually in Figure 3.2. Fluid reasoning (Gf) is a domain-general reasoning capacity. There are several acquired-knowledge capacities, including comprehension–knowledge (Gc), domain-specific knowledge (Gkn), reading and writing (Gw), and quantitative knowledge (Gq). There are several domain-specific sensory abilities corresponding to each of the major senses: visual (Gv), auditory (Ga), olfactory (Go), tactile (Gh), and kinesthetic (Gk), as well as the psychomotor ability factor (Gp). There are three factors related to memory: working memory capacity (Gwm), learning efficiency (Gl), and retrieval fluency (Gr). Finally, there are several abilities related to speed: reaction/decision time (Gt), processing speed (Gs), and psychomotor speed (Gps).

If CHC theory were simply a list of abilities, it would not be much of a theory. The broad and narrow abilities are distinct, yet form an integrated system of problem solving (Schneider & McGrew, 2013). We conclude the chapter with a model of cognition that incorporates the CHC broad abilities as parameters of information processing.

BEYOND CHC: THE EVOLUTION OF THE THEORY AND ASSESSMENT OF COGNITIVE ABILITIES

Historical accounts of the evolution of the psychometric approach to the study of human individual differences abound (Brody, 2000; Carroll, 1993; Cudeck & MacCallum, 2007; Horn & Noll, 1997). We cannot possibly convey the depth, breadth, and subtlety of thought that characterizes the work of the great early theorists. In our chapter in the third edition of this book (Schneider & McGrew, 2012), we presented a visual-graphic timeline and narrative description of significant historical CHC theory and assessment events. The timeline ended with the heading "Beyond CHC theory (the next generation?)." This final phase included the subheadings of "CHC neuropsychological assessment" and "CHC theory impact and extensions."

The question mark in the heading title was prophetic. We initially planned to update this time-

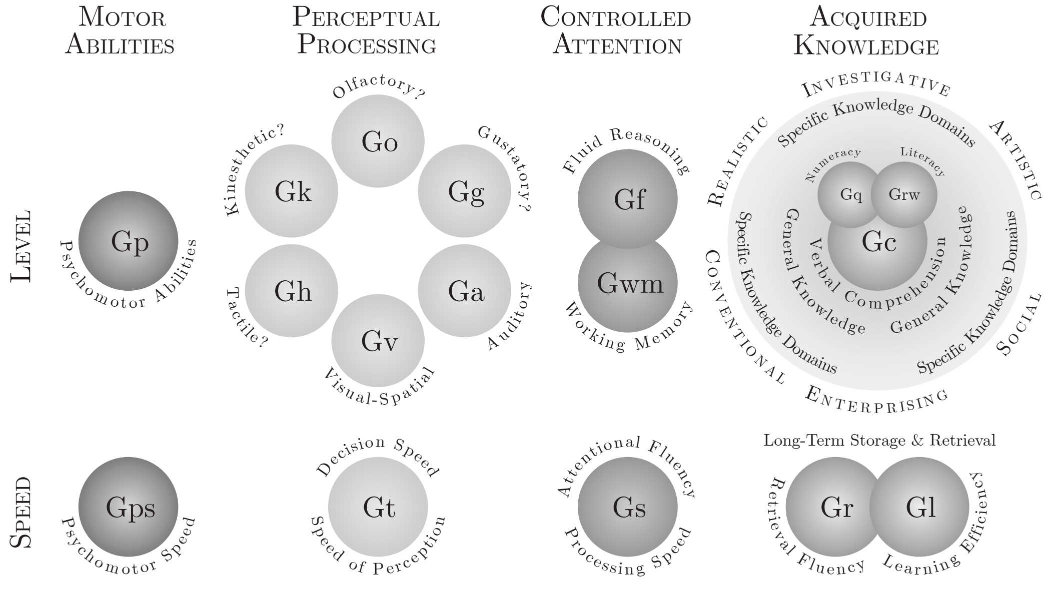


FIGURE 3.2. Conceptual groupings of CHC broad abilities.

line in the current chapter, but found, after numerous frustrating false starts, that distinct crystallized milestones were difficult to identify after 2012. Borrowing from the literature in design innovation (Norman & Verganti, 2014), we believe that CHC theory and assessment developments since 2012 have been more *incremental* or "hill-climbing" innovations (gradual improvements within a given frame of reference; "doing better what we already do") and not *radical* innovations (a novel, unique, and somewhat abrupt change in a frame of reference; "doing what we did not do before"). As a result, instead of providing a revised visual-graphic timeline, we present a list of notable CHC developments. We also provide updates on our prior discussions of CHC-based neuropsychological research and assessment, and of the theory's impact and extension. With the passage of time, we expect that a more focused picture of the "Beyond CHC theory" phase will emerge.

The reader is referred to our chapter in the third edition (Schneider & McGrew, 2012), as well as to several other excellent resources for in-depth descriptions of the history of intelligence theory and testing and CHC theory and assessment (A. S. Kaufman, 2009; Schneider & Flanagan, 2015; Wasserman, 2012 and Chapter 1, this volume).

THE CONTINUING IMPACT OF CHC THEORY

CHC's Global Reach

CHC theory is now officially a globetrotter, with a large bank of frequent flier miles. An indicator of the spread of CHC theory is reflected in the globalization of CHC assessment activities in countries beyond the United States. We highlight some of the major global developments.

The influence of CHC theory, primarily via university assessment training in the use of the CHC-based Batería III Woodcock–Muñoz (Muñoz-Sandoval, Woodcock, McGrew, Mather, & Schrank, 2005), is prominent in Spanish-speaking countries south of the U.S. border. This includes training, research or clinical use of the Batería III in Cuba, Mexico, Chile, Costa Rica, Panama, and Guatemala.[1] Farther south, researchers in Brazil were early adopters of CHC theory as a guide for intelligence test development (Primi, 2003; Wechsler & Nakano, 2016). For example, S. M. Wechsler and colleagues (Wechsler & Schelini, 2006; Wechsler et al., 2010; Wechsler, Vendramini, & Schelini, 2007) completed several studies to adapt the CHC-based Woodcock–Johnson III (WJ III) to Brazil. More recently, Wechsler and colleagues (2014) developed the Brazilian Adult Intelligence Battery, which, although only measuring Gf and Gc, is grounded in CHC theory. Other Brazilian researchers have focused on the nature and measurement of Gf (Primi, 2014; Primi, Ferrão, & Almeida, 2010), with their research couched in the context of CHC theory.

CHC influences are also present north of the U.S. border in Canada. The CHC-based WJ III has been used by practitioners in Canada, based on a U.S.–Canadian matched-sample comparison study (Ford, Swart, Negreiros, Lacroix, & McGrew, 2010). The WJ IV is now sold and used in Canada. In addition, a school-based group-administered CHC test (Insight; Beal, 2011) measuring Gf, Gc, Gv, Ga, Gwm, Glr, Gs, and CDS (Gt) is available in Canada. CHC theory and testing has a prominent place in school psychology assessment courses in several major Canadian universities (e.g., the University of British Columbia, the University of Alberta).[2]

One of the first systematic global CHC test development outreach projects was an effort, led by Richard Woodcock and the Woodcock–Muñoz Foundation, to provide Eastern European countries with cost-effective brief versions of the CHC-based WJ III Tests of Cognitive Abilities. The WJ III-IE (International Editions) project started in the early 2000s and continued until approximately 2015. WJ III-IE norming efforts occurred in the Czech Republic, Hungary, Latvia, Romania, and Slovakia. Other European efforts include the Austrian-developed computerized Intelligence Structure Battery (Arendasy et al., 2009),[3] which measures six broad CHC abilities (Gf, Gq, Gc, Gw, Gv, Glr). The spread of CHC theory has also reached France and Spain. French researchers have analyzed French versions of the various Wechsler scales from the perspective of the CHC framework (Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010). In Spain, researchers in computer science education have used the CHC taxonomy to analyze the components of the Computational Thinking Test (Román-González, Pérez-González, & Jiménez-Fernández, 2017).[4] German intelligence research has also been influenced by the CHC model (e.g., see Baghaei & Tabatabaee, 2015), as best illustrated by its incorporation in the German-based Berlin Intelligence Structure (BIS) program of research literature (Beauducel, Brocke, & Liepmann, 2001; Süß & Beauducel, 2015), In addition, the Wuerzburger Psychologische Kurz-Diagnostik

(WUEP-KD), a neuropsychological battery used in German-speaking countries, is grounded in the CHC model (Ottensmeier et al., 2015).

Additional emerging CHC outposts in northern Europe include the Netherlands and Belgium. Hurks and Bakker (2016) reviewed the influence of CHC theory, as well as the neuropsychological planning, attention, simultaneous, successive (PASS) processing theory (see Naglieri & Otero, Chapter 6, this volume), in a historical overview of intelligence testing in the Netherlands. A strong indicator of the growing interest in CHC theory was a CHC theory and assessment conference at Thomas More University, Antwerp, Belgium, in February 2015. Faculty at Thomas More University have developed a CHC assessment battery for children in Flanders that measures the CHC domains of Gf, Gc, Gv, Gwm, and Gs.[5]

Transported via the Chunnel to the United Kingdom and Northern Ireland, the CHC flame has been lit, but has not yet resulted in significant CHC test development. In the 1990s, the WJ III author team was consulted to develop Irish norms for the WJ III. One of the WJ III authors (Fredrick Schrank) visited with several universities in Ireland (University College Dublin, in particular) and continues to do so with regard to the WJ IV (F. A. Schrank, personal communication, March 2, 2017). The CHC theory is now the dominant cognitive taxonomy taught in Irish psychology departments (T. James, personal communication, March 3, 2017).

In the Middle East, known CHC activities have been occurring in Jordan and Turkey. Under the direction of Bashir Abu-Hamour (Abu-Hamour & Al-Hmouz, 2014, 2017; Abu-Hamour, Al Hmouz, Mattar, & Muhaidat, 2012),[6] the CHC-based WJ III has received considerable attention, and the WJ IV was recently translated into Arabic, adapted, and nationally normed in Jordan (Abu-Hamour & Al-Hmouz, 2017). In Turkey, the first national intelligence test (Anatolu–Sak Intelligence Scale, or ASIS)[7] was developed between 2015 and 2017. Although the ASIS composite scores are not couched in the CHC nomenclature, the theories listed as the foundation for the Turkish ASIS are general intelligence, CHC, and PASS. Also, one of us (Kevin S. McGrew) worked with two universities in 2016 in the preparation of government-sponsored grant proposals for additional national intelligence test development in Turkey; both proposals involved using the CHC taxonomy.

Pivoting toward Asia and the world "down under" reveals major CHC test development efforts. Since the publication of the CHC-based WJ III, several key universities and an Australian publisher have delved deep into CHC theory and assessment. Psychological Assessments Australia[8] has translated, adapted, and normed the CHC-based WJ III and WJ IV in Australia and New Zealand. The Melbourne area has been a flashpoint for CHC training and research. Neuropsychologist and researcher Stephen Bowden and his students at the University of Melbourne have produced a series of multiple-sample confirmatory factor analysis (CFA) studies with markers of CHC abilities to investigate the constructs measured by neuropsychological tests (see "CHC-Based Neuropsychological Research and Assessment," below). The University of Monash, initially under the direction of John Roodenburg, and subsequently by his students, placed the CHC model at the core of their assessment course sequence and have infused the CHC framework into the assessment practices of Australian psychologists (James, Jacobs, & Roodenburg, 2015).

Finally, one of the most ambitious CHC test development projects has been occurring in Indonesia since 2013.[9] Sponsored and directed by the Yayasan Dharma Bermakna Foundation, a nationally normed (over 4,000 individuals), individually administered CHC-based battery of tests for school-age children (ages 5–18) is, at this writing, nearing completion. The AJT Cognitive Assessment Test (AJT-CAT) will be one of the most comprehensive individually administered tests of cognitive abilities in the world. The AJT-CAT currently consists of 27 individual cognitive tests designed to measure 21 different narrow CHC abilities (and two psychomotor tests to screen for motor difficulties). Preliminary CFA indicated that the battery measures eight broad CHC cognitive domains (Gf, Gc, Gv, Gwm, Ga, Gs, Gl, Gr) and the Gp motor domain.[10]

CHC Cross-Disciplinary Research and Applied Impact

Not only has CHC theory had a major impact on psychological assessment contexts in the United States and abroad; the usefulness of the taxonomy has been recognized for other important research functions in psychology, other disciplines, and applied settings. The common CHC taxonomic coding system allows for a diverse variety of cognitive measures to be synthesized in theoretically meaningful comparisons across extant research literature bases. Several examples are presented.

The CHC taxonomy has been used to organize meta-analyses that investigate the relations between cognitive ability constructs and other variables across numerous studies. An excellent example is Loughman, Bowden, and D'Souza's (2014) use of the CHC taxonomy to classify a diverse set of cognitive tests used across 26 different studies of idiopathic generalized epilepsy. By classifying the dependent cognitive variables according to the CHC taxonomy, these researchers used meta-analytic methods to present cognitive deficit effect sizes associated with idiopathic generalized epilepsy by the CHC domains of Gc, Gf, Glr, Gs, Gwm, and Gv, as well as executive functions.

Other examples come from health researchers studying the links among nutritional enhancements, cancer, and prenatal exposure to pathogens and cognitive functioning. Pase and Stough (2014; Stough & Pase, 2015) have conducted and reviewed research on pharmacological methods for improving cognitive functioning with a focus on dietary and herbal supplements. Like researchers in many areas of psychology, Stough and Pase (2015) noted:

> It is often hard to compare and contrast the results of individual trials given the differences in the cognitive tests, sample demographics, and products used across studies . . . there is little consensus in the field about how to select, analyze, and report on cognitive outcomes. Cognitive composite scores are often created by combining tenuously related cognitive tasks without adequate justification. We recently suggested that the Cattell–Horn–Carroll (CHC) cognitive framework could be applied to help organize cognitive outcomes in clinical trials according to validated latent cognitive factors . . . the CHC model can help researchers measure and categorize the effects of a specific intervention against the full spectrum of human cognitive abilities. Although the utility of the CHC is in helping individuals categorize cognitive tests into broad and narrow abilities, one could also use the CHC to examine the effects of a pharmacological intervention on general intelligence (g). (p. 180)

Researchers at the U.S. National Cancer Institute have also conceptualized measures of cognitive functioning according to CHC theory in research focused on the impact of specific types of cancer on cognitive outcomes (Padgett, 2015). Finally, clinical epidemiologists have used the CHC model in causal modeling studies focused on the impact of prenatal exposure to methylmercury in cognitive functioning in young adulthood in longitudinal datasets (Debes, Weihe, & Grandjean, 2016).

Psychologists working in military and defense settings have recognized the value of the CHC taxonomy in military selection and classification research as well (Rumsey & Arabian, 2014). A National Research Council (2015) report titled *Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance for Military Accession* referenced the CHC-related research of both John Horn and John Carroll. Research on the selection and training of Canadian military pilots has also used the CHC model to categorize cognitive and aptitude measures (Forgues, 2014; Herniman, 2013). The U.S. Federal Aviation Administration has conducted research on the aptitude requirements needed for air traffic control specialists and utilized the CHC taxonomy in the analysis of the various aptitude measures (Broach, 2013).

Recognition of the CHC taxonomy has also played a role in stimulating new debates in the field of industrial–organizational (I-O) psychology. Several I-O psychologists have argued for the I-O field's need to reconsider its relatively long-standing neglect of the construct of intelligence. In the target article of a special issue of *Industrial and Organizational Psychology,* Scherbaum, Goldstein, Yusko, Ryan, and Hanges (2012) challenged I-O researchers—in part because of the progress made in building validated psychometric (CHC) and other models of intelligence—to reengage in the study of intelligence, since contemporary research on general intelligence and specific cognitive abilities has shown greater promise than the older "all there is is *g*" I-O consensus. Schneider and Newman (2015) echoed the same messages articulated by Murphy (2017) and Scherbaum and colleagues (2012) in the human resource management literature.

Moreover, the CHC model is now involved in life-or-death decisions.[11] Recognizing the variability in forensic intelligence-testing practice and resulting diagnostic chaos, several forensic psychologists have argued for the use of the CHC taxonomy to better organize forensic research and to provide stronger validity for the interpretation of intelligence test results in *Atkins* cases and other criminal procedures (Habets, Jeandarme, Uzieblo, Oei, & Bogaerts, 2015; Uzieblo, Winter, Vanderfaeillie, Rossi, & Magez, 2012). Although the latest classification manual of the American Association on Intellectual and Developmental Disabilities (AAIDD; 2010) focused primarily on IQ as a deficit in general intelligence and did not recognize the CHC model, AAIDD has subsequently published a guidebook specifically on the death penalty (Polloway, 2015) that, in the chapters on intellectual functioning (McGrew, 2015) and intelligence testing (Watson, 2015), places the CHC model front and center as the cognitive taxonomy for understanding intelligence test scores in *Atkins* cases. We believe that this publication sets the stage for AAIDD's formal recognition of the importance of understanding multiple cognitive abilities (viz., CHC theory) and reducing the exclusive focus on the global IQ score in the next revision of the official AAIDD classification manual.

Other uses of the CHC taxonomy have been reported in such diverse contexts as international Paralympic athletics; computer science education and the Internet; and the search for extraterrestrial (ET) life. Van Biesen, Mactavish, McCulloch, Lenaerts, and Vanlandewijck (2016) investigated the relation between nine CHC-classified cognitive measures and tactile proficiency in well-trained table tennis players with intellectual disabilities (members of the International Federation for Para-Athletes with Intellectual Disabilities). Tsianos and colleagues (2013) used the CHC taxonomy to organize the individual-differences user profile layer, the adaptation mapping layer, and the web content layer in an "ontological adaptation mechanism" for designing personalized web environments. Also in the field of computer science, Román-González and colleagues (2017) used the CHC model to investigate the cognitive abilities (Gf, Gv, Gwm) measured by the Computational Thinking Test, a measurement of computer science knowledge in educational settings.

CHC theory's role in the search for ET does not reference any real practical application of the CHC taxonomy in astrobiology. It is mentioned here simply to indicate that the CHC theory received honorable mention (viz., a citation of the McGrew [2009] CHC article in *Intelligence*) in the article "Astrobiology in Culture: The Search for Extraterrestrial Life as 'Science'" (Billings, 2012) when the problem of defining ET life was compared with that of the problem of defining intelligence in psychology.

Finally, the CHC model of "intelligence" is informing another important type of "intelligence"—national security intelligence. The U.S. Office of the National Director of Intelligence sponsors research programs focused on increasing the critical problem solving and analytic thinking of U.S. intelligence analysts. One is the Strengthening Human Adaptive Reasoning and Problem-Solving (SHARP) Program, focused on the advancement of the science for optimizing human adaptive reasoning and problem solving.[12] "The goal of the program is to test and validate interventions that have the potential to significantly improve these capabilities, leading to improvements in performance for high-performing adults in information-rich environments" (*https://tinyurl.com/n7ww3zc*). CHC-organized research and models have been incorporated in the conceptualization, implementation, and evaluation of the SHARP research program (Hartman et al., 2017).[13]

A CONVERGENCE IN INTELLIGENCE TEST BATTERY THEORETICAL FRAMEWORKS?

It is official. Pigs are flying. Hell has frozen over. Why? Because, as astutely pointed out by A. S. Kaufman, Raiford, and Coalson in *Intelligent Testing with the WISC-V* (2016), a form of rapprochement in theory as applied to test construction has occurred among the authors and publishers of the primary individualized intelligence batteries used in the field. There has always been a healthy degree of competition among the authors and publishers of the most frequently used intelligence batteries. Despite obvious differences in the latest versions of the competitive Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) and WJ IV, the foundational frameworks for the respective batteries may reveal more similarities than differences (A. S. Kaufman et al., 2016).

Clearly, the Sputnik-like CHC moment among intelligence test publishers was the 1989 publication of the WJ-R (A. S. Kaufman, 2009; McGrew, 2005; Schneider & McGrew, 2012). The WJ-R caught the rest of the intelligence test development community largely off guard when it appeared as an operationalization of the extended Cattell–Horn Gf-Gc theory and included John Horn and John Carroll (the HC in CHC) as intelligence theory consultants. Since that moment, the authors and publisher of the subsequent WJ III and WJ IV batteries have been the standard bearers of the CHC structural test development movement. Other intelligence batteries quickly followed suit, either by explicitly organizing revisions of intelligence batteries according to CHC theory (e.g., the Stanford–Binet Intelligence Scales, Fifth Edition [SB5] and the Kaufman Assessment Battery for Children—Second Edition [KABC-II]), or by implicitly providing CHC interpretation material in their manuals (e.g., the Differential Ability Scales–II [DAS-II]) (Keith & Reynolds, 2010).

The publisher of the traditional leaders in sales of intelligence tests (viz., the Wechsler scales) has been more cautious in its reaction to the CHC testing movement launched by the WJ-R. However, the incremental shift in recognizing the CHC model and incorporating it into the Wechslers can be seen—first in the revision of the WISC-III to the WISC-IV four-index model (Verbal Comprehension, Gc; Working Memory, Gwm; Processing Speed, Gs; and Perceptual Organization, Gv/Gf), and then in the further changes to the WISC-V, which has a much cleaner CHC structural model. Specifically, the WISC-V has Verbal Comprehension (Gc), Visual–Spatial (Gv), Fluid Reasoning (Gf), Working Memory (Gwm), and Processing Speed (Gs) index scales, together with the select ancillary index scales of Quantitative Reasoning (Gf-RQ), Naming Speed (Gr-LA), and Storage and Retrieval (Gl and Gr) (Flanagan & Alfonso, 2017). A. S. Kaufman and colleagues (2016) acknowledged that CHC theory has had a significant impact on the WISC-V. However, they noted that this theory "is not the sole or primary force behind revisions of the Wechsler scales" (p. 13). The Wechsler research directors remained true to the historical and traditional foundations of the Wechslers by also placing focus on clinical utility and other functional theories, "including processing theories from the fields of cognitive psychology, as well as various theories related to more specific aspects of working memory, attention, and executive function" (p. 13).

While the WJ III was the most structurally pure approach to measuring the CHC model, A. S. Kaufman and colleagues (2016) correctly note that the WJ IV authors added similar functional theoretical research to the rationale for some of the new WJ IV tests. The WJ IV authors refer to the incorporation of neuroscience and working memory research in the WJ IV test development framework as going "beyond CHC" (McGrew, LaForte, & Schrank, 2014). Although the publishers and authors of the WISC-V and WJ IV may not explicitly acknowledge this apparent theoretical framework convergence (WISC-V = functional theories + CHC theory; WJ IV = CHC theory + functional theories), we believe that A. S. Kaufman and colleagues are largely correct when they state that "the theoretical basis for the development of the Wechsler intelligence scales (as well as the theoretical foundations for the WJ IV) would be more correctly termed a theoretical framework, with content reflecting aspects of both structural (e.g., CHC theory) and functional (e.g., neuropsychological processing theory) theories of intelligence or cognitive ability as well as current, relevant research related to theory and practice" (p. 14). Although the Wechsler and WJ batteries have probably not arrived on the same street or block of the theoretical framework neighborhood, it does appear that a form of theoretical framework convergence is occurring in the field of intelligence test development. CHC theory has played a fundamental role in this convergence.

The major exception to the advance of CHC theory into the major test batteries is the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014), which marches to the beat of its own PASS theory drummer. We believe that PASS theory (Das, Naglieri, & Kirby, 1994) has much to add to CHC theory. For example, we believe that the simultaneous–successive distinction in PASS theory is valid, and we do not believe that it competes with the kind of distinctions made in CHC theory. Despite differences in labels and emphasis, there is already considerable overlap between PASS and CHC theory constructs (Flanagan, Alfonso, & Dixon, 2014; A. S. Kaufman & Kaufman, 2004; Keith, Kranzler, & Flanagan, 2001; Keith & Reynolds, 2010; Kranzler & Keith, 1999; Schneider & Flanagan, 2015).

CHC-BASED NEUROPSYCHOLOGICAL RESEARCH AND ASSESSMENT

In our 2012 chapter, we indicated that neuropsychological assessment was the most active CHC assessment "spillover." This judgment was based on the appearance of several texts on school neuropsychological assessment (Hale & Fiorello, 2004; Miller, 2007, 2010). We continue to believe that CHC theory has the potential to help neuropsychologists generalize their interpretations beyond specific test batteries and give them greater theoretical unity. Previously, we referenced a limited number of CHC-organized factor-analytic studies of joint neuropsychological and CHC-validated batteries (Floyd, Bergeron, Hamilton, & Parra, 2010; Hoelzle, 2008), and concluded that there was a critical need for more research. Several important CHC-based CFA analyses of neuropsychological and cognitive batteries have occurred since our last foray into the neuropsychological sandbox. A number of these analyses have focused on asking whether executive functions have a place in the CHC model.

In a sample of nondisabled school-age children, Ottensmeier and colleagues (2015) used the short version of the WUEP-KD battery to examine the CHC factor structure of common cognitive and neuropsychological specific measures. Although only using single indicators to represent Gf, Gv, Ga, and Gwm in an exploratory factor analysis, multiple subcomponent scores from a continuous-performance test (CPT) and a finger-tapping measure revealed three distinct factors the authors interpreted as Gs (CPT speed and hits), Gt (CPS visual and auditory reaction time), and Gps/Gp (finger-tapping speed and steadiness scores). The interpretation of neuropsychological tests (CPT, finger tapping) as measuring CHC constructs instead of common neuropsychological constructs (e.g., planning, attention, executive functions) suggests that the CHC and neuropsychological nomenclature systems may represent an example of the *jingle–jangle* fallacy.[14] The psychometrically driven CHC taxonomy describes abilities measured from an individual differences construct trait framework, while traditional neuropsychological assessment describes abilities as reflecting functional neuropsychological constructs (see Hoelzle, 2008, for explanation of the historically different origins of these two approaches to assessment).

The most important recent CHC/neuropsychological research has been conducted by Stephen Bowden and colleagues at the University of Melbourne (see "CHC's Global Reach," above). These researchers completed three sets of Carroll-like (1993) CHC-based mini-meta-analytic CFA studies across multiple samples and multiple assessment instruments (Jewsbury & Bowden, 2016; Jewsbury, Bowden, & Duff, 2017; Jewsbury, Bowden, & Strauss, 2016).

In the Jewsbury and Bowden (2016) study, five datasets were analyzed. These analyses demonstrated that a variety of tests of narrow CHC fluency abilities (e.g., associative fluency, expressional fluency, ideational fluency, word fluency) loaded on a single broad CHC ability (broad retrieval fluency, or Gr). This Gr factor was distinct from the current notion of Glr, which consists of both learning efficiency (associative memory or meaningful memory; Jewsbury and Bowden used Gy as per Carroll's original model) and retrieval fluency (associative fluency, expressional fluency, etc.; Gr) components (Schneider & McGrew, 2012). A possible semantic versus orthographic Gr narrow-ability substructure was also suggested. This set of integrated studies indicated that most common neuropsychological tests of fluency (e.g., Boston Naming Test, Controlled Oral Word Association Test, Letter Fluency, Category Fluency), which are typically interpreted as measuring executive functioning, are measuring constructs similar to those assessed in cognitive test batteries (e.g., WJ IV Retrieval Fluency; WISC-V Naming Speed and Naming Quantity) and should be interpreted as indicators of a broad retrieval ability (Gr) according to the CHC framework. This research also suggests that the ability of retrieval fluency may be a specific component of the construct of executive functioning.

As in Jewsbury and Bowden's (2016) Gr analysis, Jewsbury and colleagues (2017) used the CHC model as the framework for a secondary CFA of nine different datasets that included multiple measures of cognitive and neuropsychological abilities (viz., executive functions). After systematically evaluating different CFA models specified to answer their primary research questions, Jewsbury and colleagues reported that

> the CHC model fit well across all data sets . . . the CHC model is an excellent fitting model that is replicable and consistent across diverse tests and populations. In particular, the data sets together provided replicated evidence for the CHC construct validity for many of the most popular neuropsychological tests and batteries. . . . Furthermore, the CHC construct validity was supported across a range of clinically relevant populations . . . the CHC model was found to apply equally well to traditional instruments such as the WAIS and putative executive function measures that are commonly believed to measure constructs beyond the CHC constructs. (p. 557)

These researchers found that common executive function measures loaded across the CHC abilities of Gf, Gwm, Gv, or Gs. Jewsbury and colleagues (2017) concluded that executive function is not a unitary construct and should not be considered as a separate CHC domain of cognition similar to other broad CHC constructs: "Averaging or combining various executive function test scores potentially leads to results that confound cognitive constructs" (p. 560).

In a third set of analyses, Jewsbury and colleagues used CHC-organized CFA in seven datasets to evaluate the validity of the executive function subcomponents of updating (replacing old information in short-term memory span with new information), inhibition (inhibiting responses when necessary), and switching (switching between multiple tasks) (Miyake et al., 2000). The authors reported no support for a separate updating factor distinct from Gwm or a separate inhibition factor distinct from Gs. The researchers did conclude that switching may represent a new narrow ability under Gs. The authors suggested that (1) all Gs tests may involve the ability to inhibit; (2) the ability to inhibit may be a function of Gs; or (3) a third variable (e.g., maintenance of a goal state) may underlie both inhibition and Gs.

The convergence of conclusions across the collective CHC-organized research on executive functioning cannot be ignored. Executive functioning does not represent an individual differences trait construct that should be incorporated in the CHC taxonomy. Also, as noted by Salthouse (2005), "the outcomes of these analyses therefore suggest that measures currently used by neuropsychologists to assess executive functioning may not represent novel aspects of functioning in normal adults. Instead they appear to reflect the same dimensions of differences assessed by more traditional cognitive tests that may have superior psychometric properties (in terms of reliability and sensitivity)" (p. 542). The use of the term *executive functioning* in the interpretation of cognitive and neuropsychological assessments should be restricted in future research (Floyd et al., 2010).

Additional research is needed to determine whether the components of executive functions mentioned in the neuropsychological literature represent combinations, blends, or amalgams of different CHC abilities in a distributed system, or some yet-to-be-validated executive function processes that are also part of certain CHC abilities. It is also possible that executive functioning does not represent a real "thing" or psychological trait construct; instead, it may be an emergent variable reflecting "control and controlled processes [which] are colocalized within larger numbers of dispersed computation agents" (Eisenreich, Akaishi, & Hayden, 2017, p. 1684), not a factor-analytic individual-differences variable (Kovacs & Conway, 2016).

CRITERIA FOR UPDATING CHC THEORY

A theory that can only be updated by one person is more like a copyrighted work of art than a product of open science. Yet to let just anyone alter CHC theory would lead to chaos, conflict, and Wikipedia-style "edit wars," with contributors constantly overwriting each other's amendments. To keep the peace, we will continue to curate CHC theory for as long as we can, but we would like to make the updating process more open and transparent, so that any researcher can contribute to the evolution of CHC theory. Here we outline a tentative set of guidelines for making future amendments to CHC theory.

Robust Evidence Required for Overturning the Measured Opinions of Cattell, Horn, and Carroll

A theory named after Cattell, Horn, and Carroll should reflect the ideas of the source theorists except when persuasive evidence contradicts their ideas. In the initial formulation of CHC theory (McGrew, 1997), most of the nomenclature was borrowed directly from Carroll (1993) rather than from Horn and Cattell. The reason for this choice was that Carroll's work was largely congruent with or even explicitly derivative of the work of Horn and Cattell, and yet was more comprehensive and persuasive. A work that summarizes hundreds of datasets is hard to beat!

Of course, Carroll's is not the final word on the nature of intelligence, as he was careful to point out himself (Carroll, 1998). Although we do not treat Carroll's work as inerrant scripture, we nevertheless cannot think of a better starting point than his 1993 masterwork. Going forward, our default hypothesis is that Carroll's theorizing is more likely to be correct than any other single person's speculations, including our own. As various sections of this chapter indicate, we have often found it necessary to reread Carroll's original work when contemplating new questions about existing or possible new abilities. Bestowing this power on Carroll's treatise should not stop us or anyone else from speculating—but making a better theory than Carroll's requires evidence Carroll did not have. Researchers hoping to amend CHC theory should think of Carroll's work as providing a kind of "null hypothesis" to which their results can be compared.

Introducing New Ability Constructs to CHC Theory

We are not the first to wrestle with ideas about amending established taxonomies. We have borrowed liberally from criteria laid out by Gardner (1983) and Mayer, Caruso, and Salovey (1999), and we have been guided by principles laid out by

Cronbach and Meehl (1955). Using a more specific proposal by Schneider, Mayer, and Newman (2016) as our model, we suggest that the following criteria be applied:

1. *The content domain of the new ability must be laid out clearly.* Researchers who might study the new ability should have a shared understanding of the new construct, so that they can independently develop new measures of the construct.

2. *The new construct must be measurable with performance tests and with multiple test paradigms.* Although it is possible to estimate abilities with nonperformance tests (e.g., questionnaires), performance tests remain the clearest indicators of ability constructs. The new construct must also be measurable with a variety of test formats; otherwise, useless method factors would invade and overwhelm the taxonomy.

3. *Measures of the new ability construct must demonstrate convergent and discriminant validity.* In general, measures of the same ability should correlate more highly with each other than they do with measures of other abilities. A rigorous evaluation of this criterion is usually conducted via some form of covariance structure modeling, such as CFA.

4. *Measures of the new ability must demonstrate incremental validity over measures of more established constructs.* Categorizing new abilities must not be an end in itself. Ability constructs exist because they help us understand relevant outcomes. A new ability construct should improve our prediction of at least one outcome widely considered to be important (e.g., academic, occupational, or creative achievement).

5. *The new ability construct should be linked to specific neurological functions.* Evidence that a new ability is linked to specific neurological functions might include the demonstration that localized brain injuries cause selective impairments in the new ability, that distinct brain networks are distinctly associated with the ability, or that specific genetic variants are selectively associated with extremes of the ability.

6. *The new ability construct should be linked plausibly to functions that evolved to help humans survive and reproduce.* This criterion prevents an endless proliferation of theoretical entities linked to specific areas of achievement. With little effort, one could postulate a separate intelligence for each identifiable achievement domain (e.g., business intelligence, medical intelligence, gaming intelligence, sports intelligence, fashion intelligence, computer programming intelligence), until the taxonomy is so cluttered that it no longer serves its purpose. Certainly, cognitive abilities are relevant in all these domains, but most likely they require complex mixtures or amalgams of more basic cognitive abilities rather than qualitatively distinct abilities.

We acknowledge the role of subjectivity in the evaluation of each of these criteria. To satisfy each of these criteria requires many studies and a lengthy process of scientific deliberation before consensus is achieved. To wait until the evidence is so compelling that no sensible scholar could possibly disagree with the decision to include a new construct would unnecessarily hamper the development of CHC theory. To strike a balance between creative risk and proper prudence, new abilities that have reasonable but not yet compelling evidence will be given tentative status in the CHC taxonomy. When evidence comes from many high-quality studies from independent teams of scholars, the provisional status will be removed, with the proviso that in science all constructs are provisional.

Nearly all of Carroll's broad and narrow factors have met the first four criteria, some more convincingly than others. Many meet the latter two criteria as well, but many have yet to do so. In this chapter, we have not attempted a thorough review of the evidence supporting each ability construct, but such a review will be needed to give CHC a firmer scientific foundation. To date, olfactory processing (Go), tactile/haptic processing (Gh), kinesthetic processing (Gk), psychomotor abilities (Gp), and psychomotor speed (Gps) are broad abilities with tentative status (McGrew, 2009). In this chapter, we argue that emotional intelligence (Gei) also tentatively meets criteria for inclusion in CHC theory. It would satisfy our aesthetic sense if every sensory modality had its own broad factor, but we have no empirical justification for hypothesizing a broad gustatory factor (Gg).

Reorganizing CHC Theory

With each update of CHC theory, not only have new abilities been added; we have deleted, consolidated, reorganized, and renamed ability constructs. For example, short-term memory (Gsm) has been renamed as working memory capacity (Gwm), to keep pace with current theoretical developments (Schrank, McGrew, & Mather, 2014).

Some abilities have been reclassified on logical grounds. For example, geography knowledge was moved from comprehension–knowledge (Gc) to domain-specific knowledge (Gkn) because it is a specific knowledge factor (Schneider & McGrew, 2012). After a closer reading of Carroll's evidence, the language development factor was elevated from a narrow ability to a factor intermediate between Gc and specific aspects of language (Schneider & McGrew, 2012). The cloze (reading) ability was deleted because it appears to be a method factor rather than a distinct ability. For example, cloze ability is the ability to do well on tests that make use of the cloze paradigm (Schneider & McGrew, 2012).

In this chapter, we propose several further modifications to the current CHC theory. We argue for the need to split the long-term storage and retrieval (Glr) factor into separate components of long-term storage and learning (Gl) and retrieval fluency (Gr). The Gl-Gr divorce is the most dramatic change we suggest, and the one with the most solid line of historical and contemporary research evidence. We propose several other small modifications in other domains. We propose that perceptual speed (P) is most likely an intermediate-level processing speed ability, and that two distinct subtypes of narrow perceptual speed abilities exist, differentiated by degree of information-processing complexity required and stimulus content characteristics (facets). In regard to Ga, we suggest that a speech versus nonverbal facet distinction should be considered in interpreting Ga abilities and measures. In regard to Gv, we recognize the importance of efforts to add large-scale spatial navigation abilities to the wealth of small-scale Gv abilities. For some of the tentative and more or less ignored domains (Gh, Gk, Gt), we suggest that their status might be elevated when they are viewed from the relatively recent research and theorization in embodied cognition.

It is likely that future reorganizations will be needed as future evidence accumulates. Although the goal is to have an elegant, parsimonious theory, the process of tidying can itself be a bit messy. In many respects, the process of reorganizing CHC theory is like that of adding new constructs. That is, the criteria must be clear and tied to performance; the proposal must reference convergent, discriminant, and incremental validity; and there should be some reference to biological functions that have a plausible evolutionary history.

Beyond these criteria, however, there are additional considerations of parsimony, coherence, utility, and aesthetics. Here we cannot offer an enumerated list of principles to guide this process. Offering hard and fast principles would be as dangerous as being infected by *statisticism*, which is "a way of thinking in psychology that invests virtually boundless trust in the aptness of statistical concepts and methods to reveal the 'lawfulness' of human psychological functioning and behavior" (Lamiell, 2013, p. 65). Although the fathers of CHC theory were clear statistical wizards and believers in factor analysis, a thorough reading of their writings (as well as informal discussions) reveals that advocacy for a specific form of statistical factor-analytic method to support or defend CHC theory (e.g., the bifactor vs. hierarchical intelligence debate), devoid of clarity of theoretical context (e.g., genetic, neurocognitive, developmental, or outcome criterion evidence) and well-reasoned thought, will never provide a reasonable taxonomy of human intelligence that represents "intelligence in the wild." We can say that we will be as open in the future as we have in the past to persuasion via reason and evidence.

CHC MODEL REVISIONS AND EXTENSIONS

CHC Cross-Battery Research

A pivotal foundation of CHC cross-battery assessment (XBA; Flanagan, Ortiz, & Alfonso, 2013), which has contributed significantly to the extant construct validity evidence for the CHC model, has been CHC-organized XBA research. In XBA factor studies, typically two different cognitive test batteries (e.g., the WJ-R and WISC-III) are jointly factor-analyzed, and the results are interpreted according to the CHC model. CHC-organized XBA studies originated with Woodcock's (1990) seminal series of joint analyses of the WJ-R with all other major intelligence batteries at the time. This eventually resulted in the formalization of the XBA approach to assessment (for historical overviews, see McGrew, 2005; Schneider & McGrew, 2012). Although exploratory factor analysis and CFA studies of individual intelligence test batteries have continued unabated during recent years, there has been a noticeable drop since our 2012 chapter in joint intelligence test XBA studies. There have been studies that included an intelligence test (e.g., one of the Wechsler Adult Intelligence Scale [WAIS] series of tests) and neuropsychological measures (see "CHC-Based Neuropsychological Research and Assessment," above), but we could only find two CHC-organized or CHC-interpreted XBA studies of intelligence batteries.

Primi, Nakano, and Wechsler (2012) completed a Brazilian-based XBA study that included five individual tests from the Battery of Reasoning Abilities (BPR-5) and three individual tests from a Brazilian adaptation of the WJ III. The study provided validity evidence for the BPR-5 tests as measures of Gf, Gv, and Gc. The other study was an impressive demonstration of a CHC-model-driven methodological advance in XBA research. Reynolds, Keith, Flanagan, and Alfonso (2013) completed a CHC-organized, XBA, reference-variable CFA of five different cognitive or achievement test batteries (the KABC-II, WISC-III, WISC-IV, WJ III, and Peabody Individual Achievement Test—Revised/Normative Update [PIAT-R/NU]). In this study, the CHC taxonomy was married with the sophisticated reference-variable CFA methods advanced by McArdle (1994). This methodology allowed for the analysis of data from the five different test batteries in five separate KABC-II concurrent validity studies (with various KABC-II tests serving as common reference-variable links or anchors across studies) in a single combined joint analysis (as if all five batteries had been administered to all participants). That is, this CHC-organized analysis made it possible to evaluate the CHC abilities measured by 39 different tests across test batteries in a single grand analysis. The final model supported the CHC constructs of Gc, Gv, Gf, Gl-MA, and Gwm.

For the CHC XBA approach to assessment and interpretation to evolve, more CHC-designed XBA studies are needed. Stand-alone or joint intelligence battery CHC CFA studies have focused only on the broad (stratum II) CHC domain level. Yet the interpretation of CHC-based intelligence tests or XBA-organized test data has, at its core foundation, the valid classification of the individual tests at the narrow (stratum I) level. There is a serious need for CHC studies that allow for the evaluation of individual tests as indicators of CHC narrow-ability constructs. We recognize that such studies are prohibitively resource-intensive in terms of cost and time (including participant response burden) and may not be practical. Instead, we suggest designing a series of smaller XBA studies that do not attempt to include indicators of all the primary CHC domains. Instead, these more focused studies would include sufficient indicators to specify key narrow CHC abilities for two or three broad domains in a single analysis. If properly designed, a series of narrow-ability-focused studies could be linked with the reference-variable methods demonstrated by Reynolds and colleagues (2013). Such a program of research would provide important insights into the narrow CHC abilities measured by individual tests.

The Glr Divorce: Gl and Gr

In our 2012 chapter, we presided over the formal separation of Glr into Glr-learning efficiency and Glr-retrieval fluency. We recognized that the marriage of these two components of Glr was a union in name only. As we stated in that chapter, "there is a major division within Glr that was always implied in CHC theory, but we are making it more explicit here. Some Glr tests require efficient learning of new information, whereas others require fluent recall of information already in long-term memory" (p. 117). In our functional conceptualization of the CHC model, Glr-learning efficiency (hereafter labeled Gl) and Glr-retrieval fluency (hereafter labeled Gr) were considered parameters of cognitive efficiency. Gl determines how much effort is needed to store new information (of various kinds) in long-term memory, and Gr represents the speed at which information in long-term storage can be loaded into working memory structures for further cognitive processing and use. It is now clear not only that the original Glr marriage was a mistake, but that the trial separation of Gl and Gr was not the long-term solution to the marital mismatch. Therefore, we now formally grant the parties a divorce in the eyes of CHC theory. It is our belief that the separate households named Gl and Gr represent distinct broad CHC domains.

Dissecting why a marriage has gone wrong is not always beneficial. However, in this case, we briefly explain the mistake in the original Glr partnership, as it provides an important cautionary tale in the pursuit of science. The Glr marriage was presided over by McGrew (1997) in his chapter for the first edition of this book, wherein individual tests from all major intelligence batteries were first classified according to the original integration of the Cattell–Horn and Carroll models of cognitive abilities (then called a "proposed synthesized Carrell and Horn–Cattell Gf-Gc framework"). McGrew then used both logical analysis and the results from a special CFA of the WJ-R standardization battery in an attempt to resolve a number of differences between the Carroll and Cattell–Horn models: For instance, is Gq a separate domain from Gf? And should Grw be included under Gc or as a separate broad domain? The question most relevant to the current discussion was whether short-term memory, learning efficiency, and retrieval fluency abilities should be organized under Gsm (Horn's short-term acquisition and retrieval, SAR), Glr (Horn's tertiary storage and retrieval, TSR), Gy (Carroll's memory and learning), or Gr (Carroll's broad retrieval). With no fluency indicators present in the WJ-R, McGrew found support for a narrow associative memory (MM) factor that he considered a proxy for the broad Glr factor, based primarily on the belief that the Cattell–Horn conceptualization of the TSR factor was more correct than Carroll's overly broad Gy factor (which was judged more untenable). In hindsight, the lack of fluency indicators in the WJ-R analysis was a significant contributing factor to the incorrect combination of learning efficiency (e.g., associative memory, meaningful memory) and retrieval fluency (e.g., ideational fluency, naming speed). Despite appropriate caveats that his initial synthesized framework was "only an initial attempt" and only a "proposed framework," McGrew's Glr union stuck and was crystallized in all subsequent CHC articles and book chapters by multiple scholars. The weak foundation upon which it rested was not reevaluated until 15–20 years later.

With the benefit of hindsight, we can see that cracks in the Glr construct were present in the CHC-grounded norm-based CFAs reported in the WJ III technical manual (McGrew & Woodcock, 2001). In the CFAs of the WJ III, the Glr factor had moderate (.60s) to strong (.70s to .80s) loadings for four indicators of associative memory (Memory for Names, Memory for Names—Delayed; Visual–Auditory Learning, Visual–Auditory Learning—Delayed), while the new WJ III Glr fluency tests (Retrieval Fluency, Rapid Picture Naming) had relatively weak loadings (.33, .18) on the Glr factor, and stronger loadings on the Gs factor (.33, .41). Furthermore, significant correlated residuals (.23) between Retrieval Fluency and Rapid Picture Naming suggested unexplained shared variance between these two retrieval fluency measures. Finally, the Glr factor had an unexpectedly high loading (.95) on the *g* factor. Collectively, these findings could have been interpreted as indicating that the WJ III tests of associative memory and retrieval fluency did not form a cohesive Glr factor, and that separate Gl and Gr factors might be present.

The WJ III authors found similar results in the WJ IV (Schrank et al., 2014) and separated these dimensions into the Cognitive battery Glr cluster (it would have been better to call this Gl), comprising two measures of learning efficiency (Visual–Auditory Learning, associative memory; Story Recall, meaningful memory), and the Oral Language battery Speed of Lexical Access cluster (Glr-LA, which could be interpreted as a proxy for Gr; Retrieval Fluency, ideational fluency; Rapid Picture Naming, naming speed). McGrew and colleagues (2014) reported a series of alternative WJ IV broad + narrow CHC CFA models in which the speed of lexical access (LA) factor, defined by salient loadings for Retrieval Fluency, Rapid Picture Naming, and Phonological Processing, did not load on the Glr factor but instead loaded on the broad Gwm and Gc factors. A model with a stand-alone Gr factor (as defined above) was not reported by McGrew and colleagues. Subsequently, McGrew (2015) completed an unpublished CFA model where the separate broad Gr factor was specified (as per the three-indicator LA factor described above) in the WJ IV norm data for ages 6–19. Although the model fit was not significantly different from that of the broad + narrow CHC model presented in the WJ IV technical manual, the model supported the viability of separate Gl and Gr factors. Gl was defined primarily by Story Recall, Visual–Auditory Learning, Memory for Names,[9] and to a lesser extent Reading Recall and Writing Samples; Gr was defined primarily by Retrieval Fluency, and to a lesser extent Rapid Picture Naming and Phonological Processing (which also had loadings on Gs and Ga, respectively). The reduction of the original combined Glr factor loading on the *g* factor (.95) to .85 (Gl) and .56 (Gr) is more theoretically reasonable. Also, noticeably different cross-sectional growth curves for the WJ IV Glr (Gl-learning efficiency) and LA (Gr-retrieval fluency) clusters (McGrew et al., 2014) provide developmental evidence supportive of these measures as representing different CHC ability domains. The practical implications for the WJ IV Glr and Speed of Lexical Access clusters are discussed in the "CHC Theory Described and Revised" section of this chapter.

Other recent structural analysis studies support separate Gl and Gr broad abilities. In the first official recognition of the separate Gl and Gr factors in the CHC model, Affrunti, Schneider, Tobin, and Collins (2014) found support for separate Gl and Gr factors, along with validation for the broad CHC factors of Gc, Ga, Gv, Gs, and Gwm (and a narrow memory span factor) in a large clinical dataset of university students (*n* = 865) who were administered the WJ IV Cognitive and Achievement batteries and the WAIS-III. In a sample of

317 healthy adults and a separate sample of 280 adults with clinical disorders, both of which were administered a battery of neuropsychological tests, Vannorsdall, Maroof, Gordon, and Schretlen's (2012) exploratory factor analysis found a clear ideational fluency factor (Gr). Also, the WISC-V battery includes supplementary Naming Speed and Symbol Translation indexes, which, according to a recent structural analysis of the standardization data (multidimensional scaling or MDS), measure aspects of Gr and Gl (MA, associative memory), respectively (Meyer & Reynolds, 2017). A preliminary CFA of the Indonesian AJT-CAT test also provides support for separate Gl and Gr factors (see "CHC's Global Reach," above). Finally, the previously described Jewsbury and Bowden (2016) CHC-organized CFA of five datasets that included cognitive and neuropsychological tests supports the separation of Glr into distinct Gl and Gr factors.

We would be remiss if we did not recognize the decades of research in which the concept of broad retrieval fluency has always been featured as a cornerstone of creativity. Carroll (1993) included many of the extant Gr studies in his seminal survey—studies that used a variety of fluency measures to define this core feature of divergent thinking (J. C. Kaufman, 2016). Several recent creativity studies provide additional evidence for a fluency or Gr factor (Beaty, Christensen, Benedek, Silvia, & Schacter, 2017; Beaty & Silvia, 2013; Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Silvia, 2015; Silvia & Beaty, 2012), distinct from other CHC factors (Gf, Gc). Although these studies are supportive of a broad Gr ability, they are not 100% decisive, as these studies suffer from a lack of indicators that would allow for the representation of a Gl factor and comparison of this factor to the Gr factor.

Finally, clear evidence for the Gr broad ability is present in the writings of two of the fathers of the CHC model—evidence that should have been more clearly investigated in the original 1997 Glr marriage. As noted by Carroll (1993), the existence of an ability represented by various measures of fluency has been present in research traced to Spearman and others in the late 1920s to 1950s. Cattell (1987) described this broad ability as "a general fertility or facility of memory retrieval in regard to any kind of material" (p. 46). He continued: "General retrieval, g_r, is considered an *ability* concerned with the fluency-retrieval performances, and having to do with storage and accessibility facility" (p. 447; original emphasis). Carroll was correct in his seminal treatise, where he included a Gr (broad retrieval ability) in his model. He discussed this domain in depth in a separate chapter on "Abilities in the Domain of Idea Production." According to Carroll, factors representing Gr "involve the active production of ideas as opposed to the recognition, identification, selection or comparison of ideas as represented in stimuli presented to subjects" (p. 394). Furthermore, "in describing this domain as one of idea production, I mean the term idea to be taken in the broadest possible sense. An idea can be expressed in a word, a phrase, a sentence, or indeed any verbal proposition, but it may be something expressed in a gesture, a drawing, or a particular action" (p. 394).

Tentative CHC Broad Sibling: Gei

It might seem as if adding emotional intelligence to CHC theory is a radical step. It is not. In fact, emotional intelligence has been present in CHC theory under a different name since the beginning. Carroll (1993, p. 513) was convinced by Guilford's work on social intelligence (Guilford, 1967; Guilford & Hoepfner, 1971; O'Sullivan & Guilford, 1975) that comprehending the social and emotional behavior of others is an important aspect of intelligence. Though Carroll (1993, p. 513) did not find the term *knowledge of behavioral content* to be satisfactory, he retained Guilford's nomenclature to refer to sensitivity to nonverbal communication via gestures and facial expressions. Whatever the label, this ability falls squarely within the domain currently referred to as *emotional intelligence*.

The idea that intelligence involves understanding social and emotional behavior has a long history, going back to John Dewey (1909) and E. L. Thorndike (1920). Most of the early attempts to measure social intelligence failed to distinguish it clearly from verbal ability (Hunt, 1928; Landy, 2006; R. L. Thorndike, 1936; R. L. Thorndike & Stein, 1937; Walker & Foley, 1973). However, Guilford's prodigious talent for test construction allowed him to create a novel set of social intelligence measures that were clearly distinct from measures of verbal ability, convincing even previously skeptical critics like Cronbach (1970, p. 343) that social intelligence was a worthy object of study.

There have been several attempts to integrate the large literature on the "hot" intelligences (i.e., related to personality, emotion, and interpersonal processes) with the gigantic literature on more traditional "cool" intelligences, including Guilford's

(1967) structure-of-intellect model, Gardner's (1983) theory of multiple intelligences (see Chen & Gardner, Chapter 4, this volume), and Sternberg's (1985, 1997, 1999b) theory of successful intelligence (see Sternberg, Chapter 5, this volume). Here we briefly review the relevant issues involved in such an integration, but thorough discussions can be found elsewhere (Davies, Stankov, & Roberts, 1998; Matthews, Emo, Roberts, & Zeidner, 2006; Matthews, Zeidner, & Roberts, 2004, 2005; Mayer et al., 1999; Mayer, Caruso, & Salovey, 2016; Mayer, Panter, & Caruso, 2012; Mayer, Roberts, & Barsade, 2008; Mayer, Salovey, & Caruso, 2008; Mayer, Salovey, Caruso, & Sitarenios, 2001; Schneider et al., 2016). In our review, we leave aside emotional intelligence research that draws largely on self-report questionnaires (i.e., research on so-called "trait emotional intelligence") and focus entirely on ability-based measures of emotional intelligence.

The term *emotional intelligence* gained traction upon being introduced by Salovey and Mayer (1990). Operationalized in the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002), their concept has four major components: perceiving emotions, understanding emotions, facilitating thought by using emotions, and managing emotions in oneself and in others. Consistent with the criteria we have proposed for introducing new constructs into CHC theory, this model of emotional intelligence has a well-defined content domain, so that independent researchers can develop new measures of the construct that correlate with each other well (Mayer et al., 2016). This criterion is best satisfied by measures of the ability to perceive emotions (Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014), but multiple measures of emotional understanding and emotional management are now available (Allen, Weissman, Hellwig, MacCann, & Roberts, 2014; Austin, 2010; Billings, Downey, Lomas, Lloyd, & Stough, 2014; Brackett & Mayer, 2003; Krishnakumar, Hopkins, Szmerekovsky, & Robinson, 2016; MacCann, Matthews, Zeidner, & Roberts, 2003; MacCann & Roberts, 2008; Maul, 2012; Mayer et al., 2002; Mayer, Salovey, & Caruso, 2014; Mayer, Salovey, Caruso, & Sitarenios, 2003; Palmer, Gignac, Manocha, & Stough, 2005; Schultz, Izard, Ackerman, & Youngstrom, 2001). Mayer and colleagues (2016) have conceded that measures of the facilitation factor have not been found to be psychometrically distinct from the other factors, but remain hopeful that new measures can be developed to validate that component of the model.

Well-constructed ability tests of emotional intelligence have undergone a through construct validation process (Allen et al., 2014; Brackett & Mayer, 2003; Joseph & Newman, 2010; MacCann, Joseph, Newman, & Roberts, 2014; MacCann & Roberts, 2008; Maul, 2012; Mayer et al., 1999, 2002, 2016; Schneider et al., 2016). Although there are controversies about scoring procedures to be sorted out (Zeidner, Roberts, & Matthews, 2008), it appears that these measures compare favorably with measures of traditional intelligences in terms of reliability, stability, and predictive validity (Mayer et al., 1999, 2001, 2016; Mestre, MacCann, Guil, & Roberts, 2016).

Emotional intelligence ability tests correlate with traditional measures of intelligence, generally in the range of .20 to .60, with higher correlations between the component of understanding emotions and Gc (Brackett, Rivers, & Salovey, 2011). Evidence for the predictive validity of emotional intelligence has been accumulating steadily since 1990 from the findings of dozens of independent scholars (Mayer, Roberts, et al., 2008). Ability-based emotional intelligence predicts job performance, leadership ability, negotiation skill, and salary beyond the five-factor model of personality and general intelligence measures (Blickle et al., 2009; García & Costa, 2014; Iliescu, Ilie, Ispas, & Ion, 2012; Rosete & Ciarrochi, 2005; Sharma, Bottom, & Elfenbein, 2013). Ability measures of emotional intelligence predict social competence beyond verbal ability and questionnaire measures of personality (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006; Lopes et al., 2004).

For us, the tipping point for tentatively expanding the role of emotional intelligence in CHC theory was the CFA research of MacCann and colleagues (2014). In this study, emotional intelligence measures from the MSCEIT (Mayer et al., 2002, 2003) were given alongside traditional intelligence tests. The results were consistent with the hypothesis that emotional intelligence is a broad ability in the same sense as other CHC abilities.

CHC THEORY DESCRIBED AND REVISED

In the sections that follow, we have multiple aims. First, we hope to define each of the constructs in CHC theory in terms that clinicians will find useful. Second, we hope to give some guidance as to which constructs are more central to the theory or have more validity data available. Third, we wish to alert readers to existing controversies and raise some questions of our own. Fourth, we propose a number of additions, deletions, and rearrangements in the list of CHC theory abilities.

General Intelligence (*g*)

In 1904, Spearman was the first to observe what is perhaps the most consistent and frequently replicated finding in all psychological research: all mental abilities are positively correlated. The simplest explanation of this finding is that all mental abilities have a common cause: general intelligence or *g*. Spearman's (1904) efforts to understand *g* led him to invent factor analysis. To put it mildly, Spearman's explanation of the positive correlations he observed has proved controversial. When the goddess of discord attends intelligence conferences, before tossing in her golden apple, she writes on it, "Quick question. I was curious about your opinion about *g* . . ."

Many researchers have alternative explanations of the finding that mental ability tests are positively correlated, but so far no explanation has proved persuasive to our entire field. Carroll (1993, 1998, 2003) believed that the evidence for the existence of *g* is clear. As explained later, Cattell and Horn believed just as strongly that something else is going on.

We accept that a general factor exists, but we are skeptical that it is an ability. That is, there are many factors that can simultaneously influence the entire brain, such as malnutrition, exposure to toxic substances, blunt-force trauma, large strokes, infections of the brain, and thousands of genes that independently influence the functioning of neurons. These general influences make it possible to evaluate the overall level of a person's intelligence without necessarily referring to a causal force called *general intelligence*. Analogously, there are factors that influence the quality of an automobile, such as the manufacturer's commitment to excellence, the maintenance habits of the owner, the car's collision history, and so forth. We can describe the overall quality and condition of the car, but we would not say that there is a causal force called *general car quality*.

That said, the challenge for *g* skeptics is to explain why the general factor of intelligence is such a powerful predictor of life outcomes, and why the remaining factors usually explain only a few percentage points of additional variance in most outcomes (Canivez, 2013; Glutting, Watkins, Konold, & McDermott, 2006; Gottfredson, 1997; Jensen, 1998; McDermott, Fantuzzo, & Glutting, 1990; Schmidt & Hunter, 1998). We believe that meeting this challenge is possible (Ackerman, 1996b; Benson, Kranzler, & Floyd, 2016; Evans, Floyd, McGrew, & Leforgee, 2002; Horn, 1985; Horn & Cattell, 1966; Horn & McArdle, 2007; Outtz & Newman, 2010; Schneider & Newman, 2015; Wee, Newman, & Joseph, 2014), but we concede that neither side of this controversy has compelling evidence.

There is considerable debate currently about whether *g* has direct effects on performance or indirect effects through broad and narrow abilities. The bifactor model (with direct effects of *g*) usually fits better than the hierarchical model (with indirect effects of *g*). The problem is that unless the theorist has specified exactly the right model, the bifactor model fits better even when the true model is hierarchical (Murray & Johnson, 2013). We agree with Mansolf and Reise's (2017) conclusion that no conclusion is possible on this matter yet. We take solace in our observation that there are no current practical concerns that hinge on our knowing the answer to this question. We suspect that neither the bifactor nor the hierarchical model of *g* is correct, but that the true structure of intelligence is far more interesting and far weirder than any theorists, including ourselves, have imagined.

Domain-Free General Capacities

Some CHC factors (Gf, Gwm, Gl, Gr, Gs, and Gt) are not associated with specific sensory systems. These diverse factors may reflect, respectively, different parameters of brain functioning that are relevant in most or all regions of the brain (Cattell, 1987). The fact that they are grouped together does not mean that clinicians should create composite scores with names like "Domain-Free General Capacity" because this is a conceptual grouping, not an implied functional unity.

Gf-Gc Theory

If you are unfamiliar with a problem, you can apply reason to find a solution. If you have seen the problem before, you simply need to recall whatever solution was successful in the past. These two ways of solving problems, deliberate reasoning and recalling past solutions, correspond to what Raymond Cattell (1941, 1943) called *fluid* and *crystallized intelligence,* respectively.

Fluid intelligence refers to the ability to perceive conceptual relationships. People with high fluid intelligence are able to respond insightfully to unfamiliar situations by observing complex phenomena, inferring the unstated rules that govern their behavior, and exploiting this knowledge to deduce the best course of action to take (Cattell & Horn, 1978). *Crystallized intelligence* refers to the accumulated knowledge generated via fluid intelligence. Crystallized intelligence helps us find efficient solutions from the past that best match the needs of the present. People with high crystallized intelligence have broad, deep knowledge and an extensive repertoire of skills that are useful in their cultural context.

The Fluid–Crystallized Metaphor

Cattell was a creative theorist who was playful with metaphors, and his use of the terms *fluid* and *crystallized* was nuanced and multilayered. For our purposes, we can say that fluid intelligence is "fluid" in the sense that it can be applied to any situation, no matter how unfamiliar. Crystallized intelligence is so named because it consists of fixed bits of knowledge (initially acquired via fluid intelligence) that are now "crystallized" in memory.

Evidence for the Distinction between Fluid and Crystallized Intelligence

There are three primary lines of evidence that Cattell (1941, 1943, 1971, 1987) cited to support the distinction between fluid and crystallized intelligence:

1. The two types of intelligence can be distinguished with factor analysis, meaning that tests of crystallized intelligence correlate more strongly with each other than they do with tests of fluid intelligence. Likewise, tests of fluid intelligence correlate more strongly with each other than they do with tests of crystallized intelligence. This is called *structural evidence*.

2. The two abilities have different developmental growth curves. That is, fluid intelligence declines steadily starting in early adulthood, whereas crystallized intelligence increases until late adulthood. This is called *developmental evidence*.

3. Different types of brain injury have differential effects on the two types of intelligence. This has been called *neurocognitive evidence*. Cattell's evidence suggested that almost any lesion in the cortex decreases fluid intelligence in proportion to the size of the lesion, whereas most brain injuries have little effect on crystallized intelligence unless the lesions are in regions of the brain known to affect language. For this reason, Cattell believed that fluid intelligence represents the mass action of the entire cortex. Although it is true that fluid intelligence declines in response to injury in many different regions of the cortex, it appears that Cattell was not entirely correct that any lesion anywhere will decrease fluid intelligence. For example, it appears that particular regions of the frontal and parietal lobes are particularly important for fluid reasoning (Woolgar et al., 2010). Nevertheless, it is largely true that crystallized intelligence is more robust to brain injury than fluid intelligence.

Cattell's g_f and g_c as Elaborations of Spearman's g

Cattell's (1941, 1943) distinction between fluid and crystallized intelligence is sometimes seen as a challenge to Spearman's (1904) theory of general intelligence. However, it is clear from Cattell's initial formulation of his theory that he was not contradicting Spearman, but elaborating on his mentor's theory. Cattell was presenting psychometric, developmental, and neurological evidence that Spearman's *g* arose because of two different processes, and thus should be split into at least two general factors of ability, general fluid intelligence (g_f) and general crystallized intelligence (g_c). Indeed, the idea that there might be more than one general factor of intelligence comes from Spearman's (1904) first paper on general intelligence, though Spearman never developed this idea any further himself.

Note that Cattell could have chosen other ways to abbreviate his constructs (F and C, for example). In patterning his abbreviations g_f and g_c after Spearman's solitary *g*, Cattell (1971) emphasized the continuity of his thinking with that of his mentor (p. 299). The similarities are not merely typographical, though. Unlike other multifactor theories of intelligence in which there was a hierarchy of ever-narrower abilities, Cattell's g_f-g_c theory described his two constructs as true *general* factors of intelligence. In what sense are these factors general? Both factors are expected to influence performance in every mental activity, but to different degrees, depending on the task.

Cattell's g_f-g_c versus Horn's Gf-Gc

From his first publications about fluid intelligence (Cattell, 1941, 1943) to his last major work on the subject (Cattell, 1987), Cattell's thinking about fluid intelligence became increasingly elaborate, culminating in his triadic theory of intelligence (1971, 1987)—a dramatic reworking of his original g_f-g_c theory. The nuances of this underappreciated theory are not covered in this chapter, but in rough sketch Cattell believed that there are four *general capacities* (fluid reasoning, memory, retrieval fluency, and speed); a handful of *provincial powers* corresponding to sensory and motor domains (e.g., visual, auditory, olfactory, gustatory, tactile, and cerebellar); and a large number of environmentally developed capacities he called *agencies* (e.g., mechanical ability, literacy, numeracy, and so forth). One developed agency, crystallized intelligence, is so broad in scope that it behaves like a general capacity (Cattell, 1987).

Cattell's theoretical elaborations were motivated in part by the empirical results he obtained in collaboration with his student John Horn (Cattell, 1963; Cattell & Horn, 1978; Horn & Cattell, 1966, 1967). Even after Horn became a major theorist in his own right, Cattell and his former student were largely in agreement about most matters. Nevertheless, the two theorists have subtle differences of interpretation and emphasis, which can be detected in careful readings of works they wrote independently of each other (Cattell, 1980, 1987, 1998; Hakstian & Cattell, 1974, 1978; Horn, 1982, 1985; Horn & Blankson, 2005; Horn & McArdle, 2007; Horn & Noll, 1997). Whereas Cattell continued to maintain that fluid and crystallized intelligence are general abilities, Horn's description of the constructs suggested they are abilities with a narrower scope, closer to that of Thurstone's (1935, 1938) primary ability factors of reasoning and verbal ability, respectively.

The two variants of fluid and crystallized ability can be distinguished typographically, using each theorist's preferred abbreviation style (g_f-g_c vs. Gf-Gc). Horn's fluid reasoning factor (Gf) is a broad ability, but not a true general factor like Cattell's g_f. Horn's crystallized intelligence factor (Gc) is likewise broad, but not as general as Cattell's g_c. Whereas Cattell's g_f and g_c affect every aspect of intelligence, Horn's Gf and Gc affect only those aspects of intelligence that directly involve reasoning and acquired knowledge. Carroll (1993, 1998, 2003) largely drew from Horn's version of fluid and crystallized intelligence; Gf and Gc are extremely broad in scope, but not truly general in the sense that they, like Spearman's *g*, simultaneously influence all abilities to some degree.

Cattell's Investment Theory

If fluid and crystallized intelligence are different constructs, why are they correlated? In what Cattell (1987) called *investment theory*, the correlation results from the consistent application of fluid reasoning generating new knowledge, which, as it accumulates, becomes crystallized knowledge. In wealthy societies with relatively enlightened policies, people with manifest talents, whatever the circumstances of their birth, are given opportunities to develop those talents via public schooling, scholarships, informal mentoring, and so forth. All else equal, people with greater fluid ability acquire greater crystallized ability. However, all else is not equal, even in societies that aggressively promote equality. People with the same level of fluid intelligence live in different eras, cultures, regions, cities, neighborhoods, and families. Furthermore, everyone has a unique physiology and life story, giving rise to the rich array of individual differences we can observe in any population, no matter how homogeneous it might at first appear. Thus, having the same level of fluid intelligence does not imply the same opportunities to learn, the same incentives to learn, or the same drive to learn. Individuals with ambition and drive often create opportunities to maximize whatever gifts they might have. In rigidly stratified societies, talent is carefully cultivated among those at the top of the social hierarchy and lies fallow among those at the bottom. In "Elegy Written in a Country Churchyard," poet Thomas Gray (1716–1771) lamented the untapped potential and undeveloped talents of his countrymen who lived unnoticed and died unremembered:

> Full many a flow'r is born to blush unseen,
> And waste its sweetness on the desert air.
> Some village-Hampden, that with dauntless breast
> The little tyrant of his fields withstood;
> Some mute inglorious Milton here may rest,
> Some Cromwell guiltless of his country's blood.

To extend an analogy used in passing by Cattell (1943, pp. 178–179), the dynamic relationship between fluid and crystallized intelligence can be likened to the formation of a coral reef. The bulk of a coral colony's structure is not alive, but consists of the accumulated skeletal remains of long-dead coral polyps. Only the outermost layer has living polyps, which feed on small organisms in the surrounding waters. The hardened remains of coral polyps create surfaces on which new polyps can grow. Under the right conditions, many coral colonies together form coral reefs, which create sanctuary ecosystems for many organisms—some of which the coral eat, thus contributing to the growth of the reef.

To stretch the analogy just a bit further, fluid intelligence is like the outermost layer of a coral colony. Fluid reasoning creates new knowledge, and via memory, knowledge accumulates; today's fluid insights become tomorrow's crystallized knowledge (Cattell, 1987, p. 139). However, crystallized knowledge is not merely raw data stowed away somewhere in the brain. In the same way that coral skeletons create structures ideally suited for new polyps, crystallized knowledge provides conceptual structures on which new fluid insights are more likely to occur (Beier & Ackerman, 2005; Ericsson & Kintsch, 1995; Goode & Beckmann, 2010). This is one of the reasons why experts absorb new information related to their discipline more quickly than new information unrelated to their expertise (Gobet, 2005; Hambrick & Engle, 2002). Under the right conditions, complex knowledge structures distributed across individuals and institutions can create sanctuary ecosystems (e.g., universities), in which intellectual insights flourish and expertise develops to the limits of human capacity.

Like living coral polyps, fluid intelligence is fragile, vulnerable to the effects of malnutrition, injury, disease, and aging (Cattell, 1987). Like coral skeletons, crystallized intelligence is more robust. Individuals whose fluid intelligence was once high but now disrupted because of disease or injury will usually retain most of their former knowledge and skills. For this reason, knowledge tests are often used to estimate a person's intelligence before the brain injury. In Cattell's words, "[C]rystallized abilities are, as it were, a dead coral formation revealing by its outlines the limits of growth of the original living tissue" (1987, pp. 178–179).

Fluid Reasoning

When Cattell coined the term *fluid intelligence*, his definition was directly parallel to Spearman's definition of *g*: "Fluid ability has the character of a purely general ability to discriminate and perceive relationships between any fundaments, new or old" (1943, p. 178). In that same paragraph, he gave it the same role as that of *g*: It causes the intercorrelations among ability tests. Thus the difference between Spearman's and Cattell's theories is not so great, once it is understood that $g_f \approx g$.

Although Spearman is now known as a pioneering statistician and researcher, his earliest and most enduring passion was philosophy (Spearman, 1930). The question that burned brightest in his mind was that of how new knowledge was created. He objected strongly to then-prevalent ideas inspired by associationism, which held that new knowledge was created by the co-occurrence of sensations. In what he considered his best work, *The Nature of "Intelligence" and the Principles of Cognition*, Spearman (1923) outlined his philosophical argument for how sensory associations alone where insufficient for the creation of new knowledge. Spearman believed that three processes could account for knowledge creation:

- *Apprehension of experience:* Basic features of an object, the *fundaments* of apprehension, are immediately brought into awareness whenever the object is the focus of attention.
- *Eduction of relations:* When two fundaments are perceived, the relation between them tends to be evoked. For example, the juxtaposition of the words *day* and *night* tends to evoke the relation of *opposite* (see Figure 3.3).
- *Eduction of correlates:* When presented with a relation and a fundament, a correlated fundament is evoked. For example, the relation *opposite* paired with the word *friend* might evoke *enemy* (see Figure 3.3).

At the same time that Spearman was grappling with the philosophical question of how knowledge is created, he was also developing statistical procedures for studying the structure of intelligence (Spearman, 1904). After some time, it struck him that what he called *general intelligence* consists of individual differences in the ability to perform these three mental processes. Hitherto, he had considered his philosophical and scientific concerns to be separate matters. Spearman (1923, 1930) described the realization that his two main scholarly pursuits had converged on one and the same idea as the most satisfying intellectual experience of his lifetime.

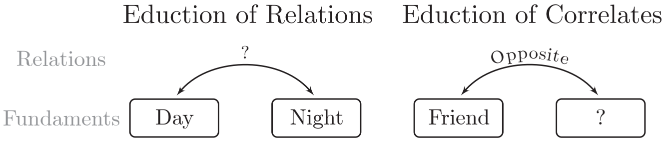


FIGURE 3.3. Illustration of the eduction of relations and correlates.

Definition of Gf

Fluid reasoning (Gf) can be defined as the use of deliberate and controlled procedures (often requiring focused attention) to solve novel, "on-the-spot" problems that cannot be solved by using previously learned habits, schemas, and scripts. Gf is considered a hallmark of intelligence, as it "is essential to learning in school, performance on the job, and life generally" (Kyllonen & Kell, 2017, p. 16). If we can translate Spearman's insights into more modern terms, fluid reasoning is the process by which we extract new knowledge from information we already have. To reason about pieces of information, we need to be able to perceive them, know what they are, and hold them in immediate awareness long enough to perceive their interrelations (Chi, Feltovich, & Glaser, 1981; Fry & Hale, 1996; Kail, 2007; Kane et al., 2004; Krumm et al., 2009). For example, one cannot detect a sequence of colors if one cannot perceive that the colors are different.

One cannot infer the relationship between two fundaments if one does not know what the fundaments are. For example, the analogy "Morning Star is to Evening Star as Venus is to ____________" is a difficult analogy, but not because the relationship between Morning Star and Evening Star is complex. The analogy is difficult because these terms are unfamiliar to most people. Once they are informed that both are alternate names for the planet Venus, the solution is easy: Venus.

Instead of talking about apprehension of experience, we would say that the ability to perceive and hold information in working memory is a precondition of being able to detect previously unnoticed patterns in the information. The question "Who is your mother's sister's husband's daughter's coach's friend's twin's boss?" is difficult not because it contains difficult words or requires complex reasoning to answer, but because it overwhelms the working memory of most listeners. Once this question is split into eight separate questions, anyone familiar with all eight people can answer each question with ease.

In everyday experience, fluid reasoning and prior knowledge are flexibly employed separately, in alternating sequences, or in tandem to meet whatever demands one faces. In contrast, tests of "fluid reasoning" developed for intellectual assessment are designed to minimize the influence of prior knowledge. The tests use simple shapes, simple words, and simple concepts familiar to most people. The difficulty in fluid reasoning tests is in detecting and working with increasingly complex relationships among the stimuli.

Narrow Abilities within Gf[15]

In Carroll's (1993) meta-analysis of over 460 datasets, he identified and classified approximately 240 instances of reasoning factors (Kyllonen & Kell, 2017). Clearly Gf is a well-established broad cognitive ability. Carroll identified several narrow abilities within the fluid reasoning domain, but made no strong claims that his list was comprehensive. It is also important to note that the separate listing of Gf narrow abilities (and narrow abilities in other CHC broad domains) does not indicate that these are highly unique and different abilities. All narrow abilities in a broad domain correlate with each other to some degree. Carroll found that the following three narrow Gf abilities are different, but they are often difficult to distinguish empirically (Kyllonen & Kell, 2017).

1. **Induction: The ability to observe a phenomenon and discover the underlying principles or rules that determine its behavior. This ability is also known as rule inference.* People good at inductive reasoning perceive regularities and patterns in things that otherwise might seem unpredictable. In most inductive reasoning tests, stimuli are arranged according to a principle, and the examinee demonstrates that the principle is understood (e.g., generating a new stimulus that also obeys the principle, identifying stimuli that do not conform to the pattern, or explaining the principle explicitly). For example, complete the next four items in this

sequence: $a^bC_De^fG_H$. Compare these figures: ╔, ╠, ╩, ╣, ╦, ╬, ╝ with these figures: ╒, ╤, ╪, ╫, ╢, ╡. With which group is ╓ most similar? What is the relationship between this group of words (*hope, dread*) and this group (*disappointed, relieved*)?

2. *General sequential reasoning: The ability to reason logically using known premises and principles. This ability is also known as deductive reasoning or rule application.* Whereas induction is the ability to use known facts to discover new principles, general sequential reasoning is the ability to use known principles to discover new facts. For example, if you know that there are no polar bears in Antarctica, and you see a picture of an iceberg with polar bears on it, you can deduce that the iceberg is not in Antarctica.

3. **Quantitative reasoning: The ability to reason with quantities, mathematical relations, and operators.* Tests measuring quantitative reasoning do not require advanced knowledge of mathematics. The computation in such tests is typically quite simple. What makes them difficult is the complexity of reasoning required to solve the problems. For example, choose from among these symbols: +–×÷= and insert them into the boxes to create a valid equation: 8 ☐ 4 ☐ 4 ☐ 8 ☐ 2.

Tentative Narrow Factors

- *Reasoning speed (RE)*: *The ability to reason quickly and correctly.* Carroll (1993) hypothesized that each of the broad factors of ability could be measured with both speed and power tests. The fact that the speed metaphor is often used in synonyms for *smart* (e.g., *quick-witted*) suggests that speed of reasoning has long been noticed as a correlate of intelligence. Unfortunately, clear evidence of a separate reasoning speed factor has not been forthcoming. One of the more rigorous studies designed specifically to distinguish this ability from other aspects of reasoning found reasons to suppose that this is not actually a distinct ability but a complex amalgamation of many influences (Danthiir, Wilhelm, & Schacht, 2005). That is, speed-of-reasoning tasks measure a complex mix of Gf, Gt, Gs, and possibly even aspects of personality such as impulsivity and self-confidence.
- *Piagetian reasoning (RP).* Carroll (1993) advanced the tentative hypothesis that the kinds of tasks used to test Piagetian theories of cognitive development form a distinct narrow factor within Gf. Such tasks are designed to measure seriation (organizing material into an orderly series that facilitates understanding of relations between events), conservation (awareness that physical quantities do not change in amount when altered in appearance), classification (ability to organize materials that possess similar characteristics into categories), and so forth. For now, the hypothesis that there is something distinctive about Piagetian reasoning does not have strong support (Carroll, Kohlberg, & DeVries, 1984; Inman & Secrest, 1981). Given that inductive and deductive reasoning can likely be sub-divided into several extremely narrow abilities, such as analogical, anomalous, antinomous, and antithetical reasoning (Alexander, Dumas, Grossnickle, List, & Firetto, 2016), it is possible that tests of Piagetian reasoning measure a diverse collection of additional narrow reasoning factors as well. Furthermore, Piaget's approach has been criticized based on research that has questioned the idea of qualitative stage transitions; suggested that his theory seriously underestimates the cognitive capabilities of infants and young children; and suggested that more detailed specifications of the mechanisms of change, beyond equilibration, are needed (Newcombe, 2013).[16]

Assessment Recommendations for Gf

Certain narrow abilities are more central to the broad factors than are others. Induction is probably the core aspect of Gf. No measurement of Gf is complete, or even adequate, without a measure of induction. If two Gf tests are given, the second should typically be a test of general sequential (deductive) reasoning. A quantitative reasoning test would be a lower priority unless there is a specific referral concern about mathematics difficulties or other clinical factors warranting such a focus. We also believe it is important that if more than one measure of Gf is administered, attempts should be made to use one measure of "Gf in the wild" (i.e., a test that is not a miniature controlled learning task with structured feedback and examiner scaffolding; Wechsler Matrix Reasoning, WJ IV Number Series) and the other should be a miniature controlled learning task (e.g., WJ IV Concept Formation or Analysis–Synthesis).

Comments and Unresolved Issues Related to Gf

- *How distinct is Gf from the g factor?* As previously noted, hierarchical factor analyses often show that Gf and *g* are perfectly correlated (Undheim & Gustafsson, 1987; Weiss, Keith, Zhu, & Chen, 2013), but not always (Carroll, 2003; Matzke, Dolan, & Molenaar, 2010). In a particularly powerful demonstration, Kvist and Gustafsson (2008) found that Gf and *g* were nearly identical when analyses considered three homogenous groups separately (native-born Swedes, immigrants to Sweden from Europe, and immigrants to Sweden from outside Europe). However, the correlation between *g* and Gf dropped to about .80 when the analyses pooled data from all three groups. The authors interpreted their results as strong support for Cattell's (1987, p. 139) hypothesis that statistical *g* represents the aggregate effect of the investment of Gf in lifelong learning. In populations with similar learning opportunities, Gf and *g* are nearly identical, but the two factors can be distinguished to the degree that learning opportunities are markedly different.

- *Should Gf be divided by content, process, or both?* The distinction between inductive reasoning and general sequential (deductive) reasoning is neat and tidy because they represent distinct kinds of reasoning processes. However, the inclusion of quantitative reasoning as a narrow ability muddies the taxonomy because it spans both inductive and deductive reasoning, but applies only to quantitative content. That is, inductive and deductive reasoning are process factors, and quantitative reasoning is a mixed process and content factor. Oliver Wilhelm (2005) pointed out that most of the tests in Carroll's (1993) induction factor were nonverbal and most of the tests in his general sequential reasoning factor were verbal, thus conflating process with content. Thus, because of this "content confound" (Kyllonen & Kell, 2017), it is not clear what the proper division of Gf should be.

Might there be other content factors in Gf beyond quantitative reasoning? It appears so. CFAs consistently show that fluid reasoning tests can be separated into verbal, spatial/figural, and numerical/quantitative factors (Lakin & Gambrell, 2012; Schroeders, Schipolowski, & Wilhelm, 2015; Schulze, Beauducel, & Brocke, 2005). Evidence from MDS studies suggest that fluid reasoning tests that use similar content, whether it be verbal, spatial, or quantitative, tend to cluster together and that content distinctions matter more in deductive reasoning tests than in inductive reasoning tests (Beauducel et al., 2001; Cohen, Fiorello, & Farley, 2006; Guttman & Levy, 1991; Marshalek, Lohman, & Snow, 1983; Meyer & Reynolds, 2017). These content dimensions are typically referred to as *facets*.[17] In the only study specifically designed to distinguish simultaneously between content (verbal, spatial, and quantitative) and process (inductive vs. deductive reasoning), fluid reasoning tests were distinguished more clearly by content than by process (Wilhelm, 2005). However, a meta-analysis of brain scan research found different patterns of cortical activation for both process (rule inference/induction vs. rule application/deduction) and content (verbal vs. visual–spatial) facets of Gf (Santarnecchi, Emmendorfer, & Pascual-Leone, 2017).

Although Gf tests cluster by content facets, it is not yet clear to us that the separate content clusters represent fundamentally distinct kinds of fluid reasoning. In many of Carroll's (1993) analyses and in more recent studies (Weiss et al., 2013), visual–spatial ability tests often load on the Gf factor, and nonverbal Gf tests often load on the Gv factor. Likewise, verbal analogy tests, markers of verbal Gf, regularly load on Gc, and quantitative reasoning tests often load on the same factor as mathematics achievement tests (McGrew et al., 2014). It seems plausible that the verbal, figural, and numeric content facets in Gf simply represent factor impurities from Gc, Gv, and Gq, respectively. That is, reasoning about words is enhanced if one has background knowledge about them; reasoning about figures is easier if one can manipulate them in the mind's eye; and reasoning about quantity is easier if one has a solid grasp of mathematics. Perhaps the content facets in Gf, and in other domains (e.g., see Gr), reflect an acquired knowledge component of the broad CHC "process" abilities. This distinction between content facet (knowledge) and process would be consistent with Ackerman's theory (see Ackerman, Chapter 8, this volume).

Despite these reservations, the evidence is clear that Gf tests cluster by content, and that the content clusters are not interchangeable in terms of their predictive validity (Gustafsson & Wolff, 2015). Inspired by pioneers who studied intelligence with MDS (Guttman, 1954; Guttman & Levy, 1991; Marshalek et al., 1983), we present both content and process facets of Gf in Figure 3.4.

In Figure 3.4, inductive processes constitute the core of Gf and are relatively independent of stimulus type. Deductive processes (general sequential reasoning and rule application) rely more heavily on content-specific mechanisms—most likely the same collection of content-specific processes that give rise to the broad factors of ability. The three types of content specified in Figure 3.4 have sufficient evidence to support their differentiation,

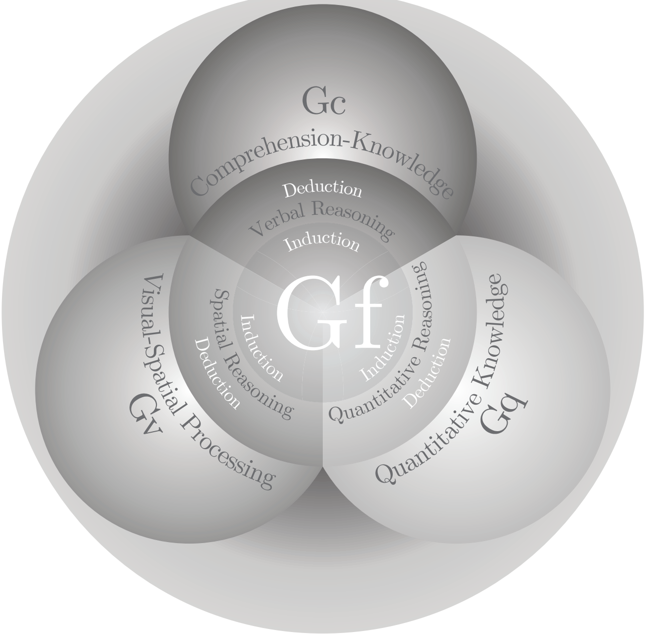


FIGURE 3.4. A conceptual map of fluid reasoning and its overlap with other broad abilities. Fluid reasoning (Gf) probably has both a process facet (inductive vs. deductive reasoning) and a content facet (verbal, spatial, quantitative, and possibly others), each of which overlaps with other broad abilities.

but there may be additional content factors corresponding to each sensory modality (auditory, tactile/kinesthetic, olfactory/gustatory, etc.).

We suspect that this faceted view of fluid reasoning may provide a conceptual bridge for reconciling CHC theory with Johnson and Bouchard's (2005) verbal–perceptual–rotation model. In the latter model, there is no fluid–crystallized distinction, but a verbal–perceptual distinction.

- *What is the relationship among working memory, attentional control, relational complexity of information processing, and Gf?* The past two decades have seen an explosion of research and spirited debate regarding the causal influence of working memory, components of executive functioning, and Gf (e.g., Gwm → Gf). Much of this research is based on the hypothesis that the amount and complexity of information processing made possible by working memory is the foundation of Gf. We touch on this research in our section on Gwm (Kyllonen & Kell, 2017).

- *What is the relation between Gf and complex problem solving and critical thinking?* Educational researchers and policymakers have been driven to embrace constructs believed to be important in the development of essential skills for learners in the 21st century (e.g., complex problem solving and critical thinking; Organization for Economic Co-Operation and Development, 2004). Existing research has indicated that Gf plays a significant role in these aptitude-like amalgams of abilities (Stadler, Becker, Gödker, Leutner, & Greiff, 2015). Research is needed to clarify the importance and role of Gf in these constructs.

- *How can we harness new developments in understanding relational cognitive complexity and cognitive strategy use in Gf test development and research?* Research and theory has indicated a strong link between fluid reasoning and degree of cognitive complexity involved in Gf tasks. Briefly, relational cognitive complexity focuses on the sheer number of interrelated elements (or pieces of information) or element relations in a task that must be processed in parallel, which in turn places a certain amount of relative cognitive load on working memory during reasoning (Birney, Halford, & Andrews, 2006; Halford, Wilson, & Phillips, 1998; Just & Carpenter, 1992). Two approaches to characterizing the cognitive complexity of tests have been employed

(Bertling, 2012). Empirical methods (e.g., MDS) have been used to analyze the relative cognitive complexity of tests after they have been developed. This is a post hoc, data-driven approach to understanding the cognitive complexity demands of Gf and other tests (Marshalek et al., 1983; McGrew et al., 2014). In contrast, theoretical models based on the constraints placed on reasoning by working memory have been used to explain relative cognitive complexity. We believe that test design strategies married with psychometric methods based on item response theory should be used to design, evaluate, and control the relative complexity of Gf items during the test development stage (e.g., see the use of a Latin square design of item types and Rasch analysis by Birney et al., 2006). Such an a priori test design strategy could result in Gf tests with better-known cognitive complexity item characteristics, as well as in the possibility of developing objective "relative cognitive complexity" indexes for comparing Gf tests. Furthermore, studies using experimental psychology methods should be employed to evaluate the amount of cognitive complexity involved in different Gf tests, as well as different cognitive strategies used by individuals as they perform on Gf tests. For example, Hayes, Petrov, and Sederberg (2015) found that eye-tracking technology revealed how cognitive strategy refinement occurred across items and test sessions on matrix reasoning tests.

- *Are we not done "sinking shafts" (Lubinski, 2000) in the domain of Gf (I, RG, RQ)?* During the past decade, there has been an increase in research across multiple domains of inquiry regarding the construct of *relational reasoning.* Relational reasoning is "the ability to recognize or derive meaningful relations between and among pieces of information that would otherwise appear unrelated" (Dumas, Alexander, & Grossnickle, 2013, p. 392). Although often considered synonymous with deductive reasoning (RG), a broader notion of relational reasoning has posited four subtypes of reasoning about relations: *analogy,* or discerning meaningful patterns with otherwise unconnected information; *anomaly,* or identifying structural similarities between two or more objects, concepts, and so on; *antinomy,* or recognizing abnormalities or deviation in an established pattern; and *antithesis,* or identifying directly oppositional relations between two ideas, concepts, or the like (Dumas et al., 2013). In a relatively large sample (*n* = 1,379), albeit limited to undergraduates, Alexander and colleagues (2016) presented structural evidence for a four-factor Test of Relative Reasoning consistent with the four subtypes briefly described above, as well as a single overarching relational reasoning factor. Given that the preponderance of research on relational reasoning has primarily used analogic reasoning tasks (e.g., matrix reasoning; semantic analogies of the form "A:B, then C:D"), this research suggests that the general sequential reasoning (RG) ability may have a substructure, which then begs for similar investigations for inductive and quantitative reasoning. Or perhaps these are just four types of method factors under RG, which could serve the valuable function of generating new test formats for measuring RG.

Memory: General Considerations

Cognitive psychologists have produced a staggeringly complex and impressive body of work on how memory works. Our model of memory draws on this field, but is necessarily much simpler than the cutting-edge models currently available today. Drawing on Unsworth and Engle (2007), we refer to short-term memory at the "trailing edge of consciousness" (Cowan, 2005) as *primary memory,* durable memories that have left consciousness as *secondary memory,* and executive processes that move and manipulate information within and across both systems as *working memory.* Baddeley (2012) calls primary memory *fluid memory systems* and secondary memory *crystallized memory systems,* which for obvious reasons is appealing to us but may well prove confusing in this context.

Primary, secondary, and working memory are not individual-difference variables. They are descriptive terms referring to cognitive structures that everyone has. The most relevant individual-difference variables are working memory capacity (Gwm), learning efficiency (Gl), and retrieval fluency (Gr). *Working memory capacity* refers to how working memory performs its functions. *Learning efficiency* refers to how much time and effort is needed to store new information in secondary memory. *Retrieval fluency* refers to the speed, ease, and accuracy of retrieval of information from secondary memory.

Working Memory Capacity (Gwm)

Definition of Gwm

Working memory capacity (Gwm) can be defined as the ability to maintain and manipulate information in active attention. In the beginning, every animal species went to the computer store to buy a brain. Different species opted for different capacities to meet their needs—dogs got extra olfactory processing, pigeons got GPS, bats got echolocation, and so forth. When it came time to select a memory system, the salesperson laid out these options:

> "Well, I can sell you two different kinds of memory. One is lavishly expensive, ridiculously fragile, achingly slow, and laughably error-prone. It holds only three or four items at a time, and is continuously overwritten every time your attention wanders, which happens pretty much all the time. The other is quick, cheap, reliable, and robust; lasts forever; and offers unlimited storage space."

Only humans were intrigued by the first option.

Compared to long-term memory, what good is a small, fragile, easily overwhelmed, temporary memory system that requires huge swaths of calorically expensive cortex? Whatever flaws working memory may have, most of what makes human thought distinctive would be impossible without it. Working memory sits at the nexus of attention, learning, language, and reasoning (Baddeley, 2002, 2003, 2012). How? Working memory is a mental workspace in which the mind can combine and reconfigure concepts and percepts.

CHC theory concerns individual differences in how well the various components of working memory function. In Baddeley's (2012) multicomponent model, there are domain-specific temporary storage structures and a domain-general control structure. The *visuospatial sketchpad* is the mind's eye, a workspace in which visual objects can be manipulated to create images never seen before. If you are told to visualize a purple hippopotamus pole-vaulting to victory at the next Olympics, your mind creates a temporary image in the visuospatial sketchpad, which you can then make do anything your creative mind can conceive. The *phonological loop* plays the same role as the visuospatial sketchpad, but for sound.

Both storage structures are severely limited in terms of how many independent pieces of information that can be maintained simultaneously. Most people can maintain about three to five "chunks" of information (Cowan, 2005), but these limitations can be circumvented in two ways. First, information can be transferred to long-term memory and then reloaded into working memory when needed (Unsworth & Engle, 2007). Second, smaller chunks can be bound together into ever more elaborate structures (Gobet, 2005; Gobet et al., 2001). For example, the letter sequence *JPGRMMLJGRHSEWAF* would be hard to remember, unless one is familiar with names of the Beatles, the Gospels, the houses of Hogwarts, and the classical elements.

The *central executive* is a hypothetical structure that, depending on the goal, focuses attention, divides attention, switches attention back and forth between objects, and interfaces with long-term memory (Baddeley, 2012). One of its principal functions, called *binding,* is to find relationships among stimuli features (e.g., remembering that a square was red and a circle was blue binds red to square and blue to circle). The limits of working memory capacity are thought to be limited by the efficiency with which new bindings are created and dissolved as one solves problems (Shipstead, Lindsey, Marshall, & Engle, 2014). If one can maintain many objects in the focus of attention for a long time, one has the possibility of discerning more complex relationships among the objects. For this reason, working memory is thought to be a core component of fluid reasoning (De Alwis, Hale, & Myerson, 2014; Duncan, Chylinski, Mitchell, & Bhandari, 2017; Fry & Hale, 1996; Kail, 2007; Unsworth, Fukuda, Awh, & Vogel, 2014).

Working memory capacity is thought to be an important precursor to all forms of academic achievement, either indirectly through fluid reasoning or directly via multistep problem solving (Bull, Espy, & Wiebe, 2008; Hall, Jarrold, Towse, & Zarandi, 2015; Lee, Ng, & Ng, 2009; St Clair-Thompson & Gathercole, 2006).

We have made several changes in our description of working memory from the 2012 version of this chapter. First, we have done what we should have done much earlier: change the name. As of the publication of the WJ IV (Schrank et al., 2014), short-term memory (Gsm) became short-term working memory (Gwm). We believe that there is no loss of information by simply shortening the term to working memory capacity (Gwm).

The second major change has been hinted at previously: Instead of having a single memory span factor, we have formally distinguished between verbal and visual–spatial short-term storage. The evidence for their dissociation is quite clear in multiple factor analyses from many independent studies (Alloway, Gathercole, & Pickering, 2006; Gilhooly, Wynn, Phillips, Logie, & Sala, 2002; Kane et al., 2004; Shah & Miyake, 1996; Swanson & Sachse-Lee, 2001). Their dissociation is also seen in the differential effects of two genetic disorders, Williams syndrome and Down syndrome (Wang & Bellugi, 1994). Verbal and visual–spatial short-term storage differentially predict verbal and visual–spatial reasoning (Kane et al., 2004; Shah & Miyake, 1996; Tanabe & Osaka, 2009). Visual–spatial measures of working memory predict math achievement in young children, but its predictive validity wanes over time. In contrast, verbal working memory becomes an increasingly powerful predictor of math achievement as children age (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; van de Weijer-Bergsma, Kroesbergen, & van Luit, 2015; Zheng, Swanson, & Marcoulides, 2011).

The reason we changed memory span's name is that the term *span* implies a certain length that can be repeated back. The number of items that can be recalled depends on many features, including whether the span test uses digits, words, sentences, nonsense words, pictures, symbols, or faces (Baddeley, 2012). Thus the term *span* is better applied to names of tests than to a psychological construct.

Narrow Abilities within Gwm[18]

1. **Auditory short-term storage (Wa): The ability to encode and maintain verbal information in primary memory.* This ability is usually measured with auditory memory span tasks such as digit span, letter span, word span, or sentence span tests that require immediately repeating back short lists of information verbatim. This ability seems to be particularly important for learning new words and understanding complex sentences (Baddeley, 2003, 2012).

2. **Visual–spatial short-term storage (Wv): The ability to encode and maintain visual information in primary memory.* This ability is usually measured with block-tapping paradigms.

3. **Attentional control (AC): The ability to manipulate the spotlight of attention flexibly to focus on task-relevant stimuli and ignore task-irrelevant stimuli. This is sometimes referred to as spotlight or focal attention, focus, control of attention, executive controlled attention, or executive attention.* It is possible to measure attentional control without a memory load, such as with the antisaccade, flankers, Stroop, and stop-signal paradigms (Miyake et al., 2000; Unsworth, Schrock, & Engle, 2004; Unsworth & Spillers, 2010). Many "processing speed" tests also require attentional control (Miyake et al., 2000).

4. **Working memory capacity (Wc): The ability to manipulate information in primary memory.* This is not technically a narrow ability:

"Working memory capacity" = "Short-term storage" + "Attentional control"

Although there are reasons to measure short-term storage and attentional control separately, working memory capacity is best measured with tests with simultaneous storage and processing demands (see Figure 3.5). After all, this is what working memory was designed to do.

Comments and Unresolved Issues Related to Gwm

- *How does working memory capacity relate to other broad abilities?* Working memory has its fingers in everything. For a while, it was plausible that it was synonymous with the general factor (Kyllonen & Christal, 1990), though that hypothesis has been put to rest (Ackerman, Beier, & Boyle, 2005). Alhough not synonymous with working memory capacity, fluid reasoning seems to depend on it heavily (Shipstead et al., 2014).

Assessment Recommendations for Gwm

For understanding academic problems, working memory capacity tests that require simultaneous storage and processing are most important. We

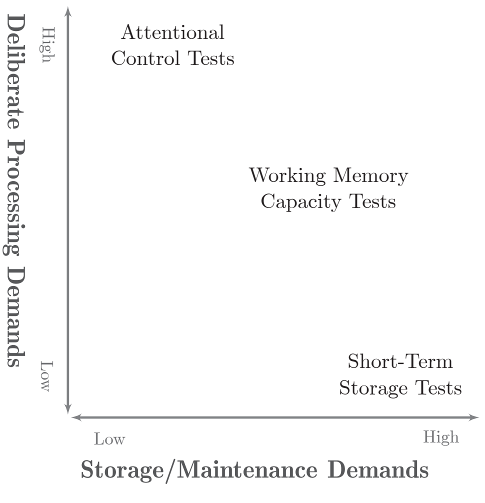


FIGURE 3.5. Working memory capacity tests have simultaneous storage and processing demands.

recommend two types of verbal short-term storage measures: those that minimize the potential for strategy use (e.g., Comprehensive Test of Phonological Processing–Second Edition [CTOPP-2] Memory for Digits), and those with high ecological validity, such as sentence span tests (e.g., WJ IV Memory for Sentences). Visual–spatial short-term storage is less important for most clinical and academic concerns.

Learning Efficiency (Gl)

Definition of Gl

Learning efficiency (Gl) can be defined as the ability to learn, store, and consolidate new information over periods of time measured in minutes, hours, days, and years. As noted previously in this chapter, we believe that long-term learning memory, storage, or learning efficiency (Gl) is distinct from retrieval fluency (Gr) on the one hand and from Gwm on the other. We recognize the risk in using the word *efficiency,* given the conceptual confusion surrounding the term—stemming from its use in a variety of disciplines and even its multiple meanings within educational psychology (Hoffman, 2012; Hoffman & Schraw, 2009, 2010). We do not mean efficiency as conveyed by the Gs + Gwm mental efficiency notion present in certain intelligence composite scores (the WJ III/WJ IV Cognitive Efficiency cluster; the Wechsler batteries' Cognitive Proficiency Index). Our definition is consistent with Hoffman's (2012) conception as related to the efficiency of learning and storing new information: "*Learning efficiency* is primarily based upon individual performance *during* learning when accounting for the incremental costs associated with the learning process. Individual performance means measurable changes in the amount, rate, frequency, or qualitative complexity of knowledge structures. Incremental costs mean factors such as time taken, effort invested, or error rates incurred" (p. 134; original emphasis). For example, to learn and retain a certain amount of information (e.g., a 16-word list), some individuals need to exert more effort than others. To achieve the same outcome, they need more learning inputs (e.g., more learning trials or more time to study).

All tests of learning efficiency must present more information than can be retained in Gwm. This can be accomplished with the repeated supra-span paradigm, where individuals are asked to remember more information than they can learn in a single exposure. Then the information is presented several more times. A paradigm that minimizes the involvement of Gwm is the structured learning task. Such tasks have a teach–test–correct structure. First, a single bit of information is taught. That item is tested, and corrective feedback is offered if required. Another item is taught, and both items are tested with corrective feedback if needed. Then another item is taught, and all three items are tested with corrective feedback if needed. The test becomes longer and longer, but short-term working memory is never overwhelmed with information. The WJ III Visual–Auditory Learning subtest is a good example of a structured learning task.

Narrow Abilities within Gl

1. **Associative memory (MA)*: The ability to form a link between two previously unrelated stimuli, such that the subsequent presentation of one of the stimuli serves to activate the recall of the other stimuli.* Pairs of stimuli (e.g., an abstract visual symbol and a word) are presented together in the teaching phase of the test. In the testing phase, one item of the pair is presented, and the individual recalls its mate. Item pairs must not have any previously established relationships (as in the word pair *table–chair*), or the test is also a measure of meaningful memory.

2. **Meaningful memory (MM)*: The ability to remember narratives and other forms of semantically related information.* Carroll (1993) allowed for tests of meaningful memory to have a variety of formats (e.g., remembering definitions to unfamiliar words), but the core of this ability is the ability to remember the gist of a narrative. After hearing a story just once, most people can retell the gist of it accurately. People who cannot do so are at a severe disadvantage in many domains of functioning. Stories are how we communicate values, transmit advice, and encapsulate especially difficult ideas. Much of the content of our interpersonal relationships consists of the stories we tell each other and the shared narratives we construct. Indeed, much of our sense of identity is the story we tell about ourselves (McAdams, Josselson, & Lieblich, 2006). Many so-called "story recall" tests are barely concealed lists of disconnected information (e.g., "Mrs. Smith and Mr. Garcia met on the corner of Mulberry Street and Vine, where they talked about the weather, their favorite sports teams, and current events. Mr. Garcia left to buy gum, shoelaces, and paper clips. Mrs. Smith left to visit with

her friends Karen, Michael, and Susan. . . ."). A good story recall test has a story that has a true narrative arc. Because story recall tasks require the listener to understand certain conventions of language and culture, many story memory tests have a moderate secondary loading on Gc.

3. *Free-recall memory (M6): The ability to recall lists in any order.* Typically, this ability is measured by having individuals repeatedly recall lists of 10–20 words. What distinguishes this ability from a method factor is that free-recall tests allow the individual to strategically maximize the primacy and recency effect by dumping the contents of primary memory first.

Assessment Recommendations for Gl

We recommend measuring learning efficiency with structured learning tasks to minimize the contaminating effects of Gwm. Structured learning tasks usually measure associative memory. We also recommend measuring meaningful memory because of its diagnostic value. Although differences in performance between tests of associative memory (e.g., WJ IV Visual–Auditory Learning) and meaningful memory (e.g., WJ IV Story Recall) may be related to differences in type of and degree of meaningfulness of the stimuli, score differences may also reflect differences in the complexity of the associative learning capacity between individuals (Cucina, Su, Busciglio, & Peyton, 2015). For example, an associative memory test like WJ IV Visual–Auditory Learning requires learning a series of one-link node pairs, with repeated cumulative study–test phases, while recalling connected discourse (meaningful memory; Story Recall) requires learning a complex network of a larger number of interconnected nodes in a single supraspan trial.

Given that all recent revisions of the major intelligence batteries were published prior to the official split between Gl and Gr, we provide additional guidance for obtaining valid broad Gl norm-based scores (and similar guidance for Gr in the next section). For the WJ IV, the Glr cluster in the cognitive battery is comprised of a measure of meaningful memory (Story Recall) and associative memory (Visual–Auditory Learning). Clinicians should simply drop the r and interpret this norm-based index as a valid proxy of broad Gl. The most recent Wechsler test (the WISC-V) includes associative memory and naming facility tests that can be combined as complementary index scales. As per CHC theory XBA test classifications (Flanagan & Alfonso, 2017), the norm-based Symbol Translation Index should be interpreted as a two-test version of associative memory (MA under Gl), but not broad Gl. The Storage and Retrieval Index score is consistent with the now historical notion of Glr. The SB5, Wechsler Preschool and Primary Scale of Intelligence–Fifth Edition (WPPSI-V), and WAIS-IV do not include Gl or Gr measures. The DAS-II includes two measures of free recall memory (Recall of Objects–Immediate; Recall of Objects–Delayed), but not enough test indicators to form a norm-based broad Gl score. The KABC-II Atlantis and Rebus tests (both first administration and delayed) are both considered measures of associative memory, and the resulting Learning/Glr composite score is best interpreted as a two-test version of associative memory under Gl, but not broad Gl. Where necessary, XBA procedures and guidelines (Flanagan et al., 2013) can be used to create a broad Gl composite by combining two or more indicators of associative memory, meaningful memory, or free-recall memory from other intelligence, neuropsychological, or memory batteries.

Comments and Unresolved Issues Related to Gl

- *Do scores on associative memory tests reflect natural variation in this ability or represent significant differences in "how" different people form associations, or both? Can scoring and interpretation procedures be developed to tease out the "how" differences between individuals?*

- *To what extent does attentional control (AC) influence performance on associative memory tests, and can scoring and interpretation systems be developed to parse out the AC component?* Some forms of associative memory learning appear influenced by attentional control—specifically, the distinction between selectively focusing on the most relevant elements of the stimuli (narrow focus) versus the entire gestalt (broad or global focus) of the stimuli (e.g., focusing on the entire visual rebus symbols vs. a focus on particularly salient parts of the visual rebus symbol, in the WJ IV Visual–Auditory Learning test) (Byrom & Murphy, 2016).

- *Can innovations in network science (e.g., latent semantic analysis) and the development of expert knowledge structure networks or concept maps (Beier, Campbell, & Crook, 2010; Day, Arthur, & Gettman, 2001) be used to develop age-based normative* *associative network maps of narratives in story recall tests, against which an individual's performance could be compared to provide a more rich and complete picture of a person's meaningful memory performance?* The use of tablets and portable computers for test administration, coupled with the high-speed computing embedded in online test scoring platforms, makes this feasible.

Retrieval Fluency (Gr)

Definition of Gr

Retrieval fluency (Gr) can be defined as the rate and fluency at which individuals can produce and selectively and strategically retrieve verbal and nonverbal information or ideas stored in long-term memory. Gr is important. "Throughout the day we are constantly being asked to retrieve facts, events from our life, names of acquaintances, and other important information. The ability to retrieve this information, generally in the absence of potent external cues, is vital for the success of many everyday tasks. As such, strategic retrieval processes are critical aspects of the overall cognitive system" (Unsworth, 2017, p. 135).

People differ in the rates at which they can access information stored in long-term memory, across different type of tasks, along a continuum from very fluent to very dysfluent. People with good Gr are often considered "quick-witted," "quick of mind," or "clever." Although related to cognitive speed (Gs), these characterizations do not reflect fundamental differences in Gs, but rather the "ease" at which mental operations are performed (Alter, 2013; Reber & Greifeneder, 2017). When people are said to be fluent in a second language, it does not mean that they talk rapidly, but that they communicate easily. This aspect of ability has been researched in many fields such as cognition, neuropsychology, and creativity. It is an ability that has become increasingly recognized as important in educational psychology (Reber & Greifeneder, 2017). There is also a long-standing line of research demonstrating that verbal fluency of recall is an important precursor to certain forms of creativity (see Silvia, 2015).[19]

Given the long history of interest in creativity and intelligence, we briefly comment on the role of Gr (and other CHC abilities) in creativity. People who can produce many ideas from memory quickly are in a good position to combine them in creative ways. Individuals high in Gr, together with good Gc and general intelligence (or Gf, depending on the model evaluated), may also be more clever and funny (Beaty et al., 2017). It is important to note that high retrieval fluency is only a facilitator of creativity, not creativity itself. Creativity has also been associated with Gf, Gc, Gs, and Gwm (Avitia & Kaufman, 2014; Silvia, 2015) and a variety of personality and dispositional characteristics. Possessing a large amount of well-organized domain-specific knowledge (Gkn) is also associated with creativity (Weisberg, 2006), although Gkn's importance is dependent to a large extent on Gr and top-down executive control processes (Silvia, 2015). The ability to leverage well-organized knowledge structures requires efficient access, fluent retrieval, management (e.g., manipulation, combination and transformation), and strategic executive control strategies (Silvia, 2015).

Narrow Abilities within Gr

We believe it is useful to categorize fluency factors according to content or stimulus facets—ideas, words, and figures.

Fluency factors in the following group are alike in that they involve the production of *ideas.*

1. **Ideational fluency (FI): The ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object.* Quantity, not quality or response originality, is emphasized. An example of such a test would be to think of as many uses of a pencil as possible in 1 minute.
2. **Expressional fluency (FE): The ability to rapidly think of different ways of expressing an idea.* For example, how many ways can you say that a person is drunk?
3. *Associational fluency (FA): The ability to rapidly produce a series of original or useful ideas related to a particular concept.* In contrast to ideational fluency (FI), quality rather than quantity of production is emphasized. Thus the same question about generating ideas about uses of pencils could be used, but credit is given for creativity and high-quality answers.
4. *Sensitivity to problems/Alternative solution fluency (SP): The ability to rapidly think of several alternative solutions to a practical problem.* For example, how many ways can you think of to get a reluctant child to go to school?
5. *Originality/creativity (FO): The ability to rapidly produce original, clever, and insightful responses (expressions, interpretations) to a given topic, situa-*

tion, or task. This factor is quite difficult to measure for a variety of reasons. Because originality manifests itself in different ways for different people, such diversity of talent does not lend itself to standardized measurement. This factor is not strictly a "retrieval" factor because it is by definition a creative enterprise. However, much of creativity is the combination of old elements in new ways. When we say that one idea sparks another, we mean that a person has retrieved a succession of related ideas from memory, and that the combination has inspired a new idea.

The next three fluency abilities are related in that they involve the fluent recall of *words*.

6. **Speed of lexical access (LA): The ability to rapidly and fluently retrieve words from an individual's lexicon. LA may also be described as verbal efficiency or automaticity of lexical access.* This is a new narrow ability proposed by the WJ IV authors (McGrew et al., 2014). Although a relatively new suggested addition to the CHC taxonomy, this narrow ability is also based on seminal research regarding the importance of verbal efficiency or automaticity of lexical access for reading that dates to the 1970s and 1980s (LaBerge & Samuels, 1974; Neely, 1977; Perfetti, 1985). See Perfetti (2007) and Perfetti and Hart (2002) for more recent information on what is now frequently referred to as the *lexical quality hypothesis* as it relates to reading. Until appropriate research studies are completed with valid indicators of these related constructs, we believe that the definition of LA is such that it subsumes most aspects of NA and FW. Thus we propose that LA is an intermediate-level ability that subsumes NA and FW.

7. **Naming facility (NA): The ability to rapidly call common objects by their names.* In contemporary reading research, this ability is called *continuous naming speed* or *rapid automatic naming* (RAN). A fair measure of this ability must include objects that are known to all examinees. Otherwise, it is a measure of lexical knowledge. The most commonly used visual stimuli are either alphanumeric (colors, objects) or nonalphanumeric (letters, digits) (Kirby, Georgiou, Martinussen, & Parrila, 2010). Naming facility (NA) and speed of lexical access (LA) are the only fluency factors in which each response is controlled by testing stimulus materials. The other fluency factors are measured by tests in which examinees generate their own answers in any order they wish. In J. P. Guilford's terms, NA is an ability involving convergent production, whereas the other fluency factors involve divergent production of ideas. In this regard, naming facility tests have much in common with Gs tests (Kirby et al., 2010): They are self-paced tests in which an easy task (naming common objects) must be done quickly and fluently in the order determined by the test developer. Deficits in this ability are known to cause reading problems (Araújo, Reis, Petersson, & Faísca, 2015; Kirby et al., 2010). In a sense, reading is the act of fluently and automatically "naming" printed words (Bowers, Sunseth, & Golden, 1999). More recently, RAN (NA) has also been linked to various math achievement skills, especially math fluency (Koponen, Georgiou, Salmi, Leskinen, & Aro, 2017).

8. **Word fluency (FW): The ability to rapidly produce words that share a phonological (e.g., fluency of retrieval of words via a phonological cue) or semantic (e.g., fluency of retrieval of words via a meaning-based representation) feature.* An example of an FW task is the ability to rapidly produce words starting with the letter *T*. This ability has been mentioned as possibly being related to the "tip-of-the-tongue" phenomenon (e.g., word-finding difficulties [WFD]; Carroll, 1993). This ability is likely to be well developed in fans of Scrabble and crossword puzzles.

The next two fluency factors are related to *figures*.

9. *Figural fluency (FF): The ability to rapidly draw or sketch as many things (or elaborations) as possible when presented with a nonmeaningful visual stimulus (e.g., a set of unique visual elements). Quantity is emphasized over quality.* For example, in one part of the Delis–Kaplan Design Fluency test, examinees must connect dots with four straight lines in as many unique ways as they can within a time limit.

10. *Figural flexibility (FX): The ability to rapidly draw different solutions to figural problems.* An example of a test that measures this ability is to draw as many ways as possible to fit several small shapes into a larger one.

Assessment Recommendations for Gr

Of the fluency measures, we recommend measures of ideational fluency and speed of lexical access (or naming facility), as the predictive validity of these factors is better understood than for the others, or research has demonstrated their importance for reading. For broad Gr scores, the following advice is offered.

For the WJ IV, the Speed of Lexical Access cluster in the Oral Language battery can be reinterpreted as a reasonable proxy for broad Gr (one indicator each of ideational fluency and naming facility). As per CHC theory XBA test classifications (Flanagan et al., 2013), the WISC-V Naming Speed Index should be interpreted as a two-test version of naming facility, and not broad Gr. The KABC-II, SB5, WPPSI-V, and WAIS-IV do not include Gr measures. The DAS-II includes a single test of naming facility, which is insufficient to form a norm-based broad Gr score. Where necessary, XBA procedures and guidelines (Flanagan et al., 2013) can be used to create a broad Gr composite by combining two or more indicators of naming facility, speed of lexical access, ideational fluency, word fluency, or figurative fluency from other intelligence, neuropsychological, or memory batteries.

Comments and Unresolved Issues Related to Gr

- *How is the RAN, NA, WFD, and LA jingle–jangle jungle to be dealt with?* Psychology has had an interest in rapid naming tasks since the late 1800s (Carroll, 1993). Despite this long history, considerable overlap and confusion exist across the similar, yet different, definitions derived from factor-analytic research (NA, naming facility; Carroll, 1993); RAN reading research (Norton & Wolf, 2012); speed of lexical access (LA) and lexical quality hypothesis reading research (Perfetti, 2007); and the WFD language research (Messer & Dockrell, 2006). A close inspection of all definitions reveals discussions and controversies regarding the role of multiple and different underlying cognitive processes (e.g., phonological access, lexical access, orthographic processing, processing speed, executive functions). We cannot resolve the similarities and differences among these related terms in this chapter; it would require a separate chapter and possibly even a book. There is a critical need for joint or XBA studies, as well as CHC-based causal modeling research, with the most commonly used and psychometrically sound measures of RAN, NA, WFD, and LA (together with other Gr fluency abilities), to carve a path through this jingle–jangle jungle.

A small step in this direction is illustrated by Decker, Roberts, and Englund's (2013) analysis of the strength of association between the CHC-based WJ III cognitive tests and a single WJ III measure of RAN (Rapid Picture Naming) in ages 5–12 of the WJ III norm sample. Contrary to some theories of RAN and reading, phonology processing (Ga) was not related to the alphanumeric WJ III RAN test (naming common objects). Multiple cognitive abilities (Gc, Gf, Gs, and Gr) were related to RAN, although there were developmental differences. The WJ III Retrieval Fluency test, a test of the narrow ability of ideational fluency (this test is also now considered one marker of LA in the WJ IV), was consistently the strongest predictor of the RAN-designated test. In the absence of definitive structural and causal evidence research, and consistent with our focus on CHC theory, we are standing pat with the separate definitions of naming facility (NA) and speed of lexical access (LA) provided above. Whether LA is a higher-level intermediate ability above NA subsumed by Gr, or whether the alphanumeric RAN tasks share significant common variance with LA (and not much with the nonalphanumeric RAN tasks), or whether LA is a narrow ability as currently listed, is currently an unresolved issue.

- *Is the long-standing hunt for practical measures of the holy grails of creativity (i.e., originality of responses, FO; alternative solution fluency, SP) doomed to failure, or are some current FO and SP measures "good enough" for understanding real-world problem solving?* Should intelligence test developers add FO and SP measures to their batteries? Research has consistently demonstrated that when both quantity and originality indexes are generated from Gr tasks, the originality indexes provide no new information beyond quantity of responses (typical correlations range from .80 to almost .90). "Clearly, there is little unique variance to be found in uniqueness scores—they are basically the same as fluency" (Silvia, 2015, p. 601). This finding is explained by the fact that as people generate more responses, the probability of unique responses increases. Despite this measurement problem, perhaps quantity indexes from FO and SP measures are sufficient and practically important.

- *Is there an option generation ability?* Del Missier, Visentini, and Mäntylä (2015) recently reported that *option generation fluency* in three simulated real-world problems (i.e., parking lot, fund-raising, and energy-saving problems) was not related to traditional verbal or word fluency measures. However, quantity scores from an alternative-uses test (a measure of FO and SP) accounted for approximately 20% of option generation fluency performance in the real-world problems. This study also suggests that a new option generation fluency ability might exist and is ripe for research. Also, in a study that used a method to control for quantity of responses to obtain scores for originality, Benedek, Jauk, Sommer, Arendasy, and Neubauer (2014) reported a latent factor correlation of .45 between divergent thinking originality (as measured by FO/SP-type tests) and Gf. Collectively this research, along with calls from experts in the intelligence–creativity nexus (J. C. Kaufman, 2016), argues for intelligence test developers to continue exploring the development of measures of CHC narrow abilities associated with creativity.

- *How can we tease out and interpret the role of top-down executive control in Gr tasks?* Neurocognitive research has demonstrated that the generation of ideas invokes top-down executive control processes (Silvia, 2015). For example, performance on ideational (verbal) fluency tasks requires the ability to strategically organize words into meaningful groups or "patches," retrieve a single bit of information (while inhibiting the interference of "persistent alternatives"), flexibly make quick shifts or transitions to new clusters or patches, retrieve from the new group, then use the immediately prior retrieved response as a cue to refine the search, and then repeat the cyclical process (Molinari & Leggio, 2015; Unsworth, 2017). It may be impossible to measure Gr in the absence of executive functioning, as it is likely to be a core component of Gr.

- *Given that retrieval tasks typically invoke executive strategies, is it possible to develop reliable and valid posttest "testing-the-limits" structured think-aloud protocols or questionnaires to identify an individual's use of typical global or specific search strategies?* For example, Unsworth (2017) presented a think-aloud system for a semantic fluency retrieval task that identified a number of different retrieval strategies (viz., visualization, link-to-previous, personal importance, general-to-specific, semantic, rhyme, alphabet, no strategy, and other). Also see Schelble, Therriault, and Miller (2012) for an example of a 16-category retrieval strategy questionnaire developed for a categorical retrieval fluency task. Build it . . . validate it . . . and they (practitioners) shall come.

- *Can new scoring methods, based on network science technology and psycholinguistic research, provide new types of information regarding the breadth, depth and organization of an individual's corpus of specific knowledge (e.g., semantic or phonological knowledge); the individual's degree of match to an age-based normative network; and, more importantly, the "how" of an individual's Gr processes?* We believe that the psycholinguistic research regarding semantic network searches via *optimal foraging* strategies (Abbott, Austerweil, & Griffiths, 2015; Hills, Todd, & Jones, 2015) may provide answers to these questions. The use of tablets and portable computers for test administration, coupled with the high-speed computing embedded in online test scoring platforms, provides the possibility of using procedures such as latent semantic analysis to develop a "Gr network retrieval efficient quotient" (and possibly other retrieval efficiency parameters; e.g., patch-to-patch transition time) from an individual's retrieved responses when compared to a normative network model (Bossomaier, Harré, Knittel, & Snyder, 2009).

- *Although a broad Gr ability exits, and most individuals who retrieve fluently from one domain tend to also excel in retrieval in other domains, what are the individual differences in retrieval processes among people and among different content domains (ideas, words, figures) (Unsworth, 2017)?* Although the extant Gr literature does not support distinct domain-specific retrieval processes, neurocognitive research suggests that differential retrieval test performance by facets (e.g., semantic vs. figural fluency; semantic vs. phonetic) may be related to impaired function in different brain networks (Schmidt et al., 2017; Vannorsdall et al., 2012).

General Cognitive Speed: Considerations

The Difference between Gs and Gt

Both processing speed (Gs) and reaction and decision speed (Gt) are general cognitive abilities related to speed. Both require speeded performance on very easy tests, although Gt tests are generally easier than Gs tests. Gt refers to the speed at which a single item can be performed, on average. That is, each item is presented singly, and the examiner or a computer controls the pace at which the next item is presented. Gs refers to the average speed at which a series of simple items is completed in succession, with sustained concentration over all items over a sustained period (e.g., 2–3 minutes). All Gs items are presented at once, and the examinee determines when the next item will be attempted. In Gt tests, the quickness of responding each time, with pauses between items, is critical. In Gs tests, there are no pauses, and the examinee must sustain mental quickness and move swiftly from item to item until told to stop. This seemingly small difference makes a big difference. In Gs tests, the examinee is constantly shifting attention from item to item. Performance can be enhanced (or hindered) by looking ahead to the next several items. In Gt tests, this is not possible because the examiner (or computer) determines when the next item is seen. Thus Gt is more purely about speed of perception or quickness of reactions, whereas Gs is more about the combination of sustained speed, fluency, and the adaptive allocation of attention. For this reason, Gs is more strongly correlated with *g* (and Gf) than is Gt. Gs requires more cognitively complex information processing than Gt (Roberts & Stankov, 1999).

What Does Recent Factor Analysis Tell Us about Gs and Gt?

The continued presentation of neat hierarchical Gs/Gt models of mental speed in the intelligence literature incorrectly suggests that the multidimensional nature of cognitive speed abilities is well known. This is wrong.[20] What is clear is that the extant factor-analytic evidence supports the conceptualization of mental speed as a multifaceted construct with at least two broad types of abilities: Gs and Gt (Carroll, 1993; Danthiir, Roberts, Schulze, & Wilhelm, 2005; Danthiir, Wilhelm, & Roberts, 2012; Danthiir, Wilhelm, et al., 2005; O'Connor & Burns, 2003; Roberts & Stankov, 1999; Stankov, 2000). Beyond that, gaps and ambiguities in the available evidence leave us with many unresolved questions about the number, nature, and structural relations of lower-level abilities (Stankov, 2000). The most recent primary mental speed structural research (Danthiir, Roberts, et al., 2005; Danthiir et al., 2012; Danthiir, Wilhelm, et al., 2005) adds little clarity. In fact, these studies have tended to muddy the waters, as both studies found Gs (defined by perceptual speed tests) and Gt to be statistically isomorphic, and the Gt narrow factors to be primarily method- or task-specific factors. The reported Gs/Gt latent correlations of 1.00 in these two studies most likely reflect shared methodological variance (the respective Gs and Gt tasks within each study were both either paper-and-pencil or computer-administered tests).

Given the current state of affairs in Gs and Gt, we went mining for slivers of gold in speed factor analyses conducted in a handful of small-scale studies of children referred for assessment (Nelson, 2009), children with known clinical disorders (Phillips, 2015), and nondisabled adolescents and adults (Feldmann, Kelly, & Diehl, 2004).[21] We also inspected the exploratory cluster analysis and MDS analysis reported in the WJ IV norm data (McGrew et al., 2014), as well as cluster analysis, MDS, and CFA of the WJ III norm data by McGrew (2010).[22] Why? First, we hoped to solidify or clarify what is currently known about Gs and Gt abilities where possible. Second, and more importantly, we wanted to continue revising a hypothesized *g*-speed hierarchical model originally presented by McGrew and Evans (2004), refined by McGrew (2005), and further refined by Schneider and McGrew (2012), in a continued effort to stimulate and "speed up" thought and research regarding human speed variables (see Figure 3.6).

Only one of these studies (Nelson, 2009) included measures of both Gs and Gt. This study reinforced the Gs and Gt distinction, as the reaction time Gt factor displayed low latent variable correlations (.14 to .20) with three Gs-type narrow abilities (naming facility, perceptual speed, and academic fluency). Combined with the Nelson (2009), Phillips (2015), and McGrew and colleagues (2014) findings, the evidence continues to support the distinct speed or fluency narrow abilities of perceptual speed (Gs-P), writing speed or fluency (Grw-WS), number facility (Gs-N), naming facility (Gr-NA), figural fluency (Gr-FF), and ideational fluency (Gr-FI).

An important conclusion from these studies is that processing speed (Gs) and fluency of retrieval (Gr) represent related but very different aspects of mental speed or fluency, which probably share an underlying speed, fluency, or automaticity dimension. This is evident in the WJ III and WJ IV cluster analysis and MDS results, where the Cognitive and Achievement speeded tests formed a distinct cluster, with subclusters reflecting an academic versus cognitive distinction. When the WJ IV Oral Language fluency tests are also included, the speed (Gs) and fluency (Gr) tests are typically found in the same MDS spatial quadrant distinct from the other three MDS quadrants, which include cognitive process or acquired knowledge tests (McGrew et al., 2014). Across four sources (McGrew, 2010; McGrew et al., 2014; Nelson, 2009; Phillips, 2015), there is consistent evidence for a distinction between cognitive and academic speed or fluency ability. This cognitive–academic distinction may reflect intermediate-stratum abilities or different facets of cognitive speed. Finally, Gs tests may be differentiated by subfactors that

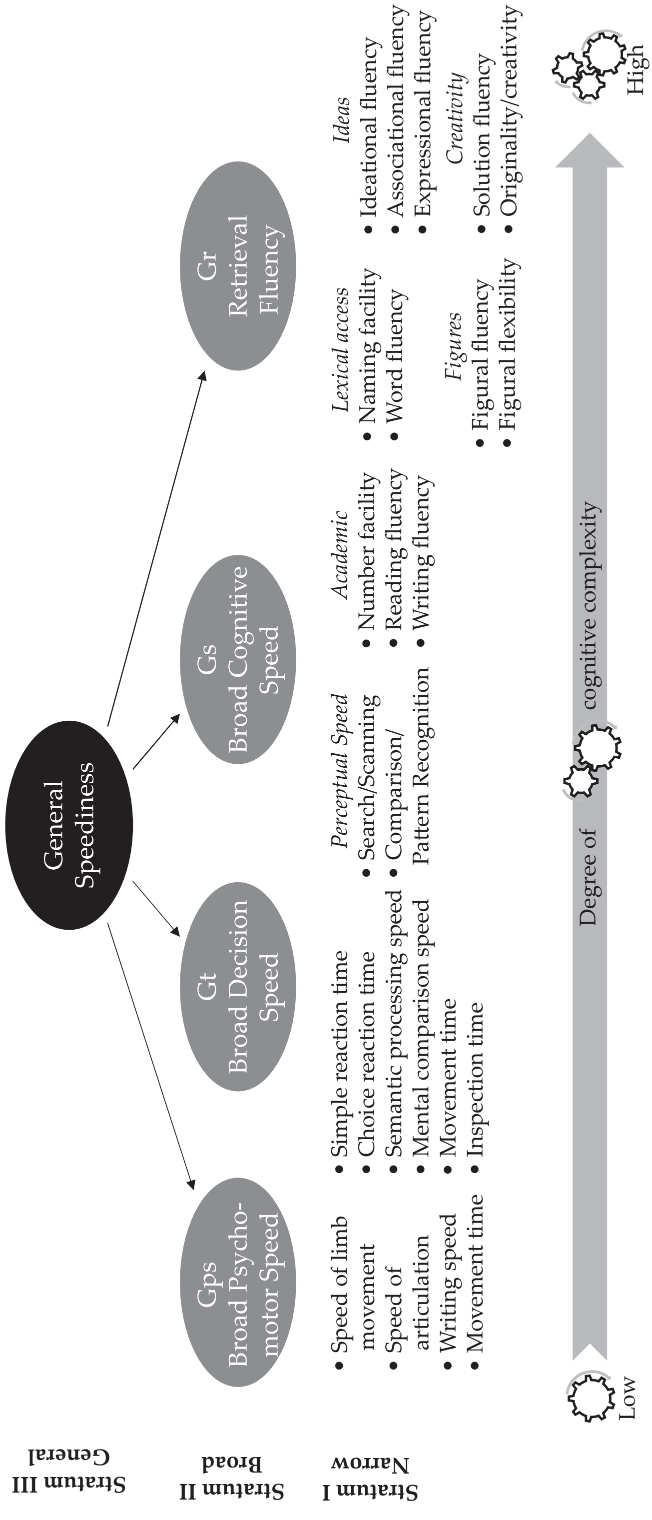


FIGURE 3.6. A hierarchy of speeded abilities.

may represent different content facets—a notion previously suggested by Carroll (1993). McGrew (2010) reported exploratory analyses of the WJ III suggesting that reading and writing (Grw), mathematic (Gq), and visual processing (Gv) stimulus content dimensions might exist under Gs.

A Hypothesized Hierarchy of Speed Abilities

In our search, we discovered that a prior unpublished white paper on structural extensions of CHC theory (McGrew & Evans, 2004), which was the basis for the proposed *g*-speed hierarchical model of human speed abilities (Schneider & McGrew, 2012), did not fully inform (as much as it should have) the most recent elaborations of the speed and fluency domains of CHC theory (McGrew, 2005; Schneider & McGrew, 2012). It is a small gold nugget that we have re-mined and extended. This valuable find was only briefly mentioned in a few sentences in McGrew (2005) and received a single paragraph of treatment in Schneider and McGrew (2012). Both sources focused on the conclusion that the speed domains of Gs and Gt might best be represented within the context of a hierarchically organized speed taxonomy with a *g*-speed factor at the apex. Yet the unpublished McGrew and Evans (2004) paper served as the basis, without the proper context and discussion (unless readers sought out this paper), for the presentation of a hypothesized speed hierarchy model. More importantly, implications for certain narrow-ability definitions and classifications were largely ignored. Where appropriate, we incorporate insights from McGrew and Evans with the factor research summarized above (in the following sections on Gs, Gt, and Gps) to clarify certain narrow speeded abilities and hypothesize possible intermediate factors in the domain of speed and fluency.

Figure 3.6 is the latest iteration of the *g*-speed model. The model is speculative and is based on information from factor studies, theory, and logic. It is not intended to be considered a formal part of the current CHC taxonomy now. It is not a strong statement based on scientific facts. It is meant to "push the edge of the envelope" regarding thinking and research regarding human speed abilities. A hypothesis presented in Figure 3.6 is that speeded and fluency abilities (and tests) might be ordered along a continuum from low to high levels of demands for cognitive complexity processing. The increase in information-processing demands would be associated with abilities and tasks that require increased use of attentional control, working memory, and executive functions.

Processing Speed (Gs)

Definition of Gs

Processing speed (Gs) is the ability to control attention to automatically, quickly, and fluently perform relatively simple repetitive cognitive tasks. Gs may also be described as attentional fluency or attentional speediness. This ability is of secondary importance (compared to Gf and Gc) in predicting performance during the learning phase of skill acquisition. However, it becomes an important predictor of skilled performance once people know how to do a task. Once people learn how to perform a task, they still differ in the speed and fluency (automaticity) with which they perform (Ackerman, 1987). For example, two people may be equally accurate in their addition skills, but one recalls math facts with ease, whereas the other has to think about the answer for an extra half-second and sometimes counts on his or her fingers.

Much as induction is at the core of Gf, perceptual speed is at the core of Gs. Previously (Schneider & McGrew, 2012), we defined perceptual speed (P) as the speed and fluency with which similarities or differences in visual stimuli (e.g., letters, numbers, patterns, etc.) can be compared and distinguished. As noted in our discussion of Gv in this chapter, this ability is also listed under Gv. One way to measure this factor is to present pairs of stimuli side by side, and have the examinees judge them to be the same or different as quickly as possible. Another method of measuring this factor is to present a stimulus to examinees, who must find matching stimuli in an array of heterogeneous figures.

We believe that it is time to acknowledge the white elephant in the perceptual speed (P) room. Does P represent an intermediate factor that includes multiple subfactors? The continued acknowledgment of possible P subfactors in the extant factor structure, test development, and test interpretation literature has been nothing more than that—scholarly recognition of the possibility that there is more to P than P itself. Although mentions are an indication of the authors scholarly awareness of the P research, the use of a single P designation to classify and interpret the wide variety of available P tests has, in our humble opinion, degraded and obscured more accurate interpretation of the factor-analytic literature and, more importantly, the interpretation of scores from P tests on intelligence batteries. It is time to "step up to the plate" and attempt to move the P research and assessment literature forward.

P has been one of the more studied cognitive abilities, starting originally with French (1951). Early factor analysis syntheses (Ekstrom, French, & Harmon, 1979; French, Ekstrom, & Price, 1963) suggested that the P factor may include multiple subfactors. Contemporary research suggests that the structure of P is not well established, but it is likely to be multidimensional (Ackerman et al., 2002; Ackerman & Cianciolo, 2000; Danthiir, Roberts, et al., 2005; French et al., 1963). P tests come in many flavors. Carroll (1993) suggested that the various tests of P consist of two types—those involving *searching* and those involving *comparisons*. Carroll characterized the myriad of possible P factors by means of a mapping statement: speed in [searching for and finding/correctly finding] [one/or more] [literal/digital/figural] stimuli in a visual field arranged [by pairs/by rows/in columns/at random] for [identity/difference/size/etc.]. The combinations and permutations are daunting. Given Carroll's recognition of the various types of P tests and factors, we admit that as the default keepers of the CHC taxonomic scrolls (McGrew, 1997, 2005, 2009; Schneider & McGrew, 2012), we failed to recognize and describe Carroll's two types of P tests and factors in prior attempts to clarify the P factor and assessment landscape.

Based on a review of the extant Gs literature (see "General Cognitive Speed: Considerations," above), we believe that P tasks differ primarily by *content facets* and type and degree of *cognitive complexity* of the cognitive *operations* involved. There is evidence that various Gs and P tests have, in addition to their primary cognitive operations or processing demand, sources of variance (often described as facets in the research) related to different stimulus content. Using the BIS model content facet dimension, which was supported and extended in the WJ IV norm data analysis via MDS, and which is like Ackerman and Cianciolo's (2000) perceptual speed content categorization, we propose that speeded tests and variables be furthered differentiated as visual-figural (Gv), numeric-quantitative (Gq), auditory (Ga), verbal or language (Gc), reading/writing (Grw), or mixed.

More important, based on our rediscovery of the McGrew and Evans (2004) paper, and our review of its original Gs and Gt source research together with newer contemporary research, we believe that Ackerman and colleagues' four-factor P substructure model (Ackerman et al., 2002; Ackerman & Cianciolo, 2000; see McGrew & Evans, 2004) is the best research-based framework that can be used in an initial attempt to further differentiate P variables and tests. Greater confidence is placed in Carroll's speculation regarding the perceptual speed distinction of *searching* and *comparison*, based on the first two subfactors identified by Ackerman and colleagues (pattern recognition and scanning).[23] We have modified the pattern recognition and scanning definitions from McGrew and Evans to be consistent with the essential features of Carroll's two types of perceptual speed abilities (Carroll, 1993). Below, we provide definitions for two types of perceptual speed narrow abilities that lie below the intermediate stratum P.

Narrow Abilities within Gs

The factors listed first are related to the ability to rapidly perform simple *cognitive* tasks (see Figure 3.6).

1. **Perceptual speed (P): An intermediate-stratum ability that can be defined as the speed and fluency with which similarities or differences in visual stimuli (e.g., letters, numbers, patterns, etc.) can be searched and compared in an extended visual field.*

2. **Perceptual speed–search (Ps): The speed and fluency of searching or scanning an extended visual field to locate one or more simple visual patterns.*

3. **Perceptual speed–compare (Pc): The speed and fluency of looking up and comparing visual stimuli that are side by side or more widely separated in an extended visual field.*

The factors listed next are related to the ability to rapidly perform basic *academic* skills or tasks (see Figure 3.6).

4. **Number facility (N): The speed, fluency, and accuracy in manipulating numbers, comparing number patterns, or completing basic arithmetic operations.* Although this factor includes recall of math facts, number facility includes speeded performance of any simple calculation (e.g., subtracting 3 from a column of two-digit numbers). Number facility does not involve understanding or organizing mathematical problems and is not a major component of mathematical/quantitative reasoning (under Gf) or higher mathematical skills. People with slow recall of math facts may be

more likely to make computational errors because the recall of math facts is more effortful (i.e., consumes attentional resources) and is thus a source of distraction.

5. *Reading speed (fluency) (RS): The speed and fluency of reading text with full comprehension.* Also listed under Grw.

6. *Writing speed (fluency) (WS): The speed and fluency of generating or copying words or sentences.* Also listed under Grw and Gps.

Assessment Recommendations for Gs

We recommend that the assessment of Gs primarily focus on perceptual speed. We suggest that examiners use both a test of perceptual speed–search (Ps) and perceptual speed–compare (Pc). Examples of Ps tests, according to Carroll (1993), have names such as "Cancellation, Finding *A's*, First Digit Cancellation, Identical Numbers, Identical Patterns, Inspection, Letter Cancellation, and Scattered *X's*. For example, in Finding *A's* the task is to look through columns of words and cross out all words that contain the letter *a*" (p. 350). Examples of Ps tests from the Ackerman research group include (1) Finding a and t: Scan for instances of "a" and "t" in text passages; (2) Finding € and ¥: Scan for instances of € and ¥ in text made up of random symbols; (3) Canceling Symbols: Scan a page for a single target figure among other simple target figures; (4) Summing to 10: Circle pairs of numbers if they sum to 10.

Examples of Pc tests, according to Carroll (1993), have names such as "Clerical Checking, Faces, Identical Forms, Name Comparisons, Number Checking, and Object Inspection. In some of these, a stimulus is presented at the left of a row of stimuli, and the task can be to find which other stimulus in the row is either identical to, or different from, the first stimulus. Sometimes the task is to find which stimulus, in a row, is different from the others" (p. 350). Examples of Pc tests from the Ackerman research group include (1) Name Comparison: Identify identical or mismatched name pairs; (2) Number Sorting: Find the largest of five large numbers; (3) Number Comparison: Identify identical or mismatched number pairs; and (4) Clerical Abilities—2: Looking up names and numbers in tables (scanning).

We also suggest that examiners classify the primary content facet of Ps and Pc tests. Thus a Ps test involving numbers would be classified as Ps-Gq. A Pc test involving visual–spatial figures would be classified as Pc-Gv. This secondary content specification might provide clues to differences in scores between various tests of Ps and Pc. The three academic fluency factors should be assessed if they are relevant to the referral concern. These abilities sometimes act as predictors of more complex aspects of academic achievement (e.g., reading comprehension and math problem solving) and sometimes are considered academic outcomes themselves, depending on the referral concern.

To facilitate an understanding of this proposed finer differentiation of P abilities, we offer a few examples from two intelligence batteries. As per our proposed scheme, we logically classify the following two WJ IV tests as Ps tests, followed by their content facet designation: Letter–Pattern Matching (Ps-Grw) and Number–Pattern Matching (Ps-Gq). We logically classify both the Wechsler Cancellation test and the old WJ III Cross Out test as Pc-Gv.

Comments and Unresolved Issues Related to Gs

The prior extensive discussion of the literature on cognitive speed and the nature and measurement of perceptual speed covers a wealth of issues. At least four issues remain.

- *Should the rate-of-test-taking (R9) ability be retired?* Yes! This "ability" has been defined as the speed and fluency with which simple cognitive tests are completed. As we noted in our 2012 chapter, Carroll (1993) indicated that these miscellaneous factors were made up of a heterogeneous group of variables (different contents, task formats, and degrees of difficulty) across 12 different studies. A careful review of the results from the 12 studies and Carroll's own statements suggest that this factor never should have been accorded serious status in the CHC framework. Carroll "provisionally" interpreted these factors and stated that "in most cases, they are contrasted with accuracy factors [based on the same speeded variables] that are also found in the same datasets" (p. 475). Carroll characterized two of the factors (from HORN01 and HORN02) as those that normally occur as a second-order speed factor (i.e., Gs). He similarly described the two factors from three studies (VERS01, VERS02, VERS03) as a general cognitive speed factor (i.e., Gs) and a general motor speed factor (i.e., Gps). Carroll further indicated that the various R9 factors he identified may be associated with the major nonspeeded or level abilities, and that their intercorrelations may be linked to a broad speediness (Gs) factor at stratum II. Carroll stated that two factors from the UNDH11 study "might well have been classified under either P (Perceptual Speed) or N (Number Facility); it was classified here [Rate-of-test-taking] because the author (Undheim, 1978) labeled it 'general speed'" (p. 475). Given the obsessive–compulsive quality of Carroll's treatise, this deference to the original source's factor label is odd. The remaining study factors are an eclectic hodgepodge of judgment tasks; verbal analogy or grammatical fluency tasks; tests using manipulatives, such as the Sequin, Manikin, and Cub Construction tests; and various other miscellaneous tasks. Furthermore, we have observed that if scholars cannot easily classify a Gs test in terms of other narrow Gs abilities (e.g., P, N), they typically assign an R9 designation. R9 has become an "I don't know" or "Other" classification. Conversely, all Gs tasks could be classified as R9. The R9 classification, as currently used, has little convergent or divergent validity.

An additional reason for removing R9 is the proverbial jingle–jangle jungle fallacy. In an important Gs structural evidence study, Roberts and Stankov (1999) differentiated cognitive speed abilities as "(a) the rate at which an individual performs complex psychometric tests, and (b) the speed of performance in which complex cognitive capabilities are only minimally involved. The terms *speed of test-taking* and *speed of information processing* are reserved (respectively) to differentiate between these constructs where necessary" (p. 13; original emphasis). Speed of test taking, in this context, is analogous to Gs and not a narrow ability such as R9, while speed of information processing is analogous to Gt. Jingle. Jangle. Jungle. Fallacy.

Furthermore, it is our conclusion that R9 was indeed a miscellaneous Gs factor bin where loosely related speeded factors were placed. The lack of common content, formats, or demands for the variables does not suggest a robust narrow factor. We have not seen any post-Carroll research that would argue for a stand-alone R9 factor. We therefore recommend that it be "voted off the CHC island." If a researcher (or author of a test or test interpretation system) is unable to classify a narrow factor (test) in terms of the remaining Gs narrow-ability factors, or as a new narrow ability with additional supporting evidence, we recommend it be labeled as "Gs-unspecified." We believe that a significant number of otherwise R9 test classifications in the literature might be classifiable with the two perceptual speed subfactors described above. Much additional work is needed to explicate narrow abilities in the Gs domain.

- *Do auditory processing speed abilities exist?* It is logical that the auditory processing channel should have a speed analogue to the visual processing channel. Preliminary evidence for the convergent, discriminant, and incremental validity of newly developed measures of auditory processing speed is emerging (Cameron, Glyde, Dillon, & Whitfield, 2014; Zajac & Nettelbeck, 2016), but it is not yet conclusive that it is distinct from traditional visual processing speed factors (Zajac, Burns, & Nettelbeck, 2012).
- *Will we ever be able to accurately measure the "speed–accuracy tradeoff" dimension on speeded tests?* The cognitive speed literature has had a lengthy and robust debate regarding the importance of the speed and level (accuracy) characteristics of performance. These two primary characteristics have produced a long-standing interest in different strategies used by individuals on speeded tests. Does an individual focus more on speed at the expense of accuracy, or vice versa? Are there validated different speed–accuracy tradeoff strategies, and if so, how do they relate to general intelligence, specific cognitive abilities, cognitive tempo styles, clinical disorders, aging patterns, and so on? The long-standing interest in measuring the speed–accuracy tradeoff has not been matched by the development of suitable measurement approaches and metrics. It is a complicated phenomenon to study (Ackerman & Ellingsen, 2016).
- *Is task switching a new potential narrow ability?* Jewsbury and colleagues' (2016) series of studies suggests that task switching might be an as-yet-unrecognized narrow ability. Task switching is one hallmark of executive control and has a demonstrated relation with working memory (Draheim, Hicks, & Engle, 2016). Typical tasks require individuals to alternate between making one of two (or more) single judgments in successive trials. The metric is the amount of slowing that occurs on trials that require switching (latency switch cost).

Reaction and Decision Speed (Gt)

Definition of Gt

Reaction and decision speed (Gt) can be defined as the speed of making very simple decisions or judgments when items are presented one at a time. Tests of Gt differ from tests of Gs in that they are not self-paced. Each item is presented singly, and there is a short period between items in which no response from the examinee is required. To date, the primary use of Gt measures has been in research settings. Researchers are interested in Gt because it may provide some insight into the nature of *g* and very basic properties of the brain (e.g., neural efficiency). One of the interesting aspects of Gt is that not only is faster reaction time (RT) in these very simple tasks associated with complex reasoning, but so is greater consistency of RT (less variability). People with more variable RTs have lower overall cognitive performance (Jensen, 2006).

Narrow Abilities within Gt[24]

1. *Simple RT (R1): RT to the onset of a single visual or auditory stimulus.* R1 is frequently divided into the phases of decision time (DT; the time to decide to make a response and the finger leaves a home button) and movement time (MT; the time to move the finger from the home button to another button where the response is physically made and recorded). MT is listed under Gps.

2. *Choice RT (R2): RT when a very simple choice must be made.* For example, examinees see two buttons and must hit the one that lights up.

3. *Inspection time (IT): The speed at which differences in visual stimuli can be perceived.* For example, two lines are shown for a few milliseconds and then are covered up. The examinee must indicate which of the two lines is longer. If given sufficient time, all examinees can indicate which is the longer line. The difficulty of the task is determined by how much time the examinees take to perceive the lines. The inspection time paradigm is noteworthy because it does not require a rapid motor response and thus has no confounds with Gps. Measures of inspection time correlate with the *g* factor at approximately .40 (Jensen, 2006). As noted by McGrew and Evans (2004), but overlooked until now, IT tasks have their origin in perceptual and psychophysical aspects of visual perception (Deary & Stough, 1996) and have consistently demonstrated stronger correlations with visual-spatial tests (e.g., Wechsler visual-perceptual organization tests; O'Connor & Burns, 2003) than its reaction/decision time (Gt-RT/DT) and movement time cousins (Gps-MT) (Burns & Nettelbeck, 2003; Luciano et al., 2004). Is this an unrecognized source of Gv variance? These findings suggest that IT may be more "cognitive" than the other Gt abilities.

4. *Semantic processing speed (R4): Reaction time when a decision requires some very simple encoding and mental manipulation of the stimulus content.*

5. *Mental comparison speed (R7): Reaction time where stimuli must be compared for a particular characteristic or attribute.*

Assessment Recommendations for Gt

Tasks measuring Gt are not typically used in clinical settings (except perhaps in CPTs). With the increasing use of low-cost mobile computing devices (i.e., smartphones and iPads/other slate notebook computers), we predict that practical measures of Gt will soon be available for clinical use. Some potential clinical applications are already apparent. We present three examples.

Gregory, Nettelbeck, and Wilson (2009) demonstrated that initial level of and rate of changes in inspection time might serve as an important biomarker of aging. Briefly, a *biomarker* for the aging process "is a biological parameter, like blood pressure or visual acuity that measures a basic biological process of ageing and predicts later functional capabilities more effectively than can chronological age . . . a valid biomarker should predict a range of important age-related outcomes including cognitive functioning, everyday independence and mortality, in that order of salience" (p. 999). In a small sample of elderly individuals, initial inspection time level and rate of slowing (over repeated testing) was related to cognitive functioning and everyday competence. Repeated, relatively low-cost assessment of adults' inspection times might serve a useful function in cognitive aging research and serve as a routine measure (much like blood pressure) to detect possible early signs of cognitive decline.

Researchers have demonstrated how to harness the typical non-normal distributions of RT as a potential aid in diagnosis of certain clinical disorders. Most RT response distributions are not normally distributed in the classic sense. They are virtually always positively skewed, with most RTs falling at the faster end of the distribution. These distributions are called *ex-Gaussian,* which is a mathematical combination of Gaussian and exponential distributions. It can be characterized by the mean (μ), the standard deviation σ, and an exponential function (τ) that reflects the mean and standard deviation exponential component (Balota & Yap, 2011). (Don't worry; one does not need to understand this statistics-as-a-second-language brief description to appreciate the potential application.) The important finding is that "individuals carry with them their own characteristic RT distributions that are relatively stable over time" (p. 162). Thus, given the ease and efficiency with which RT tests could be repeatedly administered to individuals (via smart devices and portable computers), it would be possible to readily obtain each person's RT distribution signature. Of most importance is the finding that all three RT distribution parameters are relatively stable, and τ is very stable (e.g., test–retest correlations in the high .80s to low .90s). Furthermore, there is a robust relation between τ and working memory performance that is consistent with the worst-performance rule (WPR) discovered in the intelligence literature. The WPR states that on repeated trial testing on cognitive tasks, the trials where a person does poorest (worst) are better predictors of intelligence than the best-performance trials (Coyle, 2003). It has been demonstrated, in keeping with the WPR, that the portion of each person's RT distribution representing the slowest RTs is strongly related to fluid intelligence and working memory.

In the not-too-distant future, assessment personal armed with portable smart devices or computers could test an individual repeatedly over time with RT paradigms. Then, via magical software or app algorithms, a person's RT distribution signature could be obtained (and compared against the normative distribution) to gain insights into the person's general intelligence, Gf, or working memory over time. This could have important applications in monitoring of age-related cognitive changes, responses to medication for attention-deficit/hyperactivity disorder (ADHD) or other disorders, the effectiveness of brain fitness programs, and so forth.

Finally, using the same general RT paradigms and metrics, research has indicated that it may be possible to differentiate children with ADHD from typically developing children (Kofler et al., 2013) and children with ADHD from those with dyslexia (Gooch, Snowling, & Hulme, 2012), based on the RT variability—not the mean level of performance. It is also possible that RT variability might simply be a general marker for a number of underlying neurocognitive disorders.

We have the technology.[25] We have the capability to build portable, low-cost assessment technology based on Gt assessment paradigms. With more efficient and better assessments than before, build it . . . and they (assessment professionals) will come.

Psychomotor Speed (Gps)

Definition of Gps

Psychomotor speed (Gps) can be defined as the ability to perform skilled physical body motor movements (e.g., movement of fingers, hands, legs) with precision, coordination, fluidity, or strength. The Gps domain is likely to contain more narrow abilities than are currently listed in the CHC model. In Ackerman's (1987) model of skill acquisition, Gps is the ability that determines performance differences after a comparable population (e.g., manual laborers in the same factory) has practiced a simple skill for a very long time.

Narrow Abilities within Gps[26]

1. *Speed of limb movement (R3): The speed of arm and leg movement.* This speed is measured after the movement is initiated. Accuracy is not important.
2. *Writing speed (fluency) (WS): The speed at which written words can be copied.* Also listed under Grw and Gps.
3. *Speed of articulation (PT): The ability to rapidly perform successive articulations with the speech musculature.*
4. *Movement time (MT): The time taken to physically move a body part (e.g., a finger) to make the required response, after a decision or choice has been made, in an elementary cognitive task.* MT also may measure the speed of finger, limb, or multiple-limb movements or vocal articulation (*diadochokinesis*; Greek for "successive movements"). MT is no longer listed under Gt or considered an intermediate-stratum ability (under Gt), as previously outlined in McGrew and Evans (2004) and Schneider and McGrew (2012). The cognitive complexity demands of MT are minimal to none, and are far below the simple reaction required in DT tasks, which in turn is below the demands of perceptual speed (Gs-P) tasks (Roberts & Stankov, 1999). MT is simply the quickness of initiating and producing a simple motor response after the basic and simple cognitive decision has been made.

Assessment Recommendations for Gps

Psychomotor speed is not generally used in clinical settings except for finger-tapping tests in neuropsychological settings. Although the speed of finger tapping is of some interest to neuropsychologists, they are more concerned with performance that is dramatically uneven on the right and left hands, as this may indicate in which hemisphere a brain injury may have occurred.

Acquired Knowledge

In Cattell's (1941, 1943) earliest formulation of g_f-g_c theory, some aspects of intelligence are directly shaped by cultural, familial, and personal investments in learning, whereas others are more independent of learning. In Ackerman's (1996a) terms, it is useful to distinguish between *intelligence-as-process* and *intelligence-as-knowledge*. That is, we can describe intelligence in terms of how well people processes information (e.g., perceiving, learning, remembering, reasoning, solving problems), or we can describe intelligence in terms of the breadth and depth of useful knowledge people possess. These two aspects of intelligence are a coherent system designed to work together flexibly.

Everyone has had the experience of a perplexing task's being much easier the second time it is completed. Rather than having to undergo the costly process of discovery each time the same task is encountered, humans can retrieve from memory a solution that worked well the last time. If we communicate our knowledge to others, we contribute to the advance of civilization. The human capacity to transmit knowledge permits ordinary high school students to see further than Newton and all his giants combined. Education will not make everyone a genius, but it makes everyone every genius's heir.

In CHC theory, we divide intelligence-as-knowledge into four categories: comprehension–knowledge (Gc), domain-specific knowledge (Gkn), reading and writing (Grw), and quantitative knowledge (Gq). As will be seen, the boundaries between these constructs are blurry, and their nature is such that they cannot be modeled simply as four separate latent variables.

Comprehension–Knowledge (Gc)

Definition of Gc

Comprehension–knowledge (Gc) can be defined as the ability to comprehend and communicate culturally valued knowledge. Gc includes the depth and breadth of both declarative and procedural knowledge, and skills such as language, words, and general knowledge developed through experience, learning and acculturation. Certain cultural skills such as dancing, fencing, sewing, and woodworking can be taught without words, but language allows such skills to be taught more efficiently. Those with strong language skills are better able to learn that which is transmitted via language. For this reason, it is unsurprising that individual differences in language development are closely related to individual differences in factual knowledge (Schipolowski, Wilhelm, & Schroeders, 2014). The two domains are so interconnected that it is useful to think of them as part of an integrated system for learning and transmitting knowledge. In CHC theory, this nexus of language ability and general knowledge is known as the comprehension–knowledge factor (Gc).

Narrow Abilities within Gc

1. **Language development (LD): Ability to comprehend language and use it to communicate; the general understanding of spoken language at the level of words, idioms, and sentences.* Although language development is listed as a distinct narrow ability in Carroll's model, he used the term as an intermediate category between Gc and more specific language-related abilities. It is the general term for how multiple narrow language abilities work together in concert. In its most general conceptualization, language development spans receptive and expressive aspects of both oral and written language (Carroll, 1993, p. 147). The oral–written language distinction is conceptually obvious, and has several lines of evidence supporting it. Factor-analytic evidence clearly shows that oral language and written language are distinct capacities.

In CHC theory, oral language usually refers to Gc and written language to Grw. The distinction is not of the same kind as, for example, the distinction between visual–spatial processing and auditory processing. All developmentally normal humans learn to speak. Literacy, on the other hand, only occurs with explicit instruction, usually in formal education. Specific regions of the brain evolved specifically to process oral language, but literacy has not been prevalent long enough for natural selection to optimize the brain for reading. Instead, learning to read is a painstaking process that requires repurposing capabilities evolved for other functions.

Support for these distinctions (receptive vs. expressive and oral vs. written) can be found using multidimensional scaling. Using the psych package in R (Revelle, 2015), we created distances from the correlation matrix of oral and written language-related subtests from the WJ IV (Schrank et al., 2014) standardization sample (ages 6 and higher). In Figure 3.7, the tests largely fall into the correct quadrants. On the left side of the figure,

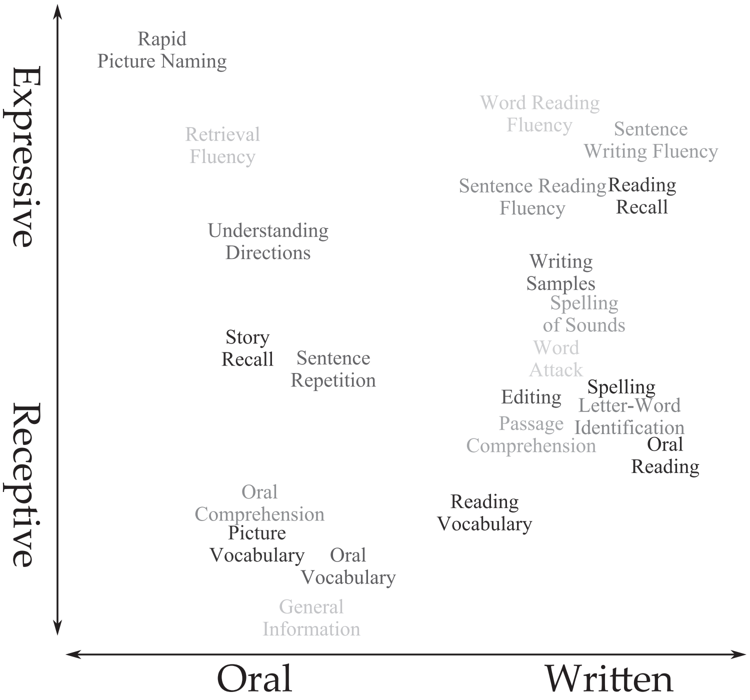


FIGURE 3.7. A multidimensional scaling (MDS) of language tests from the WJ IV standardization sample (ages 6 and higher).

the tests span a wide range of oral abilities, and the tests on the right are written (i.e., measures of Grw). Not every test is purely receptive or purely expressive. For example, Story Recall requires oral comprehension (respective) and oral production (expressive). The exceptions are the tests measuring reading speed. Adding a third dimension to the model (degree to which a test is speeded) resolves this inconsistency with Carroll's prediction. Unfortunately, the complexity of the results and the two-dimensional limitations of the page make a clear presentation of the three-dimensional results impossible to present here.

2. **Lexical knowledge (VL): Knowledge of the definitions of words and the concepts that underlie them; vocabulary knowledge.* For people with deep lexical knowledge, each word in the dictionary is a cognitive aid or tool to help them understand and talk about the world. Lexical knowledge is also an obvious precursor skill for reading decoding and reading comprehension (Storch & Whitehurst, 2002). As with language development, people who read more acquire vocabulary words that are more likely to appear in print than in speech.

3. **General knowledge (K0): The breadth and depth of knowledge considered essential, practical, or worthwhile for everyone in a culture to know.* Shared knowledge allows members of a community to communicate efficiently and come to agreements more peacefully. Deep knowledge of local culture can make the difference between success and catastrophe. For example, Sirolli (1999, p. 9) tells the sad/amusing tale of what happened when Italian aid workers tried to show Zambians how to farm a fertile river valley. Uncurious about why locals had never farmed the valley before, the aid workers invested considerable time, effort, and technical know-how to grow amazingly big tomatoes. Just before the prized tomatoes could ripen, ravenous, unstoppable hippos came out of the river and ate every last one of them.

Much of what we are calling *general knowledge* is common knowledge—familiar to most members of a society. Most adults know enough about local weather patterns to prepare for seasonal changes. They know how to find and prepare food, if needed. They may not know the finer points of etiquette or the technical aspects of law, but they have a general sense of what is polite and what is prohibited.

Beyond common knowledge, general knowledge also includes information that most adults would have been exposed to but might not necessarily remember—such as the plot of the *Iliad,* what the Pythagorean theorem is for, the difference between a monarchy and a dictatorship, the location of Tokyo, the legacy of apartheid, how supply and demand affect prices, how a combustion engine works, and the difference between bacteria and viruses. Although familiarity with these topics is not vital for day-to-day survival for any one individual, a complex society needs a large cohort of people who have a provisional understanding of what is in the world, the processes that govern its functioning, and the diverse people who inhabit it.

General knowledge is contrasted with domain-specific knowledge (Gkn), information that is valuable for members of a specific profession or prized only by enthusiasts with particular interests or hobbies. The distinction between general knowledge and domain-specific knowledge is necessarily blurry at the edges, but clear enough most of the time. If it is taught in general education, it is likely to be general knowledge. If it is taught only in specialized training programs, it is likely to be domain-specific knowledge. Some domain-specific knowledge is so narrow that it is useful for only a few people, such as knowing how to use the complex filing system at a specific branch in a midsized firm.

We have no reason to suppose that learning general knowledge and learning domain-specific knowledge require distinct neurological processes, or that specialized parts of the brain handle the two kinds of knowledge differently. Indeed, any item in a general knowledge test could be transplanted into a domain-specific knowledge test, but not all domain-specific knowledge test items should appear in general knowledge tests. What distinguishes a good general knowledge test item from a domain-specific test item is whether the knowledge is considered important enough such that most adults have been exposed to it. Examining published curricula and introductory textbooks in general education courses can give a good sense of what would be considered general knowledge in a society. An empirical index of the generality of a knowledge item would be its correlation with items in diverse domains, which would manifest as a high item discrimination parameter in an item response theory analysis, a high factor loading in an ordinal factor analysis, or high centrality in MDS.

4. *Listening ability (LS): The ability to understand speech. This ability starts with comprehending single words and increases to long complex verbal statements.* This ability is a receptive oral language ability—a particularly important precursor to reading comprehension, a receptive written language ability (Bishop & Adams, 1990; Catts, Adlof, & Weismer, 2006; Nation, Cocksey, Taylor, & Bishop, 2010; Storch & Whitehurst, 2002). Listening comprehension is also an important predictor of written expression, most likely via oral communication ability (Re & Carretti, 2016). Tests of listening ability typically have simple vocabulary, but increasingly complex syntax or increasingly long speech samples to listen to and answer questions about. As noted later in this chapter, listening ability is not to be confused with narrow abilities listed under auditory processing (Ga).

5. *Communication ability (CM): The ability to use speech to communicate effectively.* This ability is comparable to listening ability, except that it is expressive rather than receptive. Oral communication ability is an important precursor to written expression (Carretti, Motta, & Re, 2016). Carroll's factor came from studies in which people had to communicate their thoughts in nontesting situations (e.g., giving a speech). Although there are many tests in which people are asked to compose essays, we are not aware of language tests in which people are asked to communicate orally in a comparable fashion.

6. *Grammatical sensitivity (MY): Awareness of the formal rules of grammar and morphology of words in speech.* This factor is distinguished from the English usage factor (discussed in the section on Grw) in that it is manifested in oral language instead of written language, and that it measures more the *awareness* of grammar rules than correct usage. Although less is known about this factor's relationship to academic outcomes, it appears that it is an important precursor to reading comprehension (Muter, Hulme, Snowling, & Stevenson, 2004).

Assessment Recommendations for Gc

Adequate measurement of Gc should include a measure of general information and a test of either language development or lexical knowledge (which is a facet of language development). If there is time to give three Gc tests, a test of listening ability is a good choice. A measure of lexical knowledge (vocabulary) is particularly important in the assessment of young children (pre-K and kindergarten), given that a meta-analysis of vocabulary intervention studies has shown strong effects (effect size = 0.88) during this developmental period (Marulis & Neuman, 2010). That is, lexical knowledge is malleable during this formative time period.

Domain-Specific Knowledge (Gkn)

Definition of Gkn

Domain-specific knowledge (Gkn) can be defined as the depth, breadth, and mastery of specialized declarative and procedural knowledge (knowledge not all members of a society are expected to have). Specialized knowledge is typically acquired via one's career, hobby, or other passionate interests (e.g., religion, sports). Knowledge has been featured in several definitions of intelligence, particularly during adulthood. It has been described as a "central ingredient of adult intellect" (Ackerman, 1996b, p. 241). Schank and Birnbaum (1994) stated, "The bottom line is that intelligence is a function of knowledge. One may have the potentiality of intelligence, but without knowledge, nothing will become of that intelligence" (p. 102).

The G in Gkn is somewhat paradoxical. There is no general ability called Gkn because all the abilities within the Gkn domain are specific by definition. Yet when all possible specific Gkn domains are considered collectively, it is broader than Gc (Hambrick, Pink, Meinz, Pettibone, & Oswald, 2008). Ackerman and colleagues (Ackerman, 1987, 1996a, 1996b, 2000; Ackerman & Heggestad, 1997; Ackerman & Rolfhus, 1999; Beier & Ackerman, 2005; Rolfhus & Ackerman, 1996, 1999) have conducted the most systematic study of the domain of Gkn in adults. In addition to the importance of Gc and prior domain knowledge as predictors, these researchers have demonstrated that learning new domain-specific knowledge (particularly declarative knowledge) is also influenced by several non-ability-related (conative) variables. These conative variables include situational and individual interests, as well as the Big Five personality characteristics of openness to experience and typical intellectual engagement. The personality trait of need for cognition has also been implicated as a causal factor. The Ackerman intelligence-as-process, personality, interest, and intelligence-as-knowledge (PPIK) theory is the best available empirically based comprehensive explanation of the development of Gkn abilities (see Ackerman, Chapter 8, this volume). The PPIK theory has its conceptual roots in Cattell's investment hypothesis.

Gkn is unusual in that the proper reference group is not a person's same-age peers in the general population. Rather, the basis of comparison for Gkn is a group of people expected to have the same kinds of specialized knowledge. For example, when measuring an oncologist's Gkn in oncology, it only makes sense to compare the oncologist's knowledge with the average score of other oncologists. Gkn is also unusual in that there are an infinite number of possible narrow factors of specialized knowledge (i.e., one for each potential specialization).

The accumulation of specialized knowledge is often called *expertise. Passive* expertise is a type of knowledge-based specialization that arises from experiences in life and one's position in a society or culture. *Formal* expertise is the result of a self-selection of a domain of knowledge that is mastered deliberately and for which there are clear benchmarks of success (Fisher & Keil, 2016). Interestingly, there is a phenomenon known as the "curse of expertise." Briefly, as expertise develops, it is possible for experts to become overconfident and miscalibrate their ability to solve a problem; in other words, a knowledge "blind spot" may occur (Fisher & Keil, 2016).

In Gc tests, there is a sense in which people are expected to know the answers to all the test questions. In Gkn tests, there is no such expectation unless the person is a member of a certain profession or is considered an expert in a certain domain. The fact that a nurse does not know how to tune a guitar has no bearing on the evaluation of the nurse's abilities. However, if the nurse does not know how to administer a shot, the nurse would be considered incompetent, as would a guitarist who is unable to tune a guitar.

Another noteworthy distinction between Gc and Gkn is their differing relationships with working memory. When solving problems outside their expertise, most experts are unable to perform extraordinary feats of working memory. However, in a phenomenon called *expertise wide-span memory* (Horn & Blankson, 2005) or *long-term working memory* (Ericsson & Kintsch, 1995), experts seem to be able to access large amounts of specialized knowledge very quickly in long-term memory. They can hold it in immediate awareness as if it were stored in working memory, so it can be used to solve complex problems efficiently. But this phenomenon is only true of experts who are working within their areas of specialization.

In the previous version of this chapter (Schneider & McGrew, 2012), we speculated that the structure of domain-specific knowledge would be similar to Holland's (1997) theory of career choice, also known as "Holland's hexagon" and the RIASEC model of vocational interests. According to the RIASEC model, there are six broad, partially overlapping vocational interest domains: Realistic (interest in technical, mechanical, and hands-on activities), Investigative (interest in furthering knowledge), Artistic (interest in the humanities and arts), Social (interest in working with and helping people), Enterprising (interest in getting ahead in terms of profit, power, and influence), and Conventional (interest in order, structure, and business processes). The six domains have a circumplex structure, meaning that interest in adjacent domains are more correlated than interests in nonadjacent domains (e.g., people with investigative interests are most likely also to have realistic and artistic interests, less likely to have conventional and social interests, and least likely to have enterprising interests).

The prediction that knowledge domains and interest domains would have a similar structure was rooted in the straightforward finding that people learn more about things they are interested in than about things they find less interesting (Ackerman, 2000; Ackerman & Rolfhus, 1999; Hambrick, Meinz, & Oswald, 2007; Rolfhus & Ackerman, 1996, 1999; Schmidt, 2014). For this chapter, we identified two studies that would allow us to test this hypothesis. In Reeve (2004) and Rolfhus and Ackerman (1999), participants completed a diverse battery of knowledge tests from many disciplines. We used the psych package (Revelle, 2015) in the R programing environment to convert the correlation matrices published in the two articles into Euclidian distance matrices. These distances were converted into two-dimensional maps via MDS. As seen in Figures 3.8 and 3.9, the results were broadly consistent with our hypotheses. The RIASEC labels were placed by us to summarize our interpretation of the plots.

FIGURE 3.8. An MDS of a knowledge test battery correlation matrix, from Rolfhus and Ackerman (1999). Note: Overlapping labels were manually adjusted to allow for easy interpretation.

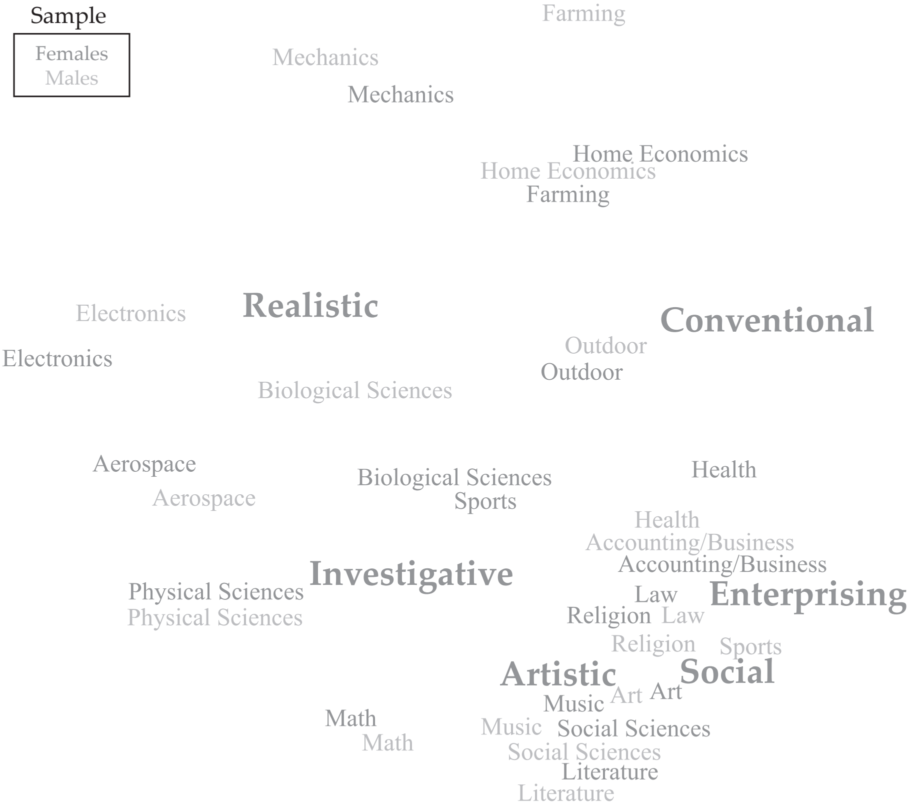


FIGURE 3.9. An MDS of a knowledge test battery correlation matrix for a sample of adults, from Reeve (2004). Results for males (light gray type) and females (darker type) are largely the same. Note: Overlapping labels were manually adjusted to allow for easy interpretation.

Our simplified conceptual structure of knowledge abilities is presented in Figure 3.10. At the center of overlapping knowledge domains is general knowledge—knowledge and skills considered important for any member of the population to know (e.g., literacy, numeracy, self-care, budgeting, civics, etiquette, and much more). The bulk of each knowledge domain is the province of specialists, but some portion is considered important for all members of society to know. Drawing inspiration from F. L. Schmidt (2011, 2014), we posit that interests and experience drive acquisition of domain-specific knowledge.

In Schmidt's model, individual differences in general knowledge are driven largely by individual differences in fluid intelligence and general interest in learning, also known as *typical intellectual engagement* (Goff & Ackerman, 1992). In contrast, individual differences in domain-specific knowledge are more driven by domain-specific interests, and also by the "tilt" of one's specific abilities (Coyle, Purcell, Snyder, & Richmond, 2014; Pässler, Beinicke, & Hell, 2015). In Figure 3.11, we present a simplified hypothetical synthesis of several ability models in which abilities, interests, and personality traits predict general and specific knowledge (Ackerman, 1996a, 1996b, 2000; Ackerman, Bowen, Beier, & Kanfer, 2001; Ackerman & Heggestad, 1997; Ackerman & Rolfhus, 1999; Fry & Hale, 1996; Goff & Ackerman, 1992; Kail, 2007; Kane et al., 2004; Rolfhus & Ackerman, 1999; Schmidt, 2011, 2014; Schneider et al., 2016; Schneider & Newman, 2015; Woodcock, 1993; Ziegler, Danay, Heene, Asendorpf, & Bühner, 2012).

Narrow Abilities within Gkn

1. *General science information (K1): Range of scientific knowledge.* This factor is quite broad, since it encompasses all disciplines within science. It is likely that this factor has many subfactors,

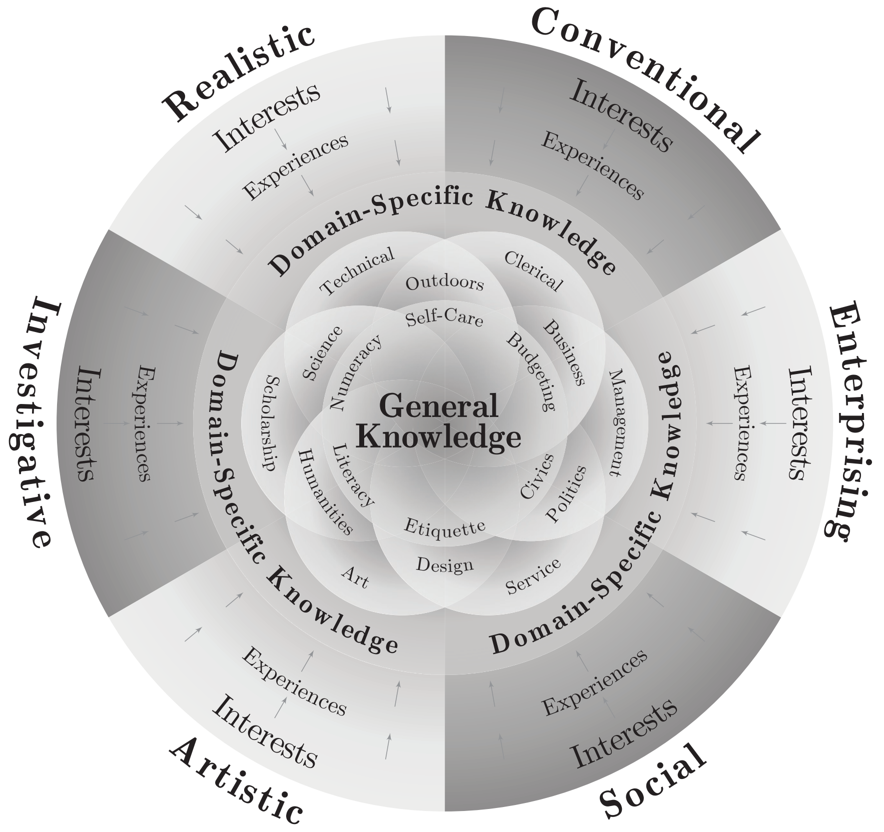


FIGURE 3.10. Conceptual structure of general and domain-specific knowledge.

which may be divided into ever-narrower areas of specialization. It is common to distinguish among the physical sciences (e.g., physics, chemistry, geology, astronomy), the life sciences (e.g., biology, ecology, medicine), and the social sciences (e.g., psychology, sociology, anthropology, archeology, political science, geography, and economics). Of course, there are many fields of study that span these artificial academic boundaries (e.g., evolutionary genetics).

2. **Knowledge of culture (K2): Range of knowledge about the humanities.* As with general science information, this factor is also quite broad and has many subfactors, which may be divided into ever-narrower areas of specialization. It is common to distinguish between artistic/aesthetic domains (e.g., art, literature, and music), and sociohistorical domains (e.g., history, law, politics, religion, language, communication studies). As with the sciences, there are many fields that deliberately cross boundaries within the humanities (e.g., cultural studies, communication). Indeed, there are many disciplines that stand astride the humanities and the sciences (e.g., philosophy, linguistics, and geography). Ultimately, scholars are working toward what E. O. Wilson (1999) calls *consilience:* one interconnected and coherent body of knowledge for all humans.

3. *Mechanical knowledge (MK): Knowledge about the function, terminology, and operation of ordinary tools, machines, and equipment.* There are many tests of mechanical knowledge and reasoning used for personnel selection (e.g., the Armed Services Vocational Aptitude Battery, the Wiesen Test of Mechanical Aptitude).

4. *Foreign-language proficiency (KL): Similar to language development, but in another language besides one's native language.* This ability is distinguished from foreign-language aptitude in that it represents achieved proficiency instead of poten-

tial proficiency. Note that this factor is unusual because it is not a single factor: There is a different foreign-language proficiency factor for every language.

5. *Knowledge of signing (KF): Knowledge of finger spelling and signing (e.g., American Sign Language).*

6. *Skill in lip reading (LP): Competence in the ability to understand communication from others by watching the movements of their mouths and expressions.*

Assessment Recommendations for Gkn

In most situations, Gkn is measured informally by peer reputation, but there are many educational tests that can serve as reasonable markers of specific Gkn domains (e.g., the WJ IV Sciences, Social Sciences, and Humanities tests). Other methods have included cataloguing the productivity and accomplishments of experts (e.g., number of patents, number of creative works, number of prestigious awards).

Unresolved Issues and Comments about Gkn

The theoretical and empirical research regarding domain-specific knowledge expertise is vast. It is so large that we provide no comments regarding issues, as this would require that we be experts in the expertise literature. Interested readers can consult a number of recent summaries and debates regarding expertise to learn more about the status and issues in the field (Ackerman, 2014; Anders Ericsson & Towne, 2010; Detterman, 2014; A. Ericsson & Pool, 2016; S. B. Kaufman & Duckworth, 2017; Sternberg, 1999a; Weiss & Shanteau, 2014).

Reading and Writing (Grw)

Definition of Grw

Reading and writing (Grw) can be defined as the depth and breadth of declarative and procedural knowledge and skills related to written language. What CHC theory calls Grw has traditionally been called *literacy*. Not long ago, literacy was a province of scribes and elites, and thus would have been considered an aspect of domain-specific knowledge (Gkn). In recent centuries, literacy has become so essential to industrialized societies that it is now a core desired skill for every citizen.

In CHC theory, Grw is not technically a separate broad ability distinct from Gc. It is best conceptualized as a facet of language development so important to modern life that it requires special

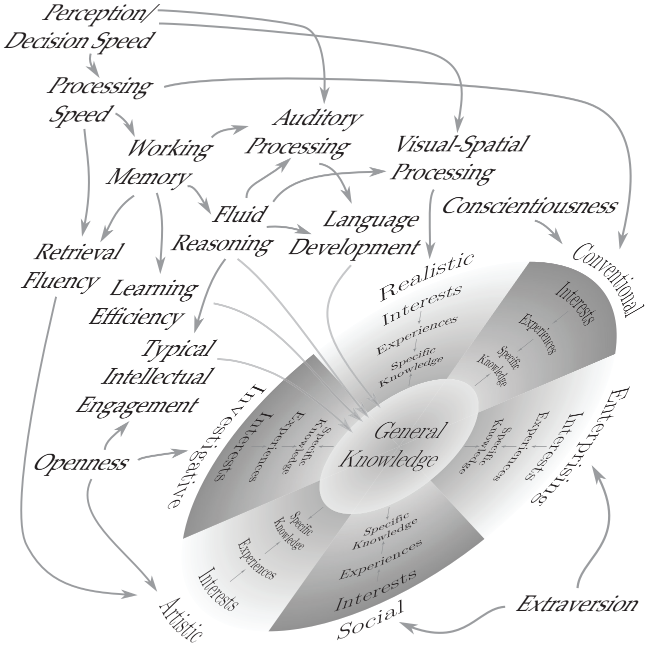


FIGURE 3.11. Ability, interest, and personality precursors of knowledge.

emphasis apart from oral language. Although oral language and Grw are clearly related, to some degree they have different antecedents and predict different outcomes.

It is important to note that when we administer tests of Grw, we are measuring much more than just literacy. Often reading comprehension tests require not just literacy skills, but verbal comprehension, general knowledge, and sometimes auditory processing, working memory, processing speed, and retrieval fluency.

Narrow Abilities within Grw

1. **Reading comprehension (RC): The ability to understand written discourse.* Reading comprehension is measured in a variety of ways. One common method is to have examinees read a short passage and then have them answer questions that they can only answer if they understood the text. A direct method of measuring Grw reading comprehension with reduced contamination from Gc (or Gf) is to ask questions about information that was stated directly in the text. However, we also wish to measure more complex aspects of reading comprehension, such as inference and sensitivity to the author's intent. Such skills draw deeply on Gc. A second method of measuring reading comprehension is the cloze technique, in which a key word has been omitted from a sentence or a paragraph. Examinees who understand what they are reading can supply the missing word.

2. **Reading decoding (RD): The ability to identify words from text.* Typically, this ability is assessed by oral reading tests with words arranged in ascending order of difficulty. Tests can consist of phonetically regular words (words that are spelled how they sound, such as *bathtub* or *hanger*), phonetically irregular words (words that do not sound how they are spelled, such as *sugar* or *colonel*), or phonetically regular pseudowords (fake words that conform to regular spelling rules, such as *gobbish* or *choggy*).

3. *Reading speed (RS): The rate at which a person can read connected discourse with full comprehension.* There are various methods of measuring reading speed, and there is no clear consensus about which method is best for which purposes. Should reading speed be measured by oral reading speed or silent reading speed? Should examinees be told to read as quickly as they can to measure maximal ability, or should they be told to read at their normal pace to measure their typical reading rate? How should the speed–accuracy (of comprehension) tradeoff be handled? Should the format be single words (to measure the efficiency of reading decoding) or full sentences or paragraphs (to measure the efficiency of reading comprehension)? We are certain that different kinds of reading speed tests measure different things that are important, but we are not sure exactly what is different about them. Clinicians are encouraged to think carefully about what exactly the test requires of examinees and to check to see whether there is a logical connection between the apparent task demands and the referral concern. Reading speed is classified as a mixed measure of Gs (broad cognitive speed) and Grw in the hierarchical speed model, although the amount of Gs and Grw measured most likely reflects the degree of difficulty of the reading involved in the task (e.g., reading lists of simple isolated words vs. reading short statements and indicating whether they are true or false).

4. *Writing speed (WS): The ability to copy or generate text quickly.* Writing speed tasks are considered to measure both Grw and Gps (broad psychomotor speed). Like measures of reading speed, the relative importance of Grw or Gps probably varies, depending on the format and level of writing skills involved.

5. *English usage (EU): Knowledge of the mechanics of writing (e.g., capitalization, punctuation, and word usage).*

Assessment Recommendations for Grw

Much more is known about reading assessment than writing assessment. For reading, it is recommended that assessments focus on the point of reading: comprehension. If a person comprehends text well, minor weaknesses in decoding and reading speed are of secondary concern (unless the assessment concerns are reading efficiency problems rather than reading comprehension problems). If there are comprehension deficits, the assessment should focus on the proximal causes of reading comprehension problems (decoding problems, slow reading speed) and then explaining the proximal causes with more distal causes (e.g., slow naming facility → slow reading speed → slow, labored, inefficient reading → comprehension problems). We recommend measuring reading decoding with both real words and pseudowords. Reading comprehension is probably best measured with a variety of methods, including the cloze method and having examinees answer both factual and inferential questions about longer passages.

Spelling ability is an important skill (especially in a phonetically irregular language like English) and is easily measured with a traditional spelling test. It is generally a good idea to select a test that allows the clinician to be able to understand the nature of spelling problems (e.g., phonetically regular misspellings?).

Writing tests are extremely varied and probably measure a wide variety of abilities other than just specific writing abilities. Observing a child's pattern of grammar, usage, and mechanics in responses to writing tests allow clinicians to distinguish between specific writing problems and more complex problems (e.g., general language difficulties). However, it is generally a good idea to examine a wide variety of samples of the examinee's writing, both from formal tests and from school writing assignments.

Quantitative Knowledge (Gq)

Definition of Gq

Quantitative knowledge (Gq) can be defined as the depth and breadth of declarative and procedural knowledge related to mathematics. Gq consists of acquired knowledge about mathematics, such as knowledge of mathematical symbols (e.g., $\int$, π, Σ, ∞, $\neq$, $\leq$, $+$, $-$, $\times$, $\div$, $\surd$, and many others), operations (e.g., addition–subtraction, multiplication–division, exponentiation–*n*th rooting, factorials, negation, and many others), computational procedures (e.g., long division, reducing fractions, the quadratic formula, and many others), and other math-related skills (e.g., using a calculator, math software, and other math aids).

At this time in history, being numerate is almost as important as being literate. Modern economies shower money on those who can use the abstract tools of mathematics to solve concrete problems (Ayres, 2007). For a complex society to function, most adults must have mastered core numeracy skills such as counting and arithmetic. To provide public services and to guide public policy, a sizeable proportion of the population must also understand the basics of algebra, geometry, and statistics. Beyond that, complex societies require many specialists who understand advanced subfields of mathematics. Thus, as with Grw, Gq is not strictly separate from Gc. It is probably best conceptualized as a construct that straddles general and domain-specific knowledge.

Generally, measures of Gq are selected as academic achievement tests and thus must be aligned with a student's curriculum for the score to be diagnostic of math difficulties. This is not the case when measures of Gq are used as aptitude tests (e.g., on the SAT, GRE, or ACT). Gq is unusual in that it consists of many subskills that are fairly well defined by curriculum guides and instructional taxonomies. Thus, metrics of Gq tests can be specified in relative terms (e.g., index scores) and in terms of absolute standards (e.g., an examinee can multiply two-digit numbers or can use the quadratic equation). We believe that both forms of description are necessary to paint a vivid picture of a person's Gq abilities.

Narrow Abilities within Gq

1. **Mathematical knowledge (KM): Range of general knowledge about mathematics, not the performance of mathematical operations or the solving of math problems.* This factor involves "what" rather than "how" knowledge (e.g., "What does π mean?" "What is the Pythagorean theorem?")

2. **Mathematical achievement (A3): Measured (tested) mathematics achievement.* This ability is measured in two ways. The first method is to administer decontexualized math calculation problems (e.g., 67 + 45 = ____). This method gets at the heart of the factor: calculation with the demands of quantitative reasoning minimized. The second method is messier, but focuses on the primary goal of mathematics: solving problems. Examinees are given a scenario and a problem, and they must use reasoning to translate the word problem into a mathematically tractable solution. Examinees then use their calculation skills to arrive at a solution. For example, how many square meters of flooring are needed to cover a 6-meter by 8-meter rectangular room? The examinee has to intuit (or use KM) that this problem is solved by setting up the equation 6 × 8 = ____. Such tests clearly draw upon quantitative reasoning, a facet of Gf.

Assessment Recommendations for Gq

As with Grw, the selection of Gq tests for assessment will depend on the question being asked. Most assessments concentrate first on calculation skills and then on math problem solving. Calculation fluency is typically of secondary concern, but can yield important information regarding the proximal causes of calculation and problem-solving difficulties (e.g., a person who must think about the answer to basic math facts can easily be distracted and make careless errors in the midst of an algebra problem). Math knowledge tests that have no calculation demands can distinguish between people who do not know how to answer the question and people who do not know what the question is.

Comments and Unresolved Issues Related to Gq

- *Are there narrower Gq abilities?* Yes. Carroll (1993) only reported the narrow KM and A3 factors, given their emergence in datasets that included mathematics measures in addition to the cognitive variables that were the primary target of Carroll's review. Carroll did not go out of his way to identify all possible datasets that included tests of mathematics. Essentially, the Gq factors he identified were "bycatch"[27] found when he cast his wide net to capture cognitive ability datasets. Thus other Gq narrow abilities most likely exist, but have yet to be validated within the context of CHC theory.
- *What is the relationship between Gq and number sense? Number sense* refers to core systems of numerical representation present in infants and other animals (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). It does not refer to the verbal counting systems in human language, but to nonverbal capacities for basic quantification. Feigenson and colleagues (2004) presented evidence for two core systems: the *parallel individuation system* for tracking and distinguishing among three to four individual objects at a time as they move over time, and the *approximate number system* for rapid nonverbal estimates of quantity. These primitive systems serve as the foundation for early mathematics skills in children as they learn to map nonverbal estimates of quantity to explicit verbal number systems (Geary, 2004, 2007, 2013; Geary, Hoard, Nugent, & Bailey, 2012). In theory, the division between number sense and Gq is clear: Number sense is fluid (inborn, not taught), and Gq is crystallized (acquired via formal instruction). In practice, however, there is no clean division between inborn number sense and developed capacities at the lower developmental end of the Gq narrow abilities of KM or A3 (or RQ in Gf). Not all measures of number sense distinguish between nonverbal/inborn and verbal/acquired quantitative abilities (e.g., Jordan & Glutting, 2012). Given the importance of number sense in understanding math development and disabilities (Geary, 2007, 2013) and predicting both future reading and math performance (Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006), we expect that number sense assessment will become increasingly common. A recent example is the Number Sense test in the WJ IV Tests of Early Cognitive and Academic Development (ECAD; Schrank et al., 2015).

Sensory- and Motor-Linked Abilities

Cattell, Horn, and Carroll all noted something different about abilities directly associated with sensory modalities. Despite the G in their abbreviation, they are not as general as Gf, Gwm, Gl, Gr, Gs, and Gt; yet they are still broad. What distinguishes these broad abilities from other broad CHC abilities is that they are linked to well-defined regions and functions of the cerebral cortex (i.e., primary regions of the cerebral cortex and their associated secondary regions).

A common theme in the discussion that follows is that these abilities are hard to define. We are not used to talking about sensory-related abilities without talking about the senses and sensory acuity. The distinction between sensation and perception is relevant here, but it is not fully adequate to describe these abilities. *Sensation* refers to the detection of a stimulus. *Perception* refers to complex processing of sensory information to extract relevant information from it (i.e., to literally to make sense of it). These abilities do encompass perception, but also refer to higher-order and goal-directed processing of sensory information (e.g., imagining how a room might look different if it were painted a darker color).

The difficulty in defining and differentiating sensory abilities is captured in a statement regarding the Gv domain, which is likely shared by each of these sensory-based domains. According to Eliot and Czarnolewski (2007),

> One difficulty with defining spatial intelligence is that it is a dimension that is so fundamental and pervasive in people's everyday lives that they take it for granted. It is fundamental and pervasive in the sense that it may operate at any given moment at several levels of human consciousness and, in combination with other cognitive functions, may contribute to the solution process in different ways for many different types of problems. (p. 362)

Well stated!

Visual Processing (Gv)

Definition of Gv

Visual processing (Gv) can be defined as the ability to make use of simulated mental imagery to solve problems—perceiving, discriminating, manipulating, and recalling nonlinguistic images in the "mind's eye." Humans do more than "act" in space; they "cognize" about space (Tommasi & Laeng, 2012). Once the eyes have transmitted visual information, the visual system of the brain automatically performs several low-level computations (e.g., edge detection, light–dark perception, color differentiation, motion detection). The results of these low-level computations are used by various higher-order processors to infer more complex aspects of the visual image (e.g., object recognition, constructing models of spatial configuration, motion prediction). Traditionally, tests measuring Gv are designed to measure individual differences in these higher-order processes as they work in tandem to perceive relevant information (e.g., a truck is approaching!) and solve problems of a visual-–spatial nature (e.g., arranging suitcases in a car trunk).

Among the CHC domains, Gv has been one of the most studied (Carroll, 1993). Yet it has long been considered a second-class citizen in psychometric models of intelligence, due in large part to its relatively weak or inconsistent prediction of important outcomes in comparison to powerhouse abilities like Gf and Gc (Lohman, 1996). But "the times they are a-changing." Carroll (1993), citing Eliot and Smith (1983), summarized three phases of research on spatial abilities, ending in large part in the late 1970s to early 1980s (Lohman, 1979). A reading of Carroll's survey conveys the impression that his synthesis reflects nothing more than what was largely known already in the 1980s. We believe that the Gv domain is entering a fourth period and undergoing a new renaissance, which will result in its increased status in CHC theory and eventually in cognitive assessment. Carroll, the oracle, provided a few hints in his 1993 Gv chapter.

Carroll (1993) was prophetic regarding two of the targets of the resurgent interest in Gv and Gv-related constellations (often broadly referred to as *spatial thinking, spatial cognition, spatial intelligence,* or *spatial expertise*; Hegarty, 2010; National Research Council, 2006).[28] In Carroll's discussion of "other possible visual perception factors" (which he did not accord formal status in his model), he mentioned "ecological" abilities (e.g., abilities reflecting a person's ability to orient the self in real-world space and maintain a sense of direction) and dynamic (vs. static) spatial reasoning factors (e.g., predicting where a moving object is moving and when it will arrive at a predicted location).

Carroll's ecological abilities are reflected in a growing body of research regarding large-scale spatial navigation. *Large-scale spatial navigation* is concerned with finding one's way, or the ability to represent and maintain a sense of direction and location, and move through the environment (Allen, 2003; Hegarty, 2010; Newcombe, Uttal, & Sauter, 2013; Wolbers & Hegarty, 2010; Yilmaz, 2009). Using a map or smartphone GPS system to find one's way to a restaurant, and then to return to one's hotel room, in an unfamiliar large city requires large-scale spatial navigation. A primary distinction between small- and large-scale spatial abilities is the use of different perspectives or frames of reference. *Small-scale spatial ability,* as represented by traditional psychometric tests on available cognitive or neuropsychological batteries, involves allocentric or object-based transformation.

Large-scale spatial ability typically involves an egocentric spatial transformation, in which the viewer's internal perspective or frame of reference changes regarding the environment, while the person's relationship with the objects do not change (Hegarty & Waller, 2004; Newcombe et al., 2013; Wang, Cohen, & Carr, 2014). Recent meta-analyses indicate that large-scale spatial abilities are clearly distinct from small-scale spatial abilities, with an overall correlation of approximately .27. In practical terms, this means that the ability to easily solve the 3D Rubik's cube may not predict the probability of getting lost in a large, unfamiliar city. Also supporting a clear distinction between the two types of spatial abilities is developmental evidence suggesting that large-scale spatial abilities show a much faster rate of age-related decline, and that the two types are most likely related to different brain networks (Newcombe et al., 2013; Wang et al., 2014).

The distinction between *static* and *dynamic* spatial abilities is typically traced to work by Pellegrino and colleagues (Hunt, Pellegrino, Frick, Farr, & Alderton, 1988; Pellegrino, Hunt, Abate, & Farr, 1987) and is now considered one of the two primary organizational facets of spatial thinking (Uttal, Meadow, et al., 2013). Static spatial abilities are well represented by standard tests of Gv (e.g., block design tests). Dynamic and static spatial tasks differ primarily by the presence or absence of movement. "Dynamic spatial ability is one's ability to estimate when a moving object will reach a destination, or one's skill in making time-to-contact (TTC) judgments" (Kyllonen & Chaiken, 2003, p. 233). The ability to catch a football, play a video game, or perform as an air traffic controller requires dynamic spatial abilities, as "one must note the position of the moving object, judge the velocity of the object, anticipate when the object will reach another point (e.g., one's hand, car, or ship), and take some motor action in response to that judgment. In the perception literature, the research surrounding this everyday human information-processing activity has been known as 'time to collision'" (Kyllonen & Chaiken, 2003, p. 233). Although the dynamic–static distinction has gained considerable traction and support (Allen, 2003; Buckley, Seery, & Canty, 2017; Contreras, Colom, Hernandez, & Santacreu, 2003), some research has questioned whether the underlying difference reflects an actual spatial ability distinction. Kyllonen and Chaiken (2003) reported research suggesting that the underlying cognitive process involved in performing dynamic spatial tasks may be a nonspatial, counting-like clock mechanism—temporal processing, not spatial.

The driving forces behind the increased interest and new conceptual developments regarding spatial thinking are threefold. First, rapid technological changes in the past decade have now made access to relatively cheap and accessible visual-graphic-based technology available to large portions of the population. Individuals can immerse themselves in 3D virtual-reality environments for pleasure or learning. Computer visualizations, often available on smartphones and computer tablets, can be used to teach medical students human anatomy and surgery. The complexities and nuances underling "big data" can now be unearthed with complex visual network models than can be rotated at will. Anyone can learn geography by zooming over the world via Google Earth to explore locations and cities. Individuals rely on car- or phone-based GPS visual navigation systems to move from point A to point B. Clearly, developing Gv abilities (or spatial thinking) is becoming simultaneously easier via technology, but also more demanding as humans must learn how to use and understand Gv graphic interface tools that present complex visual displays of multidimensional information.

Second, ever-increasing calls have been made to embed spatial thinking throughout the educational curriculum—"spatializing" the curriculum (Newcombe, 2013)—to raise the collective spatial intelligence of our children and youth (Hegarty, 2010; National Research Council, 2006). The extant research has demonstrated a significant link between spatial abilities and educational performance in the fields of science, technology, engineering, and mathematics (STEM; Buckley et al., 2017; Hegarty, 2010; Lubinski, 2010; Newcombe et al., 2013). Gv abilities and individuals with spatially oriented cognitive "tilts" (Lubinksi, 2010) are becoming increasingly valued by technologically advanced societies. More important, research has demonstrated that spatial abilities or strategies are malleable (National Research Council, 2006; Tzuriel & Egozi, 2010; Uttal, Meadow, et al., 2013; Uttal, Miller, & Newcombe, 2013).

Although many psychologists are important drivers of the renewed interest in an expanded notion of the conceptualization and measurement of Gv (e.g., Allen, 2003; Hegarty, 2010; Kyllonen & Chaiken, 2003; Kyllonen & Gluck, 2003; Lubinski, 2010; Uttal, Miller, et al., 2013; Wang et al., 2014), some of the more active research and conceptualizing are being driven by researchers in education (e.g., National Research Council, 2006; Yilmaz, 2009), cognitive neuroscience (e.g., Thompson, Slotnick, Burrage, & Kosslyn, 2009; Wolbers & Hegarty, 2010), and the STEM disciplines (Harle & Towns, 2010; Seery, Buckley, & Delahunty, 2015). Clearly the CHC model's "mind's eye" (Gv) is achieving more prominence, which needs to be supported with renewed research on yet to be identified well-supported additional narrow abilities and innovative measurement methods, particularly regarding large-scale and dynamic spatial abilities.[29]

Narrow Abilities within Gv

1. **Visualization (Vz): The ability to perceive complex patterns and mentally simulate how they might look when transformed (e.g., rotated, twisted, inverted, changed in size, partially obscured).* In the same way that induction (I) is central to Gf and language development (LD) is central to Gc, this is the core ability of Gv. It is also the Gv narrow ability that demonstrates the strongest relation with a general intelligence factor and typically the strongest loading on a well-defined broad Gv factor. These empirical characteristics of Vz most likely represent its greater degree of demand for relative cognitive complexity (compared to other Gv narrow abilities), as Vz tasks require more complicated, multistep mental transformations of the

stimuli (Lohman, 1988; Yilmaz, 2009). Almost all studies showing that Gv has predictive validity in forecasting important outcomes use measures of visualization as a proxy for broad Gv.

2. **Speeded rotation (SR): The ability to solve problems quickly by using mental rotation of simple images.* Whereas visualization is more about the difficulty of visualizing and rotating an image, speeded rotation is distinct because has more to do with the speed at which mental rotation tasks can be completed. Speeded rotation tasks typically involve simple images that can be rotated. For example, a speeded rotation test might consist of an array of letters rotated from 1 to 360 degrees. After mentally rotating the letters to an upright position, the examinee would discover that half of the letters are backward. The test measures the speed at which the correctly oriented letters can be distinguished from the backward letters.

3. **Imagery (IM):*[30] *The ability to voluntarily mentally produce very vivid images of objects, people, or events that are not actually present.* Factor evidence has indicated that visual imagery is a factor separate from visualization and other narrow Gv constructs (Burton & Fogarty, 2003). Research has suggested that mental imagery is likely to be important for a variety of skilled professions, such as surgery, the study of human anatomy, and piloting an airplane (Thompson et al., 2009). In our 2012 chapter, we proclaimed our belief "that imagery is a promising CHC ability warranting more theoretical and psychometric research attention. We would not be surprised to see multiple imagery abilities validated. More importantly, if psychometrically well-developed practical imagery measures can be constructed, there is a good chance that they will be found to have diagnostic or predictive importance in select educational and occupational domains" (Schneider & McGrew, 2012, p. 130). Although interest in imagery as a mental ability has spanned thousands of years, serious empirical work has not been done until the last 30–40 years (Pearson, 2014). Our positive assessment of the imagery ability has not been matched by significant structural evidence studies that have produced a better understanding of imagery and its place in CHC theory.

However, we remain steadfast in our belief in the eventual emergence of the importance of visual imagery ability and new measures based on promising substantive and methodological neurocognitive research (Ganis & Schendan, 2011; Pearson, 2014). Brain imaging studies have suggested that visual–spatial imagery may not be a single faculty; rather, "visualizing spatial location and mentally transforming locating rely on distinct neural networks" (Thompson et al., 2009, p. 1245). This research suggests a distinction between transformational process versus processing and memory for location substructure. An objective versus spatial imagery dichotomy has also been suggested (see Thompson et al., 2009), as well as the possibility of quality versus speed of imagery abilities (Burton & Fogarty, 2003).

4. *Flexibility of closure (CF): The ability to identify a visual figure or pattern embedded in a complex distracting or disguised visual pattern or array, when one knows in advance what the pattern is.* This factor is primarily defined by hidden-figures tests (e.g., examinees find simple figures embedded in complex backgrounds). This ability is often called *field independence* or *disembedding* by other researchers (Velez, Silver, & Tremaine, 2005). Horn (1980) considered this type of test to be the best marker of Gv, probably because it correlates less with Gf than do many visualization tests. We were unable to locate any well-designed validity studies demonstrating any outcome that flexibility of closure could predict beyond *g* or Gv. Although Carroll (1993) included CF in his model, he admitted that "the psychometric evidence for the factor is somewhat ambiguous" (p. 338).

5. *Closure speed (CS): The ability to quickly identify and access a familiar, meaningful visual object stored in long-term memory from incomplete or obscured (e.g., vague, partially obscured, disguised, disconnected) visual cues of the object, without knowing in advance what the object is.* This ability is sometimes called *gestalt perception* because it requires people to "fill in" unseen or missing parts of an image to visualize a single percept. The term *speed* does not mean speed in the classic psychometric sense, but more the mental ease or fluency of completing these tasks.[31] It is not clear that this measure specifically predicts any important life outcomes, once its associations with Gv and *g* are accounted for (Campbell & Catano, 2004). We speculate that the recall of a stored object may suggest that CS may in part represent a store of acquired Gv knowledge.

6. *Visual memory (MV): The ability to remember complex images over short periods of time (less than 30 seconds).* The tasks that define this factor involve being shown complex images and then identifying them soon after the stimulus is removed. MV is not to be confused with visual working

memory. When the stimuli in a task are simple, are numerous, and must be remembered in sequence, it becomes more of a Gwm test than a Gv test. The extent to which verbal mediation (Gc) can be employed on a specific visual memory test is important, as Carroll (1993) noted that "there is good though not abundant evidence for a visual memory factor controlling performance on tasks in which the subject must form and retain a mental image or representation of a visual configuration that is not readily encodable in some other modality" (p. 284). Also, Carroll noted that he was not aware of "any research on the usefulness of visual memory tasks in predicting educational or occupational success" (p. 284). We also have not uncovered any new convincing evidence that contradicts Carroll's judgment.

7. *Spatial scanning (SS): The ability to quickly and accurately survey (visually explore) a wide or complicated spatial field or pattern with multiple obstacles, and identify a target configuration or identify a path through the field to a target endpoint.* This factor is defined by performance on paper-and-pencil maze tasks. It is not clear whether this ability is related to complex, large-scale, real-world navigation skills.

8. *Serial perceptual integration (PI): The ability to recognize an object after only parts of it are shown in rapid succession.* Imagine that a deer is walking behind some trees, and that only a part of the deer can be seen at one time. Recognizing that this is a deer is an example of what this ability allows people to do. It is certain that this ability exists, but it is not yet certain whether individual differences in this specific ability confer any special advantages or disadvantages.

9. *Length estimation (LE): The ability to visually estimate the length of objects (without using measuring instruments).* Intuitively, this ability seems rather useful, but we could not locate any well-designed studies that have demonstrated an association of this ability with any important outcome.

10. *Perceptual illusions (IL): The ability not to be fooled by visual illusions.* It is unclear why this ability might be useful to a clinician. We could not locate any well-designed studies showing that this distinct ability is associated with any important outcome.

11. *Perceptual alternations (PN): Consistency in the rate of alternating between different visual perceptions.* It is also unclear why this ability might be useful. We could not locate any well-designed studies showing that this distinct ability is associated with any important outcome.

12. *Perceptual speed (P): The speed and fluency with which similarities or differences in visual stimuli can be distinguished.* This intermediate-stratum ability is primarily considered a Gs ability (see our discussion of Gs, above) and has not been included under Gv in all contemporary CHC definitions with a genetic heritage in McGrew (1997). However, Carroll (1993) listed and discussed P under both Gs and Gv.[32] We believe that P (and Ps and Pc) needs to be resurrected under Gs (as defined by Carroll), as various perceptual speed tests (which vary in type of visual stimulus content and required mental process—e.g., search, comparisons, matching) may include small but significant portions of Gv variance that have been ignored in contemporary CHC test classifications because of this omission.[33]

Assessment Recommendations for Gv

Adequate measurement of Gv should always include measures of visualization. If a visualization test utilizes manipulatives, it is important for it to minimize motor requirements (Gp, Gps). The physical manipulation of objects is not required to measure "in the mind's eye" visualization (see, e.g., the WJ IV Visualization test). If speeded tasks are used, they should be balanced by the inclusion of unspeeded tasks. The narrow abilities associated with Gv tend to have very low correlations, and thus one should expect uneven Gv test score profiles.

We are unaware of any psychometrically sound commercially available tests of visual mental imagery. Most measures of imagery (visual as well as other modalities) have been in the form of rather dated self-report questionnaires (Ernest, 1977), such as the Vividness of Visual Imagery Questionnaire (Marks, 1973) and the Questionnaire upon Mental Imagery (Betts, 1909). Research that has compared mental visual imagery self-reports to tests of Gv and experimental cognitive measures of quality and efficiency of image generation have found no significant relations (Poltrock & Brown, 1984).

Comments and Unresolved Issues Related to Gv

- *Is visualization part of Gf?* This question remains unanswered since our 2012 chapter. In many factor-analytic studies, Gf is defined in part by tests considered to measure visualization (e.g., Woodcock, 1990). In Carroll's (1993) analyses, visualization tests often loaded on both Gf and Gv, and about a third of the time the loadings were higher on Gf. What might be happening? Studies of visualization tests suggest that people use a variety of strategies on spatial tests (Gluck & Fitting, 2003; Hegarty, 2010; Kyllonen, Lohman, & Woltz, 1984). Hegarty (2010) has classified these strategies broadly as either using mental imagery (e.g., on the Paper Folding Test, "I imagined folding the paper, punching the hole, and unfolding the paper in my mind") or analytic strategies (e.g., "I used the number of holes/folds to eliminate some of the answer choices"). We believe that the Gv loadings for visualization tests occur because many people use imagery to complete the tests some of the time, and that the Gf loadings occur because logical/analytic strategies are also employed by some people. Furthermore, Kyllonen and colleagues (1984) found that the best performers on visualization tests were flexible in their strategy use, adapting to the task demands of a particular item. This kind of judgment is invariably associated with Gf.

- *Do large-scale spatial navigation abilities belong with Gv?* Based on our brief review above of the emerging Gv literature, we offer a tentative "yes" to this question. Large-scale spatial abilities and navigation are becoming increasingly important. This set of abilities is clearly distinct from traditional small-scale spatial abilities and measures. Tests of large-scale spatial abilities are not readily available, largely due to the historical use of paper-and-pencil or intelligence-test-like tests in the Gv research literature. With continued rapid developments in computer graphics, virtual-reality software, and hand-held technology, it is time for researchers and test developers to focus on developing new practical and clinically useful measures of large-scale spatial navigation abilities. Concurrently, the possible inclusion of these abilities in the CHC taxonomy, and their relationship to the currently identified Gv abilities (which are all static measures), need to be established through a series of CHC-designed factor-analytic studies.

- *Do dynamic spatial abilities belong with Gv?* Possibly. Like large-scale spatial abilities, the ability to comprehend visual movements is becoming more important in an increasingly complex world immersed in visual-graphic-based technology. But first research is needed to determine if what are labeled dynamic spatial abilities represent true spatial competencies, or, alternatively, if they represent a nonspatial (temporal) processing ability.

- *Do other Gv narrow abilities exist?* Of course. As with all CHC domains, the validated narrow abilities in the current taxonomy are largely the result of bottom-up programs of research predicated on developing tests for practical purposes (e.g., prediction, diagnosis). Recent conceptualizations of Gv as a broader spatial thinking construct; the dynamic versus spatial and large-scale versus small-scale conceptualizations; and other functional family conceptualizations of Gv abilities are opening a potential Pandora's box of hypothesized new Gv narrow abilities. For example, Buckley and colleagues (2017) have proposed a comprehensive Gv taxonomy that includes the current Gv abilities and posits 16 potential new narrow abilities based on either theory or research, some previously reviewed by Carroll (1993). These possible new narrow abilities are related to classic spatial tasks (spatial orientation); imagery (quality and speed); illusions (shape and direction, size contrast, overestimation and underestimation, frame of reference); judgments (direction, speed, movement); and dynamic versions of current Gv abilities (visual memory, serial perceptual integration, spatial scanning, perceptual alternations).

These new Gv conceptualizations are welcomed, but they must be studied with serious caution. All new candidates for Gv abilities will need to be validated with well-conceptualized structural validity research (see "Criteria for Updating CHC Theory," above). Also, if new Gv abilities are identified, it is important to determine whether they have any practical use or validity. An instructive example is a recent CFA CHC-designed study that provided preliminary support for a narrow ability of face recognition (called *face identification recognition* by the researchers), distinct from other Gv and CHC abilities (Gignac, Shankaralingam, Walker, & Kilpatrick, 2016). The face recognition ability may have practical usefulness, as it could facilitate measurement and research regarding the phenomenon of *prosopagnosia* (in which a cognitively capable individual is completely unable to recognize familiar faces). Although it is important to guard against premature hardening of the CHC categories (McGrew, 2005; Schneider & McGrew, 2012), we believe that even greater due diligence is necessary to prevent premature proliferation of new entries in the Gv domain in the CHC model. We don't want to be at a place soon where formal START negotiations (STrategic Ability Reduction Talks) are necessary to halt unsupported speculation about and proliferation of Gv abilities.

Is it possible to solve the "Gv mystery"? In their review of the extant CHC–achievement relations research, McGrew and Wendling (2010) found a lack of consistent relations between Gv tests on the major intelligence batteries and math achievement, while other bodies of research have consistently linked visual–spatial ability with math performance. This situation constitutes the "Gv mystery." Aside from several research methodology issues, McGrew and Wendling suggested that the specific Gv abilities measured by most intelligence tests may be "threshold" abilities (after a certain minimum level of competency, the Gv ability no longer contributes to math achievement). Given the above-referenced research that has demonstrated a consistent relation between Gv abilities (or broader notions of Gv such as spatial thinking or spatial intelligence) and performance in STEM fields, we believe that most of the mystery lies in the limitations of traditional Gv test formats. The Gv or spatial thinking discussed in the context of the STEM disciplines is, in our opinion, "Gv on steroids" compared to the abilities measured by current spatial tests in intelligence batteries. The Gv required for high math (or STEM) performance is likely to be that associated with more complex visual thought systems employed by the likes of Einstein. Perhaps Hegarty's (2010) exploration of spatial abilities "in the wild"—via the analysis of the performance of experts in different spatially based knowledge domains, to better understand their cognitive struggles and the processes they use—can produce enhanced Gv measures. Other innovative approaches to measuring Gv abilities should be incorporated in the design of the next generation of Gv tests (Jelínek, Kveton, & Voboril, 2015; Lee & Bednarz, 2012; Thissen, Koch, Becker, & Spinath, 2016). With readily available hand-held technology and portable computers for delivering test content, we believe it is now possible for researchers and test developers to harness this technology to develop measures of the elements of, and totality of, more complex Gv abilities required from success in the STEM disciplines.

- *What about spatial language (Gkn or Gc-spatial)?* As individuals explore their environments, "they build up spatial knowledge based on visual, idiothetic, and other sensory information" (Warren, Rothman, Schnapp, & Ericson, 2017, p. 152). Considerable research has revealed a relation between children's production of, and the number of, spatial words in their lexicons (e.g., *left, right, middle, on top*) and spatial ability. This has been interpreted as suggesting that children verbally encode relevant spatial information that supports performance on spatial tasks (Miller, Vlach, & Simmering, 2017). The direction of this relationship (high verbal or Gc/Gkn affects Gv performance; high Gv ability increases spatial Gc/Gkn; spatial Gc/Gkn and Gv are influenced by other variables) is unknown. Research is needed to explore these relationships and, if appropriate, to develop measures of acquired Gv knowledge, which could include spatial knowledge or vocabulary as well as learned spatial strategies. The distinction between Gv processes and Gv knowledge is consistent with the intelligence-as-process and intelligence-as-knowledge components of Ackerman's PPIK theory of intelligence.

- *Is our courtship of the imagery ability (IM) nothing more than a teenage crush, or will it eventually represent a new, mature Gv relationship?* Clearly, given our past and current treatments of this narrow ability, we believe that imagery is next in line to receive a promotion in the Gv hierarchy. Our IM crush stems in part from the encouraging neurocognitive research on mental imagery. Our IM romance is fed in large part by new interdisciplinary research on human imagination that is integrating research regarding CHC-like abilities (perceptual and motor-related mental imagery—visual imagery, mental rotation, auditory imagery, musical imagery, motor imagery), intentionality or recollective processing, novel combinatorial or generative processing, and emotion within a cognitive neuroscience brain network context (Abraham, 2016).[34] New psychometric research is necessary to better understand the nature of the IM ability in the CHC taxonomy.

Auditory Processing (Ga)

Definition of Ga

Auditory processing (Ga) is the "ability to discriminate, remember, reason, and work creatively (on) auditory stimuli, which may consist of tones, environmental sounds, and speech units" (Conzelmann & Süß, 2015, p. 28). Yes, we have borrowed and directly quoted a contemporary source for a revised definition of Ga. We previously defined auditory processing "as the ability to detect and process meaningful nonverbal information in sound" (Schneider & McGrew, 2012, p. 131). As we noted at the time, "this definition is bound to cause confusion because we do not have a well-developed vocabulary for talking about sound unless we are talking about speech sounds or music" (p. 131). We were correct, but the source of confusion was unanticipated. The previous definition seemed to confuse many assessment professionals because of the negative transfer of the historical traditional meaning of *nonverbal,* which many associate with nonverbally assessed abilities such as Gv and Gf on the Wechslers. Conzelmann and Suß (2015), after a review of the major definitions of auditory abilities and auditory intelligence (including the primary-source Ga studies of Horn and Stankov), have a much better handle on a more understandable definition of Ga, which we now use here.

There are multiple misconceptions regarding Ga. First, although Ga depends on sensory input via the momentary perturbations of air pressure in our ears, it is not sensory input itself. Ga is what the brain does with sensory information from the ear, sometimes long after a sound has been heard (e.g., after he became deaf, Beethoven composed some of his best work by imagining how sounds would blend). The second common misconception, even among professionals, is that Ga is oral or verbal language comprehension. It is true that one aspect of Ga (parsing speech sounds, or phonetic coding) is related to oral language comprehension—but this is simply a precursor to comprehension, not comprehension itself (in the same way that adequate vision is a prerequisite for playing tennis, but vision is not normally thought of as a tennis skill). This distinction was made by Carroll (1993) and recently reinforced by Conzelmann and Suß (2015), who noted that in oral or verbal comprehension tasks the entirety of the comprehension process is relevant, not just the processing of tones or speech units. Third, as also articulated by Carroll, musical comprehension, ability, aptitude, or expertise is not included in Ga, as in musical abilities the whole sequence (melody) is considered and needs to be comprehended. Certain Ga narrow abilities are important to the development of musical abilities or aptitude (e.g., maintaining and judging rhythm)—but just as mathematical aptitude (which probably requires certain Gf, Gv, and Gwm abilities and certain noncognitive characteristics) is not considered a CHC factor-based ability, musical ability, aptitude, or expertise is an amalgam of cognitive and noncognitive characteristics and should not be listed as a CHC narrow ability.

Ga has long been the Rodney Dangerfield ("I don't get no respect")[35] of CHC abilities and is often considered the "secondary" sense behind Gv (Lotto & Holt, 2011). Its second-class status is reflected in Ga's not being included in many models of intelligence and in its absence from all major individually administered intelligence batteries (save the Woodcock–Johnson series; Conzelmann & Suß, 2015). This Ga neglect is partially understandable, given that the construct has historically received much less attention than Gv abilities, in part due to the lack of reliable and valid technology for measuring Ga abilities. Ga was the least studied factor in Carroll's (1993) treatise.

This neglect is no longer scientifically sustainable. Ga serves the important function of providing perceptual and cognitive scaffolding ("auditory scaffolding") for many temporally based higher-order cognitive functions such as language (Conway, Pisoni, & Kronenberger, 2009). Ga abilities play important roles in such diverse activities as conversations, performance bottlenecks (e.g., driving a car), navigating in the dark, musical performance, foreign-language acquisition, and understanding of reading and language disorders (Conzelmann & Süß, 2015). Ga requires considerable complex perceptual–cognitive processing (e.g., attention, localization, memory, segmentation, categorization, and pattern recognition) of multiple competing, transient, and temporally ordered brief sound waves (Lotto & Holt, 2011). Clearly, Ga-related abilities are critically involved in many important human functions. Ga requires multiple cognitive processing mechanisms that are equal to, and in many cases more complex than, those involved in many Gv abilities.

These conclusions have been reinforced by a systematic program of research by Rammsayer and colleagues (Haldemann, Stauffer, Troche, & Rammsayer, 2012; Helmbold, Troche, & Rammsayer, 2006, 2007; Pahud, 2017; Rammsayer & Brandler, 2004, 2007; Rammsayer & Troche, 2016), which has demonstrated that a temporal *g*-factor demonstrates higher correlations with a psychometric *g* factor than does a classic Jensen RT *g* factor. These researchers have consistently demonstrated that temporal *g* (which they refer to as the *temporal resolution power* [TRP] hypothesis of the central nervous system) may be a causal factor for speed of information processing (reaction time *g*), which in turn affects general intelligence. Given that sound is inherently a temporal and sequential signal (Conway et al., 2009; Conzelmann & Süß, 2015; Kraus & Slater, 2016; Lotto & Holt, 2011; Pahud, 2017; Slevc, 2012; Tallal, 2004), the impressive TRP findings demand that intelligence scholars and test developers recognize the importance of Ga. This position is also supported by analysis of the only major intelligence test battery that adequately represents Ga. In the WJ III and WJ IV Cognitive batteries, the Ga cluster loaded higher than the Gv cluster on a single *g* factor extracted from the seven CHC broad cluster scores across the entire age range of the two norm samples (Ga = .72/.78; Gv = .66/.68).[36] Ga abilities should be integral to theories of intelligence and cognitive assessment, as "the ability of the auditory system to segregate, locate, and categorize events in the environment is a *remarkable accomplishment* given the complexity and transient nature of sound waves" (Lotto & Holt, 2011, p. 479; emphasis added).

The importance of Ga is now recognized by an ever-widening range of research in psychology, psychometric studies of intelligence, neuropsychology, and cognitive neuroscience (Conway et al., 2009; Conzelmann & Süß, 2015; Kraus & Slater, 2016; Lotto & Holt, 2011; Pahud, 2017; Rammsayer & Brandler, 2007; Slevc, 2012; Wolff & Gustafsson, 2015). Unfortunately, this embarrassment of riches has yet to be organized into a coherent interdisciplinary framework (or frameworks). One of the most important factor-based studies (since our 2012 chapter, which informs the current chapter) has been the Conzelmann and Suß (2015) study, which suggests that the description of Ga abilities could benefit from an auditory-*nonverbal* (e.g., tones, environmental sounds) and auditory-*speech* (e.g., language-related sound units) content facet distinction.

Narrow Abilities within Ga

The following Ga abilities are classified as auditory-*speech* abilities:

1. **Phonetic coding (PC): The ability to distinctly hear phonemes, blend sounds into words, and segment words into parts, sounds, or phonemes.*[37] This ability is also referred to as *phonological processing*, *phonological awareness*, and *phonemic awareness*.[38] People with poor phonetic coding have difficulty hearing the internal structure of sound in words. This makes sounding out unfamiliar words while reading difficult. Poor phonetic coding is one of the major risk factors in reading disorders, specifically phonological dyslexia. Most people, even with very low Ga, can understand speech and speak perfectly well without awareness of the distinct phonemes they are hearing and saying. What they lack is the ability to mentally separate phonemes and hear them in isolation.

Research continues to reinforce the prior conclusion (McGrew, 2005) that PC is a single dimension and is not multidimensional (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Nelson, Lindstrom, Lindstrom, & Denis, 2012; Wolff & Gustafsson, 2015). Of interest is the recent finding that, aside from different PC measures' being differentiated primarily along a single developmental psychological sensitivity dimension (Pufpaff, 2009), PC tasks can probably be differentiated in terms of two facets—*linguistic* (phoneme/syllable vs. morpheme) and *cognitive complexity* (blending/segmentation vs. manipulation) (Wolff & Gustafsson, 2015).

2. *Speech sound discrimination (US): The ability to detect and discriminate differences in speech sounds (other than phonemes) under conditions of little or no distraction or distortion.* The definition of this factor has been narrowed to nonphonemic aspects of speech sounds, to make it more distinct from phonetic coding. People who have poor speech sound discrimination are less able to distinguish variations in tone, timbre, and pitch in speech; this might reduce their ability to detect subtle emotional changes, or subtle changes in meaning due to differential emphasis.

3. *Resistance to auditory stimulus distortion (UR): The ability to hear words or extended speech passages correctly under conditions of distortion or background noise.* It is not yet clear to what degree this ability depends on sensory acuity. As people age, they tend to complain that they have greater difficulty understanding speech in noisy public places or on a telephone with background noise. Speaking louder usually helps them understand better.

The following Ga abilities are classified as auditory-*nonverbal* abilities:

4. **Maintaining and judging rhythm (U8): The ability to recognize and maintain a musical beat.* This may be an aspect of memory for sound patterns, as short-term memory is clearly involved. However, it is likely that something different about rhythm warrants a distinction. Important research has occurred regarding this ability since our prior chapter, pieces of which are mentioned later.

5. *Memory for sound patterns (UM): The ability to retain (on a short-term basis) auditory events such as tones, tonal patterns, voices, or speech sounds.*[39] This ability is important for musicians, who need to be able to hold in mind a musical phrase they hear so that they can reproduce it later.

6. *Musical discrimination and judgment (U1 U9): The ability to discriminate and judge tonal patterns in music with respect to melodic, harmonic, and expressive aspects (phrasing, tempo, harmonic complexity, intensity variations).*

7. *Absolute pitch (UP): The ability to perfectly identify the pitch of tones.* As a historical tidbit, John Carroll had perfect pitch.

8. *Sound localization (UL): The ability to localize heard sounds in space.*

Assessment Recommendations for Ga

Ga is unusual in CHC theory, in that psychologists have been more interested in a narrow ability (phonetic coding, PC) than in the broad ability (Ga). We believe that this focus on phonetic coding will be changing shortly. Other Ga abilities are clearly related to musical achievement and are priorities if one is attempting to assess musical aptitude, or assess impairment for a brain-injured musician. Based on the continuing explosion of interdisciplinary Ga research, we believe that several Ga abilities (and emerging abilities) warrant more attention from test developers and eventually from assessment professionals. Measures of maintaining and judging rhythm (U8) and temporal processing (temporal *g*; TRP) need to find their way into test development and assessment practices. These new assessment possibilities are briefly highlighted below.

Comments and Unresolved Issues Related to Ga

- *The CHC home of temporal tracking (UK; Carroll, 1993) is still undetermined. Does it belong under Ga or Gwm?* Previously, this factor was listed as part of Ga. *Temporal tracking* was defined as the ability to mentally track auditory temporal (sequential) events to be able to count, anticipate, or rearrange them (e.g., to reorder a set of musical tones). This factor is measured by tests that require simultaneous storage and processing; thus it appears that such tests are methods of measuring attentional control within working memory capacity (Stankov, 2000).[40]

- *Can the assessment of phonetic coding (PC) be improved?* Yes. We believe that the facets of linguistic and cognitive complexity identified by Wolff and Gustafsson (2015) should be incorporated into the design and interpretation of new and existing measures of PC.

- *Did the removal of Carroll's (1993) hearing speech threshold (UA UT UU), sound-frequency discrimination (U5), sound-intensity/duration discrimination (U6), and general sound discrimination (U3) factors from the CHC model (in our 2012 chapter) make sense?* Based on the absence of anyone's crying "foul" because of our recommendation, we believe this recommendation is solid. These are sensory acuity factors that are outside the scope of CHC theory.

- *Can intelligence researchers and applied test developers catch the beat and develop psychometrically sound measures of maintaining and judging rhythm?* Researchers at the Northwestern Auditory Neuroscience Lab[41] have published a series of studies that demonstrate significant relations between measures of beat synchronization (i.e., the coordination of movement with a pacing sound or metronome) and evoked auditory brainstem response, neural coding of speech, psychometric indicators of reading and language development, and specific reading and language disorders in children (Carr, Fitzroy, Tierney, White-Schwoch, & Kraus, 2017; Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014; Tierney, White-Schwoch, MacLean, & Kraus, 2017). Timing or temporal processing has also been linked to mathematics achievement in children (Tobia, Rinaldi, & Marzocchi, 2016). In adults, a battery of rhythm tests suggested two rhythm factors (sequencing and synchronization) that also showed significant relations with measures of brain function and verbal memory and reading (Tierney et al., 2017). We predict that practical measures of beat or rhythm production and synchronization will be part of a coming wave of new Ga-based psychometric tests.

- *It is "time" that intelligence researchers and applied test developers recognize the importance of brain clock timing mechanisms (temporal g, TPR) and develop practical measures of temporal processing and other abilities related to mental timing—which are typically administered via the auditory modality.* Given today's relatively low-cost and portable technology platforms, and a plethora of existing temporal (auditory-based) test prototypes employed in research settings (e.g., see Conzelmann & Suß, 2015—detection of repeated tones, tonal series, tonal analogies, rhythm production, recognition of familiar environmental sounds; also see Rammsayer & Brandler, 2007—rhythm perception, temporal-order judgment, auditory flutter fusion, duration discrimination, temporal generalization), the stage is set for a long-overdue tsunami of temporal test development.

Olfactory Abilities (Go)

Definition of Go

Olfactory abilities (Go) can be defined as the abilities to detect and process meaningful information in odors. Go refers not to sensitivity of the olfactory system, but to the cognition one does with whatever information the olfactory system sends via its dual-model detection system (i.e., sniffing via the external nostrils, and detecting odors arising from within the mouth via the nasopharynx) (Stevenson, 2013).

Although olfaction is the most ancient of the human senses, it is one of the least studied in relation to the practice of psychological assessment. However, odors are a big-money business. The perfume and fragrance industries invest billions of dollars in the identification of various nuances of pleasant odors. The food and beverage industries pay significant attention to the olfactory characteristics of their products. At a more primitive evolutionary level, invertebrates have evolved Go systems over time to perform three primary functions: (1) learning olfactory cues that signal nutrients or toxins when eating and drinking; (2) detecting environmental hazards, such as predators or chemical cues to diseases; and (3) the transmission of social information (Stevenson, 2013). Many of these functions are no longer evolutionarily critical to humans.

Research suggests an attentional component to olfaction. People can voluntarily selectively attend to odors in the environment, and often are involuntarily and unpleasantly forced to attend to smells in the environment (e.g., detecting the smell of natural gas or rotting food) (Stevenson, 2013). Perhaps a narrow ability of olfactory attention (OA?) will be validated in the future. Also, Engen (1982) has suggested that "Functionally, smell may be to emotion what sight or hearing are to cognition" (p. 3). Consistent with this view is research suggesting that different words have consistently been linked to different odors, with unpleasant terms outnumbering pleasant terms (Stevenson, 2013). Perhaps Gei (broad emotional intelligence) may include a narrow olfactory emotion recognition (OER?) ability. Perhaps not. Researchers willing to sniff around this topic will help answer this question.

Clearly, the Go domain includes more narrow abilities than currently listed in the CHC model. In addition to the possibilities mentioned above, a cursory skim of Go-related literature reveals reference to such abilities as olfactory memory, episodic odor memory, olfactory sensitivity, odor-specific abilities, odor identification and detection, odor naming, and olfactory imagery, to name but a few. Among the reasons why the Minnesota Multiphasic Personality Inventory has items about "peculiar odors" are that distorted and hallucinatory olfaction is a common early symptom of schizophrenia, and that poor olfaction is an associated characteristic of a wide variety of brain injuries, diseases, and disorders (Doty, 2001; Dulay, Gesteland, Shear, Ritchey, & Frank, 2008).

For Go skeptics who dislike the scent of Go's being included in CHC theory, contemporary research is increasingly suggesting that olfactory dysfunction often acts as a "canary in the coal mine" for neurological insult or decline. Olfactory dysfunction has been associated with a wide variety of disorders, including ALS, Alzheimer disease, bipolar disorder, depression, epilepsy, HIV, dementia, Huntington disease, Kallmann syndrome, motor neuron disease, multiple sclerosis, neuromyelitis optica, obsessive–compulsive disorder, Parkinson disease, and schizophrenia (Joseph & DeLuca, 2016; Martzke, Kopala, & Good, 1997; Stevenson, 2013), although some of this research has been contradictory for specific disorders (e.g., bipolar disorder, depression, epilepsy, multiple sclerosis). The most important practical implication is that impaired olfactory functioning (Go) may be a marker of disease progression (Martzke et al., 1997).

Narrow Abilities within Go

1. *Olfactory memory (OM) is the ability to recognize previously encountered distinctive odors.* The oft-noted experience of smelling a distinctive smell and being flooded with vivid memories of the last time this odor was encountered does have some basis in research. Memory for distinctive odors has a much flatter forgetting curve than many other kinds of memory (Danthiir, Roberts, Pallier, & Stankov, 2001).

Assessment Recommendations for Go

Most practical and clinical applications of smell tests are sensory acuity tests. People who work where gas leaks must be tested regularly to make sure that they can make potentially life-saving odor detections. The most common olfactory tasks used in research are olfactory acuity, identification, discrimination, recognition, and memory (Martzke et al., 1997). The University of Pennsylvania Smell Identification Test (UPSIT),[42] originally developed by Doty, Shaman, and Dann (1984), is the most recognized commercially available test of olfactory functioning (Doty, 2001). Another recent entry in the olfactory testing field is the “Sniffin’ Sticks” test, which consists of tests of odor threshold, discrimination, and identification (Hummel, Kobal, Gudziol, & Mackay-Sim, 2007). We are unable to make specific recommendations regarding the use of these tests in psychological assessment, as we are unaware of any well-designed factor-analytic studies of these measures, particularly within the CHC framework.[43] Such research is encouraged.

Comments and Unresolved Issues Related to Go

- *Are there separate short-term and long-term OM abilities, or are they part of a common single memory system?* Research regarding this question is inconclusive (Stevenson, 2013).
- *Is there a working memory for odor?* Jönsson, Møller, and Olsson (2011) demonstrated that olfactory stimuli could be retained in a short-term memory store and that this store could be continuously updated. However, they found it difficult to disentangle the influence of an individual’s odor discrimination and verbalization abilities during the task (Danthiir et al., 2001). Jönsson and colleagues concluded: “Altogether, the present study demonstrates the ability to maintain information about odorants online, updating this information in service of correctly matching odors in a series to previously presented ones. This is true for nameable odors, but also, to a lesser degree, for odors that are notoriously difficult to name. This is in line with the notion of a separate olfactory slave system, but it is premature to draw firm conclusions at this point” (p. 1030).
- *Is odor naming (ON?) a narrow ability?* If so, like face naming, is poor odor naming partially dependent on visual and verbal contextual cues? Is odor naming dependent on a different set of retrieval fluency (Gr) abilities from those associated with naming ability and retrieval fluency? Does the ability to name odors suggest a separate odor knowledge ability (OK?) and odor retrieval fluency ability (OR?)? One of the most robust findings in olfactory research is the finding that people have difficulty naming odors and rarely correctly identify over 50% of familiar odors (Stevenson, 2013).
- *Is there such an ability as odor imagery (OI)?* Although imagery has been documented in both the visual and auditory systems, research regarding whether, in the absence of immediate presence in the environment, odors can be in the “mind’s nose” (Stevenson & Case, 2005) is confusing and inconclusive (Danthiir et al., 2001; Stevenson, 2013).
- *Do the studies indicating that measures of simple olfactory functions might be an “early warning system” for underlying neurocognitive disorders reflect real Go deficits, or are these deficits due to problems with other cognitive abilities required to perform these Go tasks? Is it possible to disentangle common cognitive abilities from measures of Go, or are Go abilities by nature dependent on cognitive abilities?* In a study of olfactory sensitivity, discrimination, and identification, Hedner, Larsson, Arnold, Zucco, and Hummel (2010) found that performance on common Go measures was influenced by the participants’ executive functioning and semantic and episodic memory.
- *Does olfactory sensitivity (OS) belong in CHC theory?* This is the ability to detect and discriminate differences in odors. That is, it is a sensory acuity factor, and we believe it is thus outside the scope of CHC theory.

We await innovations in measurement and well-designed studies that will answer these questions and help determine the taxonomy of Go-related cognitive abilities. Given the Go link to a wide variety of cognitively related disorders, we hope for future psychometric research in this area. Curious noses want to know.

Tactile Abilities (Gh)

Definition of Gh

Tactile abilities (Gh) can be defined as the abilities to detect and process meaningful information in haptic (touch) sensations. This domain includes perceiving, discriminating, and manipulating touch stimuli. Gh refers not to sensitivity of touch, but to the cognition one performs with tactile sensations. Because this domain is not yet well defined and understood, it is hard to describe authoritatively. We speculate that it will include such things as tactile visualization (object identification via palpation), localization (where has one been touched), memory (remembering where one has been touched), texture knowledge (naming surfaces and fabrics by touch), and many others. Tests of Gh have long been used in neuropsychological batteries because of their ability to detect brain injury, especially to the somatosensory cortex. Attempts to develop haptic-based assessment batteries have also been made for individuals who are blind or have severe visual disabilities (e.g., the Blind Learning Aptitude Test).

Much to the disdain of new parents, our initial explorations of the world as infants is filled with touching, grabbing, and sucking almost any object within immediate reach. The human sense of touch, be it with our hands or mouth, provides perceptual information regarding our immediate environment and serves as a primary foundation for the development of many concepts. The importance of touch permeates our discussion of learning, as "we often talk about 'grasping' an idea, 'getting a handle on' a problem, or being 'touched' by a reading" (Minogue & Jones, 2006, p. 317). Driven by recent technological developments—for instance, in robotics, 3D printers (making it possible to develop precise 3D stimuli), prosthetic limbs and hands, touch screen mobile devices, haptic feedback displays, technology to aid the visually impaired, the teleoperation of remote sensing or manipulation devices (e.g., telesurgery, use of remote drones), and virtual-reality-based training and simulation (e.g., training of surgeons)—interest and research in tactile or haptic abilities are increasing (Kappers & Bergmann Tiest, 2013).

Circumscribing the emerging research on haptic perception is beyond the scope of this chapter. Haptic perceptual characteristics include material properties (e.g., roughness, compliance, viscosity, friction, temperature, density, and weight) and spatial properties (e.g., shape, curvature, length, volume, and orientation), as well as quantitative properties such as numerosity (Kappers & Bergmann Tiest, 2013). A complete understanding of haptic perception requires an understanding of the peripheral sensory receptors (in the skin, muscles, tendons, and joints); of research that has identified two channels of haptic perception ("what" and "where"); and of other issues such as vision–touch interactions, affective touch, and neural plasticity (Lederman & Klatzky, 2009). Furthermore, consumer psychologists have learned that some individuals have a "need for touch" (Peck & Childers, 2003) when evaluating products to counter common visual misperceptions (some scientists refer to touch as the "reality sense"; Nuszbaum, Voss, Klauer, & Betsch, 2010). The implications of touch for cognition are recognized by many educators who advocate the use of "hands-on" instruction (see Minogue & Jones, 2006, for emerging educational applications of haptic sensations).

Narrow Abilities within Gh

Despite the recent increase and variety of haptic perception research and applications, the limited structural evidence research does not allow us to articulate a more nuanced version of the fundamental factors of haptic abilities than the one outlined in our 2012 chapter. In other words, there are as yet no well-supported cognitive ability factors within Gh. Although Stankov, Seizova-Cajic, and Roberts (2001) identified a narrow tactile sensitivity (TS) factor, this is a sensory ability (i.e., the ability to make fine discriminations in haptic sensations) and not a cognitive ability. For example, if two caliper points are placed on the skin simultaneously, we perceive them as a single point if they are close together. Some people can make finer discriminations than others. The very narrow TS factor was found to be minimally related to higher-level broad CHC abilities (Gf, Gv, Ga; Stankov et al., 2001). Two new (or previously overlooked) Gh structural evidence studies summarized below were either inconclusive or based on samples too small to suggest revisions to the Gh domain.

In a factor study of the tactile measures from the Dean–Woodcock Sensory Motor Battery and CHC measures with a co-normed cognitive battery of measures representative of the CHC model of intelligence, Decker (2010) found that the Palm Writing and Object Identification tests did not form a distinct Gh factor and either loaded on a processing speed (Gs) factor or were factorially complex (Gs and Gv). Ballesteros, Bardisa, Millar, and Reales (2005) investigated the psychometric characteristics, including factor-analytic structure, of a psychological test battery designed to measure the perceptual and cognitive abilities of children with visual handicaps. The 20-test battery materials consist of raised-dot, raised-surface shapes and displays, and familiar and novel 3D objects requiring active touch. In a small sample, exploratory factor analysis identified six factors—spatial comprehension, short-term memory, object identification, shape identification efficient exploration, material and texture discrimination. Given the small sample size and the lack of other CHC ability indicators, the Ballesteros and colleagues study can only be considered a suggestive first step in the exploration of the structural nature of Gh.

Assessment Recommendations for Gh

Most practical and clinical applications of Gh tests actually use sensory acuity tests. There are currently no available tests of higher-order Gh processes that are clearly distinct from Gv or Gs. The Halstead–Reitan Neuropsychological Test Battery and the Dean–Woodcock Neuropsychological Battery include several Gh tests.

Comments and Unresolved Issues Related to Gh

- *How is Gh to be distinguished from Gv and Gf?* Two well-designed studies (Roberts, Stankov, Pallier, & Dolph, 1997; Stankov et al., 2001) found it difficult to distinguish between complex tests assumed to measure Gh and well-defined markers of Gv and Gf. Why might this be so? If a test involves identifying common objects (coins, keys, books, etc.) by handling them while blindfolded, the examinee is essentially using the hands instead of the eyes to visualize an object in the "mind's eyes."
- *What about "dynamic touch"?* Do abilities from the Gh and kinesthetic (Gk) domains combine to reflect individual differences in dynamic touch (Stankov et al., 2001; Turvey, 1996)?
- *Like the presence of imagery in vision and auditory abilities, does some form of haptic imagery ability exist? If so, what role would it play in Gh abilities?*

Kinesthetic Abilities (Gk)

Definition of Gk

Kinesthetic abilities (Gk) are defined as the abilities to detect and process meaningful information in proprioceptive sensations. The term *proprioception* refers to the ability to detect limb position and movement via proprioceptors (sensory organs in muscles and ligaments that detect stretching). Gk refers not to the sensitivity of proprioception, but to the cognition one does with proprioceptive sensations. Because this ability is not yet well understood, and because we have not located any significant new structural evidence literature related to the nature of this ability domain since our 2012 chapter, we can only continue to speculate that it will include such things as a dancer's ability to move into a certain position and visualize how it looks to another person (which would have Gv components), and knowledge of which body movements will be needed to accomplish a specific goal (e.g., passing through a narrow space). Such abilities are likely to be involved in Gardner's bodily–kinesthetic intelligence (see Chen & Gardner, Chapter 4, this volume). One interesting possibility is that proprioceptive receptors and other receptors in muscles are used in inferring characteristics of objects that are hefted and wielded (Turvey, 1996). That is, when an object is held and waved about (dynamic touch), one can get a sense of its length, weight, and mass distribution.

Higher-order cognition occurs when tactile information informs potential uses (affordances or "action possibilities"; Gibson, 1979) of the object (e.g., a hammer, a lever, a weapon). In our previous chapter, we hypothesized that Gk and Gp are so interconnected that they may form the same broad-ability construct. We noted that although there is a clear physiological distinction between motor abilities and kinesthetic perception, motor performance is constantly informed by sensory feedback, and thus Gk and Gp might be considered an integrated functional unit. Perhaps our speculation regarding a broad-ability Gf + Gp umbrella construct does not reflect a higher-order ability construct, but instead reflects the functional role played by Gk (as well as Gp and Gh) in understanding recent developments in the cognitive science of embodied cognition (Borghi & Cimatti, 2010; Mahon & Caramazza, 2008; Wilson, 2002).

The historical roots of embodied cognition date back to early 20th-century philosophers (Martin Heidegger, Maurice Merleau-Ponty, and John Dewey). In simple terms, *embodied cognition* means that our cognition is not restricted to the gray and white matter computing unit between our ears, and is perhaps determined by our interaction and experiences (e.g., movement and touch) in the physical world. The embodied-cognition literature, although having a relatively short history in comparison to mind-based cognitive sciences, is growing fast. This literature is somewhat construct-dense, filled with unresolved issues, and counterintuitive to the traditional view of "the mind as an abstract information processor, whose connections to the outside world were of little theoretical importance. Perceptual and motor systems, though reasonable objects of inquiry in their own right, were not considered relevant to understanding 'central' cognitive processes" (Wilson, 2002, p. 625).

The embodied-cognition perspective emphasizes the importance of sensory and motor functions for interacting with the environment. As summarized by Wilson (2002),

> there is a growing commitment to the idea that the mind must be understood in the context of its relationship to a physical body that interacts with the world. It is argued that we have evolved from crea-

> tures whose neural resources were devoted primarily to perceptual and motoric processing, and whose cognitive activity consisted largely of immediate, on-line interaction with the environment. Hence human cognition, rather than being centralized, abstract, and sharply distinct from peripheral input and output modules, may instead have deep roots in sensorimotor processing. (p. 625)

We can see the concept of embodied cognition reflected in such everyday sayings as "This is over our heads," "I've warmed up to him," "I'm on top of the situation," "They are at the height of their power," or "I could feel the electricity between us" (McNerney, 2011).[44] Finally, given that recent "learning by acting" strategies derived from generative learning theory have proven effective in enhancing learning in the STEM fields (Fiorella & Mayer, 2016), we believe that intelligence scholars and educational psychologists need to spend more time operationalizing better measures of Gk (and Gp and Gh) and completing structural validity studies to establish the possible cognitive elements that underlie embodied cognition.

Narrow Abilities within Gk

There are no well-supported cognitive ability factors within Gk yet. Kinesthetic sensitivity (KS), a sensory acuity ability, refers to the ability to make fine discriminations in proprioceptive sensations (e.g., whether and how much a limb has been moved).

Assessment Recommendations for Gk

We are unaware of commercially available measures of Gk. Very little is known about the measurement of Gk. Readers are referred to Stankov and colleagues (2001) for ideas about Gk tests.

Comments and Unresolved Issues Related to Gk

- *How separate is Gk from Gp?* We suspect that Gk and Gp (and possibly Gh) are so interconnected that they may form the same broad-ability construct or may be functional components of what is now called embodied cognition.
- *How do the CHC abilities related to immediate environment–body interaction (Gk, Gp, Gh) fit within the embodied-cognition literature? Conversely, how do psychometrically and cognitively trained psychologists wrap their heads around the concept of embodied cognition? Where does the embodied-cognition research literature fit within the CHC taxonomy—if it does at all?*
- *What about "dynamic touch"?* Do abilities from the Gh and kinesthetic (Gk) domains combine to reflect individual differences in dynamic touch (Stankov et al., 2001)?

Psychomotor Abilities (Gp)

Definition of Gp

Psychomotor abilities (Gp) can be defined as the abilities to perform physical body motor movements (e.g., movement of fingers, hands, legs) with precision, coordination, or strength. The Gp domain received little attention in our previous chapter (approximately half a page) or in Carroll's (1993) seminal treatise on the factor structure of cognitive abilities. The reasons for giving Gp the cold shoulder are threefold. First, psychomotor abilities historically have been considered to represent a noncognitive domain independent of intelligence (Chaiken, Kyllonen, & Tirre, 2000). Second, Carroll did not set sail to capture the extant psychomotor research literature—"my survey of cognitive abilities was not intended to cover the domain of physical and psychomotor abilities, but many of the datasets included measures of psychomotor abilities, with the result that a number of interpretable factors in the domain appeared" (Carroll, 1993, p. 532). Essentially, Gp was "bycatch"[45] found when he cast his wide net to capture cognitive ability datasets. Yet, largely due to the origins of the first CHC list of broad and narrow factors, the original incomplete list of Gp narrow abilities in the CHC taxonomy took on a life of its own independent of serious scientific scrutiny. Third, many educational and school psychologists (of our generation) have scars from the early heydays of special education programming, when psychomotor planning or reprogramming intervention was believed to improve academic achievement (e.g., having children relearn to crawl, walk, etc., was believed to help children learn to read). Speaking the phrase "Doman Delecato method" to many educational and school psychologists of our generation might elicit almost posttraumatic reactions.

Aside from Carroll's (1993) self-admitted "duct tape" Gp overview, which was based on Gp bycatch studies integrated with prior structural reviews of the Gp literature, little structural evidence research has been completed in the Gp domain since a relatively active period from the

1950s through the early 1980s (Fleishman, 1964, 1972; Fleishman, Quaintance, & Broedling, 1984; Guilford, 1958; Peterson & Bownas, 1982). Given the assumed disconnect between Gp and intelligence, many psychologists are at peace with this unfinished charting of the Gp landscape. Why bother learning about Gp abilities? Why be concerned with the out-of-date Gp structural evidence literature reviews? Because contemporary cognition research and theory, as well as increasing interests in newly defined motor-based disorders (e.g., developmental coordination disorder, or DCD), require us to pay closer attention to this intelligence stepchild.

As noted in our discussion of Gk, the growing embodied-cognition research is resulting, in some corners of the field of cognitive science, in a growing commitment to the idea that the mind must be understood in the context of its relationship to a physical body that interacts with the world. Hence human cognition, rather than being centralized, abstract, and sharply distinct from peripheral input and output modules, may instead have deep roots in sensory–motor processing (e.g., Gp, Gk, and Gh). For embodied-cognition research to advance and be evaluated, it is important that an empirically based, fully articulated working Gp taxonomy be available.

Structural and causal modeling research (Chaiken et al., 2000), which has confirmed the validity of a broad Gp factor, also suggests that the stereotype of Gp as distinct and independent from cognitive abilities is wrong. Chaiken and colleagues (2000) reported that general cognitive ability is a major component of general psychomotor ability: “Cognitively able individuals tend to do well on psychomotor tasks” (p. 222). These researchers also reported that processing speed (Gs) and temporal processing cognitive abilities are related to Gp abilities. Furthermore, these researchers suggested that the impact of general intelligence on Gp may be due to working memory (Gwm) capacity: “Working memory may be what limits psychomotor performance. Working memory may impact on psychomotor ability in two ways, via complexity and via novelty” (p. 222). Also, preliminary CFA of the CHC-organized Indonesian AJT-CAT (which also includes two indicators of finger dexterity, P1, and one indicator of manual dexterity, P2; see “CHC’s Global Reach,” above) supports the validity of a Gp broad ability that loads in the mid-.40s on a higher-order *g* factor.

Finally, the pseudoscience of diagnosing students with fine and gross motor disabilities to be solved by motor programming or reprogramming has been replaced by theoretically and empirically sound evidence for DCD. DCD is listed under the category of neurodevelopmental disorders in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013). DCD is a recognized as a complex neurodevelopmental disorder characterized primarily by poor Gp abilities that affect a child’s daily learning and living (Cermak & Larkin, 2002; Gillberg & Kadesjö, 2003; Henderson & Hall, 1982). “Children with DCD show (1) acquisition and execution of motor skills below what would be expected at a given chronological age and opportunity for skill learning and use; (2) motor skills deficit[s] significantly interfering with activities of daily living (ADL) appropriate to the chronological age, impacting school productivity, leisure and play; (3) early development onset of symptoms; (4) motor skills deficits that cannot be explained by intellectual disability [or] visual impairment and [are] not attributable to a neurological condition” (Bieber et al., 2016, p. 115). Not unexpectedly, given the relatively soft foundation of the current Gp taxonomy (when compared to the CHC cognitive ability taxonomy), DCD diagnostic, assessment, and intervention research is constrained by a lack of comprehensive measures of the entire range of affected motor abilities (Bieber et al., 2016; Rivilis et al., 2011).

Narrow Abilities within Psychomotor Abilities (Gp)[46]

1. *Manual dexterity (P1). The ability to make precisely coordinated movements of a hand or a hand and the attached arm.*

2. *Finger dexterity (P2): The ability to make precisely coordinated movements of the fingers (with or without the manipulation of objects).*

3. *Static strength (P3): The ability to exert muscular force to move (push, lift, pull) a relatively heavy or immobile object.*

4. *Gross body equilibrium (P4): The ability to maintain the body in an upright position in space or regain balance after balance has been disturbed.*

5. *Multilimb coordination (P6): The ability to make quick specific or discrete motor movements of the arms or legs.*

6. *Arm–hand steadiness (P7): The ability to precisely and skillfully coordinate arm–hand positioning in space.*

7. *Control precision (P8): The ability to exert precise control over muscle movements, typically in response to environmental feedback (e.g., changes in speed or position of object being manipulated).*

8. *Aiming (AI): The ability to precisely and fluently execute a sequence of eye–hand coordination movements for positioning purposes.*

Assessment Recommendations for Gp

Psychologists are not usually interested in Gp for its own sake, although this may change with the increasing interest in DCD. Neuropsychologists use measures of Gp, such as various grip tests and pegboard tests, to measure uneven performance with the right and left hands as an indicator of lateralized brain injury (Strauss, Sherman, & Spreen, 2006). I-O psychologists may use Gp measures for personnel selection in jobs that require manual dexterity. Occupational and physical therapists use measures of motor functioning with consistent regularity. For researchers or practitioners using the *International Classification of Functioning, Disability, and Health—Children and Youth Version* (World Health Organization, 2007), Bieber and colleagues' (2016) review of 19 clinical tests, 3 naturalistic observations, and 6 questionnaires is a good starting point for identifying measures of motor functioning.

It is apparent that a variety of different professionals working in different professional areas use common measures of Gp, as well as profession-specific Gp measures, with little in the way of cross-specialization communication. It is our opinion that the Gp assessment domain needs a program of research analogous to the XBA assessment work leading to the *Intelligence Test Desk Reference* (McGrew & Flanagan, 1998) and the XBA assessment system (Flanagan et al., 2013). That is, an updated empirically based Gp taxonomy needs to be established first, and then joint or XBA Gp assessment factor studies need to occur that will permit the primary Gp assessment instruments to be interpreted with a common nomenclature.

Comments and Unresolved Issues Related to Gp

- *The current Gp taxonomy, within the CHC model, is likely to be incomplete.* Efforts to complete a Carroll-like meta-analytic synthesis of the extant Gp literature are needed to revalidate existing Gp narrow abilities; determine which current Gp narrow-ability definitions should be revised; and, most importantly, identify Gp narrow abilities that have not yet been described.
- *Should psychometrically oriented CHC taxonomists monitor the embodied-cognition literature for potential new insights into the nature of the Gp domain?* We think that the answer should be yes.
- *How do the immediate environment–body interaction CHC abilities (Gk, Gp, Gh) fit within the embodied-cognition literature (Borghi & Cimatti, 2010; Mahon & Caramazza, 2008; McNerney, 2011; Wilson, 2002)?*

Emotional Intelligence (Gei)

Definition of Gei

Emotional intelligence (Gei) can be defined as the ability to perceive emotions expressions, understand emotional behavior, and solve problems using emotions. Below we present the narrow abilities from the Mayer–Salovey–Caruso four-branch model of emotional intelligence, but with the caveat that there have been few other exploratory efforts to uncover the Gei narrow abilities. Most research is focused on overall emotional intelligence and has employed only a few tests, typically from the MSCEIT. Strong tests of the existence of narrow abilities usually requires four or more indicator tests per ability, and the MSCEIT has only two indicators per factor. To explore the narrow-ability space in emotional intelligence, a wide range of new tests with diverse formats will need to be developed and delivered together.

Tentative Narrow Abilities within Gei[47]

1. **Emotion perception (Ep): The ability to accurately recognize emotions in the face, voice, and behavior.* The idea that there are important individual differences in the ability to recognize emotional expressions is probably the least controversial aspect of the emotional intelligence domain: It is a fundamental component of social competence (Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015), and may be important in academic functioning as well (Kang, Anthoney, & Mitchell, 2017; Trentacosta & Izard, 2007).

2. **Emotion knowledge (Ek): Knowledge of the antecedents of emotions and the consequences of emotional expression.* Understanding emotions is at the core of emotional intelligence (Denham, 2007) and has the highest correlations with the

"cool" intelligences, particularly with verbal ability (Roberts, Schulze, & MacCann, 2008). Low emotional knowledge is associated with many negative life outcomes, including social incompetence, social withdrawal, internalizing problems, externalizing problems, and academic difficulties (Chronaki, Garner, et al., 2015; Fine, Izard, Mostow, Trentacosta, & Ackerman, 2003; Izard et al., 2001; Schultz et al., 2001; Trentacosta & Fine, 2010; Trentacosta & Izard, 2007). Interventions designed to improve emotional knowledge in children appear to improve social competence and reduce behavioral problems (Giménez-Dasí, Fernández-Sánchez, & Quintanilla, 2015; Grazzani, Ornaghi, Agliati, & Brazzelli, 2016; Ornaghi, Brazzelli, Grazzani, Agliati, & Lucarelli, 2017).

3. *Emotion management (Em): The ability to regulate one's emotions deliberately and adaptively.* People who score higher on performance measures of emotion regulation ability are rated as more interpersonally sensitive and prosocial than people who score lower on such measures, and this effect persists even after researchers control for general intelligence and five-factor personality traits (Lopes, Salovey, Côté, & Beers, 2005).

4. *Emotion utilization (Eu): The ability to make adaptive use of emotion, especially to facilitate reasoning.* In the four-branch model of emotional intelligence (Mayer, Salovey, et al., 2008), utilizing emotion is an ability one uses consciously to facilitate reasoning and problem solving. Factor-analytic studies have not successfully isolated this factor, however. Izard (2001) believes that emotion utilization is better thought as a direct effect of emotion than as a distinct intellectual ability, but concedes that one can maximize the adaptive benefits of emotion if one understands emotions and can manage them first (Trentacosta & Izard, 2007).

Comments and Unresolved Issues Related to Gei

- *What is needed to remove the tentative status of Gei?* The scope of evidence supporting the validity of emotional intelligence is quite impressive. Although results from MacCann and colleagues (2014) are an impressive start, we would like to see more convergent and discriminant validity research showing that in a factor analysis of diverse cognitive abilities, emotional intelligence measures are distinct from other abilities and form a coherent broad ability. A related concern is whether emotional intelligence is distinct from other "hot" intelligences (Mayer et al., 2016), such as social intelligence (perceiving, understanding, and reasoning about interpersonal relationships) and personal intelligence (perceiving, understanding, and reasoning about personality). We suspect that these are not separate broad abilities and likely form their own cluster.

- *What is the structure of emotional intelligence?* We suspect that emotional intelligence is not a higher-order variable that simultaneously influences each narrow ability, but a causal system of variables arranged similarly to cascade models like those of Izard (1991, 2001) and Joseph and Newman (2010). That is, in the lower portion of Figure 3.12, emotional intelligence is not *in* the causal system; it *is* the causal system (Schneider et al., 2016).

Assessment Recommendations for Gei

Although there are many questionnaire measures of emotional intelligence, we recommend using well-normed ability tests like the MSCEIT or the WAIS-IV Advanced Clinical Solutions test. The least controversial aspects of Gei are related to emotion perception and emotion knowledge.

CHC ABILITIES AS PARAMETERS OF INFORMATION PROCESSING

We hypothesize that the various CHC broad abilities refer to parameters of information processing. New information comes into the brain via the senses. *Perception* is the process by which the brain makes sense of sensation. For example, four lines at right angles can be perceived to be a square. Gt refers to the speed at which perceptual elaboration can occur and is considered a "fundamental parameter" of information processing (Kail & Salthouse, 1994). The various perceptual processing factors of ability (Gv, Ga, Go, Gh, and Gk) govern the level of complexity at which a person can process perceptual information. That is, some people can more easily group simple percepts into more complex objects.

Attention refers to the processes that govern which percepts arise in consciousness. Some attention processes are automatic, but many require conscious control. As anyone who has tried to meditate knows, attention is only partly under our control. Attentional control refers to the steadi-

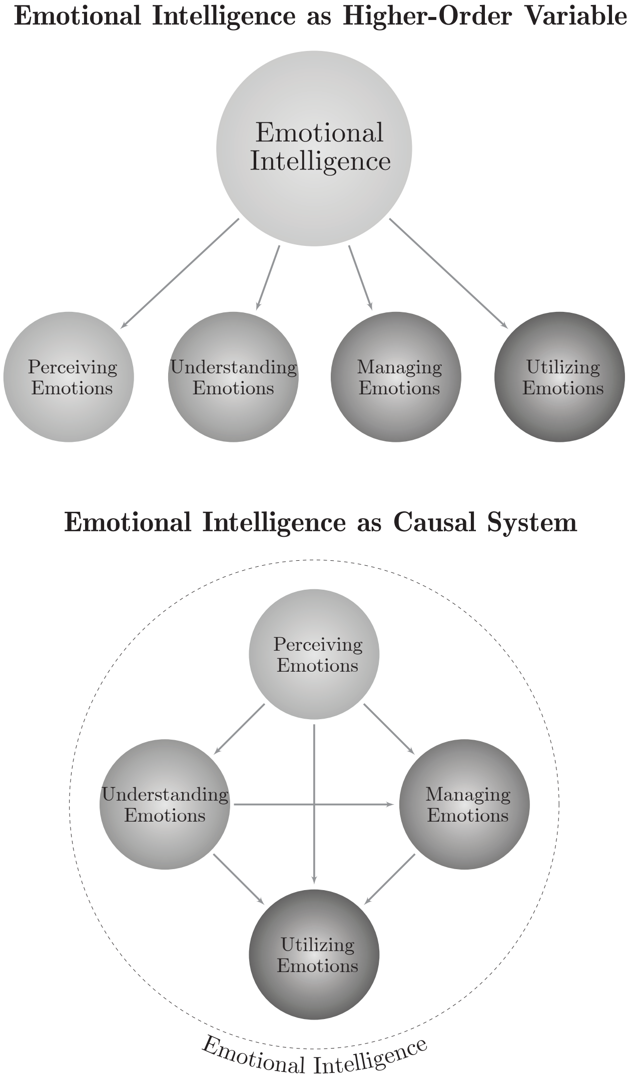


FIGURE 3.12. Two conceptualizations of emotional intelligence.

ness and precision of one's ability to control the focus of attention. Gs refers to how fluently one can shift the focus of attention from one object to the next to make a series of decisions.

Attention interfaces with many other cognitive systems (Hasson, Chen, & Honey, 2015). One can direct the focus of attention to engage in and maintain information in short-term memory, manipulate information in short-term memory, store information in long-term memory, retrieve information from long-term memory, engage in nonautomatic perceptual elaboration, reason through a novel problem, or direct motor movement (Baddeley, 2012; Unsworth, 2017; Unsworth & Engle, 2007). Because attentional control influences many tasks, we suspect that it is a partial explanation for the general factor of intelligence.

We suspect that the maturation of attentional control is the primary reason for the developmental cascade findings in which processing speed predicts working memory, which in turn predicts fluid reasoning (De Alwis et al., 2014; Fry & Hale, 1996; Kail, 2007; Nettelbeck & Burns, 2010). One explanation for the direct influence of processing speed on working memory capacity is that it allows for quicker subvocal rehearsal of information to maintain it in working memory (Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005; Kail & Salthouse, 1994). The link between working memory and fluid reasoning is that the operations of reasoning occur within working memory. Greater working memory capacity makes it possible to perceive more complex patterns and to represent more complex relationships among the fundaments of reason (Birney et al., 2006; Halford, Cowan, & Andrews, 2007; Halford et al., 1998). We also predict that the degree to which visualization tasks require attentional control determines their loadings on the fluid reasoning/general factor.

Memory refers to processes that govern the storage and retrieval of information. Gl refers to the efficiency with which information can be stored in long-term memory, and Gr refers to the fluency at which it can be retrieved. We should note that learning efficiency can be supercharged by three related abilities. First, fluid reasoning allows for faster learning via data reduction. That is, if a person can see—via fluid reasoning—that 10 seemingly separate facts are really manifestations of only two phenomena, the learning process is shortened considerably. Second, better language skills (i.e., the language development facet of Gc) allow for more efficient learning because of the conceptual tools embedded in language, especially vocabulary (Perfetti & Stafura, 2014). Third, prior knowledge about a topic allows for more efficient learning of related material (Beier & Ackerman, 2005; Hambrick & Engle, 2002). The breadth and depth of knowledge stored in memory are represented by the general knowledge facet of Gc and domain-specific knowledge (Gkn).

We hope that this presentation stimulates your thoughts and inspires you to contribute to the future development of CHC theory.

NOTES

1. Thanks to Todd Fletcher for this information.

2. Thanks to Laurie Ford and Damien Cormier for this information.

3. See *www.schuhfried.com/test/insbat.*

4. See *https://library.iated.org/view/ROMANGONZALEZ2015COM.*

5. See *https://tinyurl.com/k4qaune.*

6. See *www.time.edu.jo/our-team.html.*

7. See *www.projeiq.com/english.*

8. See *www.psychassessments.com.au/global/about-us.aspx.*

9. The Memory for Names test is not included in the WJ IV, but was part of the CFAs. It is included in the WJ IV Tests of Early Cognitive and Academic Development (ECAD; Schrank, McGrew, & Mather, 2015).

10. Kevin McGrew has served as the expert consultant for CHC and applied psychometrics on this project, and helped complete these preliminary structural analyses.

11. In the 2002 *Atkins v. Virginia* decision (see Haydel & Ferber, 2015), the U.S. Supreme Court ruled it unconstitutional to execute individuals with intellectual disabilities. Since *Atkins* cases typically rely on the definitions and guidelines of the American Association of Intellectual and Development Disabilities (2010) and the American Psychiatric Association (2013), IQ scores have become central to life-or-death Atkins decisions.

12. See *https://tinyurl.com/n7ww3zc.*

13. Kevin McGrew served as an external consultant to the SHARP project from 2012 to 2015.

14. This fallacy, which we like to call the jingle–jangle jungle, exists when erroneous assumptions are made that two different things are the same because they have the same name (the jingle fallacy), or that identical or almost identical things are different because they are labeled differently (the jangle fallacy).

15. We believe it is important, when possible, to differentiate narrow abilities within a CHC domain as major and minor. Major factors are those that represent the core characteristics of the domain, typically represent more complex cognitive processing, tend to display higher loadings on the g factor when present in the analysis, and are more predictive and clinically useful. Minor factors typically are less useful, less cognitively complex, and less g-loaded, and tend to "fall near the periphery of scaling representations, or at the bottom of a hierarchical model" (Lohman, 1979, pp. 126–127). Lohman (1979) made this clear distinction in his seminal Gv survey, which Carroll (1993) reinforced. In the discussions that follow, we designate the major abilities within a CHC broad domain with asterisks (*).

We have reordered the narrow abilities under each broad CHC domain ability so that the major abilities come first. Where possible, we have used the average g loadings for narrow factors as reported by Carroll (1993, Table 15.5, p. 597), as well as each narrow ability's average loading on its respective broad factor (e.g., average narrow factor loadings on the Gv factor as per Carroll, 1993, p. 609), to rearrange the list of narrow abilities in a rough major-to-minor continuum. In the absence of clear differences in g or broad factor loadings, we have kept functionally similar abilities adjacent to each other in our narrow-ability listings. Also, when average g loadings and average loadings within broad CHC domains differ between different narrow abilities, we have treated the within-broad-CHC-domain loadings as more important. Also, for many of the narrow abilities, no such information was summarized by Carroll; in the end, we have had to use our combined expert judgment for some of these rough ordering decisions. The reader should consider the order in which we list narrow abilities as relative, approximate, and not without possible error.

16. We would be remiss if we did not mention that a number of neo-Piagetians have advanced Piaget's original theory, especially in research related to education (e.g., mathematical thinking; see Fuson & Li, 2009).

17. Facets are based on facet theory and represent logically based classifications of test materials as per stimulus content characteristics (e.g., verbal, numerical, figures, etc.) and are not to be confused with ability factors. See Humphreys (1962).

18. Tinkering with the CHC taxonomy, and trying to stay consistent with the two-letter coding system that has evolved over time, are challenging endeavors and have required some fluid reasoning and creativity on our part. Historically, almost all narrow memory abilities were coded M__ (MV = memory visual, or visual memory; MA = memory associative, or associative memory; etc.). Using this system would not work here. Auditory short-term store would be MA, a code already used. Using W__, with the W standing for the "working" part of Gwm, might have worked better, but we could not use WA, as that stands for writing ability (WA) under Grw. The only "satisficing" solution was to use W and have the second letter be lowercase—thus eliminating any duplicate two-letter codes.

19. It should be noted that the broader concept of processing fluency in education also includes its influence in the metacognition of learning, belief formation, and affect. Also of interest is research suggesting that despite fluency's possessing a positive value connotation, sometimes cognitive dysfluency (which tends to have a negative connotation) may produce superior learning outcomes, as it results in less automatic responding and may prompt individuals to engage in more controlled, deliberate "deep" processing.

20. The concerns mentioned here apply to more than the domain of speed in the CHC model, with some ability domains more developed (e.g., Gf, Gc) and other less well developed (e.g., Ga, Gs, Gt).

21. The original Feldmann and colleagues (2004) study only presented factor-analytic evidence for a subset of the speeded variables in the study (viz., three number facility measures and three writing speed measures referenced in Carroll, 1993). We extracted the complete correlation matrix, which also included four perceptual speed tests from the DAS, WJ III, and WISC-III, and completed cluster analysis, MDS, and exploratory principal-components analysis (this was necessary due to singular matrix warnings) of all variables. Three distinct factors or dimensions emerged: perceptual speed (P), writing speed (WS), and number facility (N).

22. See *https://tinyurl.com/lyvfo9q.*

23. We believe that the Ackerman and colleagues memory and complex perceptual speed factors are not distinct narrow abilities, but represent factorially complex perceptual speed abilities that include other cognitive abilities (e.g., Gwm).

24. Given the lack of sufficient information in Carroll (1993), it was not possible to order these narrow abilities in this domain according to the major–minor distinction.

25. See King, Ryan, Kantrowitz, Grelle, and Dainis (2015); Sliwinski and colleagues (2016); and Wild, Howieson, Webbe, Seelye, and Kaye (2008).

26. Given the lack of sufficient information in Carroll (1993), it was again not possible to order the narrow abilities in this domain according to the major–minor distinction.

27. Bycatch is a term from the fishing industry for untargeted fish or marine life that are caught in fishing nets. The term also refers to untargeted material gathered in other forms of animal harvesting or collecting.

28. The broad notion of spatial thinking (or cognition, thinking or expertise) is conceptually like Snow's notion of aptitude–trait complexes. Spatial thinking includes not only the factor-analysis-based Gv psychometric abilities that have been identified (both procedural and declarative Gv abilities), but also spatial thinking dispositions (spatial habits of mind) and use of spatial strategies. "Spatial thinking is based on a constructive amalgam of three elements: concepts of space, tools of representation, and processes of reasoning. It depends on understanding the meaning of space and using the properties of space as a vehicle for structuring problems, for finding answers, and for expressing solutions. By visualizing relationships within spatial structures, we can perceive, remember, and analyze the static and, via transformations, the dynamic properties of objects and the relationships between objects" (National Research Council, 2006, p. 3). The reader is referred to the National Research Council (2006) report, which is available from the National Academy of Sciences (www.nap.edu/catalog/11019/learning-to-think-spatially).

29. The dynamic versus static and large- versus small-scale factor or facet distinctions are the two most prominent developments in the Gv domain. That does not mean that there are not other possible conceptual frameworks for organizing Gv abilities. For example, Allen (2003) presents a functional model that divides spatial tasks into "What is it?" and "Where am I?" questions, underneath which follow dynamic versus spatial and environment-based versus movement-based distinctions, respectively. The next level categorizes primarily narrow cognitive abilities.

30. Note that the designation of the imagery (IM) ability as "major" is an exception to our use of Carroll's (1993) average factor loadings and is our designation for the reasons explained in the text.

31. We debated whether to change the name of this narrow ability to closure ability or closure fluency.

Carroll (1993) himself was originally hesitant in calling this a speed factor: "It is with some hesitation that I classify this as a speed factor because some people are seemingly unable to perform very difficult items at all, even when given a very generous time-limit. . . . The factor could perhaps just as well be classified as a level factor" (p. 465). After reviewing some of the classic early sources on factor naming (Ekstrom et al., 1979) in the hopeful search for a different historical term, we decided not to disrupt this small corner of the CHC taxonomy, given the long-standing tradition of using the term closure speed and the general acceptance and understanding of the term in contemporary scholarship.

32. In fact, Carroll's (1993) full description and interpretation of the P factor occurs in his Gv chapter (Chapter 8; approximately 8 pages) and not his Gs chapter (one sentence where it is listed under "Factors in the Domain of Visual Perception [See Chapter 8]" (pp. 464–465). His intended dual Gv and Gs classification is made clear in his grand synthesis chapter (Chapter 15), where P is listed and discussed under both Gv and Gs.

33. Astute (or obsessive–compulsive) CHC literature fans may have already noticed the striking similarity in the fonts and graphic format of the model figures presented in McGrew's (1997) original synthesis of the Carroll and Cattell–Horn models (Figure 9.1, p. 155) and, in the same book, Carroll's three-stratum figure (Figure 7.1, p. 125). They are similar, as McGrew created both figures; the creation of Carroll's figure was a professional favor, to produce a higher-quality version of Carroll's somewhat crude figure in his 1993 publication (Figure 15.1, p. 626). Although created concurrently by McGrew, Carroll's figure includes perceptual speed (P) under both Gv and Gs, but McGrew's figure does not. Why? Because at the time McGrew was influenced by available factor-analytic results and common interpretation schemes. At the time, commonly used factor salience rules of thumb (e.g., considering factor loadings < .25 or .30 nonsignificant and omitting the actual values from the factor summary tables; Kline, 2005, p. 245) were being used by the primary CHC CFA investigators of the CHC-based WJ-R and WJ III, as well as the foundational CHC XBA studies (Flanagan, McGrew, & Ortiz, 2000; McGrew, 1997; McGrew & Flanagan, 1998; McGrew, Werder, & Woodcock; 1991; McGrew & Woodcock, 2001; Woodcock, 1990). In hindsight, we now believe that this practice obscured potentially important insights regarding small, yet meaningful, sources of secondary CHC ability variance present in certain tests in intelligence batteries.

For example, WJ-R and WJ III coauthors Woodcock and McGrew frequently noted (to each other) that the Cross Out test often showed small yet consistent loadings on Gv, which typically were not reported due to the rules of thumb about factor salience. The Cross Out test is not present in the WJ IV, but the Pair Cancellation test is also reported to display consistent significant loadings on Gv (median loading across age groups = .22). We (the current chapter's authors) completed a series of exploratory factor analyses of the WJ-R and WJ III norm data; we confirmed that the Cross Out test consistently displayed a lower loading on Gs (.60 to .65) than Visual Matching (mid-.80s) and displayed consistent Gv loadings (.15 to .30 range), while Visual Matching did not. Task analysis of these three mentioned tests reveals that the WJ-R/WJ III Cross Out and WJ IV Pair Cancellation tests require the processing of much more complex visual symbols than Visual Matching. Thus the Gv variance makes sense. But is this amount of variance important? Yes—and we now believe that the commonly used factor salience rules of thumb may have resulted in less-than-accurate CHC classifications of some tests in the CHC test interpretation literature.

The problem has rested in incorrect assumptions regarding intuitive judgments about smaller factor loadings. For example, let us assume that we have a standardized CFA model with latent Gs and Gv factors that correlate .63 (*g* loadings of .70 [Gs] and .90 [Gv]). We have a speed test that loads .80 on Gs, thus accounting for 64% of the test score variance (36% unexplained). If we now add a secondary Gv cross-loading of .32, the typical intuitive conclusion is that this Gv loading now explains approximately 9% of the test performance ($.32^2 = .09$). But this intuition and its underlying math are wrong. Since the latent factor correlation between Gv and Gs is .63, the model incorrectly explains more than 100% of the test score variance—which is impossible. With a .63 latent correlation, the maximum possible cross-loading for Gv is slightly less than .28. A Gv cross-loading of .20 explains not 4% additional test score variance, but 24% additional variance, leaving just 12% unexplained. A Gv cross-loading of .10 explains not 1% additional variance, but 11% additional variance in the test, with 25% unexplained (specific test variance and error variance). Thus many potentially important sources of meaningful CHC ability variance may have been missed in the extant CHC CFA-driven test interpretation and classification literature. Cattell must have recognized this issue, as he frequently considered loadings as small as .15 as salient. The resulting less-than-accurate CHC classification occurred for Cross Out. In all CHC XBA publications, when the Cross Out test was still part of the WJ-R and WJ III, it was always classified as a "strong" indicator of Gs (e.g., see Flanagan et al., 2000, p. 352)—much like its perceptual speed sister Visual Matching. We now believe that it should have not been considered a strong Gs indicator, but a moderate indicator of Gs, and more accurately a mixed Gs/Gv indicator. We have no idea of the extent to which the number of existing CHC test classifications in the literature may not be more accurate; we believe that it is not an extensive issue. However, we now believe that from this point on, CHC-organized factor studies should report the complete set of factor loadings. Retroactively reviewing the extant CHC cross-battery research data and classifications would be nearly impossible, as it would require access to the original raw output from all analyses.

34. See information regarding the recently founded Imagination Institute (http://imagination-institute.org), which is focused on the application of contemporary psychological and neurocognitive theories and methods to the study of human imagination.

35. For our younger readers, Rodney Dangerfield was a comic who, from the 1950s to the early 2000s, became known for his catchphrase "I don't get no respect."

36. Technically, these are component loadings, as the first principal component was extracted.

37. The limited factor summary data provided by Carroll (1993) made it impossible to use his average g and Ga loadings to help designate Ga narrow abilities as major or minor. Thus we have used our expert judgment and knowledge of research and theory to designate PC as major. The designation of U8 as major is based on the research we have reviewed and reported in this chapter, and is at this point a future predicted indication of the status of this narrow ability.

38. There are so many phon- words used in research in reading, psycholinguistics, psychology, cognitive neuroscience, and language that we have a clear phon- jingle–jangle jungle fallacy. Scarborough and Brady (2002) made a valiant attempt to corral and clarify many of the related terms.

39. The addition of speech sounds to UM was suggested by the authors of the WJ IV (McGrew et al., 2014), based on the analysis of the WJ IV norm data; the WJ IV includes numerous measures of different aspects of Ga. Thus the UM classification as auditory-nonverbal may be inaccurate when the memory is for patterns of sounds.

40. As noted in our 2012 chapter, Lazar Stankov has long maintained (at least from 1989) that he and John Horn (Stankov & Horn, 1980) should receive credit for the first known publication to recognize what is now known as working memory.

41. See *www.brainvolts.northwestern.edu.*

42. See *http://sensonics.com/smell-identification-test-international-versions-available.html.*

43. The UPSIT was factor-analyzed together with nine other tests of odor identification, discrimination, detection, memory, suprathershold intensity, and pleasantness perception in a small sample of 97 subjects (Doty, Smith, McKeown, & Raj, 1994). The researchers concluded that the 13 available olfactory variables suggested, "for all practical purposes, measure a common source of variance, perhaps analogous to the 'G' factor observed in intelligence measurement theory" (p. 704). Although the sample was small and there were no other broad CHC factor indicators, this study does provide some support for the broad Go domain.

44. See *https://blogs.scientificamerican.com/guest-blog/a-brief-guide-to-embodied-cognition-why-you-are-not-your-brain.*

45. Only 9+ pages in Carroll's 819-page tome were devoted to discussing the Gp bycatch findings.

46. Given the lack of sufficient information in Carroll (1993), it was not possible to order these narrow abilities in this domain according to the major–minor distinction.

47. See note 18 for a similar rationale to the one we have followed here in proposing two-letter codes for the Gei narrow abilities. EU could not be used, as that is a code already present in Grw. Thus the capital/lowercase abbreviation system is used here also.

REFERENCES

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review, 122*(3), 558–569.

Abraham, A. (2016). The imaginative mind. *Human Brain Mapping, 37*(11), 4197–4211.

Abu-Hamour, B., & Al-Hmouz, H. (2014). Special education in Jordan. *European Journal of Special Needs Education, 29*(1), 105–115.

Abu-Hamour, B., & Al-Hmouz, H. (2017). Prevalence and pattern of learning difficulties in primary school students in Jordan. *Australian Journal of Learning Difficulties, 21*(2), 99–113.

Abu-Hamour, B., Al Hmouz, H., Mattar, J., & Muhaidat, M. (2012). The use of Woodcock–Johnson tests for identifying students with special needs: A comprehensive literature review. *Procedia—Social and Behavioral Sciences, 47,* 665–673.

Ackerman, P. L. (1987). Individual differences in skill learning. *Psychological Bulletin, 102*(1), 3–27.

Ackerman, P. L. (1996a). Intelligence as process and knowledge: An integration for adult development and application. In W. Rogers & A. Fisk (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 139–156). Hillsdale, NJ: Erlbaum.

Ackerman, P. L. (1996b). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*(2), 227–257.

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 55*(2), 69–84.

Ackerman, P. L. (2014). Facts are stubborn things. *Intelligence, 45,* 104–106.

Ackerman, P. L., Beier, M. E., & Boyle, M. D. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131*(4), 567–589.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005).

Working memory and intelligence. *Psychological Bulletin, 131*(1), 30–60.

Ackerman, P. L., Bowen, K. R., Beier, M., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93*(4), 797–825.

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 6*(4), 259–290.

Ackerman, P. L., & Ellingsen, V. J. (2016). Speed and accuracy indicators of test performance under different instructional conditions: Intelligence correlates. *Intelligence, 56,* 1–9.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*(2), 219–245.

Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging, 14*(2), 314–330.

Affrunti, C. L., Schneider, W. J., Tobin, R. M., & Collins, K. D. (2014). *Predictors of academic outcomes in college students suspected of having learning disorders.* Poster presented at the annual convention of the American Psychological Association, Washington, DC.

Alexander, P. A., Dumas, D., Grossnickle, E. M., List, A., & Firetto, C. M. (2016). Measuring relational reasoning. *Journal of Experimental Education, 84*(1), 119–151.

Allen, G. L. (2003). Functional families of spatial abilities: Poor relations and rich prospects. *International Journal of Testing, 3*(3), 251–262.

Allen, V. D., Weissman, A., Hellwig, S., MacCann, C., & Roberts, R. D. (2014). Development of the Situational Test of Emotional Understanding—Brief (STEU-B) using item response theory. *Personality and Individual Differences, 65,* 3–7.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698–1716.

Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science, 22*(6), 437–442.

American Association on Intellectual and Developmental Disabilities (AAIDD). (2010). *Intellectual disability: Definition, classification and systems of supports.* Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Anders Ericsson, K., & Towne, T. J. (2010). Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 404–416.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*(4), 470–487.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance. *Journal of Educational Psychology, 107*(3), 868–883.

Arendasy, M., Hornke, L., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., et al. (2009). *Manual Intelligence Structure Battery (INSBAT).* Mödling, Austria: Schuhfried.

Austin, E. J. (2010). Measurement of ability emotional intelligence: Results for two new tests. *British Journal of Psychology, 101*(3), 563–578.

Avitia, M. J., & Kaufman, J. C. (2014). Beyond *g* and *c*: The relationship of rated creativity to long-term storage and retrieval (Glr). *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 293–302.

Ayres, I. (2007). *Super crunchers: Why thinking-by-numbers is the new way to be smart.* New York: Bantam Books.

Baddeley, A. D. (2002). Fractionating the central executive. In D. Stuss & R. Knight (Eds.), *Principles of frontal lobe function* (pp. 246–260). New York: Oxford University Press.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47–89.

Baghaei, P., & Tabatabaee, M. (2015). The C-Test: An integrative measure of crystallized intelligence. *Journal of Intelligence, 3*(2), 46–58.

Ballesteros, S., Bardisa, D., Millar, S., & Reales, J. M. (2005). The haptic test battery: A new instrument to test tactual abilities in blind and visually impaired and sighted children. *British Journal of Visual Impairment, 23*(1), 11–24.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160–166.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology, 41*(4), 579–597.

Beal, A. L. (2011). *Insight test of cognitive abilities.* Markham, ON, Canada: Canadian Test Centre.

Beaty, R. E., Christensen, A. P., Benedek, M., Silvia, P. J., & Schacter, D. L. (2017). Creative constraints: Brain activity and network dynamics underlying semantic interference during idea production. *NeuroImage, 148*(1), 189–196.

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory and Cognition, 41*(2), 255–267.

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory and Cognition, 42*(7), 1186–1197.

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences, 30*(6), 977–994.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge. *Psychology and Aging, 20*(2), 341–355.

Beier, M. E., Campbell, M., & Crook, A. E. (2010). Developing and demonstrating knowledge: Ability and non-ability determinants of learning and performance. *Intelligence, 38*(1), 179–186.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46,* 73–83.

Benson, N. F., Kranzler, J. H., & Floyd, R. G. (2016). Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models. *Journal of School Psychology, 58,* 1–19.

Bertling, J. P. (2012). *Measuring reasoning ability: Applications of rule-based item generation.* Unpublished doctoral dissertation, Westfälischen Wilhelms-Universität zu Münster, Münster, Germany.

Betts, G. H. (1909). *The distribution and functions of mental imagery.* New York: Teachers College, Columbia University.

Bieber, E., Smits-Engelsman, B. C., Sgandurra, G., Cioni, G., Feys, H., Guzzetta, A., et al. (2016). Manual function outcome measures in children with developmental coordination disorder (DCD): Systematic review. *Research in Developmental Disabilities, 55,* 114–131.

Billings, C. E., Downey, L. A., Lomas, J. E., Lloyd, J., & Stough, C. (2014). Emotional Intelligence and scholastic achievement in pre-adolescent children. *Personality and Individual Differences, 65,* 14–18.

Billings, L. (2012). Astrobiology in culture: The search for extraterrestrial life as "science. *Astrobiology, 12*(10), 966–975.

Birney, D. P., Halford, G. S., & Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square Task. *Educational and Psychological Measurement,* 66(1), 146–171.

Bishop, D. V., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31*(7), 1027–1050.

Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. R. (2009). Construct and criterion-related validation of a measure of emotional reasoning skills: A two-study investigation. *International Journal of Selection and Assessment, 17*(1), 101–118.

Borghi, A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia, 48*(3), 763–773.

Bossomaier, T., Harré, M., Knittel, A., & Snyder, A. (2009). A semantic network approach to the creativity quotient (CQ). *Creativity Research Journal, 21*(1), 64–71.

Bowers, P. G., Sunseth, K., & Golden, J. (1999). The route between rapid naming and reading progress. *Scientific Studies of Reading, 3*(1), 31–53.

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin, 29*(9), 1147–1158.

Brackett, M. A., Rivers, S. E., & Salovey, P. (2011). Emotional intelligence: Implications for personal, social, academic, and workplace success. *Social and Personality Psychology Compass, 5*(1), 88–103.

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*(4), 780–795.

Broach, D. (2013). *Selection of the next generation of air traffic control specialists: Aptitude requirements for the air traffic control tower cab in 2018.* Oklahoma City, OK: Federal Aviation Administration.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 16–33). New York: Cambridge University Press.

Buckley, J., Seery, N., & Canty, D. (2017). Spatial cognition in engineering education: Developing a spatial ability framework to support the translation of theory into practice. *European Journal of Engineering Education.* [Epub ahead of print]

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228.

Burns, N. R., & Nettelbeck, T. (2003). Inspection time in the structure of cognitive abilities: Where does IT fit? *Intelligence, 31*(3), 237–255.

Burton, L. J., & Fogarty, G. J. (2003). The factor structure of visual imagery and spatial abilities. *Intelligence, 31*(3), 289–318.

Byrom, N. C., & Murphy, R. A. (2016). Individual difference in configural associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 42*(4), 325–335.

Cameron, S., Glyde, H., Dillon, H., & Whitfield, J. (2014). Development and preliminary evaluation of a processing speed test for school-aged children utilizing auditory stimuli. *International Journal of School and Cognitive Psychology, 1*(3), 1–3.

Campbell, S., & Catano, V. M. (2004). Using measures of specific abilities to predict training performance in Canadian forces operator occupations. *Military Psychology, 16*(3), 183–201.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 84–112). New York: Oxford University Press.

Carr, K. W., Fitzroy, A. B., Tierney, A., White-Schwoch, T., & Kraus, N. (2017). Incorporation of feedback during beat synchronization is an index of neural maturation and reading skills. *Brain and Language, 164,* 43–52.

Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences USA, 111*(40), 14559–14564.

Carretti, B., Motta, E., & Re, A. M. (2016). Oral and written expression in children with reading comprehension difficulties. *Journal of Learning Disabilities, 49*(1), 65–76.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–24). Mahwah, NJ: Erlbaum.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). Amsterdam: Pergamon Press.

Carroll, J. B., Kohlberg, L., & DeVries, R. (1984). Psychometric and Piagetian intelligences: Toward resolution of controversy. *Intelligence, 8*(1), 67–91.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153–193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cattell, R. B. (1980). The heritability of fluid, Gf, and crystallised, Gc, intelligence, estimated by a least squares use of the MAVA method. *British Journal of Educational Psychology, 50*(3), 253–265.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action.* New York: Elsevier.

Cattell, R. B. (1998). Where is intelligence?: Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 29–38). Mahwah, NJ: Erlbaum.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15*(3), 139–164.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*(2), 278–293.

Cermak, S. A., & Larkin, D. (2002). *Developmental coordination disorder.* Albany, NY: Cengage Learning.

Chaiken, S. R., Kyllonen, P. C., & Tirre, W. C. (2000). Organization and components of psychomotor ability. *Cognitive Psychology, 40*(3), 198–226.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Chronaki, G., Garner, M., Hadwin, J. A., Thompson, M. J., Chin, C. Y., & Sonuga-Barke, E. J. (2015). Emotion-recognition abilities and behavior problem dimensions in preschoolers: Evidence for a specific role for childhood hyperactivity. *Child Neuropsychology, 21*(1), 25–40.

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33*(2), 218–236.

Cohen, A., Fiorello, C. A., & Farley, F. H. (2006). The cylindrical structure of the Wechsler Intelligence Scale for Children—IV: A retest of the Guttman model of intelligence. *Intelligence, 34*(6), 587–591.

Contreras, M. J., Colom, R., Hernandez, J. M., & Santacreu, J. (2003). Is static spatial performance distinguishable from dynamic spatial performance?: A latent-variable analysis. *Journal of General Psychology, 130*(3), 277–288.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science, 18*(5), 275–279.

Conzelmann, K., & Süß, H.-M. (2015). Auditory intelligence: Theoretical considerations and empirical findings. *Learning and Individual Differences, 40,* 27–40.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31*(6), 567–587.

Coyle, T. R., Purcell, J. M., Snyder, A. C., & Richmond, M. C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence, 46,* 18–24.

Cronbach, L. J. (1970). *Essentials of psychological assessment* (3rd ed.). New York: Harper & Row.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Cucina, J. M., Su, C., Busciglio, H. H., & Peyton, S. T. (2015). Something more than *g*: Meaningful memory uniquely predicts training performance. *Intelligence, 49,* 192–206.

Cudeck, R. E., & MacCallum, R. C. (2007). *Factor analysis at 100: Historical developments and future directions.* Mahwah, NJ: Erlbaum.

Danthiir, V., Roberts, R. D., Pallier, G., & Stankov, L. (2001). What the nose knows: Olfaction and cognitive abilities. *Intelligence, 29*(4), 337–361.

Danthiir, V., Roberts, R. D., Schulze, R., & Wilhelm, O. (2005). Mental speed: On frameworks, paradigms, and a platform for the future. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 27–46). Thousand Oaks, CA: SAGE.

Danthiir, V., Wilhelm, O., & Roberts, R. D. (2012). Further evidence for a multifaceted model of mental speed: Factor structure and validity of computerized measures. *Learning and Individual Differences, 22*(3), 324–335.

Danthiir, V., Wilhelm, O., & Schacht, A. (2005). Decision speed in intelligence tasks: Correctly an ability? *Psychology Science, 47*(2), 200–229.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence.* Boston: Allyn & Bacon.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*(4), 989–1015.

Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*(5), 1022–1033.

De Alwis, D., Hale, S., & Myerson, J. (2014). Extended cascade models of age and individual differences in children's fluid intelligence. *Intelligence, 46,* 84–93.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51*(6), 599–608.

Debes, F., Weihe, P., & Grandjean, P. (2016). Cognitive deficits at age 22 years associated with prenatal exposure to methylmercury. *Cortex, 74,* 358–369.

Decker, S. L. (2010). Tactile measures in the structure of intelligence. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 64*(1), 53–59.

Decker, S. L., Roberts, A. M., & Englund, J. A. (2013). Cognitive predictors of rapid picture naming. *Learning and Individual Differences, 25,* 141–149.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Del Missier, F., Visentini, M., & Mäntylä, T. (2015). Option generation in decision making: Ideation beyond memory retrieval. *Frontiers in Psychology, 5,* 1584.

Denham, S. A. (2007). Dealing with feelings: How children negotiate the worlds of emotions and social relationships. *Cognition, Brain, Behavior, 11*(1), 1–48.

Detterman, D. K. (2014, July/August). Introduction to the intelligence special issue on the development of expertise: Is ability necessary? *Intelligence, 45,* 1–5.

Dewey, J. (1909). *Moral principles in education.* Boston: Houghton Mifflin.

Doty, R. L. (2001). Olfaction. *Annual Review of Psychology, 52*(1), 423–452.

Doty, R. L., Shaman, P., & Dann, M. (1984). Development of the University of Pennsylvania Smell Identification Test: A standardized microencapsulated test of olfactory function. *Physiology and Behavior, 32*(3), 489–502.

Doty, R. L., Smith, R., McKeown, D. A., & Raj, J. (1994). Tests of human olfactory function: Principal components analysis suggests that most measure a common source of variance. *Perception and Psychophysics, 56*(6), 701–707.

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science, 11*(1), 133–155.

Dulay, M. F., Gesteland, R. C., Shear, P. K., Ritchey, P. N., & Frank, R. A. (2008). Assessment of the influence of cognition and cognitive processing speed on three tests of olfaction. *Journal of Clinical and Experimental Neuropsychology, 30*(3), 327–337.

Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review, 25*(3), 391–427.

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences USA, 114*(20), 5295–5299.

Eisenreich, B. R., Akaishi, R., & Hayden, B. Y. (2017). Control without controllers: Toward a distributed neuroscience of executive control. *Journal of Cognitive Neuroscience, 29*(10), 1684–1698.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*(2), 3–84.

Eliot, J., & Czarnolewski, M. Y. (2007). Development of an everyday spatial behavioral questionnaire. *Journal of General Psychology, 134*(3), 361–381.

Eliot, J., & Smith, I. M. (1983). *An international directory of spatial tests.* Windsor, UK: NFER-Nelson.

Engen, T. (1982). *The perception of odors.* New York: Academic Press.

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise.* New York: Houghton Mifflin Harcourt.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245.

Ernest, C. H. (1977). Imagery ability and cognition: A critical review. *Journal of Mental Imagery, 1*(2), 181–215.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*(2), 246–262.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Feldmann, G. M., Kelly, R. M., & Diehl, V. A. (2004). An interpretative analysis of five commonly used processing speed measures. *Journal of Psychoeducational Assessment, 22*(2), 151–163.

Fine, S. E., Izard, C. E., Mostow, A. J., Trentacosta, C. J., & Ackerman, B. P. (2003). First grade emotion knowledge as a predictor of fifth grade self-reported internalizing behaviors in children from economically disadvantaged families. *Development and Psychopathology, 15*(2), 331–342.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741.

Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science, 5*(40), 1251–1269.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment*. Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Dixon, S. G. (2014). Cross-battery approach to the assessment of executive functions. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 379–409). New York: Springer.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Fleishman, E. A. (1964). *The structure and measurement of physical fitness*. Oxford, UK: Prentice-Hall.

Fleishman, E. A. (1972). Structure and measurement of psychomotor abilities. In R. N. Singer (Ed.), *The psychomotor domain: Movement behaviors* (pp. 78–106). Philadelphia: Lea & Febiger.

Fleishman, E. A., Quaintance, M. K., & Broedling, L. A. (1984). *Taxonomies of human performance: The description of human tasks*. San Diego, CA: Academic Press.

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell–Horn–Carroll model?: Some evidence from a joint factor analysis of the Delis–Kaplan Executive Function System and the Woodcock–Johnson III Tests of Cognitive Abilities. *Psychology in the Schools, 47*(7), 721–738.

Ford, L., Swart, S., Negreiros, J., Lacroix, S., & McGrew, K. (2010). *Use of the Woodcock–Johnson III NU Tests of Cognitive Abilities and Tests of Achievement with Canadian populations* (Woodcock–Johnson III Assessment Service Bulletin No. 12). Rolling Meadows, IL: Riverside.

Forgues, S. (2014, October). *Aptitude testing of military pilot candidates*. Unpublished master's thesis, Queen's University, Kingston, ON, Canada.

French, J. W. (1951). *The description of aptitude and achievement tests in terms of rotated factors*. Chicago: University of Chicago Press.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10,* 29–44.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*(4), 237–241.

Fuson, K. C., & Li, Y. (2009). Cross-cultural issues in linguistic, visual-quantitative, and written-numeric supports for mathematical thinking. *ZDM, 41*(6), 793–808.

Ganis, G., & Schendan, H. E. (2011). Visual imagery. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(3), 239–252.

García, J. M. de H., & Costa, J. L. C. (2014). Does trait emotional intelligence predict unique variance in early career success beyond IQ and personality? *Journal of Career Assessment, 22*(4), 715–725.

Gardner, H. (1983). *Frames of mind*. New York: Basic Books.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4–15.

Geary, D. C. (2007). An evolutionary perspective on learning disability in mathematics. *Developmental Neuropsychology, 32*(1), 471–519.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23–27.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*(1), 206–223.

Gibson, J. J. (1979). *The ecological approach to visual perception*. New York: Psychology Press.

Gignac, G. E., Shankaralingam, M., Walker, K., & Kilpatrick, P. (2016). Short-term memory for faces relates to general intelligence moderately. *Intelligence, 57,* 96–104.

Gilhooly, K., Wynn, V., Phillips, L., Logie, R., & Sala, S. D. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking and Reasoning, 8*(3), 165–178.

Gillberg, C., & Kadesjö, B. (2003). Why bother about clumsiness?: The implications of having developmental coordination disorder (DCD). *Neural Plasticity, 10*(1–2), 59–68.

Giménez-Dasí, M., Fernández-Sánchez, M., & Quintanilla, L. (2015). Improving social competence through emotion knowledge in 2-year-old children: A pilot study. *Early Education and Development, 26*(8), 1128–1144.

Gluck, J., & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing, 3*(3), 293–308.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40*(2), 103–114.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology, 19*(2), 183–204.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236–243.

Goff, M., & Ackerman, P. L. (1992). Personality–intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology, 84*(4), 537–552.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25*(2), 496–508.

Gooch, D., Snowling, M. J., & Hulme, C. (2012). Reaction time variability in children with ADHD symptoms and/or dyslexia. *Developmental Neuropsychology, 37*(5), 453–472.

Goode, N., & Beckmann, J. F. (2010). You need to know: There is a causal relationship between structural knowledge and control performance in complex problem solving tasks. *Intelligence, 38*(3), 345–352.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132.

Grazzani, I., Ornaghi, V., Agliati, A., & Brazzelli, E. (2016). How to foster toddlers' mental-state talk, emotion understanding, and prosocial behavior: A conversation-based intervention at nursery school. *Infancy, 21*(2), 199–227.

Gregory, T., Nettelbeck, T., & Wilson, C. (2009). Within-person changes in inspection time predict memory. *Personality and Individual Differences, 46*(7), 741–743.

Guilford, J. P. (1958). A system of the psychomotor abilities. *American Journal of Psychology, 71*(1), 164–174.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Guilford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence*. New York: McGraw-Hill.

Gustafsson, J.-E., & Wolff, U. (2015). Measuring fluid intelligence at age four. *Intelligence, 50,* 175–185.

Guttman, L. (1954). A new approach to factor analysis: The Radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). New York: Free Press.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15*(1), 79–103.

Habets, P., Jeandarme, I., Uzieblo, K., Oei, K., & Bogaerts, S. (2015). Intelligence is in the eye of the beholder: Investigating repeated IQ measurements in forensic psychiatry. *Journal of Applied Research in Intellectual Disabilities, 28*(3), 182–192.

Hakstian, A. R., & Cattell, R. B. (1974). The checking of primary ability structure on a broader basis of performances. *British Journal of Educational Psychology, 44*(2), 140–154.

Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology, 70*(5), 657–669.

Haldemann, J., Stauffer, C., Troche, S., & Rammsayer, T. (2012). Performance on auditory and visual temporal information processing is related to psychometric intelligence. *Personality and Individual Differences, 52*(1), 9–14.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*(6), 236–242.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(6), 803–831.

Hall, D., Jarrold, C., Towse, J. N., & Zarandi, A. L. (2015). The developmental influence of primary memory capacity on working memory and academic achievement. *Developmental Psychology, 51*(8), 1131–1147.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*(4), 339–387.

Hambrick, D. Z., Meinz, E. J., & Oswald, F. L. (2007). Individual differences in current events knowledge: Contributions of ability, personality, and interests. *Memory and Cognition, 35*(2), 304–316.

Hambrick, D. Z., Pink, J. E., Meinz, E. J., Pettibone, J. C., & Oswald, F. L. (2008). The roles of ability, personality, and interests in acquiring current events knowledge: A longitudinal study. *Intelligence, 36*(3), 261–278.

Harle, M., & Towns, M. (2010). A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education, 88*(3), 351–360.

Hartman, R., Sprenger, A., Weeks, J., Kyllonen, P. C., Bertling, M., McGrew, K., et al. (2017). *General fluid/inductive reasoning battery for a high-ability population.* Manuscript submitted for publication.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*(6), 304–313.

Haydel, J., & Ferber, Y. V. (2015). Atkins v. Virginia (2002). In D. Schultz & J. R. Vile (Eds.), *Encyclopedia of civil liberties in America* (Vol. 1, p. 50). New York: Taylor & Francis.

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48,* 1–14.

Hedner, M., Larsson, M., Arnold, N., Zucco, G. M., & Hummel, T. (2010). Cognitive factors in odor detection, odor discrimination, and odor identification tasks. *Journal of Clinical and Experimental Neuropsychology, 32*(10), 1062–1067.

Hegarty, M. (2010). Components of spatial intelligence. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 265–297). San Diego, CA: Academic Press.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*(2), 175–191.

Helmbold, N., Troche, S., & Rammsayer, T. (2006). Temporal information processing and pitch discrimination as predictors of general intelligence. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 60*(4), 294–306.

Helmbold, N., Troche, S., & Rammsayer, T. (2007). Processing of temporal and nontemporal information as predictors of psychometric intelligence: A structural-equation-modeling approach. *Journal of Personality, 75*(5), 985–1006.

Henderson, S. E., & Hall, D. (1982). Concomitants of clumsiness in young schoolchildren. *Developmental Medicine and Child Neurology, 24*(5), 448–460.

Herniman, D. (2013). *Investigating predictors of primary flight training in the Canadian forces.* Unpublished master's thesis, Carleton University, Ottawa, ON, Canada.

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science, 7*(3), 513–534.

Hoelzle, J. B. (2008). Neuropsychological assessment and the Cattell–Horn–Carroll (CHC) cognitive abilities model. *Dissertation Abstracts International: Section B: The Sciences and Engineering,* 69(9B). (UMI No. AAI 3328214)

Hoffman, B. (2012). Cognitive efficiency: A conceptual and methodological comparison. *Learning and Instruction, 22*(2), 133–144.

Hoffman, B., & Schraw, G. (2009). The influence of self-efficacy and working memory capacity on problem-solving efficiency. *Learning and Individual Differences, 19*(1), 91–100.

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45*(1), 1–14.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments.* Odessa, FL: Psychological Assessment Resources.

Horn, J. L. (1980). Concepts of intellect in relation to learning and adult development. *Intelligence, 4*(4), 285–317.

Horn, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In R. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 237–278). New York: Plenum Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 41–68). New York: Guilford Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107–129.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 767–782). Mahwah, NJ: Routledge.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hummel, T., Kobal, G., Gudziol, H., & Mackay-Sim, A. (2007). Normative data for the "Sniffin'Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: An upgrade based on a group of more than 3,000 subjects. *European Archives of Oto-Rhino-Laryngology, 264*(3), 237–243.

Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist, 17*(7), 475.

Hunt, E., Pellegrino, J. W., Frick, R. W., Farr, S. A., & Alderton, D. (1988). The ability to reason about movement in the visual field. *Intelligence, 12*(1), 77–100.

Hunt, T. (1928). The measurement of social intelligence. *Journal of Applied Psychology, 12*(3), 317–334.

Hurks, P. P. M., & Bakker, H. (2016). Assessing intelligence in children and youth living in the Netherlands. *International Journal of School and Educational Psychology, 4*(4), 266–275.

Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2012). Emotional intelligence in personnel selection: Applicant reactions, criterion, and incremental validity. *Inter-*

national Journal of Selection and Assessment, 20(3), 347–358.

Inman, W. C., & Secrest, B. T. (1981). Piaget's data and spearman's theory—An empirical reconciliation and its implications for academic achievement. *Intelligence, 5*(4), 329–344.

Izard, C. E. (1991). *The psychology of emotions.* New York: Plenum Press.

Izard, C. E. (2001). Emotional intelligence or adaptive emotions? *Emotion, 3*(1), 249–257.

Izard, C. E., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science, 12*(1), 18–23.

James, L., Jacobs, K. E., & Roodenburg, J. (2015). Adoption of the Cattell–Horn–Carroll model of cognitive abilities by Australian psychologists. *Australian Psychologist, 50*(3), 194–202.

Jelínek, M., Kveton, P., & Voboril, D. (2015). Innovative testing of spatial ability: Interactive responding and the use of complex stimuli material. *Cognitive Processing, 16*(1), 45–55.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier Science.

Jewsbury, P. A., & Bowden, S. C. (2016). Construct validity of fluency and implications for the factorial structure of memory. *Journal of Psychoeducational Assessment, 35*(5), 460–481.

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2017). The Cattell–Horn–Carroll model of cognition for clinical assessment. *Journal of Psychoeducational Assessment, 35*(6), 547–567.

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell–Horn–Carroll model. *Journal of Experimental Psychology: General, 145*(2), 220–245.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*(4), 393–416.

Jönsson, F. U., Møller, P., & Olsson, M. J. (2011). Olfactory working memory: Effects of verbalization on the 2-back task. *Memory and Cognition, 39*(6), 1023–1032.

Jordan, N. C., & Glutting, J. J. (2012). *The Number Sense Screener.* Baltimore: Brookes.

Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153–175.

Joseph, A., & DeLuca, G. C. (2016). Back on the scent: The olfactory system in CNS demyelinating diseases. *Journal of Neurology, Neurosurgery and Psychiatry, 87*(10), 1146–1154.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology, 95*(1), 54–78.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review,* 99(1), 122–149.

Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science, 18*(4), 312–313.

Kail, R. V., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica,* 86(2), 199–225.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217.

Kang, K., Anthoney, L., & Mitchell, P. (2017). Seven- to 11-year-olds' developing ability to recognize natural facial expressions of basic emotions. *Perception, 46*(9), 1077–1089.

Kappers, A. M., & Bergmann Tiest, W. M. (2013). Haptic perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(4), 357–374.

Kaufman, A. S. (2009). *IQ testing 101.* New York: Springer.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition (KABC-II).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V.* Hoboken, NJ: Wiley.

Kaufman, J. C. (2016). *Creativity 101.* New York: Springer.

Kaufman, S. B., & Duckworth, A. L. (2017). World-class expertise: A developmental model. *Wiley Interdisciplinary Reviews: Cognitive Science,* 8(1–2), 1–7.

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure?: Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability. *School Psychology Review, 30*(1), 89–119.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650.

King, D. D., Ryan, A. M., Kantrowitz, T., Grelle, D., & Dainis, A. (2015). Mobile Internet testing: An analysis of equivalence, individual differences, and reactions. *International Journal of Selection and Assessment, 23*(4), 382–394.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*(3), 341–362.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: SAGE.

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., et al. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review, 33*(6), 795–811.

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology, 109*(7), 977–992.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177.

Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review, 28*(1), 117–144.

Kraus, N., & Slater, J. (2016). Beyond words: How humans communicate through sound. *Annual Review of Psychology, 67,* 83–103.

Krishnakumar, S., Hopkins, K., Szmerekovsky, J. G., & Robinson, M. D. (2016). Assessing workplace emotional intelligence: Development and validation of an ability-based measure. *Journal of Psychology, 150*(3), 371–404.

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37*(4), 347–364.

Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36*(5), 422–436.

Kyllonen, P. C., & Chaiken, S. (2003). Dynamic spatial ability and psychomotor performance. *International Journal of Testing, 3*(3), 233–249.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389–433.

Kyllonen, P. C., & Gluck, J. (2003). Spatial ability: Introduction to the special issue. *International Journal of Testing, 3*(3), 215–217.

Kyllonen, P., & Kell, H. (2017). What is fluid intelligence?: Can it be improved? In M. Rosen, K. Y. Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes* (pp. 15–37). Cham, Switzerland: Springer .

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76*(6), 1325–1345.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323.

Lakin, J. M., & Gambrell, J. L. (2012). Distinguishing verbal, quantitative, and figural facets of fluid intelligence in young students. *Intelligence, 40*(6), 560–570.

Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology, 31*(1), 65–71.

Landy, F. J. (2006). The long, frustrating, and fruitless search for social intelligence: A cautionary tale. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 81–124). Mahwah, NJ: Erlbaum.

Lecerf, T., Rossier, J., Favez, N., Reverte, I., & Coleaux, L. (2010). The four- vs. alternative six-factor structure of the French WISC-IV. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie,* 69(4), 221–232.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, and Psychophysics, 71*(7), 1439–1459.

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*(1), 15–26.

Lee, K., Ng, E. L., & Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of Educational Psychology, 101*(2), 373–387.

Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Stanford University Technical Report No. 8). Stanford, CA: Stanford University.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Hillsdale, NJ: Erlbaum.

Lohman, D. F. (1996). Spatial ability and *g*. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Hillsdale, NJ: Erlbaum.

Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I., & Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin, 30*(8), 1018–1034.

Lopes, P. N., Salovey, P., Côté, S., & Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion, 5*(1), 113–118.

Lotto, A., & Holt, L. (2011). Psychology of auditory perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 479–489.

Loughman, A., Bowden, S., & D'Souza, W. (2014). Cognitive functioning in idiopathic generalised epilepsies: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews, 43,* 20–34.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: Sinking shafts at a few critical points. *Annual Review of Psychology, 51*(1), 405–444.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49*(4), 344–351.

Luciano, M., Wright, M. J., Geffen, G. M., Geffen, L. B., Smith, G. A., & Martin, N. G. (2004). A genetic investigation of the covariation among inspection time, choice reaction time, and IQ subtest scores. *Behavior Genetics, 34*(1), 41–50.

MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion, 14*(2), 358–374.

MacCann, C., Matthews, G., Zeidner, M., & Roberts, R. D. (2003). Psychological assessment of emotional intelligence: A review of self-report and performance-based testing. *International Journal of Organizational Analysis, 11*(3), 247–274.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*(4), 540–551.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris), 102*(1), 59–70.

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence, 61,* 120–129.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*(2), 107–127.

Martzke, J. S., Kopala, L. C., & Good, K. P. (1997). Olfactory dysfunction in neuropsychiatric disorders: Review and methodological considerations. *Biological Psychiatry, 42*(8), 721–732.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335.

Matthews, G., Emo, A. K., Roberts, R. D., & Zeidner, M. (2006). What is this thing called emotional intelligence? In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 3–36). Mahwah, NJ: Erlbaum.

Matthews, G., Zeidner, M., & Roberts, R. D. (2004). *Emotional intelligence: Science and myth.* Cambridge, MA: MIT Press.

Matthews, G., Zeidner, M., & Roberts, R. D. (2005). Emotional intelligence: An elusive ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 79–100). Thousand Oaks, CA: SAGE.

Matzke, D., Dolan, C. V., & Molenaar, D. (2010). The issue of power in the identification of "g" with lower-order factors. *Intelligence, 38*(3), 336–344.

Maul, A. (2012). The validity of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) as a measure of emotional intelligence. *Emotion Review, 4*(4), 394–402.

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*(4), 267–298.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review, 8*(4), 290–300.

Mayer, J. D., Panter, A. T., & Caruso, D. R. (2012). Does personal intelligence exist?: Evidence from a new ability-based measure. *Journal of Personality Assessment, 94*(2), 124–140.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59,* 507–536.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) user's manual.* Toronto: Multi-Health Systems.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist, 63*(6), 503–517.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2014). *Mayer–Salovey–Caruso Emotional Intelligence Test: Youth Version—Research Edition (MSCEIT-YV).* Toronto: Multi-Health Systems.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1*(3), 232–242.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*(1), 97–105.

McAdams, D. P., Josselson, R., & Lieblich, A. (2006). *Identity and story: Creating self in narrative.* Washington, DC: American Psychological Assocation.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*(4), 409–454.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8*(3), 290–302.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Ed.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 136–181). New York: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.

McGrew, K. S. (2010). *Pushing the edge of the contem-*

porary cognitive (CHC) theory: New directions for neuropsychologists. Workshop presented at the 16th Annual APS College of Clinical Neuropsychologists Conference, Notre Dame University, Fremantle, Western Australia.

McGrew, K. S. (2015). Intellectual functioning. In E. A. Polloway (Ed.), *The death penalty and intellectual disability* (pp. 85–111). Washington, DC: American Association on Intellectual and Developmental Disabilities.

McGrew, K. S., & Evans, J. J. (2004). *Carroll Human Cognitive Abilities Project: Research Report No. 2. Internal and external factorial extensions to the Cattell–Horn–Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise.* St. Cloud, MN: Institute for Applied Psychometrics.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Needham Heights, MA: Allyn & Bacon.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock–Johnson IV technical manual.* Rolling Meadows, IL: Riverside.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ-R technical manual.* Chicago: Riverside.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock–Johnson III.* Itasca, IL: Riverside.

McNerney, S. (2011). A brief guide to embodied cognition: Why you are not your brain. *Scientific American Blogs.* Retrieved from *https://blogs.scientificamerican.com/guest-blog/a-brief-guide-to-embodied-cognition-why-you-are-not-your-brain.*

Messer, D., & Dockrell, J. E. (2006). Children's naming and word-finding difficulties: Descriptions and explanations. *Journal of Speech, Language, and Hearing Research, 49*(2), 309–324.

Mestre, J. M., MacCann, C., Guil, R., & Roberts, R. D. (2016). Models of cognitive ability and emotion can better inform contemporary emotional intelligence frameworks. *Emotion Review, 8*(4), 322–330.

Meyer, E. M., & Reynolds, M. R. (2017). Scores in space: Multidimensional scaling of the WISC-V. *Journal of Psychoeducational Assessment.* [Epub ahead of print]

Miller, D. C. (2007). *Essentials of school neuropsychology.* Hoboken, NJ: Wiley.

Miller, D. C. (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Miller, H. E., Vlach, H. A., & Simmering, V. R. (2017). Producing spatial words is not enough: Understanding the relation between language and spatial cognition. *Child Development, 88*(6), 1966–1982.

Minogue, J., & Jones, M. G. (2006). Haptics in education: Exploring an untapped sensory modality. *Review of Educational Research, 76*(3), 317–348.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

Molinari, M., & Leggio, M. (2015). Cerebellum and verbal fluency (phonological and semantic). In P. Mariën & M. Manto (Eds.), *The linguistic cerebellum* (pp. 63–80). San Diego, CA: Elsevier/Academic Press.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2005). *Batería III Woodcock–Muñoz: Suplemento diagnóstico para las pruebas de habilidades cognitivas.* Itasca, IL: Riverside.

Murphy, K. (2017). What can we learn from "Not much more than g"? *Journal of Intelligence, 5*(1), 8.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition.* Austin, TX: PRO-ED.

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031–1039.

National Research Council. (2006). *Learning to think spatially: GIS as a support system in the K–12 curriculum.* Washington, DC: National Academies Press.

National Research Council. (2015). *Measuring human capabilities: An agenda for basic research on the assessment of individual and group performance potential for military accession.* Washington, DC: National Academies Press.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*(3), 226–254.

Nelson, J. M., Lindstrom, J. H., Lindstrom, W., & Denis, D. (2012). The structure of phonological processing and its relationship to basic reading. *Exceptionality, 20*(3), 179–196.

Nelson, M. A. (2009). *Exploring the relationships among various measures of processing speed in a sample of children referred for psychological assessments.* Unpublished doctoral dissertation, University of Virginia.

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences, 48*(4), 379–384.

Newcombe, N. S. (2013). Seeing relationships: Using spatial thinking to teach science, mathematics, and social studies. *American Educator, 37*(1), 26–40.

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). Spatial development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology* (Vol. 1, pp. 564–590). New York: Oxford University Press.

Norman, D. A., & Verganti, R. (2014). Incremental and radical innovation: Design research vs. technology and meaning change. *Design Issues, 30*(1), 78–96.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427–452.

Nuszbaum, M., Voss, A., Klauer, K. C., & Betsch, T. (2010). Assessing individual differences in the use of haptic information using a German translation of the Need for Touch Scale. *Social Psychology, 41*(4), 263–274.

O'Connor, T. A., & Burns, N. R. (2003). Inspection time and general speed of processing. *Personality and Individual Differences, 35*(3), 713–724.

Organisation for Economic Co-operation and Development. (2004). *Innovation in the knowledge economy: Implications for education and learning.* Paris: Author.

Ornaghi, V., Brazzelli, E., Grazzani, I., Agliati, A., & Lucarelli, M. (2017). Does training toddlers in emotion knowledge lead to changes in their prosocial and aggressive behavior toward peers at nursery? *Early Education and Development, 28*(4), 396–414.

O'Sullivan, M., & Guilford, J. (1975). Six factors of behavioral cognition: Understanding other people. *Journal of Educational Measurement, 12*(4), 255–271.

Ottensmeier, H., Zimolong, B., Wolff, J. E., Ehrich, J., Galley, N., van Hoff, K., et al. (2015). Neuropsychological short assessment of disease-and treatment-related intelligence deficits in children with brain tumours. *European Journal of Paediatric Neurology, 19*(3), 298–307.

Outtz, J. L., & Newman, D. A. (2010). A theory of adverse impact. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 53–94). New York: Routledge/Taylor & Francis.

Padgett, L. S. (2015). *Looking at cancer related cognitive impairment through the lens of CHC theory: Does it improve our vision?* Paper presented at the annual meeting of the International Neuropsychological Society, Denver, CO.

Pahud, O. (2017). *The influence of attention on the relationship between temporal resolution power and general intelligence.* Unpublished doctoral dissertation, University of Bern, Bern, Switzerland.

Palmer, B. R., Gignac, G., Manocha, R., & Stough, C. (2005). A psychometric evaluation of the Mayer–Salovey–Caruso Emotional Intelligence Test Version 2.0. *Intelligence, 33*(3), 285–305.

Pase, M. P., & Stough, C. (2014). An evidence-based method for examining and reporting cognitive processes in nutrition research. *Nutrition Research Reviews, 27*(2), 232–241.

Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence, 50,* 30–51.

Pearson, J. (2014). New directions in mental-imagery research: The binocular-rivalry technique and decoding fMRI patterns. *Current Directions in Psychological Science, 23*(3), 178–183.

Peck, J., & Childers, T. L. (2003). To have and to hold: The influence of haptic information on product judgments. *Journal of Marketing, 67*(2), 35–48.

Pellegrino, J. W., Hunt, E. B., Abate, R., & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behavior Research Methods, Instruments, and Computers, 19*(2), 231–236.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. T. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 67–86). Philadelphia: Benjamins.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37.

Peterson, N. G., & Bownas, D. A. (1982). Skill, task structure, and performance acquisition. *Human Performance and Productivity, 1,* 49–105.

Phillips, J. E. (2015). *Analysis of the factor structure of processing speed.* Unpublished doctoral dissertation, Texas Woman's University.

Polloway, E. A. (Ed.). (2015). *The death penalty and intellectual disability.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence, 8*(2), 93–138.

Primi, R. (2003). Inteligência: Avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação Psicológica, 2*(1), 67–77.

Primi, R. (2014). Developing a fluid intelligence scale through a combination of Rasch modeling and cognitive psychology. *Psychological Assessment, 26*(3), 774–788.

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences, 20*(5), 446–451.

Primi, R., Nakano, T. de C., & Wechsler, S. M. (2012). Cross-battery factor analysis of the Battery of Reasoning Abilities (BPR-5) and Woodcock–Johnson Tests of Cognitive Ability (WJ-III): Bateria de Habilidades de Raciocinio (BPR-5) e Bateria de Habilidades Cognitivas Woodcock–Johnson III. *Temas em Psicologia, 20*(1), 121–132.

Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools, 46*(7), 679–691.

Rammsayer, T. H., & Brandler, S. (2004). Aspects of temporal information processing: A dimensional analysis. *Psychological Research, 69*(1–2), 115–123.

Rammsayer, T. H., & Brandler, S. (2007). Performance on temporal information processing as an index of general intelligence. *Intelligence, 35*(2), 123–139.

Rammsayer, T. H., & Troche, S. J. (2016). Validity of the worst performance rule as a function of task complexity and psychometric *g*: On the crucial role of *g* saturation. *Journal of Intelligence, 4*(1), 5.

Re, A. M., & Carretti, B. (2016). Further evidence of poor comprehenders' difficulty with expressive writing: Exploring the role of inferences. *Research in Developmental Disabilities, 51,* 145–152.

Reber, R., & Greifeneder, R. (2017). Processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist, 52*(2), 84–103.

Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than *g*. *Intelligence, 32*(6), 621–652.

Revelle, W. (2015). *psych: Procedures for psychological, psychometric, and personality research.* Evanston, IL: Northwestern University. Retrieved from *http://CRAN.R-project.org/package=psych.*

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51*(4), 535–555.

Rivilis, I., Hay, J., Cairney, J., Klentrou, P., Liu, J., & Faught, B. E. (2011). Physical activity and fitness in children with developmental coordination disorder: A systematic review. *Research in Developmental Disabilities, 32*(3), 894–910.

Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 2, pp. 461–482). Thousand Oaks, CA: SAGE.

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences, 11*(1), 1–120.

Roberts, R. D., Stankov, L., Pallier, G., & Dolph, B. (1997). Charting the cognitive sphere: Tactile-kinesthetic performance within the structure of intelligence. *Intelligence, 25*(2), 111–148.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*(1), 174–188.

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91*(3), 511–526.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking?: Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72,* 678–691.

Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership and Organization Development Journal, 26*(5), 388–399.

Rumsey, M. G., & Arabian, J. M. (2014). Military enlistment selection and classification: Moving forward. *Military Psychology, 26*(3), 221–251.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality,* 9(3), 185–211.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*(4), 532–545.

Santarnecchi, E., Emmendorfer, A., & Pascual-Leone, A. (2017). Dissecting the parieto-frontal correlates of fluid intelligence: A comprehensive ALE meta-analysis study. *Intelligence, 63,* 9–28.

Scarborough, H. S., & Brady, S. A. (2002). Toward a common terminology for talking about speech and reading: A glossary of the "phon" words and some related terms. *Journal of Literacy Research, 34*(3), 299–336.

Schank, R., & Birnbaum, L. (1994). Enhancing intelligence. In J. Khalfa (Ed.), *What is intelligence?* (pp. 72–106). Cambridge, UK: Cambridge University Press.

Schelble, J. L., Therriault, D. J., & Miller, M. D. (2012). Classifying retrieval strategies as a function of working memory. *Memory and Cognition, 40*(2), 218–230.

Scherbaum, C. A., Goldstein, H. W., Yusko, K. P., Ryan, R., & Hanges, P. J. (2012). Intelligence 2.0: Reestablishing a research program on *g* in I–O psychology. *Industrial and Organizational Psychology,* 5(2), 128–148.

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence, 46,* 156–168.

Schmidt, C. S., Schumacher, L. V., Römer, P., Leonhart, R., Beume, L., Martin, M., et al. (2017). Are semantic and phonological fluency based on the same or distinct sets of cognitive processes?: Insights from factor analyses in healthy adults and stroke patients. *Neuropsychologia, 99,* 148–155.

Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspectives on Psychological Science,* 6(6), 560–573.

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A commentary on four recent articles. *Perspectives on Psychological Science,* 9(2), 211–218.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274.

Schneider, W. J., & Flanagan, D. P. (2015). The relationship between theories of intelligence and intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 317–340). New York: Springer.

Schneider, W. J., Mayer, J. D., & Newman, D. A. (2016). Integrating hot and cool intelligences: Thinking broadly about broad abilities. *Journal of Intelligence, 4*(1), 1–25.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schneider, W. J., & McGrew, K. S. (2013). Cognitive performance models: Individual differences in the ability to process information. In B. Irby, G. Brown, R. Laro-Alecio, & S. Jackson (Eds.), *Handbook of educational theories* (pp. 767–782). Charlotte, NC: Information Age.

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review, 25*(1), 12–27.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2015). *Woodcock–Johnson IV Tests of Early Cognitive and Academic Development.* Rolling Meadows, IL: Riverside.

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence, 48,* 15–29.

Schultz, D., Izard, C. E., Ackerman, B. P., & Youngstrom, E. A. (2001). Emotion knowledge in economically disadvantaged children: Self-regulatory antecedents and relations to social difficulties and withdrawal. *Development and Psychopathology, 13*(1), 53–67.

Schulze, D., Beauducel, A., & Brocke, B. (2005). Semantically meaningful and abstract figural reasoning in the context of fluid and crystallized intelligence. *Intelligence, 33*(2), 143–159.

Seery, N., Buckley, J., & Delahunty, T. (2015). *Developing a spatial ability framework to support spatial ability research in engineering education.* Paper presented at the 6th Research in Engineering Education Symposium, Dublin Institute of Technology, Dublin, Ireland.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4–27.

Sharma, S., Bottom, W. P., & Elfenbein, H. A. (2013). On the role of personality, cognitive ability, and emotional intelligence in predicting negotiation outcomes: A meta-analysis. *Organizational Psychology Review, 3*(4), 293–336.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72,* 116–141.

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review, 27*(4), 599–606.

Silvia, P. J., & Beaty, R. E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence, 40*(4), 343–351.

Sirolli, E. (1999). *Ripples from the Zambezi: Passion, entrepreneurship, and the rebirth of local economies. Community works!* Gabriola Island, BC, Canada: New Society.

Slevc, L. R. (2012). Language and music: Sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(4), 483–492.

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2016). Reliability and validity of ambulatory cognitive assessments. *Assessment, 25*(1), 14–30.

Spearman, C. E. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*(2), 201–292.

Spearman, C. E. (1923). *The nature of "intelligence" and the principles of cognition.* London: Macmillan.

Spearman, C. E. (1930). Autobiography. In C. Murchison (Ed.), *History of psychology in autobiography* (Vol. I, pp. 299–333). Worcester, MA: Clark University Press.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745–759.

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence, 53,* 92–101.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*(1), 35–51.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology, 72*(1), 21–44.

Stankov, L., Seizova-Cajic, T., & Roberts, R. D. (2001). Tactile and kinesthetic perceptual processes within the taxonomy of human cognitive abilities. *Intelligence, 29*(1), 1–29.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1997). *Successful intelligence.* New York: Plume.

Sternberg, R. J. (1999a). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*(4), 359–375.

Sternberg, R. J. (1999b). The theory of successful intelligence. *Review of General Psychology, 3*(4), 292–316.

Stevenson, R. J. (2013). Olfactory perception, cognition, and dysfunction in humans. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 273–284.

Stevenson, R. J., & Case, T. I. (2005). Olfactory imagery: A review. *Psychonomic Bulletin and Review, 12*(2), 244–264.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947.

Stough, C., & Pase, M. P. (2015). Improving cognition in the elderly with nutritional supplements. *Current Directions in Psychological Science, 24*(3), 177–183.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Süß, H.-M., & Beauducel, A. (2015). Modeling the construct validity of the Berlin Intelligence Structure model. *Estudos de Psicologia (Campinas), 32*(1), 13–25.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*(3), 294–321.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience, 5*(9), 721–728.

Tanabe, A., & Osaka, N. (2009). Picture span test: Measuring visual working memory capacity involved in remembering and comprehension. *Behavior Research Methods, 41*(2), 309–317.

Thissen, A., Koch, M., Becker, N., & Spinath, F. M. (2016). Construct your own response: The cube construction task as a novel format for the assessment of spatial ability. *European Journal of Psychological Assessment*. [Epub ahead of print]

Thompson, W. L., Slotnick, S. D., Burrage, M. S., & Kosslyn, S. M. (2009). Two forms of spatial imagery: Neuroimaging evidence. *Psychological Science, 20*(10), 1245–1253.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine, 40,* 227–235.

Thorndike, R. L. (1936). Factor analysis of social and abstract intelligence. *Journal of Educational Psychology, 27*(3), 231–233.

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin, 34*(5), 275–285.

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* Chicago: University of Chicago Press.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience, 29,* 855–868.

Tobia, V., Rinaldi, L., & Marzocchi, G. M. (2016). Time processing impairments in preschoolers at risk of developing difficulties in mathematics. *Developmental Science*. [Epub ahead of print]

Tommasi, L., & Laeng, B. (2012). Psychology of spatial cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(6), 565–580.

Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development, 19*(1), 1–29.

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7*(1), 77–88.

Tsianos, N., Germanakos, P., Belk, M., Lekkas, Z., Samaras, G., & Mourlas, C. (2013). An individual differences approach in designing ontologies for efficient personalization. In I. E. Anagnostopoulos, M. Bielikova, P. Mylonas, & N. Tsapatsoulis (Eds.), *Semantic hyper/multimedia adaptation* (pp. 3–21). Berlin: Springer.

Turvey, M. T. (1996). Dynamic touch. *American Psychologist, 51*(11), 1134–1152.

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81*(5), 1417–1430.

Undheim, J. O. (1978). Broad ability factors in 12- to 13-year-old children, the theory of fluid and crystallized intelligence, and the differentiation hypothesis. *Journal of Educational Psychology, 70*(3), 433–443.

Undheim, J. O., & Gustafsson, J.-E. (1987). The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research, 22*(2), 149–171.

Unsworth, N. (2017). Examining the dynamics of strategic search from long-term memory. *Journal of Memory and Language, 93,* 135–153.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1–26.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1302–1321.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both?: A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392–406.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science, 22*(5), 367–373.

Uzieblo, K., Winter, J., Vanderfaeillie, J., Rossi, G., & Magez, W. (2012). Intelligent diagnosing of intellectual disabilities in offenders: Food for thought. *Behavioral Sciences and the Law, 30*(1), 28–48.

Van Biesen, D., Mactavish, J., McCulloch, K., Lenaerts, L., & Vanlandewijck, Y. C. (2016). Cognitive profile of young well-trained athletes with intellectual disabilities. *Research in Developmental Disabilities, 53,* 377–390.

van de Weijer-Bergsma, E., van de, Kroesbergen, E. H., & van Luit, J. E. (2015). Verbal and visual–spatial working memory and mathematical ability in different domains throughout primary school. *Memory and Cognition, 43*(3), 367–378.

Vannorsdall, T. D., Maroof, D. A., Gordon, B., & Schretlen, D. J. (2012). Ideational fluency as a domain of human cognition. *Neuropsychology, 26*(3), 400–405.

Velez, M. C., Silver, D., & Tremaine, M. (2005, October). *Understanding visualization through spatial ability differences.* Paper presented at the IEEE Xplore Conference: Visualization, 2005, Minneapolis, MN. Retrieved from *http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=10269.*

Vernon, P. E. (1950). *The structure of human abilities.* New York: Wiley.

Walker, R. E., & Foley, J. M. (1973). Social intelligence: Its history and measurement. *Psychological Reports, 33*(3), 839–864.

Wang, L., Cohen, A. S., & Carr, M. (2014). Spatial ability at two scales of representation: A meta-analysis. *Learning and Individual Differences, 36,* 140–144.

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual–spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology, 16*(2), 317–322.

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition, 166,* 152–163.

Wasserman, J. D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 3–55). New York: Guilford Press.

Watson, D. (2015). Intelligence testing. In E. A. Polloway (Ed.), *The death penalty and intellectual disability* (pp. 113–140). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Wechsler, S. M., & Nakano, T. de C. (2016). Cognitive assessment of Brazilian children and youth: Past and present perspectives and challenges. *International Journal of School and Educational Psychology, 4*(4), 215–224.

Wechsler, S. M., Nunes, C. S., Schelini, P. W., Pasian, S. R., Homsi, S. V., Moretti, L., et al. (2010). Brazilian adaptation of the Woodcock–Johnson III cognitive tests. *School Psychology International, 31*(4), 409–421.

Wechsler, S. M., & Schelini, P. W. (2006). Bateria de habilidades cognitivas Woodcock–Johnson III: Validade de construto. *Psicologia: Teoria e Pesquisa, 22*(3), 287–296.

Wechsler, S. M., Vendramini, C. M. M., & Schelini, P. W. (2007). Adaptação brasileira dos testes verbais da Bateria Woodcock–Johnson III. *Interamerican Journal of Psychology, 41*(3), 285–294.

Wechsler, S. M., Vendramini, C. M. M., Schelini, P. W., Lourençoni, M. A., Ferreira de Souza, A. A., & Mundim, M. C. B. (2014). Investigating the factorial structure of the Brazilian Adult Intelligence Battery. *Psychology and Neuroscience, 7*(4), 559–566.

Wee, S., Newman, D. A., & Joseph, D. L. (2014). More than *g*: Selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology,* 99(4), 547–563.

Weisberg, R. W. (2006). Expertise and reason in creative thinking: Evidence from case studies and the laboratory. In J. C. Kaufman & J. Baer (Eds.), *Creativity and reason in cognitive development* (pp. 7–42). New York: Cambridge University Press.

Weiss, D. J., & Shanteau, J. (2014). Who's the best?: A relativistic view of expertise. *Applied Cognitive Psychology, 28*(4), 447–457.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31*(2), 114–131.

Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's and Dementia,* 4(6), 428–437.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). Thousand Oaks, CA: SAGE.

Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology,* 5(1), 404.

Wilson, E. O. (1999). *Consilience: The unity of knowledge.* New York: Vintage Books.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review,* 9(4), 625–636.

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences, 14*(3), 138–146.

Wolff, U., & Gustafsson, J.-E. (2015). Structure of phonological ability at age four. *Intelligence, 53,* 108–117.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*(3), 231–258.

Woodcock, R. W. (1993). An information processing view of the Gf-Gc theory. In *Journal of Psychoeducational Assessment, Monograph Series: Woodcock–Johnson Psycho-Educational Assessment Battery–Revised* (pp. 80–102). Cordova, TN: Psychoeducational Corporation.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., et al. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences USA, 107*(33), 14899–14902.

World Health Organization. (2007). *International Classification of Functioning, Disability, and Health: Children and Youth Version: ICF-CY.* Geneva, Switzerland: Author.

Yilmaz, H. B. (2009). On the development and measurement of spatial ability. *International Electronic Journal of Elementary Education, 1*(2), 83–96.

Zajac, I. T., Burns, N. R., & Nettelbeck, T. (2012). Do purpose-designed auditory tasks measure general speediness? *International Journal of Intelligence Science, 2*(2), 23–31.

Zajac, I. T., & Nettelbeck, T. (2016). Auditory speed tasks as potential candidates for the study of cognitive ageing. *Aging, Neuropsychology, and Cognition,* 1–19.

Zeidner, M., Roberts, R. D., & Matthews, G. (2008). The science of emotional intelligence: Current consensus and controversies. *European Psychologist, 13*(1), 64–78.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology, 110*(4), 481–498.

Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality, 46*(2), 173–183.

CHAPTER 4

Assessment from the Perspective of Multiple-Intelligences Theory

Principles, Practices, and Values

Jie-Qi Chen
Howard Gardner

Howard Gardner of Harvard University introduced the theory of multiple intelligences (MI) in his book *Frames of Mind*, first published in 1983. The theory has touched numerous branches of psychology and education, including counseling psychology, developmental therapy, teacher training, gifted education, programs for at-risk children, and museum education, to name a few. As the book has been translated into more than 20 languages, MI-inspired educational practices have made noticeable impacts in many countries and regions around the world (Chen, Moran, & Gardner, 2009).

A central issue in MI-inspired educational practice is assessment. With MI's focus on individual differences, educators need assessment tools designed to detect these differences. This chapter addresses the questions of what to assess, as well as how to assess individuals' diverse intellectual abilities, from an MI-based perspective. We begin the chapter with an overview of MI theory. Moving from theory to practice, we identify distinctive features of the MI approach to child assessment. We then introduce several instruments that incorporate MI assessment features. Empirical studies based on these instruments are reported as well. We conclude with a discussion of the role that values play in the assessment process—specifically, in determining both what is assessed and how it is measured.

UNDERSTANDING MI THEORY

A significant departure from the traditional view of intelligence as IQ, MI theory (Gardner, 1983, 1993a, 1999, 2006, 2011) fundamentally changes what intelligence is by (1) redefining what constitutes intelligence, (2) redesigning the methods used to study intelligence, and (3) identifying intelligences beyond the verbal and quantitative abilities that IQ measures.

Gardner (1999) defines *intelligence* as "a biopsychological potential to process information that can be activated in a cultural setting to solve problems or create products that are of value in a culture" (p. 33). By considering intelligence a potential, Gardner asserts its emergent and responsive nature, thereby differentiating his theory from traditional ones that view human intelligence as fixed and innate. Whether a potential can be activated depends on the values of the culture in which an individual grows up; the opportunities available in that culture; and personal choices made by individuals, their families, and others in their lives. These activating forces shape the development and expression of a range of intelligences from culture to culture and from individual to individual. Gardner's definition of intelligence also differs from other formulations in that it considers the creation of products, such as musical scores, sculptures, or computer programs, to be as important an expression of intelligence as abstract problem solving. Traditional theories do not recognize created artifacts as manifestations of intelligence and therefore are limited in both conceptualization and measurement.

The methodology Gardner used to develop MI theory represents a major departure from the psychological testing approach that has been used to study intelligence. The distinctive method Gardner used in studying intelligence is inseparable from his groundbreaking view of human intelligence. In the process of developing MI theory, Gardner considered the range of adult end states that are valued in diverse cultures around the world. To identify the abilities supporting these end states, he examined empirical data from disciplines that had not been considered previously for the purpose of defining human intelligence. His examination of these datasets yielded eight criteria for identifying an intelligence. The criteria took into consideration brain function, evolutionary history, special human populations, end-state adult performances, skill training, correlation of intelligence test results, development of symbol system, and core operations of intelligence (Gardner, 1993a). Of principal importance to Gardner in developing the criteria was capturing the range of purposes and processes entailed in human cognitive functioning.

Gardner (1993a, 2006, 2011) has argued that standardized intelligence tests typically probe a limited number of intelligences, such as linguistic, logical–mathematical, and certain forms of spatial intelligences. MI theory has added five more candidates to the list: musical, bodily–kinesthetic, naturalistic, interpersonal, and intrapersonal intelligences (see Table 4.1). According to Gardner, all human beings possess all of the intelligences, but differ in relative strengths and weaknesses—an important source of individual differences. The eight identified intelligences, according to Gardner, cannot be viewed merely as a group of raw computational capacities. They are subject to encoding in varied symbol systems created by various cultures. It is through symbol systems that intelligences are applied in specific domains or bodies of knowledge within a culture, such as mathematics, art, basketball, and medicine (Gardner, 1993a, 1999). As well, the world is wrapped in meanings. Intelligences can be implemented only to the extent that they partake of these meanings and enable individuals to develop into functioning, symbol-using members of their community. An individual's intelligences, to a great extent, are shaped by cultural influences and refined by educational processes. It is through the process of education that "raw" intellectual competencies are

TABLE 4.1. Identified Multiple Intelligences

Intelligence	Sample adult role	Definition
Linguistic intelligence	Writers and poets	The ability to perceive and generate spoken or written language
Logical–mathematical intelligence	Mathematicians and computer programmers	The ability to appreciate and utilize numerical, abstract, and logical reasoning to solve problems
Musical intelligence	Musicians and composers	The ability to create, communicate, and understand meanings made out of sound
Spatial intelligence	Graphic designers and architects	The ability to perceive, modify, transform, and create visual or spatial images
Bodily–kinesthetic intelligence	Dancers and athletes	The ability to use all or part of one's body to solve problems or fashion products
Naturalistic intelligence	Archaeologists and botanists	The ability to distinguish among, classify, and use features of the environment
Interpersonal intelligence	Leaders and teachers	The ability to recognize, appreciate, and contend with the feelings, beliefs, and intentions of other people
Intrapersonal intelligence	Apparent when individuals pursue a particular interest or choose a field of study or work	The ability to understand oneself—including emotions, desires, strengths, and vulnerabilities—and to use such information effectively in regulating one's own life

developed and individuals are prepared to assume mature cultural roles. Rich educational experiences are essential for the development of each individual's particular configuration of interests and abilities (Gardner, 1993b, 2006).

MI-BASED ASSESSMENT PRINCIPLES

Assessment based on MI theory calls for a significant departure from traditional approaches to assessment. From the start, a distinctive hallmark of MI theory has been its spurning of one-shot, decontextualized, paper-and-pencil tests to rank an individual's "smartness" based on a single score (Gardner, 1993b, 2006). MI theory presents several basic principles for the assessment of intelligences: (1) sampling intellectual capacities in a wide range of domains; (2) using media appropriate to each assessed domain; (3) choosing assessment materials that are meaningful to students; (4) attending to the ecological validity of assessment contexts; and (5) portraying complete intellectual profiles to support learning and teaching (Chen, 2004; Krechevsky, 1998) (see Table 4.2).

TABLE 4.2. MI Principles for the Assessment of Intellectual Profiles

- *Sample intellectual capacities in a range of domains* that include both traditionally defined academic areas such as reading, literacy, math, and science, as well as nonacademic areas such as visual arts, performing arts, movement, and understanding of others and self.
- *Use media appropriate to each domain of assessment* to engage in an intelligence-fair assessment process by looking at the problem-solving features and operational mechanisms of particular intelligences; such a process allows one to look directly at the functioning of each intellectual capacity.
- *Choose assessment materials that are meaningful to students* by supporting thinking, inviting questions, stimulating curiosity, facilitating discovery, and encouraging the use of imagination and multiple symbol systems in the students' problem-solving processes.
- *Attend to the ecological validity of assessment contexts* to ensure that the assessment environments are natural, familiar, and ongoing; use multiple samples of a child's performance; and incorporate clinical judgments from those who are knowledgeable about the child being assessed and directly responsible for using the results.
- *Portray complete intellectual profiles* that focus on students' strengths and include concrete, practical suggestions to help educators understand each child as completely as possible, and then mobilize the child's intelligences to achieve specific educational goals.

Sampling Intellectual Capacities in a Wide Range of Domains

Because the fundamental principle of MI is that human intelligence is pluralistic, assessment based on MI theory incorporates a range of domains to tap different facets of each intellectual capacity. In addition to language, literacy, and mathematics—the primary foci of traditional intelligence tests and school achievement tests—MI-based assessment also looks at children's performance in areas often viewed as nonacademic, such as music, arts, movement, and understanding of self as well as others. The MI approach to assessment recognizes students who excel in linguistic and/or logical pursuits, as well as those who have cognitive and personal strengths in other intelligences. By virtue of the wider range they measure, MI types of assessment identify more students who are "smart," albeit in different ways (Gardner, 1993b, 2000).

It has been well documented that students who have trouble with some academic subjects, such as reading or math, are not necessarily inadequate in all areas (Diaz-Lefebvre, 2009; Levin, 2005). The challenge is to provide comparable opportunities for these students to demonstrate their strengths and interests. When students recognize that they are good at something, and their accomplishment is acknowledged by teachers, parents, and peers, they are far more likely to feel valued in the classroom and to experience further school success. In some instances, the sense of success in one area may make students more likely to engage in areas where they feel less comfortable. When that occurs, the systematic use of multiple measures goes beyond its initial purpose of identifying diverse cognitive abilities and becomes a means of bridging students' strengths from one area to other areas of learning (Chen, Krechevsky, & Viens, 1998; Dweck, 2007; Gardner, 1998).

Using Media Appropriate to Each Domain Assessed

On the basis of its contention that each intelligence exhibits particular problem-solving features and operational mechanisms, MI theory argues for *intelligence-fair* instruments to assess the unique capacities of each intelligence. Too often, language is the gatekeeper, forcing individuals to reveal their intelligence through the customary lens of linguistic ability; or logical analysis serves as a route to, or an obstacle thwarting, the measurement of nonscholastic abilities. In contrast, intelligence-fair instruments engage the key abilities of particular intelligences, allowing one to look directly at the functioning of each intellectual capacity.

When intelligence-fair instruments are used, bodily intelligence can be assessed by recording how a person learns a new dance or physical exercise. To consider a person's interpersonal intelligence, it is necessary to observe how he or she interacts with and influences others in different social situations. One situation might be the individual's interacting with a friend to offer extra support when the friend loses an art contest. Another relevant situation is observing an individual giving advice to a friend who is the target of a rumor. It is important to note that what is assessed is never an intelligence in pure form. Intelligences are always expressed in the context of specific tasks, domains, and disciplines. For example, there is no "pure" spatial intelligence; instead, there is spatial intelligence as expressed in a child's puzzle solution, route finding, block building, or basketball passing (Gardner, 1993b).

Choosing Assessment Materials Meaningful to Students

Materials are important in assessment because intelligence is manifested through a wide variety of artifacts. To be meaningful, the assessment materials first need to be familiar to children. Assessment based on MI theory is responsive to the fact that students' prior experience with assessment materials directly affects their performance on tasks. For example, children who have little experience with blocks are less likely to do well on a block design task. Likewise, it would be unfair to assess a child's musical ability by asking the child to play a music instrument that he or she has never experienced. In recognition of the role that experience plays, the MI approach to assessment emphasizes using materials that are familiar to children. If children are not familiar with materials, they are given ample opportunities to explore materials prior to any formal assessment.

The term *meaningful* also signifies the role of assessment materials in supporting a student's problem-solving process. To be fair, materials used in many current intelligence tests, such as pictures, geometric shapes, and blocks, are not unfamiliar to children in industrial societies. Yet such materials provide little intrinsic attraction because they have little meaning in children's daily lives. For assessment to be meaningful for students, the selection of materials must be a careful and deliberate process. Materials ought to be an integral part of students' problem-solving processes, supporting thinking, inviting questions, and stimulating curiosity. Meaningful materials also facilitate students' discovery and encourage the use of imagination and multiple symbol systems (Rinaldi, 2001).

Attending to the Ecological Validity of Assessment Contexts

In the traditional intelligence assessment situation, a psychologist works with one child at a time, preferably in a relatively quiet room away from any distractions, including regular classroom activities. MI theory emphasizes the ecological validity of assessment contexts (Gardner, 1993b); that is, the assessment environments must be natural, familiar, and ongoing. When a child's ability is measured through a one-shot test using decontextualized tasks, the child's profile of abilities is often incomplete and may be distorted. In contrast, when assessment is naturally embedded in learning environments, it allows psychologists and educators to observe children's abilities in various situations over time. Such observations generate multiple samples of a child's ability that can be used to document variations of the child's performances within and across domains, and so to portray the child's intellectual profile more accurately.

Integrating authentic activities and observations over time, assessment based on MI theory does not typically function as a norm-referenced instrument does. Intelligence in the MI framework is defined as a potential, exhibiting various possible forms and subject to continuous changes in expression and strength. MI-based assessment involves performance standards or criterion references that educators can use to guide and evaluate their observations. In contrast to norm-referenced tests, which feature decontextualized and seemingly impartial judgments of students' performance, MI-based assessment is open to incorporating the clinical judgments of classroom teachers. In so doing, MI-based assessment places greater value on the experience and expertise of educators who are knowledgeable about the child being assessed and are directly responsible for using the assessment results (Chen, 2004; Darling-Hammond & Ancess, 1996; Linn, 2000).

Portraying Complete Intellectual Profiles to Support Learning and Teaching

Traditional tests—of achievement, readiness, intelligence, and the like—are often used to rank-order and sort students on the basis of a single quantitative score. Seemingly objective scores on these standardized tests disguise the complex nature of human intelligence. In the process, the scores also limit a child's range of learning potentials and may narrow opportunities for success in school. Instead of ranking and labeling, the purpose of MI types of assessment is to support students on the basis of their complete intellectual profiles. Such an intellectual profile portrays a child's strengths, interests, and weaknesses. It also includes concrete, practical suggestions to the student, such as how to build on the identified strengths, work on areas that need attention or intervention, and develop approaches to learning that are conducive to productive work (Chen & McNamee, 2011; Krechevsky, 1998).

It is important to note that the identification of intellectual strengths and weaknesses of individuals is not the endpoint of MI types of assessment. The purpose of portraying a complete intellectual profile is to help educators understand each child as completely as possible and then mobilize his or her intelligences to achieve specific educational goals. MI-based assessments promote achievement of these goals by assisting educators in selecting appropriate instructional strategies and pedagogical approaches, based on a comprehensive and in-depth understanding of each child.

MI-BASED ASSESSMENT PRACTICES

Since MI theory was introduced, educators have looked for an assessment that could be used to better understand students' diverse intellectual abilities and inform instructional practice. We introduce three tools that are designed for such purposes: the Spectrum battery, the Bridging assessment, and the Multiple Intelligences Developmental Assessment Scales (MIDAS™). Empirical work associated with the three instruments is also reported.

The Spectrum Battery

The Spectrum battery, designed by the staff of Project Spectrum at the Harvard Graduate School of Education, is the only MI-based assessment instrument developed with Gardner's direct involvement. Designed for preschool children, the Spectrum battery is composed of 15 activities in seven domains of knowledge: language, math, music, art, social understanding, sciences, and movement (Chen, 2004; Krechevsky, 1998) (see Table 4.3). The Spectrum's name reflects its mission of recognizing diverse intellectual strengths in children.

TABLE 4.3. Spectrum Assessment Activities

Spectrum areas	Spectrum activities	Measured abilities
Language	1. Storyboard activity	Invented narrative
	2. Reporter activities	Descriptive narrative
Visual Arts	3. Art portfolio	Visual arts skills
Mathematics	4. Dinosaur game	Counting strategies
	5. Bus game	Calculating and notation
Sciences	6. Discovery area	Naturalist observation
	7. Treasure hunt game	Logical inference
	8. Sink and float activity	Hypothesis testing
	9. Assembly activity	Mechanical abilities
Movement	10. Movement with music	Creative movement
	11. Obstacle course	Athletic movement
Social Understanding	12. Classroom model	Social analysis
	13. Peer interaction checklist	Social roles
Music	14. Montessori bells	Music perception
	15. Singing songs	Music production

During the assessment process, an assessor or a teacher works with children either individually or in small groups. Children engage in a range of activities, such as disassembling and assembling several house gadgets in the science domain; playing Montessori bells in the music domain; keeping track of passengers getting on and off a toy bus in the mathematics domain; and manipulating figures in a scaled-down, three-dimensional replica of the children's classroom to assess social understanding. Fun and challenging, these activities invite children to engage in problem-solving tasks. They are intelligence-fair, using materials appropriate to particular domains rather than relying only on language to assess multiple forms of competence and ability. They help to tap key abilities—abilities that are essential to the operation of particular intellectual domains in children's task performance. Each activity is accompanied by written instructions for task administration. These instructions include a score sheet that identifies and describes different levels of the key abilities assessed in the activity, making a child's performance on many activities quantifiable.

Spectrum assessment results are presented in the form of a Profile—a narrative report based on the information obtained from the assessment process (Chen, 2004; Krechevsky, 1998). Using nontechnical language, the report focuses on the range of cognitive abilities examined by the Spectrum battery. It describes each child's relative strengths and weaknesses in terms of that child's own capacities, and only occasionally in relation to peers. Strengths and weaknesses are described in terms of the child's performance in different content areas. For example, a child's unusual sensitivity to different kinds of music may be described in terms of facial expressions, movement, and attentiveness during and after listening to various music pieces. It is important to note that the child's intellectual profile is described not only in terms of capacities, but also in terms of the child's preferences and inclinations. Furthermore, the Profile is not a static image, but a dynamic composition that reflects a child's interests, capabilities, and experiences at a particular point in time. The profile changes as the child's life experience changes. The conclusion of the Profile typically includes specific recommendations to parents and teachers about ways to support identified strengths and improve weak areas (Krechevsky, 1998).

Using the Spectrum battery, Chen studied a group of first graders in four public classrooms (Chen & Gardner, 2005). Among those who had been identified as at risk for school failure, Chen found out that they demonstrated identifiable strengths on the Spectrum assessment. Also noteworthy, these students showed more strengths in nonacademic areas such as mechanical, movement, and visual arts than in academic areas such as language and math. Had the assessment been limited to academic areas, these at-risk children's strengths would have gone undetected and could not have served as bridges for extending the children's interest and learning to other curricular areas.

Bridging

Bridging was developed by Chen and McNamee (2007) to help teachers document and describe intellectual profiles of children between the ages of 3 and 8. It shares certain features with the Spectrum assessment, including the identification of children's diverse cognitive strengths, the use of engaging activities, and a focus on guided observation and careful documentation. It differs from the Spectrum assessment by focusing on the operation of intellectual abilities in school curricular areas, such as language and literacy; number sense and geometry; physical, mechanical, and natural sciences; and performing and visual arts. Bridging is organized in terms of school subject areas rather than intellectual domains, for several reasons: (1) Intelligences never function in abstract form, but rather are used in the context of specific disciplinary tasks; (2) school subject areas reflect intellectual abilities valued in our society; (3) children mobilize their intelligences in the pursuit of studying subject areas; and (4) aligning assessment areas with school subject areas facilitates teachers' incorporation of the assessment results into curriculum planning. The name of the instrument, Bridging, signifies its goal of building a bridge from assessment to teaching (Chen & McNamee, 2008).

Bridging includes a total of 15 regular classroom activities, such as reading a child's favorite book, constructing a model car with recycled materials, and experimenting with light and shadows (see Table 4.4). Children's performance in each of the 15 activities is scored according to a 10-level, criterion-referenced rubric developed specifically for 3- to 8-year-olds. As an example, the rubric used to measure performance on the "reading books" activity was based on the stages of pretend reading developed by Sulzby (1985) and work in guided reading by Fountas and Pinnell (1996). The ru-

TABLE 4.4. Bridging Assessment Areas and Activities

Bridging areas	Bridging activities
Language Arts and Literacy	1. Reading books (child's choice, teacher's choice) 2. Dictating a story 3. Acting out stories
Visual Arts	4. Experimenting crayon technique 5. Drawing a self-portrait 6. Making pattern block pictures
Mathematics	7. Creating pattern block pinwheels 8. Solving pattern block puzzles 9. Exploring number concepts
Sciences	10. Exploring shadows and light 11. Assembling a nature display 12. Building a model car
Performing Arts	13. Moving to music 14. Playing an instrument 15. Singing a song

bric progresses from attending to pictures without forming stories at level 1 to reading for meaning independently at level 10 (Chen & McNamee, 2007).

Empirical data support the Bridging approach—both its conceptualization of children's abilities and the effective design of its assessment activities. In one study using Bridging, Chen, McNamee, and McCray (2011) examined the intellectual profiles of 92 preschool and kindergarten children. Results indicated that *within a child's profile,* levels of competence varied as a function of content area. A child's competence level was higher in some areas and lower in others. *Among children's profiles,* the patterns of their performance levels were distinctive. That is, the pattern in each child's profile differed from the pattern found in other children's profiles. Children's competence is thus domain-specific.

In terms of Bridging's utility, over 400 preservice and inservice teachers from preschool through third grade have integrated it into their classrooms under the direct supervision of the instrument's developers. An implementation study of 75 preservice teachers revealed that the construction of intellectual profiles for individual students using the Bridging assessment process was a key component in these student teachers' understanding of diverse learners in their classrooms (Chen & McNamee, 2006). In addition, teachers' understanding of individual students and content knowledge increased as the result of implementing Bridging in their classrooms. This increased understanding contributed to their ability to be more effective in curriculum planning and teaching (McNamee & Chen, 2005).

The MIDAS Battery

Developed by Branton Shearer, the MIDAS battery (*www.MIResearch.org*) consists of eight main scales that correspond to the eight intelligences defined by Gardner. It is designed to portray the profile of an individual's intellectual dispositions, both quantitatively and qualitatively (Shearer, 2007). Covering ages from preschool through adulthood in six different forms, the MIDAS can be completed online in less than 30 minutes or administered as a structured interview for respondents with reading limitations. Young children complete the assessment process through the help of parents in collaboration with the teacher. Responding to the criticism that the MIDAS results are questionable because they rely primarily on self-report, Shearer (2007) argues that one of the hallmarks of the MIDAS is attention to the intrapersonal intelligence reflected in the self-report process and profile interpretation. For the MIDAS, intrapersonal intelligence is the "royal road to learning, achievement, and personal growth" (Shearer, 2007).

For each respondent, a MIDAS Profile Report is constructed. This report describes the respondent's specific skills, capacities, and enthusiasms in terms of Gardner's eight intelligences. Depending on the level of detail and descriptive information needed about the respondent, the MIDAS can be used for three purposes: as a brief Screening instrument, to generate an Individual Profile with a brief learning summary, or to produce a more comprehensive Personalized Assessment Plan. An ultimate goal for MIDAS interpretation is to build a communicating *bridge* among important stakeholders in a student's education, so that strength-based plans may be effectively conceived and then implemented. The Profile Report provides instructional information to teachers. It also suggests practical activities for each student and parents at home.

The MIDAS has been translated into numerous foreign languages for use in regions and countries such as Taiwan, Korea, the Netherlands, Iran, Singapore, India, and Chile. It has also been tested for its psychometric properties in the United States as well as internationally, including test–retest reliability, interrater reliability, concurrent validity, and level of independence among scales (Shearer, 2007). In an analysis of 23,000 individuals who completed the MIDAS profile, for example, Shearer (2005) found nine factors corresponding with the MIDAS main scales. These factor-analytic results have been replicated with the Chinese, Farsi, and Korean translations, providing additional cross-cultural empirical validity. Shearer thus concluded that the MIDAS provides a reasonable estimate of the respondents' multiple intellectual dispositions.

The MIDAS has been used with diverse populations, including elementary, secondary, and university students; students with attention-deficit/hyperactivity disorder (ADHD) and learning disabilities; and adults seeking career help. In studies of 116 elementary children diagnosed with ADHD, for example, Shearer and colleagues found that, compared to typically developing peers, these students scored lower on the math, linguistic, and intrapersonal scales of the MIDAS, but higher on the naturalistic, spatial, and kinesthetic scales (Proulx-Schirduan, Shearer, & Case, 2009). The MIDAS profiles provided unique descriptions of the students with ADHD, whose interests and strengths included artistic design, craft work, and recognizing different kinds of plants.

ASSESSMENT AS A VALUES-BASED PROCESS

Over the decades since the inception of MI theory, many MI-based assessments have been developed (Chen & McNamee, 2007; Krechevsky, 1998; Lazear, 1998, 1999; Maker, 2005; New City School, 1994; Niccolini, Alessandri, & Bilancioni, 2010; Shearer, 2007), and several dozen schools in the United States and across the world have adopted the MI approach to assessment (Chen et al., 2009; Kornhaber, Fierros, & Veenema, 2004). However, neither the assessment nor the approach has been widely implemented. In critiquing MI's approach to assessment, no one has asserted that individuals do not have distinctive sets of intellectual abilities or argued that individual differences in abilities are not important in providing instruction. It is our contention that the primary reason the MI approach has not been widely used is related to the role that values play in assessment.

Many educators and parents tend to think of assessment as objective and regard assessment measures as accurate indicators of the intellectual abilities they test. There is an untested assumption of agreement about both what is assessed and how it is measured. Measuring intellectual abilities is likened to the process of measuring characteristics like height and weight. MI theory challenges this assumption. Standardized tests are not measuring tools like tape measures and scales. An individual's intelligences are not amounts of ability that can be quantified in numerical terms, the way height and weight can be measured in inches and pounds. There is no established scale of measure or unit of measurement that objectively describes intelligences.

Because MI theory redefines what intelligences are, it requires us to acknowledge that *intelligence* is a construct. We, as psychologists and educators, *decide* what counts as intelligence, what to measure, and how to measure it. Our purpose in discussing this point is to call attention to the fact that assessment *is* a value-laden process and to question the adequacy of traditional standardized tests, including intelligence tests, aptitude tests, and achievement tests. We briefly describe four aspects of assessment that are affected by values.

1. *Values define the purpose of assessment.* Standardized test scores are used to rank-order, label, and categorize individuals. This stands in contrast to the value that underlies MI assessment—namely, to learn about every individual's competencies and interests, and use this knowledge to design educational experiences that maximize each student's opportunities to succeed in school and in life.

2. *Values determine the content of assessment.* Traditional school achievement tests focus largely on math and reading. MI-based approaches to assessment look across a wide range of domains, including art music, dance, science, competencies in relating to others, and understanding of self. Attention to this wide range of intellectual abilities increases the likelihood of identifying each student's strengths and interests, which they can pursue further and use to develop skills in other areas of learning.

3. *Values influence the selection of assessment methods and materials.* Paper-and-pencil tests are used in standardized school testing of math and reading. Assessing abilities and interests in a wide range of domains, MI-based approaches are based on diverse materials specific to the domains being assessed. Furthermore, MI-based approaches rely

on students' use of these materials in contexts that enable them to demonstrate their competencies through applying them to solve problems and create products.

4. *Values shape how assessment results are used.* Students' performance on standardized end-of-year tests are used to make decisions about the students' promotion and to evaluate their schools' effectiveness. Scores also affect federal funding decisions. In contrast, MI approaches focus on directly linking assessment results to teachers' ongoing design and delivery of instruction.

Assessment, as much current literature points out, provides major leverage for affecting educational practice. Many people criticize current educational practices; a significant part of our educational malaise, however, lies in our current methods of assessing student learning and abilities—methods that, not incidentally, signal restricted views of what learning is and what cognitive abilities constitute. The assumption that assessment is an objective process, based on measuring quantifiable intellectual abilities, produces an approach that is inflexible and impervious to review. In contrast, MI theory gives us the basis for a more fluid understanding of what constitutes *intelligence,* a more flexible approach to assessing how it is applied in authentic educational settings, and more attention to the values that underlie educators' decisions about assessment. Recognizing that values affect assessment makes it possible to continually examine what is assessed, including its relevance to preparing students to be the citizens of the 21st century. Whether or not one agrees with its assertions, MI theory calls us to be explicit about the subjective nature of the assessment process and to be objective about the adequacy of the current assessment methods used. Nothing less than the future of students' lives depends on our willingness to question the relevance and usefulness of traditional approaches, rather than simply to continue their use because they are accepted.

REFERENCES

Chen, J.-Q. (2004). The Project Spectrum approach to early education. In J. Johnson & J. Roopnarine (Eds.), *Approaches to early childhood education* (4th ed., pp. 251–179). Upper Saddle River, NJ: Pearson.

Chen, J., & Gardner, H. (2005). Assessment based on multiple-intelligences theory. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 77–102). New York: Guilford Press.

Chen, J.-Q., Krechevsky, M., & Viens, J. (1998). *Building on children's strengths: The experience of Project Spectrum.* New York: Teachers College Press.

Chen, J.-Q., & McNamee, G. (2006). Strengthening early childhood teacher preparation: Integrating assessment, curriculum development, and instructional practice in student teaching. *Journal of Early Childhood Teacher Education, 27,* 109–128.

Chen, J.-Q., & McNamee, G. (2007). *Bridging: Assessment for teaching and learning in early childhood classrooms.* Thousand Oaks, CA: Corwin Press.

Chen, J.-Q., & McNamee, G. (2008). From Spectrum to Bridging: Approaches to integrating assessment with curriculum and instruction in early childhood classrooms. In J. Johnson & J. Roopnarine (Eds.), *Approaches to early childhood education* (5th ed., pp. 251–279). Upper Saddle River, NJ: Pearson.

Chen, J.-Q., & McNamee, G. (2011). Positive approaches to learning in the context of preschool classroom activities. *Early Childhood Education Journal, 39*(1), 71–78.

Chen, J.-Q., McNamee, G., & McCray, J. (2011). The learning profile: A construct to understand learning and development of the whole child in content areas. *International Early Learning Journal, 1*(1), 1–24.

Chen, J.-Q., Moran, S., & Gardner, H. (Eds.). (2009). *Multiple intelligences around the world.* San Francisco: Jossey-Bass.

Darling-Hammond, L., & Ancess, L. (1996). Authentic assessment and school development. In J. B. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities (95th yearbook of the National Society for the Study of Education)* (pp. 52–83). Chicago: University of Chicago Press.

Diaz-Lefebvre, R. (2009). What if they learn differently? In J.-Q. Chen, S. Moran, & H. Gardner (Eds.), *Multiple intelligences around the world* (pp. 317–328). San Francisco: Jossey-Bass.

Dweck, C. (2007). *Mindset: The new psychology of success.* New York: Ballantine.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children.* Portsmouth, NH: Heinemann.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, H. (1993a). *Frames of mind: The theory of multiple intelligences* (10th-anniversary ed.). New York: Basic Books.

Gardner, H. (1993b). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H. (1998). The bridges of Spectrum. In J.-Q. Chen, M. Krechevsky, & J. Viens (Eds.), *Building on children's strengths: The experience of Project Spectrum* (pp. 138–145). New York: Teachers College Press.

Gardner, H. (1999). *The disciplined mind.* New York: Simon & Schuster.

Gardner, H. (2000). *The disciplined mind: Beyond facts* *and standardized tests, the K–12 education that every child deserves.* New York: Penguin Books.

Gardner, H. (2006). *Multiple intelligences: New horizons.* New York: Basic Books.

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences* (30th-year ed.). New York: Basic Books.

Kornhaber, M. L., Fierros, E., & Veenema, S. (2004). *Multiple intelligences: Best ideas from research and practice.* New York: Pearson.

Krechevsky, M. (1998). *Project Spectrum preschool assessment handbook.* New York: Teachers College Press.

Lazear, D. (1998). *The rubrics way: Using MI to assess understanding.* Tucson, AZ: Zephyr Press.

Lazear, D. (1999). *Multiple intelligence approaches to assessment: Solving the assessment conundrum.* Tucson, AZ: Zephyr Press.

Levin, H. M. (2005). *Accelerated schools: A decade of evolution.* New York: Springer.

Linn, R. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–15.

Maker, C. J. (2005). *The DISCOVER Project: Improving assessment and curriculum for diverse gifted learners.* Storrs, CT: National Research Center on the Gifted and Talented.

McNamee, G., & Chen, J.-Q. (2005). Dissolving the line between assessment and teaching. *Educational Leadership, 63*(3), 72–77.

New City School. (1994). *Multiple intelligences: Teaching for success.* St. Louis, MO: Author.

Niccolini, P., Alessandri, G., & Bilancioni, G. (2010). Web-Ob for multiple intelligences observation. *Procedia—Social and Behavioral Sciences, 2,* 728–732.

Proulx-Schirduan, V., Shearer, C. B., & Case, K. I. (2009). *Mindful education for ADHD students: Differentiating curriculum and instruction using multiple intelligences.* New York: Teachers College Press.

Rinaldi, C. (2001). Introduction. In Project Zero & Reggio Children, *Making learning visible: Children as individual and group learners* (pp. 28–31). Reggio Emilia, Italy: Reggio Children.

Shearer, C. B. (2005). *Large scale factor analysis of the Multiple Intelligences Developmental Assessment Scales.* Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada.

Shearer, C. B. (2007). *The MIDAS: Professional manual* (rev. ed.). Kent, OH: MI Research and Consulting.

Sulzby, E. (1985, Summer). Children's emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly, 20*(4), 458–481.

CHAPTER 5

The Triarchic Theory of Successful Intelligence

Robert J. Sternberg

Some people seem always to be getting lost, even if they have a map; other people never seem to get lost, even if they don't have a map (or GPS!). In general, some of us seem to do what we do better than others; as a result, various cultures have created roughly comparable psychological constructs to try to explain, or at least to describe, this fact. The construct Western cultures have created is called *intelligence* (in English or French, or *inteligencia* in Spanish, or *Intelligenz* in German). Whatever the language, it is our way of saying that some of us seem to adapt to the environments we both create and confront better than do others.

There have been numerous approaches to understanding the construct of intelligence, based on somewhat different metaphors for understanding the construct (Sternberg, 1985c, 1986, 1990; see also essays in Sternberg, 2000; Sternberg & Kaufman, 2011). For example, some investigators seek to understand intelligence via what I have referred to as a geographic model, in which intelligence is conceived as a map of the mind. Such researchers have used psychometric tests to uncover the latent factors alleged to underlie intellectual functioning (Cooper, 2015; Willis, Dumont, & Kaufman, 2011). Other investigators have used a computational metaphor, viewing intelligence in much the way they view the symbolic processing of a computer (Conway, Getz, Macnamara, & Engel de Abreu, 2011; Nettelbeck, 2011). Still others have followed an anthropological approach, viewing intelligence as a unique cultural creation (Ang, Van Dyne, & Tan, 2011; Saklofske, van de Vijver, Oakland, Mpofu, & Suzuki, 2015). The approach I take in the triarchic theory proposed here can be viewed as a systems approach, in which many different aspects of intelligence are interrelated to each other in an attempt to understand how intelligence functions as a system.

The *triarchic theory of successful intelligence* (Sternberg, 1984, 1985a, 1988, 1997a, 1999, 2004, 2005b, 2005c, 2008, 2012b, 2015; Sternberg, Jarvin, & Grigorenko, 2011) explains in an integrative way the relationship between intelligence and (1) the internal world of the individual, or the mental mechanisms that underlie intelligent behavior; (2) experience, or the mediating role of the individual's passage through life between his or her internal and external worlds; and (3) the external world of the individual, or the use of these mental mechanisms in everyday life in order to attain an intelligent fit to the environment. The theory has three subtheories, one corresponding to each of the three relationships mentioned in the preceding sentence.

A crucial difference between this theory and many others is that the operationalizations (measurements) follow rather than precede the theory. Thus, rather than the theory's being derived from factor or other analyses of tests, the tests are chosen on the basis of the tenets of the theory. My colleagues and I have used many different kinds of tests (see Sternberg, 1985a, 1988, 2003a, 2003b, 2003c for reviews), such as analogies, syllogisms, verbal comprehension, prediction of future outcomes, and decoding of nonverbal cues. In every case, though, the choice of tasks has been dictated by the aspects of the theory that are being investigated, rather than the other way around.

DEFINITION OF SUCCESSFUL INTELLIGENCE

According to the proposed theory, *successful intelligence* is (1) the use of an integrated set of skills needed to attain success in life, however an individual defines it, within his or her sociocultural context. People are successfully intelligent by virtue of (2) recognizing their strengths and making the most of them, at the same time that they recognize their weaknesses and find ways to correct or compensate for them. Successfully intelligent people (3) adapt to, shape, and select environments through (4) finding a balance in their use of analytical, creative, practical, and wisdom-based skills (Sternberg, 1985e, 1997a, 1997b, 1997c, 1999, 2002, 2003c). Let us consider each element of the theory in turn.

According to the first element, there is no one definition of success that works for everyone. For some people, success is brilliance as lawyers; for others, it is originality as novelists; for others, it is caring for their children; for others, it is devoting their lives to God. For many people, it is some combination of things. Because people have different life goals, education needs to move away from single targeted measures of success, such as grade point average (GPA) or standardized test scores. Education should be geared toward the goals of each individual, rather than toward one predefined goal that may be relevant to some students but not to many others.

In considering the nature of intelligence, we need to consider the full range of definitions of success by which children can be intelligent. For example, in research we have done in rural Kenya (Sternberg et al., 2001), we have found that children who may score quite high on tests of an aspect of practical intelligence—knowledge of how to use natural herbal medicines to treat parasitic and other illnesses—may score quite poorly on tests of IQ and academic achievement. Indeed, we found an inverse relationship between the two skill sets, with correlations reaching the –.30 level. For these children, time spent in school takes away from time in which they learn the practical skills that they and their families view as needed for success in life. The same might be said, in the Western world, for many children who want to enter careers in athletics, theater, dance, art, music, carpentry, plumbing, entrepreneurship, and so forth. They may see time spent developing academic skills as time taken away from the time they need to develop practical skills relevant to meeting their goals in life.

The second element asserts that there are different paths to success, no matter what goal one chooses. Some people achieve success in large part through personal charm; others through brilliance of academic intellect; others through stunning originality; and yet others through working extremely hard. For most of us, there are at least a few things we do well, and our successful intelligence is dependent in large part upon making these things “work for us.” At the same time, we need to acknowledge our weaknesses and find ways either to improve upon them or to compensate for them. For example, we may work hard to improve our skills in an area of weakness, or work as part of a team so that other people compensate for the kinds of things we do not do particularly well.

The third element asserts that success in life is achieved through some balance of adapting to existing environments, shaping those environments, and selecting new environments. Often when we go into an environment—as do students and teachers in school—we try to modify ourselves to fit that environment. In other words, we adapt. But sometimes it is not enough to adapt. We are not content merely to change ourselves to fit the environment; rather, we also want to change the environment to fit us. In this case, we shape the environment in order to make it a better one for us, and possibly for others as well. But there may come times when our attempts to adapt and to shape the environment lead us nowhere—when we simply cannot find a way to make the environment work for us. In these cases, we leave the old environment and select a new environment. Sometimes the smart thing is to know when to get out.

Finally, we balance four kinds of skills in order to achieve these ends: analytical skills, creative skills, practical skills, and—in the augmented version of the theory (Sternberg, 2015)—wisdom-based skills. We need creative skills to generate ideas, analytical skills to determine whether they are good ideas, practical skills to implement the ideas and to convince others of the value of our ideas, and wisdom-based skills to help achieve a common good that goes beyond just our own self-interest. Most people who are successfully intelligent are not equally endowed with these diverse skills, but they find ways of making the four skills work harmoniously together.

We have used five kinds of converging operations to test the theory of successful intelligence: cultural studies, factor-analytic studies, information-processing analyses, correlational analyses, and instructional studies (some of which are described below). Much of this work is summarized elsewhere (e.g., Sternberg, 1985a, 1997b, 2003a, 2003c, 2015). Examples of kinds of evidence in this work supporting the theory are the factorial separability of analytical, creative, and practical skills; the substantial incremental validity of measures of practical intelligence over the validity of measures of academic (general) intelligence in predicting school and job performance; the usefulness of instruction based on the theory of successful intelligence, in comparison with other forms of instruction; and differences in the nature of what constitutes practical intelligence across cultures (see Sternberg et al., 2000; Sternberg & Hedlund, 2002; Sternberg & Smith, 1985; Wagner, 2011).

INTELLIGENCE AND THE INTERNAL WORLD OF THE INDIVIDUAL

Psychometricians, Piagetians, and information-processing psychologists have all recognized the importance of understanding the mental states or processes that underlie intelligent thought. In the triarchic theory of successful intelligence, they seek this understanding by identifying and understanding three basic kinds of information-processing components, referred to as *metacomponents*, *performance components*, and *knowledge acquisition components*.

Metacomponents

Metacomponents are higher-order, executive processes used to plan what one is going to do, to monitor it while one is doing it, and evaluate it after it is done. These metacomponents include (1) recognizing the existence of a problem, (2) deciding on the nature of the problem confronting one, (3) selecting a set of lower-order processes to solve the problem, (4) selecting a strategy into which to combine these components, (5) selecting a mental representation on which the components and strategy can act, (6) allocating one's mental resources, (7) monitoring one's problem solving as it is happening, and (8) evaluating one's problem solving after it is done. Let us consider some examples of these higher-order processes (see Sternberg & Smith, 1988, for further details on executive processes).

Deciding on the nature of a problem plays a prominent role in intelligence. For example, the difficulty for young children as well as older adults in problem solving often lies not in actually solving a given problem, but in figuring out just what the problem is that needs to be solved (see, e.g., Flavell, 1977; Sternberg & Rifkin, 1979). A major feature distinguishing people with intellectual disabilities from persons with typical functioning is the need of the former to be instructed explicitly and completely as to the nature of the particular task they are solving and how it should be performed (Butterfield, Wambold, & Belmont, 1973; Campione & Brown, 1979; Hodapp, Griffin, Burke, & Fisher, 2011). The disability is probably partially genetic (Lombroso, Sternberg, & Grigorenko, 1999).

Selection of a strategy for combining lower-order components is also a critical aspect of intelligence. In early information-processing research on intelligence, including my own (e.g., Sternberg, 1977, 1985b; Sternberg & Weil, 1980), the primary emphasis was simply on figuring out what study participants do when confronted with a problem. What components do participants use, and into what strategies do they combine these components?

Soon information-processing researchers began to ask why study participants use the strategies they choose. For example, Siegler (1986) proposed a model of strategy selection in arithmetic computation problems that links strategy choice to both the rules and the mental associations participants have stored in long-term memory (see also Siegler, 2007). MacLeod, Hunt, and Mathews (1978) found that study participants with high spatial abilities tend to use a spatial strategy in solving sentence–picture comparison problems, whereas study participants with high verbal abilities are more likely to use a linguistic strategy.

In my own work, I have found that study participants tend to prefer strategies for analogical reasoning that place fewer demands on working memory (Sternberg & Ketron, 1982). They learn better when taught in ways that fit their preferred strategies (Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999). In such strategies, study participants encode as few features as possible of complex stimuli, trying to disconfirm incorrect multiple-choice options on the basis of these few features, and then choosing the remaining answer as the correct one. Similarly, study participants choose different strategies in linear–syllogistic reasoning (spatial, linguistic, mixed spatial–linguistic), but in this task, they do not always capitalize on their ability patterns to choose the strategy most suitable to their respective levels of spatial and verbal abilities (Sternberg & Weil, 1980). In sum, the selection of a strategy seems to be at least as important for understanding intelligent task performance as the efficacy with which the chosen strategy is implemented.

Intimately tied up with the selection of a strategy is the selection of a mental representation for information. Investigators have realized that people are quite flexible in their representations of information. The most appropriate question to ask seems to be not how such information is represented, but which representations are used in what circumstances. For example, I (Sternberg, 1977) found that analogy problems using animal names can draw on either spatial or clustering representations of the animal names. In the studies of strategy choice mentioned earlier, it was found that study participants can use either linguistic or spatial representations in solving sentence–picture comparisons (MacLeod et al., 1978) or linear syllogisms (Sternberg & Weil, 1980). We (Sternberg & Rifkin, 1979) found that the mental representation of certain kinds of analogies can be either more or less holistic, depending on the ages of the study participants. Younger children tend to be more holistic in their representations.

As important as any other metacomponent is the ability to allocate one's mental resources. Different investigators have studied resource allocation in different ways. I have found that better problem solvers tend to spend relatively more time in global strategy planning (Sternberg, 1981b, 1981c). Similarly, in solving analogies, better analogical reasoners seem to spend relatively more time encoding the terms of the problem than do poorer reasoners, but relatively less time in operating on these encodings (Sternberg, 1977; Sternberg & Rifkin, 1979). In reading as well, superior readers are better able than poorer readers to allocate their time across reading passages as a function of the difficulty of the passages to be read and the purpose for which the passages are being read (see Brown, Bransford, Ferrara, & Campione, 1983; Spear-Swerling & Sternberg, 1994; Wagner & Sternberg, 1987).

Finally, monitoring one's thinking and solution processes is a key aspect of intelligence (see also Brown, 1978; Conway et al., 2011; Otero, 2015). Consider, for example, the "missionaries and cannibals" problem, in which the study participants must "transport" a set of missionaries and cannibals across a river in a small boat without allowing the cannibals an opportunity to eat the missionaries—an event that can transpire only if the cannibals are allowed to outnumber the missionaries on either side of the river bank. The main kinds of errors that can be made are either to return to an earlier state in the problem space for solution (i.e., the problem solver goes back to where he or she was earlier in the solution process) or to make an impermissible move (i.e., the problem solver violates the rules, as in allowing the number of cannibals on one side to exceed the number of missionaries on that side) (Simon & Reed, 1976; see also Sternberg, 1982). Neither of these errors will result if a given participant closely monitors his or her solution processes. For young children, learning to count, a major source of errors in counting objects is to count a given object twice; again, such errors can result from failures in solution monitoring (Gelman & Gallistel, 1978). The effects of solution monitoring are not limited, of course, to any one kind of problem. One's ability to use the strategy of means–ends analysis (Newell & Simon, 1972)—that is, reduction of differences between where one is solving a problem and where one wishes to get in solving that problem—depends on one's ability to monitor just where one is in problem solution.

Performance Components

Performance components are lower-order processes that execute the instructions of the metacomponents. These lower-order components solve the problems according to the plans laid out by the metacomponents. Whereas the number of metacomponents used in the performance of various tasks is relatively limited, the number of performance components is probably quite large. Many of these performance components are relatively specific to narrow ranges of tasks (Sternberg, 1979, 1983, 1985a).

One of the most interesting classes of performance components is that found in inductive reasoning of the kind measured by tests such as matrices, analogies, series completions, and classifications. These components are important because of the importance of the tasks into which they enter: Induction problems of these kinds show the highest loading on the so-called *g*, or general intelligence factor (Jensen, 1980, 1998; Ortiz, 2015; Snow & Lohman, 1984; Sternberg & Gardner, 1982; see essays in Sternberg & Grigorenko, 2002). Thus identifying these performance components can give us some insight into the nature of the general factor. I am not arguing for any one factorial model of intelligence (i.e., one with a general factor) over others; on the contrary, I believe that most factor models are mutually compatible, differing only in the form of rotation that has been applied to a given factor space (Sternberg, 1977). The rotation one uses is a matter of theoretical or practical convenience, not of truth or falsity.

Knowledge Acquisition Components

Knowledge acquisition components are used to learn how to do what the metacomponents and performance components eventually do. Three knowledge acquisition components appear to be central in intellectual functioning: (1) selective encoding, (2) selective combination, and (3) selective comparison.

Selective encoding involves sifting out relevant from irrelevant information. When new information is presented in natural contexts, relevant information for one's given purpose is embedded in the midst of large amounts of purpose-irrelevant information. A critical task for the learner is that of "sifting the wheat from the chaff," or recognizing just what among all the pieces of information is relevant for one's purposes (see Schank, 1980).

Selective combination involves combining selectively encoded information in such a way as to form an integrated, plausible whole. Simply sifting out relevant from irrelevant information is not enough to generate a new knowledge structure. One must know how to combine the pieces of information into an internally connected whole (see Lohman & Lakin, 2011; Mayer & Greeno, 1972).

Selective comparison involves discovering a nonobvious relationship between new information and already acquired information. For example, analogies, metaphors, and models often help individuals solve problems. The solver suddenly realizes that new information is similar to old information in certain ways, and then uses this information to form a mental representation based on the similarities. Teachers may discover how to relate new classroom material to information that students have already learned. Relating the new to the old can help students learn the material more quickly and understand it more deeply.

My emphasis on components of knowledge acquisition differs somewhat from the focus of some theorists in cognitive psychology, who emphasize what is already known and the structure of this knowledge (e.g., Chase & Simon, 1973; Chi, 1978; Keil, 1984). These various emphases are complementary. If one is interested in understanding, for example, differences in performance between experts and novices, clearly one would wish to look at the amount and structure of their respective knowledge bases. But if one wishes to understand how these differences come to be, merely looking at developed knowledge would not be enough. Rather, one would have to look as well at differences in the ways in which the knowledge bases were acquired. It is here that understanding of knowledge acquisition components will prove to be most relevant.

We have studied knowledge acquisition components in the domain of vocabulary acquisition (e.g., Sternberg, 1987; Sternberg & Powell, 1983). Difficulty in learning new words can be traced, at least in part, to the application of components of knowledge acquisition to context cues stored in long-term memory. Individuals with higher vocabularies tend to be those who are better able to apply the knowledge acquisition components to vocabulary-learning situations. Given the importance of vocabulary for overall intelligence, almost without respect to the theory or test one uses, utilization of knowledge acquisition components in vocabulary-learning situations would appear to be critically important for the development of intelligence.

Effective use of knowledge acquisition components is trainable. I have found, for example, that just 45 minutes of training in the use of these components in vocabulary learning can significantly and fairly substantially improve the ability of adults to learn vocabulary from natural language contexts (Sternberg, 1987). This training involves teaching individuals how to learn meanings of words presented in context. The training consists of three elements. The first is teaching individuals to search out certain kinds of contextual cues, such as synonyms, antonyms, functions, and category memberships. The second is teaching mediating variables. For example, cues to the meaning of a word are more likely to be found close to the word than at a distance from it. The third is teaching process skills—encoding relevant cues, combining them, and relating them to knowledge one already has.

To summarize, then, the components of intelligence are important parts of an individual's intelligence. The various kinds of components work together. Metacomponents activate performance and knowledge acquisition components. These latter kinds of components in turn provide feedback to the metacomponents. Although one can isolate various kinds of information-processing components from task performance through experimental means, in practice the components function together in highly interactive, and not easily isolable, ways. Thus diagnoses as well as instructional interventions need to consider all three types of components in interaction, rather than any one kind of component in isolation. But understanding the nature of the components of intelligence is not in itself sufficient to understand the nature of intelligence because there is more to intelligence than a set of information-processing components. One could scarcely understand all of what it is that makes one person more intelligent than another by understanding the components of processing on, say, an intelligence test. The other aspects of the triarchic theory address some of the other aspects of intelligence that contribute to individual differences in observed performance, outside testing situations as well as within them.

INTELLIGENCE AND EXPERIENCE

Components of information processing are always applied to tasks and situations with which one has some level of prior experience (even if it is minimal experience). Hence these internal mechanisms are closely tied to one's experience. According to the experiential subtheory, the components are not equally good measures of intelligence at all levels of experience. Assessing intelligence requires one to consider not only components, but the level of experience at which they are applied.

According to the experiential subtheory, intelligence is best measured at those regions of the experiential continuum involving tasks or situations that are either relatively novel on the one hand, or in the process of becoming automatized on the other. Totally novel tasks and situations provide poor measures of intelligence: One would not want to administer, say, trigonometry problems to a first grader roughly 6 years old. But one might wish to administer problems that are just at the limits of the child's understanding, in order to test how far this understanding extends. Related is Vygotsky's (1978) concept of the *zone of proximal development,* in which one examines a child's ability to profit from instruction to facilitate his or her solutions of novel problems. To measure automatization skill, one might wish to present a series of problems—mathematical or otherwise—to see how long it takes for their solution to become automatic, and to see how automatized performance becomes. Thus both the slope and the asymptote (if any) of automatization are of interest.

Ability to Deal with Novelty

Several sources of evidence converge on the notion that the ability to deal with relative novelty is a good way of measuring intelligence. Consider three such sources of evidence. First, we have conducted several studies on the nature of insight, both in children and in adults (Davidson & Sternberg, 1984; Sternberg & Davidson, 1982). In the studies with children (Davidson & Sternberg, 1984), we separated three kinds of insights: insights of selective encoding, insights of selective combination, and insights of selective comparison. Use of these knowledge acquisition components is referred to as *insightful* when they are applied in the absence of existing scripts, plans, or frames. In other words, one must decide what information is relevant, how to put the information together, or how new information relates to old, in the absence of any obvious cues on the basis of which to make these judgments. A problem is insightfully solved at the individual level when a given individual lacks such cues (Sternberg & Zhang, 1995). A problem is insightfully solved at the societal level when no one else has these cues, either. In our studies, we found that children who are intellectually gifted are so in part by virtue of their insight abilities, which represent an important part of the ability to deal with novelty (Sternberg, 1981a; Sternberg et al., 2011).

The critical finding was that providing insights to the children significantly benefited the nongifted, but not the gifted, children. (None of the children performed anywhere near ceiling level, so that the interaction was not due to ceiling effects.) In other words, the gifted children spontaneously had the insights and hence did not benefit from being given these insights. The nongifted children did not have the insights spontaneously and hence did benefit. Thus the gifted children were better able to deal with novelty spontaneously.

Ability to Automatize Information Processing

Several converging lines of evidence in the literature support the claim that automatization ability is a key aspect of intelligence. For example, I (Sternberg, 1977) found that the correlation between people–piece (schematic picture) analogy performance and measures of general intelligence increased with practice, as performance on these items became increasingly automatized. Skilled reading is heavily dependent on automatization of bottom-up functions (basic skills such as phonetic decoding), and the ability to read well is an essential part of crystallized ability—whether it is viewed from the standpoint of theories such as Cattell's (1971), Carroll's (1993), or Vernon's (1971), or from the standpoint of tests of crystallized ability, such as the verbal portion of the SAT. Poor comprehenders often are those who have not automatized the elementary, bottom-up processes of reading and hence do not have sufficient attentional resources to allocate to top-down comprehension processes.

Theorists such as Jensen (1982) and Hunt (1978, 1980) have attributed the correlation between such tasks as choice reaction time and letter matching to the relation between speed of information processing and intelligence (see Nettelbeck, 2011). Indeed, there is almost certainly some relation, although I believe it is much more complex than these theorists seem to allow for. But a plausible alternative hypothesis is that at least some of that correlation is due to the effects of automatization of processing: Because of the simplicity of these tasks, they probably become at least partially automatized fairly rapidly, and hence can measure both rate and asymptote of automatization of performance. In sum, then, although the evidence is far from complete, there is at least some support for the notion that rate and level of automatization are related to intellectual skill.

The ability to deal with novelty and the ability to automatize information processing are interrelated, as shown in the example of the automatization of reading described in this section. If one is well able to automatize, one has more resources left over for dealing with novelty. Similarly, if one is well able to deal with novelty, one has more resources left over for automatization. Thus performances at the various levels of the experiential continuum are related to one another.

These abilities should not be viewed in a vacuum with respect to the componential subtheory. The components of intelligence are applied to tasks and situations at various levels of experience. The ability to deal with novelty can be understood in part in terms of the metacomponents, performance components, and knowledge acquisition components involved in it. *Automatization* refers to the way these components are executed. Hence the two subtheories considered so far are closely intertwined. Now we need to consider the application of these subtheories to everyday tasks, in addition to laboratory ones.

INTELLIGENCE AND THE EXTERNAL WORLD OF THE INDIVIDUAL

According to the contextual subtheory, intelligent thought is directed toward one or more of three behavioral goals: *adaptation to an environment, shaping of an environment,* or *selection of an environment.* These three goals may be viewed as the functions toward which intelligence is directed. Intelligence is not aimless or random mental activity that happens to involve certain components of information processing at certain levels of experience. Rather, it is purposefully directed toward the pursuit of these three global goals, all of which have more specific and concrete instantiations in people's lives (Cianciolo & Sternberg, 2018; Sternberg et al., 2000).

Adaptation

Most intelligent thought is directed toward attempts to adapt to one's environment. The requirements for adaptation can differ radically from one environment to another—whether environments are defined in terms of families, jobs, subcultures, or cultures. Hence, although the components of intelligence required in these various contexts may be the same or quite similar, and although all of them may involve (at one time or another) dealing with novelty and automatization of information processing, the concrete instantiations that these processes and levels of experience take may differ substantially across contexts. This fact has an important implication for our understanding of the nature of intelligence. According to the triarchic theory in general, and the contextual subtheory in particular, the processes, experiential facets, and functions of intelligence remain essentially the same across contexts, but the particular instantiations of these processes, facets, and functions can differ radically. Thus the content of intelligent thought and its manifestations in behavior will bear no necessary resemblance across contexts. As a result, although the mental elements that an intelligence test should measure do not differ across contexts, the vehicle for measurement may have to differ. A test that measures a set of processes, experiential facets, or intelligent functions in one context may not provide equally adequate measurement in another context. To the contrary, what is intelligent in one culture may be viewed as unintelligent in another.

Different contextual milieus may result in the development of different mental abilities. For example, Puluwat navigators must develop their large-scale spatial abilities for dealing with cognitive maps to a degree that far exceeds the adaptive requirements of contemporary Western societies (Gladwin, 1970). Similarly, Kearins (1981) found that Australian Aboriginal children probably develop their visual–spatial memories to a greater degree than do Australian children of European descent. The latter are more likely to apply verbal strategies to spatial memory tasks than are the Aboriginal children, who employ spatial strategies. This greater development is presumed to be due to the greater need the Aboriginal children have for using spatial skills in their everyday lives. In contrast, members of Western societies probably develop their abilities for thinking abstractly to a greater degree than do members of societies in which concepts are rarely dealt with outside their concrete manifestations in the objects of the everyday environment.

One of the most interesting differences among cultures and subcultures in the development of patterns of adaptation is in the matter of time allocation, a metacomponential function. In Western cultures in general, careful allocation of time to various activities is a prized commodity. Our lives are largely governed by careful scheduling at home, school, work, and so on. There are fixed hours for certain activities, and fixed lengths of time within which these activities are expected to be completed. Indeed, the intelligence tests we use show our prizing of time allocation to the fullest. Almost all of them are timed in such a way as to make completion of the tests a nontrivial challenge. A slow or cautious worker is at a distinct disadvantage.

Not all cultures and subcultures view time in the same way that we do. For example, among the Kipsigi, schedules are much more flexible; hence these individuals have difficulty understanding and dealing with Western notions of the time pressure under which people are expected to live (Super & Harkness, 1982). In Hispanic cultures, such as Venezuela, my own personal experience indicates that the press of time is taken with much less seriousness than it is in typical North American cultural settings. Even within the continental United States, though, there can be major differences in the importance of time allocation (Heath, 1983; Suzuki, Short, & Lee, 2011).

The point of these examples has been to illustrate how differences in environmental press and people's conception of what constitutes an intelligent response to it can influence just what counts as adaptive behavior. To understand intelligence, one must understand it not only in relation to its internal manifestations in terms of mental processes and its experiential manifestations in terms of facets of the experiential continuum, but also in terms of how thought is intelligently translated into action in a variety of different contextual settings. The differences in what is considered adaptive and intelligent can extend even to different occupations within a given cultural milieu. For example, I (Sternberg, 1985d) have found that individuals in different fields of endeavor (art, business, philosophy, physics) view intelligence in slightly different ways that reflect the demands of their respective fields.

Shaping

Shaping of the environment is often used as a backup strategy when adaptation fails. If one is unable to change oneself to fit the environment, one may attempt to change the environment to fit oneself. For example, repeated attempts to adjust to the demands of one's romantic partner may eventually lead to attempts to get the partner to adjust to oneself. But shaping is not always used in lieu of adaptation. In some cases, shaping may be used before adaptation is ever tried, as in the case of the individual who attempts to shape a romantic partner with little or no effort to shape him- or herself so as to suit the partner's wants or needs better.

In the laboratory, examples of shaping behavior can be seen in strategy selection situations where one essentially molds the task to fit one's preferred style of dealing with tasks. For example, in comparing sentence statements, individuals may select either a verbal or a spatial strategy, depending on their pattern of verbal and spatial ability (MacLeod et al., 1978). The task is "made over" in conformity to what they do best.

Selection

Selection involves renunciation of one environment in favor of another. In terms of the rough hierarchy established so far, selection is sometimes used when both adaptation and shaping fail. Sometimes one attempts to shape an environment only after attempts to leave it have failed. Other times, one may decide almost instantly that an environment is simply wrong and feel that one need not or should not even try to fit into or to change it. For example, every now and then we get a new graduate student who realizes almost immediately that he or she came to graduate school for the wrong reasons, or who finds that graduate school is nothing at all like the continuation of undergraduate school he or she expected. In such cases, the intelligent thing to do may be to leave the environment as soon as possible, to pursue activities more in line with the student's goals in life.

To conclude, adaptation, shaping, and selection are functions of intelligent thought as it operates in context. They may (although they need not) be employed hierarchically, with one path followed when another one fails. It is through adaptation, shaping, and selection that the components of intelligence, as employed at various levels of experience, become actualized in the real world. In this section, it has become clear that the modes of actualization can differ widely across individuals and groups, so that intelligence cannot be understood independently of the ways in which it is manifested.

INSTRUCTIONAL INTERVENTIONS BASED ON THE THEORY

The triarchic theory has been applied to instructional settings in various ways, with considerable but not total success (Sternberg, 2010b; Sternberg & Grigorenko, 2004, 2007; Sternberg, Grigorenko, & Zhang, 2008; Sternberg, Jarvin, & Grigorenko, 2009; Sternberg et al., 2014). The componential subtheory has been applied in teaching the learning of vocabulary from context to adult study participants (Sternberg, 1987), as mentioned earlier. Experimental study participants were taught components of decontextualization. There were three groups, corresponding to three types of instruction that were based on the theory (see Sternberg, 1987, 1988). Control study participants either received no relevant material at all, or else received practical items but without theory-based instruction. Improvement occurred only when study participants were given the theory-based instruction, which involved teaching them how to use contextual cues, mediating variables such as matching parts of speech, and processes of decontextualization.

The experiential subtheory was the basis for the program (Davidson & Sternberg, 1984) that successfully taught insight skills (selective encoding, selective combination, and selective comparison) to children roughly 9–11 years of age. The program lasted 6 weeks and involved insight skills as applied to a variety of subject matter areas. An uninstructed control group received a pretest and a posttest, like the experimental group, but no instruction. We found that the experimental study participants improved significantly more than the controls, both when participants were previously identified as gifted and when they were not so identified. Moreover, we found durable results that lasted even 1 year after the training program, and we found transfer to types of insight problems not specifically used in the program.

The contextual subtheory served as the basis for a program called Practical Intelligence for Schools, developed in collaboration with a team of investigators from Harvard (Gardner, Krechevsky, Sternberg, & Okagaki, 1994; Okagaki & Sternberg, 1993; Sternberg, Okagaki, & Jackson, 1990) and based on Gardner's (2006) theory of multiple intelligences as well as on the triarchic theory. The goal of this program is to teach practical intellectual skills to children roughly 9–11 years of age in the areas of reading, writing, homework, and test taking. The program is completely infused into existing curricula. Over a period of years, we studied the program in a variety of school districts and obtained significant improvements for experimental versus uninstructed control study participants in a variety of criterion measures, including study skills measures and performance-based measures of performance in the areas taught by the program. The program has been shown to increase practical skills, such as those involved in doing homework, taking tests, or writing papers, as well as school achievement (Williams et al., 2002; see Detterman & Sternberg, 1982, for a discussion of attempts to increase intelligence).

We have sought to test the theory of successful intelligence in the classroom. In a first set of studies, we explored the question of whether conventional education in school systematically discriminates against children with creative and practical strengths (Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg et al., 1999). Motivating this work was the belief that the systems in most schools strongly tend to favor children with strengths in memory and analytical abilities. However, schools can be unbalanced in other directions as well.

One school we visited in Russia in 2000 placed a heavy emphasis on the development of creative abilities—much more so than on the development of analytical and practical abilities. While on this trip, we were told of yet another school (catering to the children of Russian businessmen) that strongly emphasized practical abilities, and in which children who were not practically oriented were told that eventually they would be working for their classmates who were practically oriented.

To validate the relevance of the theory of successful intelligence in classrooms, we have carried out a number of instructional studies. In one study, we used the Sternberg Triarchic Abilities Test (Sternberg, 1993). The test was administered to 326 children around the United States and in some other countries who were identified by their schools as gifted by any standard whatsoever (Sternberg et al., 1999). Children were selected for a summer program in (college-level) psychology if they fell into one of five ability groupings: *high-analytical, high-creative, high-practical, high-balanced* (high in all three abilities), or *low-balanced* (low in all three abilities). Students who came to Yale were then assigned at random to four instructional groups, with the constraint that roughly equal numbers with each ability pattern be assigned to each group. Students in all four instructional groups used the same introductory psychology textbook (a preliminary version of Sternberg, 1995) and listened to the same psychology lectures. What differed among them was the type of afternoon discussion section to which they were assigned. They were assigned to an instructional condition that emphasized either memory, analytical, creative, or practical instruction. For example, in the memory condition, they might be asked to describe the main tenets of a major theory of depression. In the analytical condition, they might be asked to compare and contrast two theories of depression. In the creative condition, they might be asked to formulate their own theory of depression. In the practical condition, they might be asked how they could use what they had learned about depression to help a friend who was depressed.

Students in all four instructional conditions were evaluated in terms of their performance on homework, a midterm exam, a final exam, and an independent project. Each type of work was evaluated for memory, analytical, creative, and practical quality. Thus all students were evaluated in exactly the same way. Our results suggested the utility of the theory of successful intelligence. This utility showed itself in several ways.

First, we observed when the students arrived at Yale that the students in the high-creative and high-practical groups were much more diverse in terms of racial, ethnic, socioeconomic, and educational backgrounds than were the students in the high-analytical group, suggesting that correlations of measured intelligence with status variables such as these may be reduced by using a broader conception of intelligence. Thus the kinds of students identified as strong differed in terms of the populations from which they were drawn, in comparison with students identified as strong solely by analytical measures. More importantly, just by expanding the range of abilities measured, we discovered intellectual strengths that might not have been apparent through a conventional test.

Second, we found that all three ability tests—analytical, creative, and practical—significantly predicted course performance. When multiple-regression analysis was used, at least two of these ability measures contributed significantly to the prediction of each of the measures of achievement. In particular, for homework assignments, significant beta weights were obtained for analytical (.25) and creative (.16) ability measures; for the independent project, significant weights were obtained for the analytical (.14), creative (.22), and practical (.14) measures; for the exams, significant weights were obtained for the analytical (.24) and creative (.19) measures (Sternberg et al., 1999). Perhaps as a reflection of the difficulty of deemphasizing the analytical way of teaching, one of the significant predictors was always the analytical score. (However, in a replication of our study with low-income African American students from New York, Deborah Coates of the City University of New York found a different pattern of results. Her data indicated that the practical tests were better predictors of course performance than were the analytical measures, suggesting that which ability test predicts which criterion depends on population as well as mode of teaching.)

Third and most important, there was an aptitude–treatment interaction, whereby students who were placed in instructional conditions that better matched their pattern of abilities outperformed students who were mismatched. In particular, repeated-measures analysis revealed statistically significant effects of match for analytical and creative tasks as a whole. Three of five practical tasks also showed an effect. In other words, when students are taught in a way that fits how they think, they do better in school (see Cronbach & Snow, 1977, for a discussion of the difficulties in eliciting aptitude–treatment interactions). Children who have high levels of creative and practical abilities, but who are almost never taught or assessed in a way that matches their pattern of abilities, may be at a disadvantage in course after course, year after year.

A follow-up study (Sternberg, Torff, & Grigorenko, 1998) examined learning of social studies and science by third graders and eighth graders. The 225 third graders were students in a very low-income neighborhood in Raleigh, North Carolina. The 142 eighth graders were largely middle- to upper-middle-class students studying in Baltimore, Maryland, and Fresno, California; these children were part of a summer program sponsored by the Johns Hopkins University for gifted students. In this study, students were assigned to one of three instructional conditions. Randomization was by classroom. In the first condition, they were taught the course that basically they would have received had there been no intervention. The emphasis in the course was on memory. In a second condition, students were taught in a way that emphasized critical (analytical) thinking. In the third condition, they were taught in a way that emphasized analytical, creative, and practical thinking. All students' performance was assessed for memory learning (through multiple-choice assessments), as well as for analytical, creative, and practical learning (through performance assessments).

As expected, students in the successful-intelligence (analytical, creative, practical) condition outperformed the other students in terms of the performance assessments. For the third graders, respective means were highest for the triarchic (successful-intelligence) condition; second highest for the critical-thinking condition; and lowest for the memory condition for memory, analytical, and creative performance measures. For practical measures, the critical-thinking mean was insignificantly higher than the triarchic mean, but both were significantly higher than the memory mean. For the eighth graders, the results were similar. One could argue that this pattern of results merely reflected the way students were taught. Nevertheless, the result suggested that teaching for these kinds of thinking succeeded. More important, however, was the result that children in the successful-intelligence condition outperformed the other children even on the multiple-choice memory tests. In other words, to the extent that the goal is just to maximize children's memory for information, teaching for successful intelligence is still superior. It enables children to capitalize on their strengths and to correct or to compensate for their weaknesses, and it allows them to encode material in a variety of interesting ways.

We extended these results to reading curricula at the middle school and high school levels (Grigorenko, Jarvin, & Sternberg, 2002). In a study of 871 middle school students and 432 high school students, we taught reading either triarchically or through the regular curriculum. Classrooms were assigned randomly to treatments. At the middle school level, reading was taught explicitly. At the high school level, reading was infused into instruction in mathematics, physical sciences, social sciences, English, history, foreign languages, and the arts. In all settings, students who were taught triarchically substantially outperformed students who were taught in standard ways. Effects were statistically significant at the .001 level for memory, analytical, creative, and practical comparisons.

Thus the results of three sets of studies suggest that the theory of successful intelligence is valid as a whole. Moreover, the results suggest that the theory can make a difference not only in laboratory tests, but in school classrooms and even the everyday lives of adults as well. At the same time, the studies have weaknesses that need to be remedied in future studies. The samples were relatively small and not fully representative of the entire U.S. population. Moreover, the studies have examined a limited number of alternative interventions. All interventions were of relatively short duration (up to a semester-long course). In addition, future studies should look at durability and transfer of training.

In one study (Sternberg et al., 2014), we attempted to upscale our efforts to hundreds of teachers and many thousands of fourth-grade students located in elementary schools across the United States. The results were not particularly encouraging. Although triarchic instruction was better than alternative instruction in some conditions, it was not better in other conditions. When one obtains null results, of course there can be many interpretations. But one thing was clear: Our small staff lost control of fidelity of implementation. That is, we did not always have good control over how well the programs were implemented. Unfortunately, we had a similar experience in an upscaled teaching-for-wisdom study (Sternberg, Jarvin, & Reznitskaya, 2008; Sternberg, Reznitskaya, & Jarvin, 2007). The bottom line is that if one is seeking to implement an intervention program on a large scale, one needs sufficient monitors of the implementation to ensure fidelity to the pedagogical principles of the program.

In sum, the triarchic theory serves as a useful basis for educational interventions; in our own work, it has shown itself to be a basis for interventions that improve students' performance relative to that of controls who do not receive the theory-based instruction.

ASSESSMENT STUDIES

One of the primary venues for assessing abilities is university admissions. When universities make decisions about selective admissions, the main quantitative data they have available to them are typically (1) GPA in high school or its equivalent, and (2) scores on standardized tests (Lemann, 2000). Is it possible to create assessments that are psychometrically sound and that provide incremental validity over existing measures, without destroying the cultural and ethnic diversity that makes a university environment a place in which students can interact with and learn from others who are different from themselves?

The Rainbow Project

The Rainbow Project (for details, see Sternberg, 2009, 2010a; Sternberg, Bonney, Gabora, Karelitz, & Coffin, 2010; Sternberg, Bonney, Gabora, & Merrifield, 2012; Sternberg & the Rainbow Project Collaborators, 2005, 2006) was a first project designed to enhance university admissions procedures at the undergraduate level. The Rainbow measures were intended to supplement the SAT in the United States, but they can supplement any conventional standardized test of abilities or achievement. In the theory of successful intelligence, abilities and achievement are viewed as being on a continuum—abilities are largely achieved (Sternberg, 1998, 1999)—so it is not clear that it matters greatly exactly what test is used, given that most of the tests used are highly *g*-loaded.

The SAT is a comprehensive examination currently measuring verbal comprehension and mathematical thinking skills, with a writing component recently added. A wide variety of studies have shown the utility of the SAT and similar tests as predictors of university and job success, with success in college typically measured by GPA (Schmidt & Hunter, 1998). Taken together, these data suggest reasonable predictive validity for the SAT in predicting undergraduate performance. Indeed, traditional intelligence or aptitude tests have been shown to predict performance across a wide variety of settings. But as is always the case for a single test or type of test, there is room for improvement. The theory of successful intelligence provides one basis for improving prediction and possibly for establishing greater equity and diversity, which is a goal of most higher-educational institutions (Bowen, Kurzweil, & Tobin, 2006). It suggests that broadening the range of skills tested to go beyond analytic skills, to include practical and creative skills as well, might significantly enhance the prediction of undergraduate performance beyond current levels. Thus the theory does not suggest *replacing,* but rather *augmenting,* the SAT and similar tests (such as the ACT or, in the United Kingdom, the A-levels) in the undergraduate admissions process. Our collaborative team of investigators sought to study how successful such an augmentation could be. Even if we did not use the SAT, ACT, or A-levels in particular, we still would need some kind of assessment of the memory and analytical abilities these tests measure.

Methodological Considerations

In the Rainbow Project, data were collected at 15 schools across the United States, including 8 four-year undergraduate institutions, 5 community colleges, and 2 high schools.

The participants were 1,013 students predominantly in their first year as undergraduates or their final year of high school. In this chapter, analyses only for undergraduate students are discussed because they were the only ones for whom my colleagues and I had data available regarding undergraduate academic performance. The final number of participants included in these analyses was 793.

Baseline measures of standardized test scores and high school GPAs were collected to evaluate the predictive validity of current tools used for undergraduate admission criteria, and to provide a contrast for the current measures. Students' scores on standardized university entrance exams were obtained from the College Board.

The measure of analytical skills was provided by the SAT, plus multiple-choice analytical items we added to measure inference of meanings of words from context, number series completions, and figural matrix completions.

Creative skills were measured by multiple-choice items and by performance-based items. The multiple-choice items were of three kinds. In one, students were presented with verbal analogies preceded by counterfactual premises (e.g., money falls off trees). They had to solve the analogies as though the counterfactual premises were true. In a second, students were presented with rules for novel number operations—for example, *flix,* which involves numerical manipulations differing as a function of whether the first of two operands is greater than, equal to, or less than the second. Participants had to use the novel number operations to solve presented math problems. In a third, participants were first presented with a figural series involving one or more transformations; they then had to apply the rule of the series to a new figure with a different appearance, and complete the new series. These measures are not typical of assessments of creativity and were included for relative quickness of participants' responses and relative ease of scoring (cf. Niu & Sternberg, 2003). Also, they measured various types of creativity (Sternberg, 2005a).

Creative skills were also measured with open-ended measures. One measure required writing two short stories with a selection from among unusual titles, such as "The Octopus's Sneakers"; one required orally telling two stories based on choices of picture collages; and the third required captioning cartoons from among various options. Open-ended performance-based answers were rated by trained raters for novelty, quality, and task-appropriateness. Multiple judges were used for each task, and satisfactory reliability was achieved (Sternberg & the Rainbow Project Collaborators, 2005, 2006).

Multiple-choice measures of practical skills were of three kinds. In one, students were presented with a set of everyday problems in the life of an adolescent and had to select the option that would best solve each problem. In another, students were presented with scenarios requiring the use of math in everyday life (e.g., buying tickets for a ball game) and had to solve math problems based on the scenarios. In a third, students were presented with a map of an area (e.g., an entertainment park) and had to answer questions about navigating effectively through the area depicted by the map.

Practical skills were also assessed with three situational-judgment inventories: the Everyday Situational Judgment Inventory (Movies), the Common Sense Questionnaire, and the College Life Questionnaire, each of which taps different types of tacit knowledge. The general format of tacit-knowledge inventories has been described elsewhere (Sternberg et al., 2000), so only the contents of the inventories used in this study are described here. The movies presented everyday situations that confront undergraduates, such as a student's asking for a letter of recommendation from a professor who shows, through nonverbal cues, that he does not recognize the student very well. Participants then had to rate various options for how well they would work in response to each situation. The Common Sense Questionnaire provided everyday business problems, such as being assigned to work with a coworker whom one cannot stand. The College Life Questionnaire provided everyday university situations for which a solution was required.

Unlike the creativity performance tasks, in the practical performance tasks the participants were not given a choice of situations to rate. For each task, participants were told that there was no "right" answer, and that the options described in each situation represented variations on how different people approach different situations.

Consider examples of the kinds of items participants might find on the Rainbow assessment. An example of a creative item might be to write a story using the title "3516" or "It's Moving Backward." Another example might show a collage of pictures in which people are engaged in a wide variety of activities helping other people. A participant would then orally tell a story based on the collage. An example of a practical item might show a movie in which a student has just received a poor grade on a test. His roommate has had a health crisis the night before, and he has been up all night helping him. His professor hands him back the test paper, with a disappointed look on her face, and suggests to the student that he study harder next time. The movie then stops. Participants would then have to describe how the student might handle the situation. Or the participants might receive a written problem describing a conflict with another individual with whom a student is working on a group project. The project is getting mired down in the interpersonal conflict. The participants had to indicate how the student might resolve the situation to get the project done.

All materials were administered in either of two formats. A total of 325 of the university students took the test in paper-and-pencil format, whereas a total of 468 students took the test on the computer via the World Wide Web. No strict time limits were set for completing the tests, although the instructors were given rough guidelines of about 70 minutes per session. The time taken to complete the battery of tests ranged from 2 to 4 hours.

As a result of the lengthy nature of the complete battery of assessments, participants were administered parts of the battery in an intentionally incomplete overlapping design. The participants were randomly assigned to the test sections they were to complete. Details about the use of this procedure are given in Sternberg and the Rainbow Project Collaborators (2006).

Creativity in the Rainbow Project (and the subsequent Project Kaleidoscope) was assessed on the basis of the novelty and quality of responses. Practicality was assessed on the basis of the feasibility of the products with respect to human and material resources.

Findings

The analysis described below is a conservative one that does not correct for differences in the selectivity of the institutions at which the study took place. In a study across so many undergraduate institutions differing in selectivity, validity coefficients will seem to be lower than is typical because an A at a less selective institution counts the same as an A at a more selective institution. When we corrected for institutional selectivity, the results described below became stronger. But correcting for selectivity has its own problems (e.g., on what basis does one evaluate selectivity?), and so uncorrected data are used in this chapter. We also did not control for university major: Different universities may have different majors, and the exact course offerings, grading, and populations of students entering different majors may vary from one university to another, rendering control difficult.

When we examined undergraduate students alone, the sample showed slightly higher mean SAT scores than those found in undergraduate institutions across the United States. The standard deviation was above the normal 100-point standard deviation, meaning that we did not suffer from restriction of range. Our means, although slightly higher than typical, were within the range of average undergraduate students.

Another potential concern was pooling data from different institutions. We pooled data because in some institutions we simply did not have large enough numbers of cases for the data to be meaningful.

Some scholars believe that there is only one set of skills that is highly relevant to school performance—what is sometimes called *general ability*, or *g* (e.g., Jensen, 1998). These scholars believe that tests may appear to measure different skills, but when statistically analyzed, show themselves just to be measuring the single general ability. Did the Rainbow tests actually measure distinct analytical, creative, and practical skill groupings? Factor analysis addressed this question. Three meaningful factors were extracted from the data: practical performance tests, creative performance tests, and multiple-choice tests (including analytical, creative, and practical). In other words, multiple-choice tests, regardless of what they were supposed to measure, clustered together. Thus method variance proved to be very important. The results show the importance of using multiple formats to measure skills, precisely because method is so important in determining factorial structure. The results show the limitations of exploratory factor analysis in analyzing such data, and also of dependence on multiple-choice items outside the analytical domain. In the ideal situation, one wishes to ensure that one controls for method of testing in designing aptitude and other test batteries.

Undergraduate admissions offices are not interested, exactly, in whether these tests predict undergraduate academic success. Rather, they are interested in the extent to which these tests predict school success *beyond* those measures currently in use, such as the SAT and high school GPA. In order to test the incremental validity provided by Rainbow measures above and beyond the SAT in predicting GPA, we conducted a series of hierarchical regressions that included the items analyzed above in the analytical, creative, and practical assessments.

If one looks at the simple correlations, the SAT (both verbal and math), high school GPA, and the Rainbow measures all predicted first-year undergraduate GPA. But how did the Rainbow measures fare on incremental validity? In one set of analyses, the SAT (both verbal and math) and high school GPA were included in the first step of the prediction equation because these are the standard measures used today to predict undergraduate performance. Only high school GPA contributed uniquely to prediction of undergraduate GPA. Inclusion of the Rainbow measures roughly doubled prediction (percentage of variance accounted for in the criterion) over that obtained with the SAT alone.

These results suggest that the Rainbow tests add considerably to the predictive power of the SAT alone. They also suggest the power of high school GPA in prediction, particularly because it is an atheoretical composite that includes within it many variables, including motivation and conscientiousness.

Studying group differences requires careful attention to methodology and sometimes has led to erroneous conclusions (Hunt & Carlson, 2007). Although one important goal of the Rainbow Project was to predict success in the undergraduate years, another important goal involved developing measures that would reduce ethnic group differences in mean levels. There has been a lively debate as to why there are socially defined racial group differences, and as to whether scores for members of underrepresented minority groups are over- or underpredicted by SATs and related tests (see, e.g., Bowen & Bok, 2000; Rushton & Jensen, 2005; Sternberg, Grigorenko, & Kidd, 2005; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). There are a number of ways one can test for group differences in these measures, each of which involves a test of the size of the effect of ethnic group. Two different measures were chosen: ω^2 (omega squared) and Cohen's *d*.

There were two general findings. First, in terms of overall differences, the Rainbow tests appeared to reduce ethnic group differences, relative to traditional assessments of abilities like the SAT. Second, in terms of specific differences, it appears that the Hispanic American students benefited the most from the reduction of group differences. The African American students, too, seemed to show a reduction in difference from the European American mean for most of the Rainbow tests, although a substantial difference appeared to be maintained with the practical performance measures.

Although the group differences were not perfectly reduced, these findings suggest that measures can be designed that reduce ethnic and racial group differences on standardized tests, particularly for historically disadvantaged groups such as African American and Hispanic American students. These findings have important implications for reducing adverse impact in undergraduate admissions.

The SAT is based on a conventional psychometric notion of cognitive skills. Using this notion, it has had substantial success in predicting undergraduate academic performance. The Rainbow measures alone roughly doubled the predictive power of undergraduate GPA when compared to the SAT alone. In addition, the Rainbow measures predicted substantially beyond the contributions of the SAT and high school GPA. These findings, combined with encouraging results regarding the reduction of between-ethnicity differences, make a compelling case for furthering the study of the measurement of analytic, creative, and practical skills for predicting success at a university.

One important goal for this research was, and for future studies still is, the creation of standardized assessments that reduce the different outcomes between different groups as much as possible to maintain test validity. The measures described here suggest results toward this end. Although the group differences in the tests were not reduced to zero, the tests did substantially attenuate group differences relative to other measures such as the SAT. This finding could be an important step toward ultimately ensuring fair and equal treatment for members of diverse groups in the academic domain.

The principles behind the Rainbow Project apply at other levels of admissions as well. For example, we (Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006) have shown that the same principles can be applied in admissions to business schools, also with the result of increasing prediction and decreasing ethnic (as well as gender) group differences. Another study (Stemler, Grigorenko, Jarvin, & Sternberg, 2006) has found that including creative and practical items in augmented Advanced Placement psychology and statistics examinations can reduce ethnic group differences on the tests. Comparable results were found for the Advanced Placement physics examination (Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009). And the same principles were employed in a test for assessing the abilities of students in elementary school (Chart, Grigorenko, & Sternberg, 2008).

It is one thing to have a successful research project, and another actually to implement the procedures in a high-stakes situation. We have had the opportunity to do so. The results of a second project, Project Kaleidoscope, are reviewed here.

Project Kaleidoscope

Tufts University in Medford, Massachusetts, has strongly emphasized the role of active citizenship in education. It has put into practice some of the ideas from the Rainbow Project. In collaboration with former Dean of Admissions Lee Coffin, my colleagues and I instituted Project Kaleidoscope, which represents an implementation of the ideas of the Rainbow Project, but goes beyond that project to include in its assessment the construct of wisdom (for more details, see Sternberg, 2007, 2010a, 2010b, 2010c; Sternberg et al., 2010, 2012).

For all of the over 15,000 students applying to the School of Arts and Sciences and the School of Engineering at Tufts, we placed on the 2006–2007 application questions designed to assess wisdom (analytical and practical), intelligence, and creativity synthesized (WICS)—an extension of the theory of successful intelligence (Sternberg, 2003c). The program still continues, but the data reported here are for the first year, for which we have more nearly complete data (see Sternberg, 2010a).

The WICS theory extends the theory of successful intelligence on the basis of the notion that some people may be academically and even practically intelligent, but unwise—as in the case of numerous corporate and political scandals in which the perpetrators were smart, well educated, and foolish. The conception of wisdom used here is the balance theory of wisdom (Sternberg, 2003c), according to which wisdom is the application of intelligence, creativity, and knowledge for the common good, by balancing intrapersonal, interpersonal, and extrapersonal interests over the long and short terms, through the infusion of positive ethical values. Unwise people often fail on the ethical dimension, or skirt its edges (Sternberg, 2012a).

The questions are optional. Whereas the Rainbow Project was done as a separate set of high-stakes tests administered with a proctor, Project Kaleidoscope was (and continues to be) done as a section of the Tufts-specific supplement to the Common Application. It just was not practical to administer a separate high-stakes test battery such as the Rainbow measures for admission to one university. Moreover, the advantage of Project Kaleidoscope is that it got us away from the high-stakes testing situation in which students must answer complex questions in very short amounts of time under incredible pressure.

Students were encouraged to answer just a single question, so as not overly to burden them. Tufts University competes for applications with many other universities, and if the Tufts application had been substantially more burdensome than those of competitor schools, it would have put Tufts at a real-world disadvantage in attracting applicants. In the theory of successful intelligence, individuals with such intelligence capitalize on strengths and compensate for or correct weaknesses. Our format gave students a chance to capitalize on a strength.

As examples of items, a creative question asked students to write stories with titles such as "The End of MTV" or "Confessions of a Middle-School Bully." Another creative question asked students what the world would be like if some historical event had come out differently—for example, if Rosa Parks had given up her seat on the bus. Yet another creative question, a nonverbal one, gave students an opportunity to design a new product or an advertisement for a new product. A practical question queried how students had persuaded friends of an unpopular idea they held. A wisdom question asked students how a passion they had could be applied toward a common good.

Creativity and practicality were assessed in the same way as in the Rainbow Project. Analytical quality was assessed by the organization, logic, and balance of the essay. Wisdom was assessed by the extent to which the response represented the use of abilities and knowledge for a common good by balancing one's own, others', and institutional interests over the long and short terms through the infusion of positive ethical values.

Note that the goal was (and still is) not to replace the SAT and other traditional admissions measurements (e.g., GPA and class rank) with some new test. Rather, it was to reconceptualize applicants in terms of academic/analytical, creative, practical, and wisdom-based abilities, using the essays as one but not the only source of information. For example, highly creative work submitted in a portfolio could also be entered into the creativity rating, as could evidence of creativity through winning of prizes or awards. The essays were major sources of information, but if other information was available, the trained admissions officers used it.

Among the applicants who were evaluated as being academically qualified for admission, approximately half completed an optional essay in the first year and two-thirds in later years. Doing these essays had no meaningful effect on chances of admissions. However, *quality* of essays or other evidence of creative, practical, or wisdom-based abilities did have an effect. For those applicants given an A (top rating) by a trained admission officer in any of these three categories, average rates of acceptance were roughly double those for applicants not getting an A. Because of the large number of essays (over 8,000), only one rater rated applicants except for a sample to ensure that interrater reliability was sufficient, which it was.

Many measures did not look like conventional standardized tests, but had statistical properties mimicking them. We were therefore interested in convergent–discriminant validation of our measures. The correlation of our measures with a rated academic composite that included SAT scores and high school GPA were modest but significant for creative thinking, practical thinking, and wise thinking. The correlations with a rating of quality of extracurricular participation and leadership were moderate for creative, practical, and wise thinking. Thus the pattern of convergent–discriminant validation was what we had hoped for.

The average academic quality of applicants in the Tufts School of Arts and Sciences rose slightly in 2006–2007, the first year of the project, in terms of both SAT and high school GPA. In addition, there were notably fewer students in what before had been the bottom third of the pool in terms of academic quality. Many of those students, seeing the new application, seem to have decided not to bother to apply. Many more strong applicants applied.

Thus adopting these new methods does not seem to result in less qualified applicants applying to the institution and being admitted. Rather, the applicants who are admitted are *more* qualified, but in a broader way. Perhaps most rewarding were the positive comments from large numbers of applicants that they felt our application gave them a chance to show themselves for who they were. Of course, many factors are involved in admissions decisions, and Project Kaleidoscope ratings were only one small part of the overall picture.

We did not get meaningful differences across ethnic groups—a result that surprised us, given that the earlier Rainbow Project reduced but did not eliminate differences. And after a number of years in which applications by underrepresented minorities were relatively flat in terms of numbers, during 2006–2007 they went up substantially. In the end, applications from African Americans and Hispanic Americans increased significantly, and admissions of African Americans were up 30% and of Hispanic Americans up 15%. So the Project Kaleidoscope results, like those of the Rainbow Project, showed that it is possible to increase academic quality and diversity simultaneously, and to do so for an entire undergraduate class at a major university, not just for small samples of students at some scattered schools. Most importantly, we sent a message to students, parents, high school guidance counselors, and others that we believe there is more to a person than the narrow spectrum of skills assessed by standardized tests, and that these broader skills can be assessed in a quantifiable way.

The Panorama Project

When I went to Oklahoma State University in 2010 as provost and senior vice president, Vice President for Enrollment Management Kyle Wray and his team of admissions officers instituted the Panorama Project, which was loosely based on Kaleidoscope but was oriented toward the very different group of students applying to Oklahoma State University as opposed to Tufts University. I left Oklahoma State before formal results were collected, but the admissions office was pleased with the results, and the project continues today.

BEYOND TRADITIONAL THEORIES OF INTELLIGENCE

The triarchic theory consists of three interrelated subtheories that attempt to account for the bases and manifestations of intelligent thought; as such, it represents an expanded view of intelligence that departs from traditional, general, and dichotomous theoretical perspectives. The componential subtheory relates intelligence to the internal world of the individual. The experiential subtheory relates intelligence to the experience of the individual with tasks and situations. The contextual subtheory relates intelligence to the external world of the individual. The elements of the three subtheories are interrelated: The components of intelligence are manifested at different levels of experience with tasks, and in situations of varying degrees of contextual relevance to a person's life. The components of intelligence are posited to be universal to intelligence; thus the components that contribute to intelligent performance in one culture do so in all other cultures as well. Moreover, the importance of dealing with novelty and the automatization of information processing to intelligence are posited to be universal. But the manifestations of these components in experience are posited to be relative to cultural contexts. What constitutes adaptive thought or behavior in one culture is not necessarily adaptive in another culture. Moreover, thoughts and actions that would shape behavior in appropriate ways in one context might not shape them in appropriate ways in another context. Finally, the environment one selects will depend largely on the available environments and on the fit of one's cognitive abilities, motivation, values, and affects to the available alternatives.

REFERENCES

Ang, S., Van Dyne, L., & Tan, M. L. (2011). Cultural intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 582–602). New York: Cambridge University Press.

Bowen, W. G., & Bok, D. (2000). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.

Bowen, W. G., Kurzweil, M. A., & Tobin, E. M. (2006). *Equity and excellence in American higher education*. Charlottesville: University of Virginia Press.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77–165). Hillsdale, NJ: Erlbaum.

Brown, A. L., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering, and understanding. In P. H. Mussen (Series Ed.) & J. Flavell & E. Markman (Vol. Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 77–166). New York: Wiley.

Butterfield, E. C., Wambold, C., & Belmont, J. M. (1973). On the theory and practice of improving short-term memory. *American Journal of Mental Deficiency, 77,* 654–669.

Campione, J. C., & Brown, A. L. (1979). Toward a theory of intelligence: Contributions from research with retarded children. In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 139–164). Norwood, NJ: Ablex.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Chart, H., Grigorenko, E. L., & Sternberg, R. J. (2008). Identification: The Aurora Battery. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education* (pp. 281–301). Waco, TX: Prufrock.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chi, M. T. H. (1978). Knowledge structure and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Cianciolo, A. T., & Sternberg, R. J. (2018). Practical intelligence and tacit knowledge: A prototype view of expertise. In K. A. Ericsson, R. Hoffman, A. Kozbelt, & M. Williams (Eds.), *The Cambridge handbook on expertise and expert performance* (2nd ed.). New York: Cambridge University Press.

Conway, A. R. A., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 394–418). New York: Cambridge University Press.

Cooper, C. (2015). *Intelligence and human abilities: Structure, origins, and applications.* New York: Routledge.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

Davidson, J. E., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly, 28,* 58–64.

Detterman, D. K., & Sternberg, R. J. (Eds.). (1982). *How and how much can intelligence be increased?* Norwood, NJ: Ablex.

Flavell, J. H. (1977). *Cognitive development.* Englewood Cliffs, NJ: Prentice-Hall.

Gardner, H. (2006). *Multiple intelligences: New horizons.* New York: Basic Books.

Gardner, H., Krechevsky, M., Sternberg, R. J., & Okagaki, L. (1994). Intelligence in context: Enhancing students' practical intelligence for school. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 105–127). Cambridge, MA: Bradford Books.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gladwin, T. (1970). *East is a big bird.* Cambridge, MA: Harvard University Press.

Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2002). School-based tests of the triarchic theory of intelligence: Three settings, three samples, three syllabi. *Contemporary Educational Psychology, 27,* 167–208.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* New York: Cambridge University Press.

Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: A supplement to the Graduate Management Admissions Test. *Learning and Individual Differences, 16,* 101–127.

Hodapp, R. M., Griffin, M. M., Burke, M. M., & Fisher, M. H. (2011). Intellectual disabilities. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 193–209). New York: Cambridge University Press.

Hunt, E. B. (1978). Mechanics of verbal ability. *Psychological Review, 85,* 109–130.

Hunt, E. B. (1980). Intelligence as an information-processing concept. *British Journal of Psychology, 71,* 449–474.

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science, 2,* 194–213.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1982). Reaction time and psychometric *g.* In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer-Verlag.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Kearins, J. M. (1981). Visual spatial memory in Australian Aboriginal children of desert regions. *Cognitive Psychology, 13,* 434–460.

Keil, F. C. (1984). Transition mechanisms in cognitive development and the structure of knowledge. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 81–99). San Francisco: Freeman.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy.* New York: Farrar, Straus & Giroux.

Lohman, D. F., & Lakin, J. M. (2011). Intelligence and reasoning. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 419–441). New York: Cambridge University Press.

Lombroso, P. J., Sternberg, R. J., & Grigorenko, E. L. (1999). Genetics of childhood disorders: I. Genetics and intelligence. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 486–488.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence–picture relationships. *Journal of Verbal Learning and Verbal Behavior, 17,* 493–507.

Mayer, R. E., & Greeno, J. G. (1972). Structural differences between learning outcomes produces by different instructional methods. *Journal of Educational Psychology, 63,* 165–173.

Nettelbeck, T. (2011). Basic processes of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 371–393). New York: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Niu, W., & Sternberg, R. J. (2003). Societal and school influences on student creativity: The case of China. *Psychology in the Schools, 40*(1), 103–114.

Okagaki, L., & Sternberg, R. J. (1993). Putting the distance into students' hands: Practical intelligence for school. In R. R. Cocking & K. A. Renninger (Eds.), *The development and meaning of psychological distance* (pp. 237–253). Hillsdale, NJ: Erlbaum.

Ortiz, S. O. (2015). CHC theory of intelligence. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 209–228). New York: Springer.

Otero, T. M. (2015). Intelligence: Defined as neurocognitive processing. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 193–208). New York: Springer.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294.

Saklofske, D. H., van de Vijver, F. J. R., Oakland, T., Mpofu, E., & Suzuki, L. A. (2015). Intelligence and culture: History and assessment. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 341–366). New York: Springer.

Schank, R. C. (1980). How much intelligence is there in artificial intelligence? *Intelligence, 4,* 1–14.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Siegler, R. S. (1986). Unities across domains in children's strategy choices. In M. Perlmutter (Ed.), *The Minnesota Symposia on Child Psychology: Vol. 19. Perspectives on intellectual development* (pp. 1–48). Hillsdale, NJ: Erlbaum.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10,* 104–109.

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem solving task. *Cognitive Psychology, 8,* 86–97.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347–376.

Spear-Swerling, L., & Sternberg, R. J. (1994). The road not taken: An integrative theoretical model of reading disability. *Journal of Learning Disabilities, 27*(2), 91–103.

Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a basis for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology, 31*(2), 344–376.

Stemler, S., Sternberg, R. J., Grigorenko, E. L., Jarvin, L., & Sharpes, D. K. (2009). Using the theory of successful intelligence as a framework for developing assessments in AP physics. *Contemporary Educational Psychology, 34,* 195–209.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1979). The nature of mental abilities. *American Psychologist, 34,* 214–230.

Sternberg, R. J. (1981a). A componential theory of intellectual giftedness. *Gifted Child Quarterly, 25,* 86–93.

Sternberg, R. J. (1981b). Intelligence and nonentrenchment. *Journal of Educational Psychology, 73,* 1–16.

Sternberg, R. J. (1981c). Testing and cognitive psychology. *American Psychologist, 36,* 1181–1189.

Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225–307). New York: Cambridge University Press.

Sternberg, R. J. (1983). Components of human intelligence. *Cognition, 15,* 1–48.

Sternberg, R. J. (1984). What should intelligence tests test?: Implications of a triarchic theory of intelligence for intelligence testing. *Educational Researcher, 13,* 5–15.

Sternberg, R. J. (1985a). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1985b). Componential analysis: A recipe. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 1, pp. 179–201). Norwood, NJ: Ablex.

Sternberg, R. J. (1985c). Human intelligence: The model is the message. *Science, 230,* 1111–1118.

Sternberg, R. J. (1985d). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49,* 607–627.

Sternberg, R. J. (1985e). Teaching critical thinking: Part 1. Are we making critical mistakes? *Phi Delta Kappan, 67,* 194–198.

Sternberg, R. J. (1986). Inside intelligence. *American Scientist, 74,* 137–143.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–105). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Sternberg, R. J. (1990). *Metaphors of mind.* New York: Cambridge University Press.

Sternberg, R. J. (1993). *Sternberg Triarchic Abilities Test.* Unpublished test.

Sternberg, R. J. (1995). *In search of the human mind.* Fort Worth, TX: Harcourt Brace.

Sternberg. R. J. (1997a). Managerial intelligence: Why IQ isn't enough. *Journal of Management, 23*(3), 463–475.

Sternberg, R. J. (1997b). *Successful intelligence.* New York: Plume.

Sternberg, R. J. (1997c). What does it mean to be smart? *Educational Leadership, 54*(6), 20–24.

Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 17–20.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology, 3,* 292–316.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (2002). *Why smart people can be so stupid.* New Haven, CT: Yale University Press.

Sternberg, R. J. (2003a). Construct validity of the theory of successful intelligence. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 55–80). Washington, DC: American Psychological Association.

Sternberg, R. J. (2003b). Driven to despair: Why we need to redefine the concept and measurement of intelligence. In L. G. Aspinwall & U. M. Staudinger (Eds.), *A psychology of human strengths: Fundamental questions and future directions for a positive psychology* (pp. 319–330). Washington, DC: American Psychological Association.

Sternberg, R. J. (2003c). *Wisdom, intelligence, and creativity synthesized.* New York: Cambridge University Press.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*(5), 325–338.

Sternberg, R. J. (2005a). Creativity or creativities? *International Journal of Human Computer Studies, 63,* 370–382.

Sternberg, R. J. (2005b). The theory of successful intelligence. *Interamerican Journal of Psychology, 39*(2), 189–202.

Sternberg, R. J. (2005c). The triarchic theory of successful intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 103–119). New York: Guilford Press.

Sternberg, R. J. (2007). Finding students who are wise, practical, and creative. *Chronicle of Higher Education, 53*(44), B11.

Sternberg, R. J. (2008). Successful intelligence as a framework for understanding cultural adaptation. In S. Ang & L. Van Dyne (Eds.), *Handbook on cultural intelligence* (pp. 306–317). New York: Sharpe.

Sternberg, R. J. (2009). The theory of successful intelligence as a basis for new forms of ability testing at the high school, college, and graduate school levels. In J. C. Kaufman (Ed.), *Intelligent testing: Integrating psychological theory and clinical practice* (pp. 113–147). New York: Cambridge University Press.

Sternberg, R. J. (2010a). *College admissions for the 21st century.* Cambridge, MA: Harvard University Press.

Sternberg, R. J. (2010b). WICS: A new model for cognitive education. *Journal of Cognitive Education and Psychology, 9,* 34–46.

Sternberg, R. J. (2010c). WICS: A new model for school psychology. *School Psychology International, 31*(6), 599–616.

Sternberg, R. J. (2012a). Ethical drift. *Liberal Education, 98*(3), 60.

Sternberg, R. J. (2012b). The triarchic theory of successful intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 156–177). New York: Guilford Press.

Sternberg, R. J. (2015). Successful intelligence: A new model for testing intelligence beyond IQ tests. *European Journal of Education and Psychology, 8,* 76–84.

Sternberg, R. J., Bonney, C. R., Gabora, L., Karelitz, T., & Coffin, L. (2010). Broadening the spectrum of undergraduate admissions. *College and University, 86*(1), 2–17.

Sternberg, R. J., Bonney, C. R., Gabora, L, & Merrifield, M. (2012). WICS: A model for college and university admissions. *Educational Psychologist, 47*(1), 30–41.

Sternberg, R. J., & Clinkenbeard, P. R. (1995). The triarchic model applied to identifying, teaching and assessing gifted children. *Roeper Review, 17*(4), 255–260.

Sternberg, R. J., & Davidson, J. E. (1982, June). The mind of the puzzler. *Psychology Today,* pp. 37–44.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P. R., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly, 40,* 129–137.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Snook, S., Williams, W. M., et al. (2000). *Practical intelligence in everyday life.* New York: Cambridge University Press.

Sternberg, R. J., & Gardner, M. K. (1982). A componential interpretation of the general factor in human intelligence. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 231–254). Berlin: Springer-Verlag.

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Erlbaum.

Sternberg, R. J., & Grigorenko, E. L. (2004). Successful intelligence in the classroom. *Theory into Practice, 43*(4), 274–280.

Sternberg, R. J., & Grigorenko, E. L. (2007). *Teaching for successful intelligence* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment, 15,* 1–11.

Sternberg, R. J., Grigorenko, E. L., & Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist, 60*(1), 46–59.

Sternberg, R. J., Grigorenko, E. L., & Zhang, L.-F. (2008). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science, 3*(6), 486–506.

Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, *g*, and work psychology. *Human Performance, 15*(1–2), 143–160.

Sternberg, R. J., Jarvin, L., Birney, D., Naples, A., Stemler, S., Newman, T., et al. (2014). Testing the theory of successful intelligence in teaching grade 4 language arts, mathematics, and science. *Journal of Educational Psychology, 106,* 881–899.

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L. (2009). *Teaching for wisdom, intelligence, creativity, and success.* Thousand Oaks, CA: Corwin Press.

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L. (2011). *Explorations of giftedness.* New York: Cambridge University Press.

Sternberg, R. J., Jarvin, L., & Reznitskaya, A. (2008). Teaching of wisdom through history: Infusing wise thinking skills in the school curriculum. In M. Ferrari & G. Potworowski (Eds.), *Teaching for wisdom* (pp. 37–57). New York: Springer.

Sternberg, R. J., & Kaufman, S. B. (Eds.). (2011). *The Cambridge handbook of intelligence.* New York: Cambridge University Press.

Sternberg, R. J., & Ketron, J. L. (1982). Selection and implementation of strategies in reasoning by analogy. *Journal of Educational Psychology, 74,* 399–413.

Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., et al. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence, 29,* 401–418.

Sternberg, R. J., Okagaki, L., & Jackson, A. (1990). Practical intelligence for success in school. *Educational Leadership, 48,* 35–39.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38,* 878–893.

Sternberg, R. J., & the Rainbow Project Collaborators. (2005). Augmenting the SAT through assessments of analytical, practical, and creative skills. In W. Camara & E. Kimmel (Eds.), *Choosing students: Higher education admission tools for the 21st century* (pp. 159–176). Mahwah, NJ: Erlbaum.

Sternberg, R. J., & the Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence, 34*(4), 321–350.

Sternberg, R. J., Reznitskaya, A., & Jarvin, L. (2007). Teaching for wisdom: What matters is not just what students know, but how they use it. *London Review of Education, 5*(2), 143–158.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology, 27,* 195–232.

Sternberg, R. J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. *Social Cognition, 2,* 168–192.

Sternberg, R. J., & Smith, E. E. (Eds.). (1988). *The psychology of human thought.* New York: Cambridge University Press.

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998). Teaching triarchically improves school achievement. *Journal of Educational Psychology, 90,* 374–384.

Sternberg, R. J., & Weil, E. M. (1980). An aptitude–strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology, 72,* 226–234.

Sternberg, R. J., & Zhang, L. F. (1995). What do we mean by "giftedness"?: A pentagonal implicit theory. *Gifted Child Quarterly, 39*(2), 88–94.

Super, C. M., & Harkness, S. (1982). The infants' niche in rural Kenya and metropolitan America. In L. L. Adler (Ed.), *Cross-cultural research at issue* (pp. 47–55). New York: Academic Press.

Suzuki, L. A., Short, E. L., & Lee, C. S. (2011). Racial and ethnic group differences in intelligence in the United States. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 273–292). New York: Cambridge University Press.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623–628.

Vernon, P. E. (1971). *The structure of human abilities.* London: Methuen.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds., & Trans.). Cambridge, MA: Harvard University Press.

Wagner, R. K. (2011). Practical intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 550–563). New York: Cambridge University Press.

Wagner, R. K., & Sternberg, R. J. (1987). Executive control in reading comprehension. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 1–21). Hillsdale, NJ: Erlbaum.

Williams, W. M., Blythe, T., White, N., Li, J., Gardner, H., & Sternberg, R. J. (2002). Practical intelligence for school: Developing metacognitive sources of achievement in adolescence. *Developmental Review, 22,* 162–210.

Willis, J. O., Dumont, R., & Kaufman, A. S. (2011). Factor-analytic models of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39–57). New York: Cambridge University Press.

CHAPTER 6

Redefining Intelligence with the Planning, Attention, Simultaneous, and Successive Theory of Neurocognitive Processes

Jack A. Naglieri
Tulio M. Otero

Practitioners and test authors have become increasingly conscious of the need for theory-based intelligence tests. Although several theories of intelligence have been attached to traditional ability tests such as the Wechsler scales (Plucker & Esping, 2014), one theory, first described by Das, Kirby, and Jarman (1979), was used explicitly to develop a new way to construct an intelligence test. In 1997, Naglieri and Das (1997a) published the Cognitive Assessment System (CAS), which was based on a neurocognitive theory called *planning, attention, simultaneous, and successive* (PASS) processing. These authors argued that a neurocognitive theory of intelligence provides the foundation necessary for test construction and is equally important for test interpretation. They also suggested that traditional IQ tests, which were based largely on the work of the U.S. military (see Naglieri, 2015), were too limited and could be improved if the constructs that were measured were related to brain functions. Naglieri and Das anticipated that the PASS neurocognitive approach would yield better diagnostic information, have relevance to instructional decision making, and be more appropriate for diverse populations (Naglieri & Otero, 2011, 2017).

The PASS neurocognitive theory is a concept of intelligence and human learning that is best measured with tests devised explicitly to measure the four PASS processes. PASS theory has been most recently operationalized in the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014a), the CAS2: Español (Naglieri, Moreno, & Otero, 2017), the CAS2: Brief (Naglieri, Das, & Goldstein, 2014b), and the CAS2: Rating Scale (Naglieri, Das, & Goldstein, 2014c). We describe these measures comprehensively in Chapter 15 of this book. In this chapter, we focus on the PASS theory upon which all of these measures are based.

The PASS theory and the CAS2 provide a neurocognitive perspective on ability that differs from that of traditional batteries (those including, in part, subtests requiring verbal and quantitative knowledge). These batteries have been used since the Army mental testing program described by Yoakum and Yerkes (1920) almost 100 years ago. The PASS theory, as operationalized by the CAS and CAS2, has created an opportunity to move the field of intelligence and ability testing forward by emphasizing (1) that a *test* of intelligence should be based on a *theory* of intelligence and (2) that the test should measure basic neurocognitive processes defined by the intellectual demands of the test, not the content of the questions. Naglieri and colleagues further assert that a test of intelligence should be designed explicitly to measure specific constructs defined by the theory. This assertion raises two important questions: First, what is a theory? Second, what is a neurocognitive process? Both questions are answered next.

A *theory* is an organized set of concepts that explains a phenomenon or set of phenomena, preferably in the most parsimonious manner. Theories are concise, coherent, systematic, predictive, and broadly applicable, often integrating and generalizing many hypotheses. As such, a test of neurocognitive functions should measure psychological processes on the basis of a coherent theory that provides useful information in a concise and systematic way. There are different definitions in the literature for the term *psychological process*. However, all definitions share the notion that a psychological process involves the performance of some composite cognitive activity. Moreover, a test of neurocognitive processing should measure *thinking* apart from *knowing*.

Consider the four test questions that appear in Figure 6.1. The verbal analogy requires *knowledge* of verbal concepts (*girl, woman, boy*, etc.) and relationships among those concepts. To answer the first question in Figure 6.1 (top left), the child needs to understand that a girl becomes a woman, and similarly that a boy becomes a man. The relationships between the younger and older persons need to be comprehended to arrive at the correct answer. The second question (middle left) requires *knowledge* of a number series: The child must examine the series and detect that the numbers double from 3 to 6 and 6 to 12. The third question (bottom left) requires that the relationship between a pair of chords (C^7 and F major) is understood to figure out that an E^7 would be followed by an A major chord. In these examples, the examinee must know certain facts to understand the relationships among the words, numbers, and musical chords. In the fourth example—the drawing at right, in which only shapes are provided—the relationships among the shapes must be understood to answer the question (small oval becomes big; small rectangle also becomes big), but knowledge of the names of the shapes is not needed. In these examples, the solution is based on the examinee's ability to *recognize and understand the relationships* between the words, numbers, musical notations, or shapes. Despite these differences in content, the thinking is the same (Simultaneous processing, in PASS theory).

As noted above, a theory of intelligence did not influence the content of traditional IQ tests first introduced more than 100 years ago (Naglieri, 2015). In recent years, considerable efforts have been made to reconceptualize already published tests within some theoretical model. For example, the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014) has been linked to the Cattell–Horn–Carroll (CHC) view of intelligence (called a *theory of cognitive abilities* in Chapter 3 of this volume), but was never developed on the basis of that view. We suggest that applying a view of intelligence to already entrenched tests does not help the field advance. Instead, we advocate that a modern test of intelligence should be based on a clearly defined theory. Our choice is the PASS theory. In this chapter, we discuss the origins of PASS theory, describe three functional

Girl is to woman as boy is to ____?
a. Sister
b. Woman
c. Man
d. Brother

Which number is next in the series:
3, 6, 12, ____?

Which chord comes next?
C^7, F, E^7, ____?

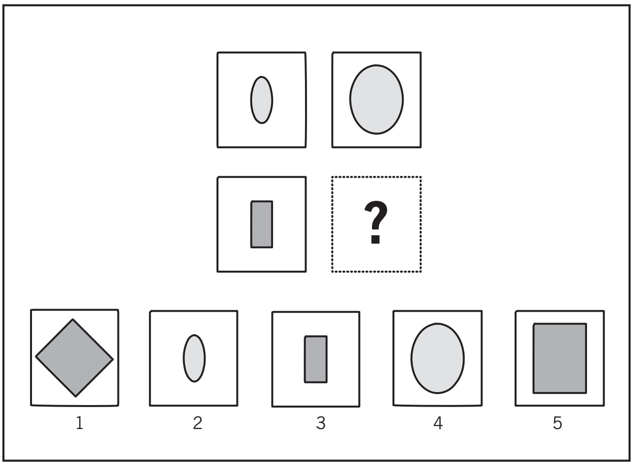


FIGURE 6.1. An illustration of the same kind of thinking applied to different content.

units of the brain as originally articulated by Luria, explain how the PASS processes were operationalized, and summarize the current empirical support for the theory.

ORIGINS OF THE PASS THEORY

The functional aspects of brain structures as delineated in the work of Alexander Luria (1966, 1973a, 1973b, 1980) were the underpinnings of PASS theory (Naglieri & Das, 1997a). Naglieri and Das used Luria's work as a blueprint for defining the important components of a neurocognitive view of intelligence (Das, Naglieri, & Kirby, 1994; Naglieri & Das, 1997a; Naglieri et al., 2014a) because they believed strongly that a test of intelligence should be based on a *theory of intelligence,* and that a working theory of intelligence should be based on an understanding of basic neurocognitive processes and consistent with neuropsychology and neuroscience. Their efforts represented the first time that a specific researched neurocognitive theory was used to reconceptualize the concept of intelligence and develop a specific tool to measure that theory.

Luria theorized that human cognitive functions can be conceptualized within a framework of three separate but related brain systems that provide four basic psychological processes. The three brain systems are referred to as *functional units* because the neurocognitive mechanisms work in separate but interrelated systems. Although Luria lacked the sophisticated neuroscientific resources that exist today, his conceptualization of how the brain functions still stands as valid. For example, studies using functional imaging technology (Avram et al., 2013; Yeo et al., 2011; Zaytseva et al., 2014) have shown that each area of the brain participates in numerous large- and small-scale functional systems within and across cortical and subcortical brain structures (for supportive research in the neuroscience literature, see Koziol, Barker, Joyce, & Hrin, 2014; Koziol, Beljan, Bree, Mather, & Barker, 2016,).

Luria (1973b) stated that "each form of conscious activity is always a complex functional system and takes place through the combined working of all three brain units, each of which makes its own contribution" (p. 99). Luria's assertion remains fundamentally true. Cognition and behavior *are products of functional brain networks,* and these networks have a profound impact on constructs such as attention, executive function, learning and memory, and information processing. Much of early neuropsychology has traditionally interpreted tests within a serial-order processing paradigm (Koziol et al., 2014): First we perceive, then we think, and then we react.

Children, adolescents, and adults are continuously responding to an ever-changing, dynamic environment, however. This makes a static paradigm insufficient for understanding and measuring neurocognitive processes, as well as for interpreting test performance. As we are constantly bombarded by external and internal stimuli, there must be some brain mechanisms that allow us to funnel and direct information by activating and inhibiting different brain regions dynamically. Once information is selected for further processing, there must be some way for different brain regions to communicate to integrate the required information. Whether in our daily lives or during psychological assessment, there must also be a mechanism for allowing the brain to shift from the demands of one task to the demands of another. In other words, there must be a process allowing different parts of the brain to communicate and interact as we continuously adapt to ever-changing demands of different tasks.

Brain regions never function in seclusion, but any given cortical region is characterized by a certain degree of information-processing specificity for a cognitive ability or facet of cognitive operations (Friston, 2002; Johnson, 2005). This specificity is referred to as *functional specialization.* As originally put forth by the work of Luria, effective performance on any given task is characterized by the functional integration of distal brain regions. This integration represents the momentary, dynamic, context-specific communications that convey information via subsets of anatomical connections among a limited number of brain regions engaged by a cognitive process (Koziol & Stevens, 2012).

The functional architecture of the brain is characterized by reciprocal connections across several brain profiles of the cerebro-cortical, cortical–basal ganglia, cerebro-cerebellar, and basal ganglia–cerebellar circuitry systems (Bostan, Dum, & Strick, 2010, 2013; Bostan & Strick, 2010; Koziol, Budding, & Chidekel, 2013). As an example, Yeo and colleagues (2011) have consistently observed seven patterns of cortical networks in adults, adolescents, children, and infants, as assessed through resting-state neuroimaging technologies (Uddin, Supekar, & Menon, 2013). These networks are fundamental for adapting to the rapidly changing demands of our environments, including undergoing assessment of neurocognitive functions. These networks are the fronto-parietal network, ventral attention network, dorsal attention network, visual network, limbic network, sensory–motor network, and default mode network. Examples of some of these networks and how they add to our understanding of PASS processes are provided below.

The *fronto-parietal network* is the "cognitive control" network, which includes working memory functions. It is typically engaged when information or rules need to be kept in mind to guide behavior during effortful cognitive task performance. The network consists of the dorsolateral prefrontal cortex, the anterior cingulate cortex, the anterior prefrontal cortex, the lateral cerebellum, the anterior insula, the caudate nucleus, and the inferior parietal lobule.

The *ventral attention network* provides salience information and allows for the identification of objects and of what these objects are used for. It includes the temporo-parietal junction, the supramarginal gyrus, the frontal operculum, and the anterior insula. The *dorsal attention network* is within the intraparietal sulcus and frontal eye fields; it participates in goal-directed executive control processes by managing spatial attention and attentional shifting, in conjunction with identifying where objects are and knowing how to use them. The interaction of the ventral and dorsal networks guides purposeful behavior as we constantly interact with our dynamically changing environmental events. After we become aware of something we need to orient to, dorsal fronto-parietal regions become activated, and the dorsal network is central to selective attention (Corbetta, Patel, & Shulman, 2008). When we attend to a constantly changing environment, however, both ventral and dorsal networks become activated.

The *visual network* is made up of the occipital lobe and lateral temporal and superior parietal regions; it connects with the superior parietal lobe and intraparietal sulcus, both of which are linked to the dorsal attention network. The visual network is involved in sustaining attention, suppressing attention to irrelevant stimuli, and interacting with these control systems to help direct attention. Other neural networks include the limbic, sensory–motor, and default mode networks. The *limbic network* acts together with other systems to provide motivational and reward influences. This network consists of the dorsal anterior cingulate and the bilateral insulae, and it provides a cortical signal of salient events, including errors. The *motor network* is composed of the primary, supplementary, and premotor cortex, along with the sensory cortex, putamen, thalamus, and cerebellum. The *default mode network* includes the anterior medial prefrontal cortex, the posterior cingulate, and the dorsomedial prefrontal and medial temporal systems. This network is active when external stimuli are at a minimum.

The four basic neurocognitive processes responsible for cognitive activity and behavior represent a "working constellation" (Luria, 1966, p. 70) of networks. Just as different neural networks come into play in an integrative fashion for a particular task, a person may perform the same task with different contributions of the PASS processes, along with the application of the person's knowledge and skills. The central assumption that underlies cognitive interventions is based on this interplay between cognitive activity and behavior (Naglieri & Pickering, 2010). This interplay also helps to explain how different PASS profiles may explain the same profile of academic strengths and weaknesses.

Although effective functioning is accomplished through the integration of all processes as demanded by the particular task, not every process is involved equally in every task. In addition, a task may be approached via a different combination of processes, depending on how the task was initially taught or learned. For example, tasks like math calculation may be dominated by a single process (e.g., planning), while tasks such as reading decoding may be strongly related to another process (e.g., successive), while also recruiting other neurocognitive processes. Reading comprehension of familiar text may, for example, recruit both simultaneous and successive processes, while reading something composed of unfamiliar content may require an additional process to be recruited.

Description of Luria's Three Functional Units

The function of Luria's first functional unit provides regulation of cortical arousal and attention; the second codes information, using simultaneous and successive processes; and the third provides for strategy development, strategy use, self-monitoring, and control of cognitive activities. These functional units also intersect with functional networks. The functional units and networks of the brain provide the infrastructure necessary to interact with the environment, acquire knowledge, and learn.

First Functional Unit

The attention–arousal system is the first of these three functional units of the brain, and is located primarily in the brainstem, the diencephalon, and the medial regions of the cortex (Luria, 1973b). This unit provides the brain with the appropriate level of arousal or cortical tone, as well as directive and selective attention. When many stimuli are presented to a person who is then required to pay attention to only one stimulus, the inhibition of responding to other (often more salient) stimuli and the focusing of attention to the target stimulus depend on the first functional unit. Luria (1973b) stated that optimal conditions of arousal are needed before the more complex forms of attention, involving "selective recognition of a particular stimulus and inhibition of responses to irrelevant stimuli" (p. 271), can occur. Moreover, only when individuals are aroused sufficiently and their attention is focused adequately can they utilize processes in the second and third functional units.

Contemporary neuroscience literature continues to support Luria's initial description of the three functional units and offers additional observations. The brain's default mode network (resting state) becomes increasingly active as goal-directed cognitive tasks and behavior are required. This network is anchored in two brain regions referred to as *hubs*, or centers of primary neural activity. The anterior medial prefrontal cortex and the posterior cingulate cortex are the two central hubs. These regions relate to two subsystems: the dorsomedial prefrontal subsystem and the medial temporal lobe subsystem. The arousal–attention system described by Luria activates these higher brain centers by initial suppression of the default mode network and activation of the ventral and dorsal attention networks.

The ventral attention network informs other brain regions about the importance of what is being attended to externally. The dorsal attention network's role is to shift the focus of attention. Because we are constantly interacting and adapting to demands in the world, this system plays a critical role in that process. The dorsal attention network, in essence, reorients attention to what is relevant to the demands of the task or situation, and specifies the parameters for action by informing other parts of the brain (i.e., frontal systems) about "how to do" something (Koziol et al., 2013). Thus the first functional network, along with its related networks, allows for orientating, sustaining, and reorienting attention to what has relevance at any moment in time. It also activates the fronto-parietal system and facilitates simultaneous and successive processes.

Second Functional Unit

The second functional unit provides for simultaneous and successive processing though the activation and coactivation of the fronto-parietal network and the temporo-parietal junctions of both the right and left hemispheres. Activation of the parietal regions is key to both simultaneous and successive processing, as this region is considered the association cortex—a zone in which many related functions (such as attention, spatial representation, working memory, eye movements, an assortment of other sensory information, and the guidance of actions) come together.

Simultaneous processing involves integrating stimuli into groups so that the interrelationships among the components are understood. For example, for a person to produce a diagram correctly when given the instruction "Draw a triangle above a square that is to the left of a circle under a cross," the relationships among the different shapes must be comprehended correctly. Another example is comprehending the main idea of a story or movie. In short, simultaneous processing involves *understanding and appreciating how the separate parts of a task result in a final product*. Whereas simultaneous processing involves working with stimuli that are interrelated, successive processing is important whenever actions or information form a chain-like progression.

Successive processing is the primary neurocognitive process used in the production of sequences of sounds used to make words, decoding of unfamiliar words, production of syntactic aspects of language, and speech articulation. Other examples of successive processing include following a sequence such as the order of operations in a math problem, and learning a new series of physical and cognitive actions (such as in a sport, a dance, or a board game). Initial learning of almost any new activity or task often requires the use of successive processing. Whereas simultaneous processing involves integration of separate elements into a cohesive whole, successive processing allows the learner to acquire the steps needed to solve a task.

Third Functional Unit

The third functional unit is associated with the prefrontal areas of the frontal lobes of the brain (Luria, 1980) and interacts with the networks already mentioned, as well as the fronto-parietal and the somatosensory networks. Luria stated that "the frontal lobes synthesize the information about the outside world . . . and are the means whereby the behavior of the organism is regulated in conformity with the effect produced by its actions" (1980, p. 263). This functional unit provides for the programming, regulation, and verification of behavior, and is responsible for behaviors such as asking questions, solving problems, and self-monitoring (Luria, 1973b). Other responsibilities of the third functional unit include the regulation of voluntary activity, conscious impulse control, and various linguistic skills such as spontaneous conversation. The third functional unit provides for the most complex aspects of behavior, including personality and consciousness (Das, 1980). The frontal lobes interact with posterior areas of the brain, establishing the fronto-parietal network. This network consists of the dorsolateral prefrontal cortex, anterior cingulate, anterior insula, caudate nucleus, and inferior parietal lobe. The left hemisphere's fronto-parietal network is responsible for internally guided behavior; the right hemisphere's is activated by external influences when situations or information are unfamiliar and require problem solving. From a network perspective, the frontal systems of the brain need to have reciprocal interactions with posterior cortices and subcortical regions to produce the most complex of human behaviors.

Additional support for Luria's initial conceptualizations is provided by research examining the PASS processes and brain functions. For example, Luria initially described simultaneous processing as a function of the occipito-parietal region, whereas he described successive processing as a function of a fronto-temporal region (each region with a bilateral location). Researchers from Japan (Okuhata, Okazaki, & Maekawa, 2009) studied the two processes via electroencephalography. They investigated patterns during six tasks of the CAS (Naglieri & Das, 1997a), three from the Simultaneous scale and three from the Successive scale. The results showed two significantly distinguishable patterns corresponding to the two types of processing. Both processes are localized in the posterior part of the brain, as Luria suggested. Similarly, McCrea (2007) showed that simultaneous processing is strongly dependent on occipito-parietal activity, whereas successive processing shows fronto-temporal specificity, with some evidence of interhemispheric coordination across the prefrontal cortex. McCrea's results provide support for the validity of two of the PASS processes. In addition, Christensen, Goldberg, and Bougakov (2009) provided a substantive summary of brain imaging research that supports both Luria's conceptualizations and the PASS processes.

Functional Units: Influences and Issues

In simplest terms, the three functional units involve four PASS processes (see Figure 6.2). The first functional unit involves attention, which assists with focus and resistance to distractions; the second functional unit involves simultaneous processing, which is used when thinking about how ideas or things go together, and successive processing, which is used to manage information or actions in a specific order; and the third functional unit involves planning, which is used when thinking about how to do something before or during an action.

Luria's organization of the brain into functional units also accounts for the interaction of brain structures with the environment. He stated that "perception and memorizing, gnosis and praxis, speech and thinking, writing, reading and arithmetic, cannot be regarded as isolated or even indivisible faculties" (1973b, p. 29). That is, it is not possible to identify a reading or writing spot in the brain; instead, a consideration of the concept of units of the brain that provide a function is necessary. Luria described the advantage of this approach:

> It is accordingly our fundamental task not to "localize" higher human psychological processes in limited areas of the cortex, but to ascertain by careful analysis which groups of concertedly working zones of the brain are responsible for the performance of complex mental activity; when contributions [are] made by each of these zones to the complex functional system; and how the relationship between these concertedly working parts of the brain in the performance of complex mental activity changes in the various stages of its development. (p. 34)

Activities such as reading and writing can be analyzed and linked as constellations of activities to specific working zones of the brain that support them (Luria, 1979, p. 141). Because the brain oper-

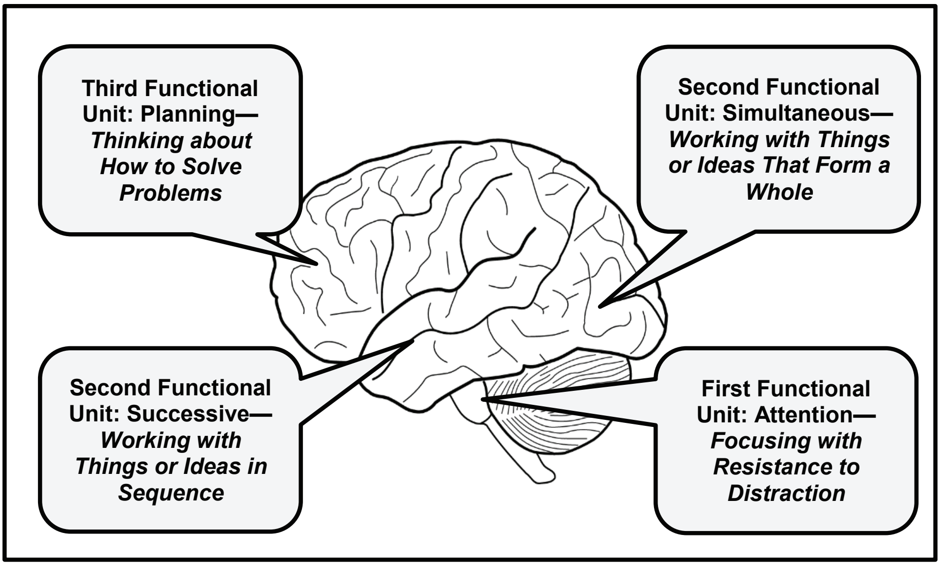


FIGURE 6.2. Luria's three functional units, the four PASS processes, and areas of the brain. Graphic by Jack A. Naglieri.

ates as an integrated functional system, dedicated networks have not been identified for domains typically assessed in psychological assessments. In fact, what we find are ever-changing patterns of dynamic brain network interactions. Disturbances in interaction patterns may cause disorganization in the entire functional system (Das & Varnhagen, 1986).

Luria's concept of dynamic functional units provides the foundation for the PASS processes. These basic neurocognitive processes are firmly based on biological correlates, yet develop within a sociocultural milieu. In other words, they are influenced in part by a person's cultural experiences. Luria (1979) noted that "the child learns to organize his memory and to bring it under voluntary control through the use of the mental tools of his culture" (p. 83). Kolb, Gibb, and Robinson (2003) have also noted that although "the brain was once seen as a rather static organ, it is now clear that the organization of brain circuitry is constantly changing as a function of experience" (p. 1). Similarly, Stuss and Benson (1990) recognized this interplay and especially the use of speech as a regulatory function when they stated:

> The adult regulates the child's behavior by command, inhibiting irrelevant responses. The child learns to speak, the spoken instruction shared between the child and adult are taken over by the child, who uses externally stated and often detailed instructions to guide his or her own behavior. By the age of 4 to 4½, a trend towards internal and contract speech (inner speech) gradually appears. The child begins to regulate and subordinate his behavior according to his/her speech. Speech, in addition to serving communication thought, becomes a major self-regulatory force, creating systems of connections for organizing active behavior inhibiting actions irrelevant to the task at hand. (p. 34)

Luria stressed the role of the frontal lobes in language, organization, and direction of behavior and speech as cultural tools that further the development of the frontal lobes and self-regulation. Cultural experiences thus actually help to accelerate the *utilization* of planning and self-regulation, as well as the other PASS processes. It follows, then, that lack of appropriate experiences that foster the development of speech and language can interfere with the use of planning and self-regulation (see Abdul Aziz, Fletcher, & Bayliss, 2017).

Luria (1979) also pointed out that abstraction and generalizations are themselves products of the cultural environment. Children learn, for example, to attend selectively to relevant objects through playful experiences and conversations with adults. Even simultaneous and successive processes are influenced by cultural experiences (e.g., learning songs, poems, or rules of games). Naglieri (2003) has summarized the influence of social interaction on children's use of plans and strategies, and the resulting changes in performance on classroom tasks.

The relationship between the third and first functional units is particularly strong. The first functional unit works in cooperation with, and is regulated by, higher systems of the cerebral cortex, which receive and process information from the external world and determine an individual's dynamic activity (Luria, 1973b). In other words, the first functional unit has a reciprocal relationship with the cortex. It influences the tone of the cortex and is itself influenced by the regulatory effects of the cortex. These influences are made possible through the ascending and descending systems of the reticular formation, which transmit impulses from lower parts of the brain to the cortex and vice versa (Luria, 1973b).

Functional connectivity (communication) among the functional units described by Luria can be further explained in terms of the current neuroscience literature. The term *functional integration* describes communication across distal brain regions (Koziol et al., 2016). This refers to "transient, dynamically changing, ongoing, and context specific regional interactions that convey information though a subset of anatomical connections among a limited handful of brain regions engaged by a particular cognitive process" (Koziol & Stevens, 2012). Therefore, the third and first functional units are characterized by regional functional segregation and specialization, as well as functional integration.

Facilitating communication across regions of the brain is achieved by the functional hubs, or brain regions that play a critical role in information sharing among distributed brain networks within all sensory and motor systems. Hwang, Hallquist, and Luna (2013) studied the development of functional hubs in 99 children, adolescents, and young adults. Connections among cerebellar, subcortical, and cortical regions increased with age. These increased connections highlight the reciprocal connectivity originally discussed by Luria. For PASS theory, this means that attention and planning are necessarily related strongly because attention is often under the conscious control of planning. That is, planning of behavior dictates the allocation of limited attentional resources. The interplay between attention and planning helps explain how these two components of PASS theory are related to executive function.

Although definitions of *executive function* vary considerably (Goldstein, Naglieri, Princiotta, & Otero, 2014), most theorists agree that it is necessary for purposeful behavior so that goals are achieved. The frontal lobes (especially the dorsolateral and ventromedial regions), in combination with midbrain structures in the basal ganglia and the cerebellum, are key to efficient executive functioning (Otero & Barker, 2014). A task that measures executive function should (1) be relatively unfamiliar, so that the examinee must develop a way to solve the problem; (2) require self-monitoring and error correction; (3) involve selective attention in settings where a well-learned response must be inhibited; and (4) draw upon methods of working with information that needs to be remembered over a short period of time. The CAS2 Planning and Attention tests include many of these attributes of executive function.

OPERATIONALIZATION OF PASS THEORY

The initial operationalization of PASS theory in the CAS (Naglieri & Das, 1997a) has undergone considerable experimental examination (for summaries of research, see Naglieri & Conway, 2009; Naglieri & Otero, 2011, 2017). To provide more ways to measure PASS constructs, several new measures were included in the CAS2 (Naglieri et al., 2014a), CAS2: Español (Naglieri et al., 2017), CAS2: Brief (Naglieri et al., 2014b), and CAS2: Rating Scale (Naglieri et al., 2014c) (see Chapter 15, Figure 15.1). These new measures take different amounts of time to administer, can be used in a wide variety of settings by professionals of different qualification levels, and can be employed for different purposes. The CAS2 (English and Spanish) are intended for highly trained assessment professionals (e.g., psychologists, school psychologists); the CAS2: Brief is appropriate for a wider range of assessment professionals; and the CAS2: Rating Scale can be used by assessment professionals as well as teachers who have been provided with appropriate training.

The main goal in the development of the CAS2 (English) and CAS2: Español, the CAS2: Brief, and the CAS2: Rating Scale is to provide several ways of applying PASS theory to gain a better understanding of students' learning and learning problems. The CAS2 provides the most complete examination of the four PASS neurocognitive abilities, as well as additional scales such as Executive Function, Working Memory, and others. This version will be used most often in comprehensive evaluations, typically in response to a referral for determining whether a student has a disorder in basic psychological processing that is affecting academic or social performance and for selecting interventions that address cognitive and academic needs. The CAS2: Brief is intended to be used in situations when a fast measure of PASS is needed for screening or reevaluation. The CAS2: Rating Scale provides a way for a user of the CAS2 or the CAS2: Brief to determine whether behaviors related to PASS processes have been observed by a teacher. This rating scale also provides a way to inform teachers about the PASS theory and the behaviors associated with the four neurocognitive abilities. We describe all of these measures in detail in Chapter 15 of this volume.

Operationalization of PASS theory in the CAS2 suite of measures was based on a careful analysis of the cognitive processing demands of the prospective subtests and behavioral items. Several guiding principles were established. First, the kind of thinking (i.e., planning, attention, simultaneous, successive) required to solve each subtest had to match the theoretical description of the respective neurocognitive construct. That is, the Planning subtests had to evaluate *how* the student completed the relatively simple task; the Attention subtests had to measure the extent to which the examinee could focus on the appropriate part of the stimulus and ignore distractions; the Simultaneous subtests had to require understanding of the way information presented in each item was related to information in that item; and the Successive subtests had to be sensitive to the student's ability to work with information arranged in a specific sequence.

Second, the PASS subtests had to require a minimum amount of knowledge. This meant that traditional IQ subtests such as Vocabulary, Information, Similarities, and Arithmetic on the WISC-V, for example, would not be included in the CAS2. The idea that a test of ability should include questions that are not contaminated with knowledge is not new. The originators of traditional IQ tests recognized the obstacles verbal tests can present for those with limited educational backgrounds. In fact, Yoakum and Yerkes (1920) stated that test questions that do not rely on knowledge are necessary, "in order that injustice by reason of relative unfamiliarity with English may be avoided" (p. 19).

Third, the PASS subtests had to be administered in a sequence that would maximize validity. For example, the CAS and CAS2 subtest administration follows the sequence of Planning, Simultaneous, Attention, and Successive. The Planning subtests come first in the sequence because the examinee is given freedom to complete the task in any way that seems best, as long as the basic instructions are followed (e.g., write XO under the number 1, OO under the number 2, and so on). The examiner's directions explicitly state, "You can do it any way you want," so that the student is given the opportunity to initiate a strategy to complete the task. This, according to Goldberg (2009), is an essential aspect of frontal lobe functioning, which the Planning tasks are intended to measure. The remaining sequence of CAS and CAS2 subtests was also carefully determined. Simultaneous subtests come second because, unlike the Planning subtests, the first two of these subtests do not involve paper and pencil and contain multiple-choice items. The Attention subtests are very structured and follow the Planning subtests, so that the overall test sequence moves from less to more structured. Finally, the Successive subtests come last because these take a short amount of time and require no paper-and-pencil activities. Fourth, the PASS subtests should vary based on their content. This improves interest, but also allows for assessing each PASS process across modalities (see Table 6.1).

Operationalization of Planning

Planning is a neurocognitive ability used to determine, select, and apply strategies to solve problems where initiation, self-monitoring, and self-correction are especially important (Naglieri & Otero, 2017). Planning is essential to success on tasks that require an awareness of the need for a solution, monitoring of how well things are going, consideration of alternative solutions, and judging whether continuing with a behavior or changing to a different one is most beneficial (Shadmehr, Smith, & Krakauer, 2010). To measure planning, a test question *must* allow a student to solve a novel problem for which there is no previously acquired strategy, and minimal constraints should be placed on the way the student chooses to complete the task. For example, all the Planning subtests on the CAS2 and CAS2: Brief (Naglieri et al., 2014a, 2014b) allow the examinee to decide how to complete the tasks, using whatever methods seem best. For this reason, the test scores reflect efficiency, measured by how long it takes to complete the task with the highest number of correct responses.

Observable behaviors included in the CAS2: Rating Scale can provide insight into a student's

TABLE 6.1. Content or Requirement of CAS2 and CAS2: Brief Subtests

	Visual	Auditory	Numbers	Letters/words	Motor	Memory
			CAS2			
Planned Codes	×			×	×	
Planned Connections	×		×	×	×	
Planned Number Matching	×		×		×	
Expressive Attention				×		
Number Detection	×		×		×	
Receptive Attention	×			×	×	
Matrices	×					
Verbal–Spatial Relations	×			×		
Figure Memory	×				×	×
Word Series		×				×
Sentence Repetition		×		×		×
Sentence Questions		×		×		×
Visual Digit Span	×		×			×
			CAS2: Brief			
Planned Codes			×		×	
Expressive Attention				×		
Simultaneous Matrices						
Successive Digits		×	×			×

use of planning. Observations of how well a student can solve new problems, and especially how well a student can think of several ways to solve the same problem, can give insight into how well a child is using planning processing. Good planning involves having a goal in mind when various strategies are being considered, applying a strategy, and deciding whether the result is consistent with the intention. The use of planning, however, can be influenced by the nature of the instruction. For example, the role of planning will be encouraged if instruction allows the child to consider multiple ways to solve problems. If, however, classroom instruction is very structured and each student is taught to use the same method of solving problems, then the role of planning will be suppressed.

Operationalization of Attention

Operationalizing the measurement of attention calls for tasks requiring a child to focus selectively on a particular stimulus while inhibiting responses to competing stimuli presented over time (Naglieri et al., 2014a). Luria stated that optimal conditions of arousal are needed for the more complex forms of attention involving "selective recognition of a particular stimulus and inhibition of responses to irrelevant stimuli" (1973b, p. 271).

The Attention subtests on the CAS2 and CAS2: Brief were constructed so that each stimulus is multidimensional and the task requires responding to the most salient part. For example, on one Attention subtest, the student is instructed to tell the color a word is printed in (red) and resist responding to the word (blue). This task is known as the *Stroop test,* named after John Ridley Stroop, who first published a paper about it in English over 80 years ago (Stroop, 1935). This kind of a task requires selective focus of attention over time and resistance to distraction. The CAS2 Attention subtests we describe in Chapter 15 of this volume demand focused, selective, sustained, and effortful activity.

Classroom behaviors can yield insights into a student's ability to attend to instruction and resist distractions over time. For this reason, the CAS2: Rating Scale includes items for teachers regarding how well a student can stay focused in the classroom. Poor attention is inferred when a teacher endorses items suggesting that a student can only work for a short period of time, has difficulty listening and following directions, and cannot concentrate except when distractions are minimal. It is important to note, however, that off-task behavior at home or the classroom can also indicate a failure in control of attention, which can be detected by a low score on the Planning scale of the CAS2.

Operationalization of Simultaneous Processing

The development of subtests to measure simultaneous processing called for test items requiring an examinee to integrate separate stimuli into a single whole or interrelated group (Naglieri et al., 2014a). Some of these tasks include visual–spatial demands; others require comprehension of grammatical relationships. The spatial aspect of simultaneous processing involves both the perception of stimuli as a group or whole and the formation of visual images. The grammatical dimension of simultaneous processing allows for the integration of words into ideas through the comprehension of word relationships, prepositions, and inflections, which are important for deriving meaning. The diversity of the stimuli in the Simultaneous processing subtests of the CAS2 and CAS2: Brief was important, to demonstrate that test question content (verbal vs. visual–spatial) and test requirement (memory vs. little memory) is secondary to the main goal—organization of item content into a coherent and meaningful whole. The Simultaneous processing subtests of the CAS2 and CAS2: Brief are described in Chapter 15 in this volume.

Classroom behaviors can also provide insight into a student's Simultaneous processing. For example, the CAS2: Rating Scale items ask if the student likes (1) to work with "hands-on" materials and visual–spatial tasks; (2) to draw designs, especially three-dimensional ones; and (3) to work with patterns. Simultaneous processing is also involved in reading a whole word and in understanding grammar, verbal concepts, and patterns in language. It is additionally involved in reading comprehension and following a discussion to obtain the overall concept or main point. Overall, whenever any task requires integrating many parts into an organized whole, simultaneous processing is involved.

Operationalization of Successive Processing

The development of the CAS2 Successive processing subtests required tasks involving working with information that is arranged in a specific serial order (Naglieri et al., 2014a). These subtests involve the perception of sequences, as well as the repetition of words, sentences, and numbers in a specific order. Successive processing is necessary for recall of information verbatim, as well as for phonological analysis and the syntax of language (Das, Naglieri, & Kirby, 1994). To ensure that the successive processing is measured across the auditory and visual modalities, a visual subtest has been added to the CAS2. All Successive processing tasks demand working with information in order; however, the sequencing of information is most critical to success on these subtests.

Classroom behaviors can provide information about a student's ability to work with information in order or sequentially. For example, the student's success with blending sounds in sequence (e.g., decoding unfamiliar words, spelling) is an indication of good successive processing. The CAS2: Rating Scale includes items that assist in understanding whether a child has difficulty with successive processing by asking questions about classroom and learning activities that are based on a strict order.

EMPIRICAL SUPPORT FOR THE THEORY

The fundamental validity of PASS theory is rooted in the neuropsychological work of Luria (1966, 1973a, 1973b, 1980), who associated areas of the brain with basic psychological processes as described earlier in this chapter. Luria's research was based on an extensive combination of his and other researchers' understanding of brain functions, amply documented in his book *The Working Brain* (1973b). Using Luria's three functional units as a foundation, Das and colleagues (Das, 1972; Das, Kirby, & Jarman, 1975, 1979; Das, Naglieri, & Kirby, 1994) initiated the task of finding ways to measure the processes associated with these brain areas. These efforts included extensive analysis of the methods used by Luria, related procedures used within neuropsychology, experimental research in cognitive and educational psychology, and related areas. The initial operationalization of Luria's conceptualization of basic psychological processes led to the research summarized in several books (e.g., Das, Naglieri, & Kirby, 1994; Kirby, 1984; Kirby & Williams, 1991; Naglieri, 1999; Naglieri & Das, 1997b, Naglieri et al., 2014a; Naglieri & Otero, 2011, 2017), demonstrating that the PASS processes associated with Luria's concept of the three functional units could be measured. This work also illustrated that the theoretical conceptualization of basic psychological processes had considerable potential for application. The publication of the CAS2 has provided additional evidence for PASS theory (Naglieri & Otero, 2017).

Relationships between PASS and Achievement

Explaining current academic strengths and needs, as well as predicting future achievement, is a critically important use of intelligence tests. The examination of the relationship between ability and achievement is, therefore, an essential aspect of validity. Examining the relationship between ability and achievement, however, is complicated by the fact that traditional ability tests often have content similar to that of achievement tests. It would seem reasonable that an ability test should measure something different than an academic achievement test does, but this is not always true. The verbal and quantitative portions of some ability tests are remarkably like questions found in achievement tests used to measure knowledge and skills. The similarity between ability and achievement test questions has been amply documented by Naglieri and Bornstein (2003) and Naglieri (2008).

All traditional IQ tests include measures of word knowledge, just as tests of achievement do. For example, examinees are required to define a word like *bat* on subtests included in the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003) and the WISC-IV and WISC-V (Wechsler, 2003, 2014), just as they are required to do in the Woodcock–Johnson Tests of Achievement (WJ III; Woodcock, McGrew, & Mather, 2001). The WJ III Cognitive battery contains a Verbal Comprehension subtest that has an item like this: "Tell me another word for *small*," and the WJ III Achievement battery contains this Reading Vocabulary question: "Tell me another word for *little*." In addition, an item on the WJ III Reading Vocabulary achievement test is "Tell me another word for [examiner points to the word *big*]," and in the Cognitive battery the examiner asks something like "Tell me another word for *tiny*." Additionally, both the WJ III Cognitive and Achievement batteries contain vocabulary tests. The WJ III Cognitive tests also require the subject to name as many examples as possible from a given category in a 1-minute time period, and the same question appears on the Oral Expression subtest of the Wechsler Individual Achievement Test—Second Edition (WIAT-II; Psychological Corporation, 2001). Although these examples do not constitute a complete list of item overlap, they do represent the most strikingly similar questions.

Test questions requiring math achievement also appear on ability and achievement tests. For example, the SB5 contains Quantitative Reasoning items, one of which requires the child to calculate the total number of stars on a page (e.g., two stars in one box plus four in a second box plus one in a third box). Similarly, the WISC-V Arithmetic subtest requires the child to count the number of butterflies pictured on a page, and the Information subtest asks questions like "How many days are there in a week?" Although the scores these test items yield are used to determine a child's level of intelligence, very similar items appear on the WIAT-II. For example, Numerical Operations on the WIAT-II requires the child to determine the total number of marbles shown (e.g., 3 + 5). Similarly, Applied Problems in the WJ III Achievement battery asks the child to count the number of crayons pictured on the stimulus book (e.g., 4). Additionally, an SB5 Quantitative Reasoning item requires the child to complete a simple math problem (e.g., 4 + 2 = ?), just as the WJ III Math Fluency (e.g., 5 + 2 = ?) and the WIAT-II Numerical Operations (e.g., 2 + 2 = ?) achievement tests do. There is an obvious problem when questions on tests of achievement and intelligence are so similar, yet the interpretations of the scores on these tests are considerably different. A student's knowledge of math or verbal skills should be used to understand academic achievement, but not to determine level of intelligence.

Using items with similar content across achievement and ability tests is ill advised, for several reasons. First, because the correlation between ability scores (especially verbal scores) and achievement scores has been considered a source of evidence for the validity of IQ tests, these correlations should be considered overestimates of the relationship, due to overlapping content. The authors and/or publishers of ability tests with this content should justify how similar questions can be used across tests that were designed for different purposes. Second, having ability test questions that require knowledge very similar to that required by achievement tests may increase their predictive validity artificially, but at a cost to those with limited educational backgrounds as well as those who are learning English.

Ability and Achievement Test Correlations

Therefore, it is important to understand how well traditional intelligence tests and the more modern tests designed to measure ability differently correlate with achievement. This question was examined by Naglieri (1999), who first reported that the correlations of achievement test scores with scores on the CAS (see Naglieri & Rojahn, 2004) and the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983b) were as high as or higher than those found for the WISC-III (Wechsler, 1991) and the WJ-R (Woodcock & Johnson, 1989). More recent findings are provided next.

Table 6.2 provides a summary of the correlations between several ability and achievement tests, based on published data. The method used to summarize the correlations was simple. The respective test manuals were consulted, and the average correlation (using Fisher *z* transformations) between the scale scores yielded by an ability test and a total achievement score was computed. To examine the findings for each ability test with and without the scales that clearly have academic content, two average correlations were obtained: one with all the scales of the particular ability test, and one excluding the scales with obvious academic content.

The data for the WISC-V and WIAT-III came from the WISC-V's technical and interpretive

TABLE 6.2. Average Correlations between Ability and Achievement Tests

Ability and achievement tests	Scales on each ability test	Scale correlation with total achievement	Average of all scales	Average of scales without achievement
WISC-V and WIAT-II (*n* = 201)	Verbal Comprehension	.74	.53	
	Visual Spatial	.46		.47
	Fluid Reasoning	.40		
	Working Memory	.63		
	Processing Speed	.34		
WJ IV Cognitive and WJ IV Achievement (*n* = 825)	Comprehension–Knowledge	.50	.54	
	Fluid Reasoning	.71		
	Auditory Processing	.52		
	Short-Term Working Memory	.55		.50
	Long-Term Retrieval	.43		
	Visual Processing	.45		
KABC-II and WJ III Achievement (*n* = 167)	Sequential/Gsm	.43	.53	.48
	Simultaneous/Gv	.41		
	Learning/Glr	.50		
	Planning/Gf	.59		
	Knowledge/GC	.70		
CAS and WJ III Achievement (*n* = 1,600)	Planning	.57	.59	
	Simultaneous	.67		
	Attention	.50		
	Successive	.60		

Note. WJ-IV scales: Comprehension–Knowledge = Vocabulary and General Information; Fluid Reasoning = Number Series and Concept Formation; Auditory Processing = Phonological Processing. All average correlations were obtained by using Fisher *z* transformations.

manual (Wechsler, 2014, Table 5.13). The relationship between ability and achievement with and without the influence of those portions of the WISC-V that clearly require verbal knowledge was examined by using the two procedures just mentioned. The average correlation of all five WISC-V scales with the WIAT-III Total Achievement score was .53, and the average of the WISC-V scales with the Verbal Comprehension Index excluded was .47. The latter correlation is likely a more accurate assessment of the relationship between ability and achievement because it eliminated overlapping content to the extent possible. The same approach was taken with data from the WJ IV Tests of Cognitive Abilities and Achievement (Schrank, McGrew, & Mather, 2014, Table 5.7) and the KABC-II (Kaufman & Kaufman, 2004b). The results were similar, regardless of the ability test or achievement test used (see Table 6.2).

The results reveal a clear pattern across the WISC-V, WJ III, and KABC-II: The correlation between scores on each of these tests and achievement scores was higher when the scales that demand verbal knowledge were included. The best explanation for why this pattern emerged is the similarity in content across the two kinds of tests. Some (e.g., Lohman & Hagan, 2001) argue that this is evidence of validity. It seems clear, however, that the correlations between achievement test scores and ability tests that demand knowledge of words and arithmetic are *artificially inflated* because of the shared content. The correlations between the scores from ability scales that do not require knowledge and the total achievement test score are more accurate estimates of the relationship.

There was no need to eliminate a scale from the CAS for this analysis because it does not include tests that demand so much knowledge. As seen in Table 6.2, the average correlation between the PASS scales and achievement was .59—higher than the correlations obtained both when achievement was excluded from all the other ability tests and when it was not excluded. The results are clear: Academic content can be substantially reduced in ability tests without compromising validity.

PASS Profiles

One of the most important uses of an ability test that has scales organized on a conceptual or theoretical basis is to detect a pattern of cognitive strengths and weaknesses that can help explain academic success and failure and contribute to a diagnosis of specific learning disability (SLD) (Naglieri & Otero, 2017) or some other disability. Practitioners have the choice to analyze the scales that represent conceptual or theoretical constructs, as well as the subtests that make up the scales. The analysis of subtest and scale variation on, for example, the Wechsler scales has been advocated by Kaufman (1994) and others (e.g., Sattler, 1988) as a way to identify a child's intellectual strengths and/or weaknesses. Information about strengths and weaknesses is then used to generate hypotheses that are integrated with other information, so that decisions can be made regarding eligibility, diagnosis, and treatment. This approach has been widely used, even though some have argued that subtest profile analysis does not provide useful information (e.g., Dombrowski & Watkins, 2013; McDermott, Fantuzzo, & Glutting, 1990). Naglieri (1999) and Naglieri and Otero (2017) have further proposed that subtest analysis is problematic because of limitations in subtest reliability and validity, as well as the paucity of research supporting the many interpretations which have been proposed for subtests (see WISC-V review by Naglieri, 2016). Scale-level, rather than subtest-level, analysis that is based on a theory can be used to identify a specific pattern of strengths and weaknesses to understand a student's learning difficulty. The pattern can then be used to guide eligibility decisions (see Naglieri, 2011) and interventions (Naglieri & Pickering, 2010).

In order to examine the patterns of strengths and weaknesses at the scale level on several measures of ability for students with SLD, attention-deficit/hyperactivity disorder (ADHD), and autism spectrum disorder (ASD), mean scores found in the technical manuals of the WISC-V, WJ III, KABC-II, and CAS have been summarized and are reported in Table 6.3. The findings must be considered with recognition that the samples were not matched on demographic variables across the various studies, the accuracy of the diagnoses may not have been verified, and some of the sample sizes were small. Notwithstanding these limitations, the findings provide important insights into the extent to which these various tests are likely to yield scalelevel profiles that are distinctive, theoretically logical, and relevant to instruction.

The results of this analysis are presented in two ways. First, a graphic display of the mean scores is provided in Figure 6.3, so that the patterns of strengths and weaknesses can be clearly understood. Second, the difference between each scale

TABLE 6.3. Scale Variability within Each Ability Test by Diagnostic Group

Test	Index or scale	ASD	SLD	ADHD
WISC-V	Verbal Comprehension	0.2	–2.0	1.5
	Visual Spatial	2.6	2.2	1.0
	Fluid Reasoning	4.1	1.4	1.3
	Working Memory	–2.6	–3.3	–1.5
	Processing Speed	–4.4	1.9	–2.1
WJ III	Comprehension–Knowledge	1.7	–0.7	–0.7
	Long-Term Retrieval	1.7	–4.7	–4.7
	Visual–Spatial Thinking	3.7	5.3	2.3
	Auditory Processing	4.7	4.3	3.3
	Fluid Reasoning	–1.3	2.3	5.3
	Processing Speed	–9.3	–3.7	–4.7
	Short-Term Memory	–1.3	–2.7	–0.7
KABC-II	Sequential/Gsm	1.6	–0.5	–1.0
	Simultaneous/Gv	–2.4	2.2	–1.9
	Learning/Glr	5.4	–1.6	1.5
	Planning/Gf	0.0	0.9	–0.3
	Knowledge/Gc	–4.6	–1.1	1.5
CAS	Planning	5.9	2.5	–7.4
	Simultaneous	3.0	1.9	2.2
	Attention	–9.0	2.9	1.3
	Successive	0.1	–7.4	3.9

and the average standard score for all the scales in each ability test was computed and is presented in Table 6.3. The results of the summary of scale profiles for the WISC-V, WJ III, KABC-II, and CAS suggest that some of these tests yield more distinct profiles than others do across the groups of children with SLD in reading decoding, ADHD, and ASD.

Specific Learning Disability

The scores for students with SLD in reading decoding across all scales on the WISC-V profile was lowest for the Working Memory Index, but there was little variability among the scores. The Working Memory Index was only 3.3 standard score points lower than the average or this group. The WJ III mean scores for students with SLD in reading decoding were all within the average range and showed some variability. The lowest score was for Long-Term Retrieval (a knowledge-dense scale). All the KABC-II scores were in the 80s, and there was little variability among the scales. The CAS profile showed the most variability across the four PASS scales, with a very low score of 82.9 on the Successive scale. These findings are consistent with the view that students with SLD in reading decoding also have considerable difficulty with tasks that involve sequencing of information (Das, Janzen, & Georgiou, 2007).

Reading researchers generally agree that phonological skills play an important role in early reading, and some have suggested this to be the major cause of reading disability in children (Stanovich, 1988; Wagner, Torgesen, & Rashotte, 1994). One of the most frequently cited articles in the field, by Torgesen, Wagner, and Rashotte (1994), proposes that phonological skills are causally related to normal acquisition of reading skills. Support for this claim can also be found in the relationship between prereaders' phonological scores and their reading development 1–3 years later (e.g., Bradley

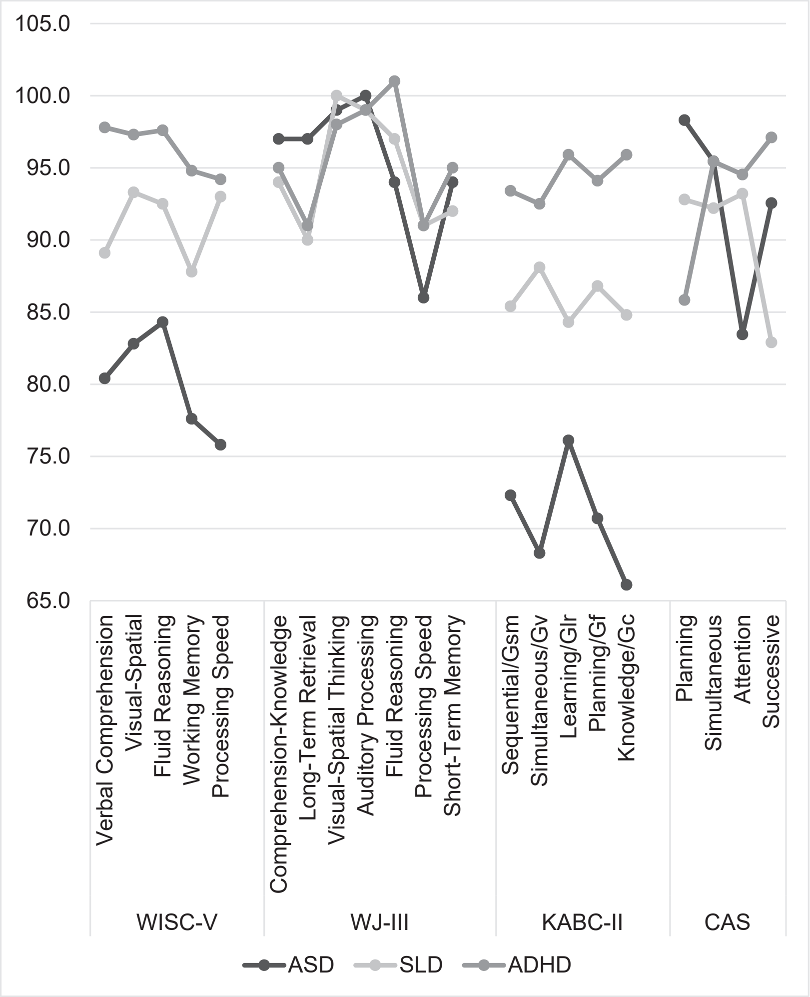


FIGURE 6.3. A graphic display of the mean scores on the WISC-V, WJ III, KABC-II, and CAS for groups of children with SLD in reading decoding, ADHD, or ASD.

& Bryant, 1985). Moreover, Share and Stanovich (1995) concluded that there is strong evidence that poor readers, as a group, are impaired in a very wide range of basic tasks in the phonological domain. Das, Naglieri, and Kirby (1994) suggested, however, that underlying a phonological skills deficit is a specific deficit in successive processing that leads to word-reading deficits.

The results of the current summary are consistent with previous research. Das, Mishra, and Kirby (1994) found that the Successive scale scores from the CAS were better than a test of phonemic segmentation at distinguishing normal readers from children with dyslexia. Additional studies have since supported the hypothesis that PASS processes, especially successive processing, are as important as phonological skills (Das, Parrila, & Papadopoulos, 2000). Several recent studies of Canadian First Nations children are particularly important. Das and colleagues (2007) reported that successive processing made a unique contribution to predicting both word identification and reading pseudowords (word attack). Furthermore, the poor readers demonstrated a significant weakness on the CAS Successive scale, both in relation to the norm and in relation to their scores on the other three CAS scales. Similarly, Naglieri, Otero, DeLauder, and Matto (2007) reported Successive scale deficits for bilingual children with reading disabilities who were administered the CAS in

English and Spanish. In addition, 90% of the children with reading disabilities had cognitive weaknesses on both the English and Spanish versions of the CAS.

In contrast to the relationship between reading decoding and successive processing, disability in reading comprehension has been shown to be primarily related to deficits in simultaneous processing (Das, Kar, & Parrila,1996; Das, Naglieri, & Kirby, 1994; Naglieri & Das, 1997b). In a study conducted with English-speaking children in India, Mahapatra, Das, Stack-Cutler, and Parrila (2010) found that children with comprehension problems had a substantially lower mean score on the CAS Simultaneous scale. These studies further suggest that PASS profiles could have utility for diagnosis of reading disabilities, as suggested by Naglieri (1999, 2011).

Attention-Deficit/Hyperactivity Disorder

The intelligence test profiles for students with ADHD showed that all the scores for the scales on the WISC-V, WJ III, and KABC-II were within the average range. There was some variability in the scores on the WJ III, with the lowest scores being on Long-Term Retrieval and Processing Speed tests (like the findings for the students with SLD). These findings suggest that none of these tests provided evidence of a cognitive problem related to ADHD. However, on the CAS, children with ADHD showed a low score on the Planning scale. Difficulty with planning (i.e., executive function) for children with ADHD is consistent with the view that ADHD is a disorder related to frontal lobe functioning (Goldberg, 2009). A low score on the CAS Planning, not Attention, scale reported in this summary may seem illogical, but it is consistent with Barkley's (1997) view that ADHD is a failure of self-control (i.e., planning in the PASS theory), rather than a failure of attention. The research summarized here confirms Barkley's view.

Naglieri, Salter, and Edwards (2004) were the first to report a weakness in planning, not attention, for children with ADHD. Canivez and Gaboury (2016) found that the CAS was accurate at distinguishing between students meeting *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, text revision (DSM-IV-TR) criteria for ADHD and matched controls. PASS scores were related significantly to ADHD characteristics, demonstrating both distinct group differences and diagnostic utility. Importantly, Naglieri and Conway (2009) found that PASS profiles for children with ADHD were different from those for children with SLD, as well as those with anxiety disorders. In summary, the findings of profile analysis suggest that determining whether a child with ADHD has a deficit in planning as measured by the CAS/CAS2 may be important for both diagnosis and intervention (Goldstein & Naglieri, 2006; Naglieri & Pickering, 2010).

Autism Spectrum Disorder

The results for individuals with ASD show that Processing Speed scores on the WISC-V and WJ III were relatively low for this group. The low Processing Speed scores provide little insight into the cognitive characteristics of these students because the interpretability of these scores on these batteries is unclear. For example, the WISC-V manual lists more than 40 abilities that may be involved in completing the three subtests that make up the Processing Speed Index. In contrast, the PASS profile, which shows a low Attention scale score, is consistent with the conceptualization of individuals with ASD as having "difficulties in disengaging and shifting attention" (Klinger, O'Kelley, & Mussey, 2009, p. 214). The findings for those with ASD, like the results for those with SLD and ADHD, show that the PASS processes as measured by the CAS result in cognitive profiles that may inform differential diagnosis in these three groups.

The research on scale-level profiles across the several ability tests has suggested that most tests' profiles look similar, except for the PASS scores from the CAS. As a group, children with SLD in reading decoding had a specific weakness on the Successive processing scale. This is consistent with the view of Das (see Das, 2009; Das, Naglieri, & Kirby, 1994) that reading failure results from a deficit in sequencing of information (successive processing). Those with ADHD performed poorly on the Planning scale, but adequately on the remaining PASS constructs. This finding is consistent with Barkley's (1997) and Goldberg's (2009) views of ADHD as a failure of self-control related to frontal lobe function. Finally, the PASS profile for individuals with ASD is consistent with Klinger and colleagues' (2009) description of these individuals as having difficulty with disengaging and shifting attention. Together, these findings support the view that PASS has relevance for understanding the neurocognitive processing components of these disorders, which in turn has implications for diagnosis and intervention.

It is important to note that Glutting, McDermott, Konold, Snelbaker, and Watkins (1998) suggested some time ago that research concerning profiles for children with specific disorders is confounded because the "use of subtest profiles for both the initial formation of diagnostic groups and the subsequent search for profiles that might inherently define or distinguish those groups" (p. 601). They suggested that researchers should "begin with unselected cohorts (i.e., representative samples, a proportion of which may be receiving special education), identify children with and without unusual subtest profiles, and subsequently compare their performance on external criteria" (p. 601). Naglieri (2000) followed this methodology, using PASS theory and the concepts of *relative weakness* and *cognitive weakness* (Naglieri, 1999). A relative weakness is found when a score is significantly lower than a student's mean PASS score; this is determined by using the ipsative methodology originally proposed by Davis (1959) and modified by Silverstein (1982, 1993). In contrast, a cognitive weakness is found when a child has a significantly low score relative to the student's average PASS score, *and* the low score also falls below some cutoff designed to indicate what is typical or average, perhaps 85 (16th percentile rank). The difference between a relative weakness and a cognitive weakness, therefore, is that the determination of a cognitive weakness is based on dual criteria (a low score relative to the child's mean, *and* a low score relative to the norm group). It is also important that a cognitive weakness be accompanied by an achievement test weakness comparable to the level of the PASS scale cognitive weakness. Children who have both a PASS cognitive weakness and an achievement test weakness should be considered candidates for special educational services if other appropriate conditions are met.

The utility of PASS profiles was studied by Naglieri (2000), using a nationally representative sample of 1,597 students ages 5–17 years. The sample included students in regular (*n* = 1,453) and special (*n* = 144) educational settings. Because these data were obtained before the original CAS was published, there was no way the results of the test influenced the groups in which these students were placed. Analysis of the PASS profiles suggested that the relative-weakness method identified children who earned average scores on the CAS as well as on achievement, and that approximately equal percentages of children from regular and special education classes had a relative weakness. That is, the concept of relative weakness did *not* identify children who achieved differently from children in regular education. By contrast, children with a PASS score significantly below the students' PASS average *and* below the average classification (i.e., a cognitive weakness) earned lower achievement scores. In addition, students with a PASS scale cognitive weakness were more likely to have been identified and placed previously in special education. Finally, the presence of a cognitive weakness was significantly related to achievement, whereas the presence of a relative weakness was not.

The findings related to relative weaknesses partially support arguments against the use of profile analysis (see Glutting et al., 1998, for a summary). However, the findings related to cognitive weaknesses support the PASS-theory-driven approach that includes the dual criteria of a relative and a normative weakness in a PASS process. These criteria are part of the discrepancy–consistency method, which is used to inform the use of the CAS/CAS2 to make a diagnosis of SLD (for details, see Naglieri, 1999, 2011; Naglieri & Otero, 2017). It is important to note that this approach is different from a subtest analysis approach because the method uses the PASS-theory-based scales included in the CAS/CAS2.

Racial/Ethnic Differences

The need for tests of ability to be appropriate for diverse populations has been recognized since Yoakum and Yerkes (1920) description of the Army Alpha and Beta tests. Fair assessment has become progressively more important as the characteristics of the U.S. population have changed. The issue is more than an aspiration with the requirement in the Individuals with Disabilities Education Improvement Act of 2004 that assessments must be nondiscriminatory. It is therefore critical that any measures used for evaluation of ability be evaluated for test bias. The psychometric analysis should include internal evidence such as reliability, item difficulty, and factor structure, as well as the presence of construct-irrelevant influences (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The issue of fairness is related most clearly to the theoretical perspective taken by the test authors.

Efforts to improve test fairness have led some researchers to suggest that conceptualizing intelligence in terms of neuropsychological abilities would make tests more appropriate for diverse populations (Fagan, 2000; Naglieri, 2005; Suzuki & Valencia, 1997). These authors have stressed the construct-irrelevant influences of verbal and quantitative tests included in traditional IQ tests, and have suggested that measures of cognitive processes that do not rely on tests with language and quantitative content are more appropriate for assessment of culturally and linguistically diverse populations. Although there is considerable evidence for the validity of general intelligence as measured by traditional IQ tests (see Jensen, 1980), researchers have traditionally found a mean difference of about 12–15 points between blacks and whites on measures of IQ that include verbal, quantitative, and nonverbal tests (Kaufman & Lichtenberger, 2006). Results for newer intelligence tests have been different.

The first evidence of smaller race differences for a test of ability that did not include traditional measures of vocabulary and math was reported in the original K-ABC interpretive manual (Kaufman & Kaufman, 1983a). For children ages 2.5 to 12.5, without controls for background variables, whites (*n* =1,569) scored 7 points higher than blacks (*n* = 807) and 3 points higher than Hispanics (*n* = 160) on the Mental Processing Index (MPI) (i.e., the total test score). These differences are considerably smaller than the differences of 16 points and 11 points, respectively, reported for the WISC-R (Wechsler, 1974) Full Scale IQ (Kaufman & Kaufman, 1983a, Tables 4.36 and 4.37; Kaufman, Lichtenberger, Fletcher-Janzen & Kaufman, 2005, Table 6.7). Naglieri (1986) examined this question further in a study of 172 fifth-grade students (86 whites and 86 blacks, matched on basic demographic variables) who were administered the K-ABC and the WISC-R. The difference between the groups on WISC-R Full Scale IQ was 9.1, but the difference for the K-ABC was 6.0. Results for the KABC-II (Kaufman & Kaufman, 2004a) showed a similar reduction in race/ethnic differences. When gender and mothers' education were controlled for, black children at ages 3–18 years earned mean MPIs that were only 5 points lower than the means for white children (Kaufman & Kaufman, 2004a, Tables 8.7 and 8.8; Kaufman et al., 2005, Table 6.7). Similar findings have been reported for the CAS.

Naglieri, Rojahn, Matto, and Aquilino (2005) compared PASS scores on the CAS for 298 black children and 1,691 white children. They found a CAS Full Scale mean score difference of 4.8 points in favor of white children when demographic variables were controlled for via regression analyses. In a similar examination of test fairness, Naglieri, Rojahn, and Matto (2007) compared PASS scores from the CAS of Hispanic and white non-Hispanic children. The study showed that the two groups differed by 6.1 points when unmatched samples were used, 5.1 when samples were matched on basic demographic variables, and 4.8 points when demographic differences were statistically controlled for. Researchers have also examined children with limited English-language skills.

The examination of PASS scores when the CAS: English and CAS: Spanish were administered to the same students provides an important complement to the large-scale studies of test fairness. Naglieri, Otero, and colleagues (2007) compared scores on the CAS when it was administered in English and Spanish to bilingual children (*n* = 40) referred for reading difficulties. They found a 3.0-point difference between the CAS Full Scale scores for the two versions, and these scores were highly correlated (.96). Otero, Gonzales, and Naglieri (2013) replicated that study with another group of students of varying levels of English-language proficiency referred for reading problems; they found CAS Full Scale scores that differed by less than 1 point, and a high correlation between the scores (.94).

Results for the CAS2 Full Scale scores were reported in the test manual (Naglieri et al., 2014a). Standard scores for African American and non–African American children and adolescents ages 5–18 years differed by 6.3 points without controls for demographic variables, and 4.5 points with such controls. Similarly, without controls for demographic variables, Hispanics and non-Hispanics differed on the CAS Full Scale scores by 4.5 points; with controls for demographic characteristics, the difference was 1.8.

The findings presented above for racial/ethnic differences are best understood when placed within the context of differences found on traditional intelligence tests. Table 6.4 provides a summary of standard score differences by race for the Stanford–Binet IV (SB-IV; Thorndike, Hagen, & Sattler, 1986), WJ III, WISC-V, K-ABC, KABC-II, CAS, and CAS2. The results for the WISC-V were reported by Kaufman, Raiford, and Coalson (2016); the results for the SB-IV were reported by Wasserman and Becker (2000); and the results for the WJ-III were reported by Edwards and Oakland (2006). The race differences for the K-ABC normative sample were reported by Kaufman and

TABLE 6.4. Mean Score Differences in Total Scores by Race by Intelligence Test

Test	Difference
SB-IV (matched samples)	12.6
WISC-V (normative sample)	11.6
WISC–IV (normative sample)	11.5
WJ- III (normative sample)	10.9
WISC–IV (matched samples)	10.0
WISC-V (statistical controls normative sample)	8.7
K-ABC (normative sample)	7.0
K-ABC (matched samples)	6.1
KABC-2 (matched samples)	5.0
CAS-2 (normative sample)	6.3
CAS (statistical controls normative sample)	4.8
CAS-2 (statistical controls normative sample)	4.3

Note. The data for these results are reported for the SB-IV from Wasserman and Becker (2000); for the Woodcock–Johnson III from Edwards and Oakland (2006); for the K-ABC from Naglieri (1986); for the KABC-II from Lichenberger, Sotelo-Dynega, and Kaufman, (2009); for the CAS from Naglieri, Rojahn, Matto, and Aquilino (2005); for the CAS2 from Naglieri, Das, and Goldstein (2014a); for the WISC-IV from O'Donnell (2009); and for the WISC-V from Kaufman, Raiford, and Coalson (2016).

Kaufman (1983a), and similar race differences for the KABC-II were summarized by Lichtenberger, Sotelo-Dynega, and Kaufman (2009). Differences for the CAS and CAS2 were reported by Naglieri and colleagues (2005) and Naglieri and colleagues (2014a), respectively.

The results provided in Table 6.4 illustrate that measuring ability as a cognitive process, in contrast to traditional concepts of IQ, provides a more equitable way to assess diverse populations. The findings suggest that as a group, traditional IQ tests showed differences in ability scores between the races that were about twice as large as those found for cognitive processing tests such as the K-ABC, CAS, and CAS2. All the tests with the largest differences between the racial/ethnic groups included in Table 6.4 have verbal and quantitative content demanding knowledge that is very similar to the knowledge required by standardized achievement tests (see Naglieri & Bornstein, 2003). The results suggest that using a cognitive processing approach to measure ability results in smaller racial/ethnic differences without a loss of (1) capacity to predict achievement or (2) in the case of PASS theory, sensitivity to learning problems—both of which are critical components of validity.

CLOSING STATEMENTS

In this chapter, we have described the neuropsychological and neuroscientific foundations of PASS theory. We have reviewed Luria's original conceptualizations of how the brain functions, and have integrated that understanding with current knowledge in neuroscience, which continues to validate Luria's original conceptualizations. Our underlying aim has been to illustrate through logic and research findings that PASS theory is parsimonious, yet powerful for assessing diverse populations. Our hope is to attract both new and seasoned practitioners by presenting a coherent theory that provides useful information about the neurocognitive abilities being measured to explain why people fail or succeed because we focus on how one thinks versus what one knows.

We have provided the rationale and empirical support for a revolutionary step in the field of intellectual assessment: an effort to move from IQ to a theory of human neurocognitive functioning. We see our approach as one that works *because* it departs considerably from traditional IQ, is grounded in a theory of how the brain works, and takes a neurocognitive approach instead of building upon the outdated foundation of the U.S. Army mental tests. We have presented scientific evidence that the PASS scores derived from the CAS/CAS2 (1) are more predictive of achievement test scores than the scores from any other ability test; (2) yield distinctive profiles for different children with different disabilities; (3) can be used for SLD eligibility determination in a manner consistent with federal law; (4) offer the most equitable way to measure diverse populations; and (5) can be readily used for instructional planning and interventions. But the most important advantage of PASS theory is that it assists us in understanding the underlying reasons for success and failure in school and life.

Our overarching goal has been to encourage practitioners to embrace the conceptualization of intelligence in terms of neurocognitive abilities, which can be instrumental in helping children achieve their greatest potential. PASS theory is an innovative way of thinking about intelligence, and the CAS2 is a way to measure PASS neurocognitive processes (as we describe in Chapter 15, this volume).

We suggest that practitioners manage the transition from traditional IQ tests to those that measure neurocognitive abilities with the assurance that an evolutionary step in our field is needed, given all that has been learned in the past 100 years. As a founding father of the United States, Thomas Jefferson, once said, "I am not an advocate for frequent changes in laws. But laws must go hand in hand with the progress of the human mind. As that becomes more developed, more enlightened, as new discoveries are made, new truths discovered and opinions change, institutions must advance also to keep pace with the times." Only though revolutionary change can we improve the evaluation of intelligence and better meet the needs of the children and adolescents we serve.

ACKNOWLEDGMENT

We acknowledge Luremy Fraga for her excellent feedback on selected portions of this chapter.

REFERENCES

Abdul Aziz, S., Fletcher, J., & Bayliss, D. M. (2017). Self-regulatory speech during planning and problem-solving in children with SLI and their typically developing peers. *International Journal of Language and Communication Disorders, 52*(3), 311–322.

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

Avram, M., Gutyrchik, E., Bao, Y., Pöppel, E., Reiser, M., & Blautzik, J. (2013). Neurofunctional correlates of esthetic and moral judgments. *Neuroscience Letters, 534,* 128–132.

Barkley, R. A. (1997). *ADHD and the nature of self-control.* New York: Guilford Press.

Bostan, A. C., Dum, R. P., & Strick, P. L. (2010). The basal ganglia communicate with the cerebellum. *Proceedings of the National Academy of Sciences USA, 107*(18), 8452–8456.

Bostan, A. C., Dum, R. P., & Strick, P. L. (2013). Cerebellar networks with the cerebral cortex and basal ganglia. *Trends in Cognitive Sciences, 17*(5), 241–254.

Bostan, A. C., & Strick, P. L. (2010). The cerebellum and basal ganglia are interconnected. *Neuropsychology Review, 20*(3), 261–270.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling.* Ann Arbor: University of Michigan Press.

Canivez, G. L., & Gaboury, A. R. (2016). Construct validity and diagnostic utility of the Cognitive Assessment System for ADHD. *Journal of Attention Disorders, 20*(6), 519–529.

Christensen, A., Goldberg, E., & Bougakov, D. (2009). *Luria's legacy in the 21st century.* New York: Oxford University Press.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58*(3), 306–324.

Das, J. P. (1972). Patterns of cognitive ability in nonretarded and retarded children. *American Journal of Mental Deficiency, 77,* 6–12.

Das, J. P. (1980). Planning: Theoretical considerations and empirical evidence. *Psychological Research, 41*(2), 141–151.

Das, J. P. (2009). *Reading difficulties and dyslexia: An interpretation for teachers* (rev. ed.). Thousand Oaks, CA: SAGE.

Das, J. P., Janzen, T., & Georgiou, G. (2007). Correlates of Canadian native children's reading performance: From cognitive styles to cognitive processes. *Journal of School Psychology, 45,* 589–602.

Das, J. P., Kar, B. C., & Parrila, R. K. (1996). *Cognitive planning: The psychological basis of intelligent behavior.* Thousand Oaks, CA: SAGE.

Das, J. P., Kirby, J. R., & Jarman, R. F. (1975). Simultaneous and successive syntheses: An alternative model for cognitive abilities. *Psychological Bulletin, 82,* 87–103.

Das, J. P., Kirby, J. R., & Jarman, R. F. (1979). *Simultaneous and successive cognitive processes.* New York: Academic Press.

Das, J. P., Mishra, R. K., & Kirby, J. R. (1994). Cognitive patterns of dyslexics: Comparison between groups with high and average nonverbal intelligence. *Journal of Learning Disabilities, 27,* 235–242.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn & Bacon.

Das, J. P., Parrila, R. K., & Papadopoulos, T. C. (2000). Cognitive education and reading disability. In A. Kozulin & Y. Rand (Eds.), *Experience of mediated learning* (pp. 276–291). Amsterdam: Pergamon Press.

Das, J. P., & Varnhagen, C. K. (1986). Neuropsychological functioning and cognitive processing. In J. E. Obzrut & G. W. Hynd (Eds.), *Child neuropsychology: Vol. 1. Theory and research* (pp. 117–140). New York: Academic Press.

Davis, F. B. (1959). Interpretation of differences among averages and individual test scores. *Journal of Educational Psychology, 50,* 162–170.

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school-aged analysis. *Psychological Assessment, 25*(2), 442–455.

Edwards, O. W., & Oakland, T. D. (2006). Factorial invariance of Woodcock–Johnson III scores for African Americans and Caucasian Americans. *Journal of Psychoeducational Assessment, 24,* 358–366.

Fagan, J. R. (2000). A theory of intelligence as processing: Implications for society. *Psychology, Public Policy, and Law, 6*(1), 168–179.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience, 25,* 221–250.

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. L. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 599–612.

Goldberg, E. (2009). *The new executive brain: Frontal lobes in a complex world.* New York: Oxford University Press.

Goldstein, S., & Naglieri, J. A. (2006). The role of intellectual processes in the DSM-V diagnosis of ADHD. *Journal of Attention Disorders, 10,* 3–8.

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 1–12). New York: Springer.

Hwang, K., Hallquist, M. H., & Luna, B. (2013). The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *NeuroImage, 82,* 208–225.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6,* 766–774.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983a). *Interpretive manual for the Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1983b). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Interpretive manual for the Kaufman Assessment Battery for Children—Second Edition (KABC-II).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Assessment Battery for Children—Second Edition (KABC-II).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Essentials of WAIS-III assessment* (3rd ed.) Hoboken, NJ: Wiley.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment.* Hoboken, NJ: Wiley.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V.* Hoboken, NJ: Wiley.

Kirby, J. R. (1984). *Cognitive strategies and educational performance.* New York: Academic Press.

Kirby, J. R., & Williams, N. H. (1991). *Learning problems: A cognitive approach.* Toronto: Kagan & Woo.

Klinger, L., O'Kelley, S. E., & Mussey, J. L. (2009). Assessment of intellectual functioning in autism spectrum disorders. In S. Goldstein, J. A. Naglieri, & S. Ozonoff (Eds.), *Assessment of autism spectrum disorders* (pp. 209–252). New York: Guilford Press.

Kolb, B., Gibb, R., & Robinson, T. E. (2003). Brain plasticity and behavior. *Current Directions in Psychological Science, 12,* 1–4.

Koziol, L. F., Barker, L. A., Joyce, A. W., & Hrin, S. (2014). Structure and function of large-scale brain systems. *Applied Neuropsychology: Child, 3*(4), 236–244.

Koziol, L. F., Beljan, P., Bree, K., Mather, J., & Barker, L. (2016). *Large-scale brain systems and neuropsychological testing: An effort to move forward.* Cham, Switzerland: Springer International.

Koziol, L. F., Budding, D. E., & Chidekel, D. (2013). *ADHD as a model of brain–behavior relationships.* New York: Springer.

Koziol, L. F., & Stevens, M. C. (2012). Neuropsychological assessment and the paradox of ADHD. *Applied Neuropsychology: Child, 1*(2), 79–89.

Lichtenberger, E. O., Sotelo-Dynega, M., & Kaufman, A. S. (2009). The Kaufman Assessment Battery for Children—Second Edition. In J. A. Naglieri & S. Goldstein (Eds.), *A practitioner's guide to assessment of intelligence and achievement.* Hoboken, NJ: Wiley.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test.* Itasca, IL: Riverside.

Luria, A. R. (1966). *Human brain and psychological processes.* New York: Harper & Row.

Luria, A. R. (1973a). The origin and cerebral organization of man's conscious action. In S. G. Sapir & A. C. Nitzburg (Eds.), *Children with learning problems* (pp. 109–130). New York: Brunner/Mazel.

Luria, A. R. (1973b). *The working brain.* New York: Basic Books.

Luria, A. R. (1979). *The making of mind: A personal account of Soviet psychology.* Cambridge, MA: Harvard University Press.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

Mahapatra, S., Das, J. P., Stack-Cutler, H., & Parrila, R. (2010). Remediating reading comprehension difficulties: A cognitive processing approach. *Reading Psychology, 31*(5), 428–453.

McCrea, S. M. (2007). Measurement of recovery after traumatic brain injury: A cognitive-neuropsychological comparison of the WAIS-R with the Cognitive Assessment System (CAS) in a single case of atypical language lateralization. *Applied Neuropsychology, 14,* 296–304.

McDermott, P., Fantuzzo, J., & Glutting, J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8,* 280–302.

Naglieri, J. A. (1986). WISC-R and K-ABC comparison for matched samples of black and white children. *Journal of School Psychology, 24,* 81–88.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.

Naglieri, J. A. (2000). Can profile analysis of ability test scores work?: An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly, 15,* 419–433.

Naglieri, J. A. (2003). Current advances in assessment and intervention for children with learning disabilities. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities: Vol. 16. Identification and assessment* (pp. 163–190). Greenwich, CT: JAI Press.

Naglieri, J. A. (2005). The Cognitive Assessment System. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 441–460). New York: Guilford Press.

Naglieri, J. A. (2008). Traditional IQ: 100 years of misconception and its relationship to minority representation in gifted programs. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 67–88). Waco, TX: Prufrock Press.

Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 145–172). Hoboken, NJ: Wiley.

Naglieri, J. A. (2015). One hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 295–316). New York: Springer.

Naglieri, J. A. (2016). Independent WISC-V test review: Theoretical and practical considerations. In A. S. Kaufman, D. Coalson, & S. Engi Raiford (Eds.), *Intelligent testing with the WISC-V* (pp. 663–668) Hoboken, NJ: Wiley.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psycheducational Assessment, 21,* 244–260.

Naglieri, J. A., & Conway, C. (2009). The Cognitive Assessment System. In J. A. Naglieri & S. Goldstein (Eds.), *A practitioner's guide to assessment of intelligence and achievement* (pp. 3–10). Hoboken, NJ: Wiley.

Naglieri, J. A., & Das, J. P. (1997a). *Das–Naglieri Cognitive Assessment System*. Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (1997b). *Das–Naglieri Cognitive Assessment System: Interpretive handbook*. Itasca, IL: Riverside.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014a). *Cognitive Assessment System—Second Edition*. Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014b). *Cognitive Assessment System—Second Edition: Brief*. Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014c). *Cognitive Assessment System—Second Edition: Rating Scale*. Austin, TX: PRO-ED.

Naglieri, J. A., Moreno, M., & Otero, T. M. (2017). *Cognitive Assessment System—Second Edition: Spanish Edition*. Austin, TX: PRO-ED.

Naglieri, J. A., & Otero, T. M. (2011). Cognitive Assessment System: Redefining intelligence from a neuropsychological perspective. In A. Davis (Ed.), *The handbook of pediatric neuropsychology* (pp. 320–333). New York: Springer.

Naglieri, J. A., & Otero, T. M. (2017). *Essentials of CAS2 assessment*. Hoboken, NJ: Wiley.

Naglieri, J. A., Otero, T., DeLauder, B., & Matto, H. (2007). Bilingual Hispanic children's performance on the English and Spanish versions of the Cognitive Assessment System. *School Psychology Quarterly, 22,* 432–448.

Naglieri, J. A., & Pickering, E. (2010). *Helping children learn: Intervention handouts for use in school and at home* (2nd ed.). Baltimore: Brookes.

Naglieri, J. A., & Rojahn, J. R. (2004). Validity of the PASS theory and CAS: Correlations with achievement. *Journal of Educational Psychology, 96,* 174–181.

Naglieri, J. A., Rojahn, J., & Matto, H. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence, 35,* 568–579.

Naglieri, J. A., Rojahn, J. R., Matto, H. C., & Aquilino, S. A. (2005). Black–white differences in intelligence: A study of the PASS theory and Cognitive Assessment System. *Journal of Psychoeducational Assessment, 23,* 146–160.

Naglieri, J. A., Salter, C. J., & Edwards, G. (2004). Assessment of children with ADHD and reading disabilities using the PASS theory and Cognitive Assessment System. *Journal of Psychoeducational Assessment, 22,* 93–105.

O'Donnell, L. (2009). The Wechsler Intelligence Scale for Children—Fourth Edition. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 153–190). New York: Wiley.

Okuhata, S., Okazaki, S., & Maekawa, H. (2009). EEG coherence pattern during simultaneous and successive processing tasks. *International Journal of Psychophysiology, 72*(2), 89–96.

Otero, T. M., & Barker, L. A. (2014). The frontal lobes and executive functioning. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 29–44). New York: Springer.

Otero, T. M., Gonzales, L., & Naglieri, J. A. (2013). The neurocognitive assessment of Hispanic English language learners with reading failure. *Applied Neuropsychology: Child, 2*(1), 24–32.

Plucker, J., & Esping, A. (2014). *Intelligence 101*. New York: Springer.

Psychological Corporation. (2001). *Wechsler Individual Achievement Test—Second Edition*. San Antonio, TX: Author.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Author.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33,* 89–108.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education, 1,* 1–57.

Silverstein, A. B. (1982). Pattern analysis as simultaneous statistical inference. *Journal of Consulting and Clinical Psychology, 50,* 234–240.

Silverstein, A. B. (1993). Type I, type II, and other types of errors in pattern analysis. *Psychological Assessment, 5,* 72–74.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological–core variable–difference model. *Journal of Learning Disabilities, 21,* 590–604, 612.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.

Stuss, D. T., & Benson, D. F. (1990). The frontal lobes and language. In E. Goldberg (Ed.), *Contemporary psychology and the legacy of Luria* (pp. 29–50). Hillsdale, NJ: Erlbaum.

Suzuki, L. A., & Valencia, R. R. (1997). Race/ethnicity and measured intelligence. *American Psychologist, 52,* 1103–1114.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27,* 276–286.

Uddin, L. Q., Supekar, K., & Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience, 7*(7), 458.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Wasserman, J. D., & Becker, K. A. (2000). Racial and ethnic group mean score differences on intelligence tests. In J. A. Naglieri (Chair), *Making assessment more fair: Taking verbal and achievement out of ability tests.* Symposium conducted at the annual meeting of the American Psychological Association, Washington, DC.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery—Revised.* Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement.* Itasca, IL: Riverside.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106,* 1125–1165.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

Zaytseva, Y., Gutyrchik, E., Bao, Y., Pöppel, E., Han, S., Northoff, G., et al. (2014). Self processing in the brain: A paradigmatic fMRI case study with a professional singer. *Brain and Cognition, 87,* 104–108.

CHAPTER 7

The Parieto-Frontal Integration Theory

Assessing Intelligence from Brain Images

Richard J. Haier
Rex E. Jung

No one argues that a student's intelligence is the only important factor for predicting academic success, but consistent empirical research shows that it is more important than family resources, motivation, or quality of schools and teachers. These other factors are not unimportant—but, for example, earnest and expensive efforts to improve schools and overcome other disadvantages have had no appreciable effect on academic achievement gaps. We believe that an understanding of the neuroscience of intelligence offers new insights and a new basis for discussing education reforms that include what we know about intelligence. That is the rationale for this chapter. We have not written a review of relevant studies, but we provide this overview of the relevance of brain imaging studies of intelligence to education.

HOW WE THINK ABOUT AND ASSESS INTELLIGENCE

Before getting to brain imaging, we start with a brief review of how intelligence is defined and assessed, and of the evidence that intelligence is the best predictor of academic success. Most intelligence research is based on the empirical observation that scores on all tests of mental abilities correlate positively with each other. This finding suggests that there is a common general mental ability underlying all mental abilities. This common ability is called the *g* factor, or *g* for short, and was first observed by Spearman (1904) more than 100 years ago. There is not universal agreement that *g* is the most important intelligence factor or the only way to account for common variance among different tests. Nonetheless, *g* is still the basis for most intelligence research, especially in the neurosciences.

The *g* factor typically is estimated, not measured. This distinction is a key point (Haier, 2014). There is no assessment of intelligence as a quantity like weight or distance. Intelligence quotient (IQ) points and other mental ability test scores have meaning only in relation to other individuals. Whereas 2 inches are exactly twice the length of 1 inch, and 20 pounds is exactly twice the weight of 10 pounds, a person with an IQ of 140 is not twice as smart as a person with an IQ of 70. That's why we most accurately interpret intelligence test scores as percentiles. The best intelligence tests provide a good estimate of the *g* factor, and the best research design includes multiple mental tests across a range of abilities, so a *g* factor can be extracted statistically. A *g* factor extracted from one battery of mental ability tests correlates almost perfectly with a *g* factor extracted from another battery, as long as both batteries sample a wide range of abilities and include a wide range of test scores (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004).

INTELLIGENCE AS THE STRONGEST PREDICTOR OF EDUCATIONAL ACHIEVEMENT

The strong evidence for the predictive validity of intelligence test scores and the *g* factor has been reviewed in many places (Haier, 2017; Hunt, 2011; Mackintosh, 2011). Here we are concerned with the question of how much variance in educational achievement can be attributed to intelligence. The answer depends somewhat on the age of students, but overall the weight of evidence from the most well-designed research studies shows that a high proportion of variance in educational achievement is due to intelligence. No single other variable accounts for more of the variance. For example, one classic study examined a battery of cognitive test scores for over 70,000 English school children at age 11, and their educational achievement in 25 subject areas 5 years later at age 16 (Deary, Strand, Smith, & Fernandes, 2007). The researchers used multiple measures of both mental ability and achievement, so they were able to extract a latent *g* score and a latent achievement score, based on the common variance among the respective batteries of tests. From a research design view, this is a powerful assessment method, and it is combined with the also powerful elements of an extremely large sample and a prospective timeline. Results were clear. The correlation between the two latent variables was .81, which accounts for about 66% of the variance (correlations squared estimate the amount of variance accounted for in one variable by the other). This estimate leaves a hefty 34% of variance for the influence of all other variables (effort, school quality, teacher effectiveness, family resources, etc.), but it verifies that a student's intelligence, defined as a general ability factor, is the single strongest predictor of academic achievement.

A similar study looked at two large samples (*N*'s = 2,520 and 4,969) that included both young and teenage students (ages 4–6 and 14–19). Different cognitive test batteries were used in each sample, and achievement tests in reasoning, math, and writing were administered (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). Overall, the correlation between a latent *g* factor and a latent achievement factor was .83, or 69% of the variance; but when analyzed by age, the correlations were even larger in the older students (ranging from .77 in the young to .94 in the teens; this range translates to 59–88% of the variance). These results essentially replicate those of the Deary and colleagues (2007) English student study with respect to the large amount of variance in achievement explained by a general intelligence factor.

Socioeconomic status (SES) is also widely recognized as a key variable for educational achievement, but SES measures typically are confounded with intelligence scores because the kinds of variables that contribute to SES (such as education and income) also are highly related to intelligence. There is a complex literature on this topic, but when both intelligence and SES variables are used in the same study to predict academic performance in large, diverse samples, the weight of evidence clearly favors the central role of intelligence. For example, a study of 155,191 college students showed that SAT scores, which are good estimates of the *g* factor (Frey & Detterman, 2004), predicted academic performance about the same with or without SES; SES added no additional predictive power (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009).

THE PARIETO-FRONTAL INTEGRATION THEORY

What about the variables that influence intelligence? There is a vast research literature on this question, but the weight of evidence clearly shows that genes play a major role (Haier, 2017), accounting for up to about 80% of variance by the teenage years. The overwhelming evidence for this finding has not been embraced or even much acknowledged in the field of education. This omission may be because of the erroneous belief that genes are completely deterministic and create absolute limitations for students. In fact, genes are probabilistic, and this fact creates positive and optimistic opportunities for education reform (Asbury & Plomin, 2014). One aspect of emerging findings from large, multinational cohorts is that there are genes in common for intelligence and for specific brain characteristics. A review of the genetic literature here is not necessary for understanding the promise of neuroimaging for assessing intelligence and identifying its neurobiological elements. Our contribution is a model of brain relationships to intelligence, the parieto-frontal integration theory, and this theory is the focus for the remainder of this chapter.

In 2007, we published a review of 37 imaging studies that correlated a variety of intelligence measures with functional brain variables like regional cerebral glucose metabolic rate and regional blood flow, and with structural variables like gray and white matter volumes (Jung & Haier, 2007). This review resulted in a model based on the brain areas most closely associated with intelligence across studies. Most of the salient areas were in the parietal and frontal lobes, so we called the model the *parieto-frontal integration theory* or PFIT. The model identified 14 areas labeled according to Brodmann area (BA) nomenclature (see Figure 7.1), and we discussed how efficient information flow among these areas might be related to individual differences in intelligence. Our model emphasized that brain areas relevant to intelligence are distributed throughout the brain, contrary to a popular view at the time that the frontal lobes alone are responsible for intelligence. Moreover, the fact that quantifiable neuroimaging variables correlated with intelligence test scores demonstrated that the scores are not meaningless artifacts, as some anti-IQ advocates had argued.

The PFIT regions suggest distinguishable information processing stages that contribute to intelligence. This is a summary of the proposed stages (Colom, Karama, Jung, & Haier, 2010):

1. Occipital and temporal areas process sensory information in the first processing stage: the extrastriate cortex (BAs 18 and 19) and the fusiform gyrus (BA 37), involved with recognition, imagery and elaboration of visual inputs, as well as Wernicke's area (BA 22) for analysis and elaboration of syntax of auditory information.
2. Integration and abstraction of the sensory information by parietal BAs 39 (angular gyrus), 40 (supramarginal gyrus), and 7 (superior parietal lobule) correspond to the second processing stage.
3. The parietal areas interact with the frontal lobes in the third processing stage, and this interaction underlies problem solving, evaluation, and hypothesis testing. Frontal BAs 6, 9, 10, 45, 46, and 47 are underscored by the model.
4. The anterior cingulate (BA 32) is implicated for response selection and inhibition of alternative responses, once the best solution is determined in the previous stage.
5. Finally, white matter, especially the arcuate fasciculus, is thought to play a critical role in reliable and efficient communication of information across these brain processing networks.

At the time we published this review, the PFIT was recognized as the most comprehensive model of brain–intelligence correlates (Hunt, 2007) and as a source of testable hypotheses about the putative brain networks involved. For example, the information-processing sequence from stages 1 to 4 might repeat itself many times over seconds or minutes during reasoning, and the sequence and/or the speed of progression through the sequence might differ among people of different intellectual abilities. Whether the same networks and sequences described men and women, or the young and the old, were also open questions. Research continues on these and other questions. Perhaps the most interesting question for this chapter is

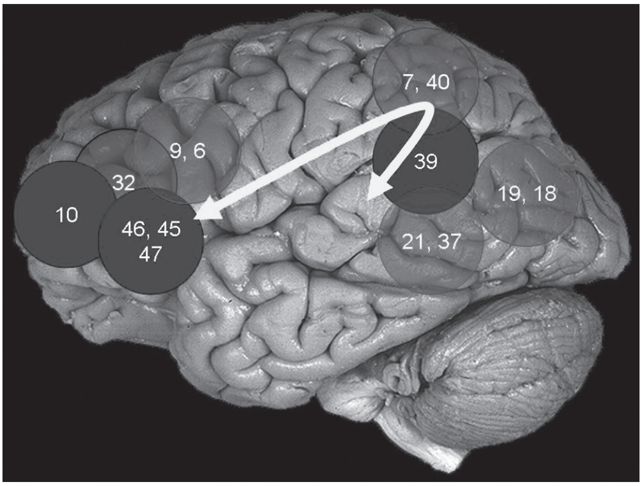


FIGURE 7.1. Brain regions by Brodmann area (BA) associated with better performance on measures of intelligence and reasoning that define the PFIT model. Numbers denote specific BAs; dark circles show left-hemisphere findings; light circles show bilateral findings. The arrow indicates the arcuate fasciculus white matter pathway. From Jung and Haier (2007). Copyright © Cambridge University Press. Reprinted by permission.

whether the PFIT could be used to predict IQ from brain images. There is a long, unsuccessful history of trying to do this, but there is new optimism, as we explain shortly below.

We have described the neuroimaging studies of intelligence from the first one in 1988 (Haier et al., 1988) up to the 2007 PFIT as Phase One in the attempt to link brain variables to intelligence (Haier, 2009). Phase Two has seen approximately 100 additional studies since 2007, and these have included much larger samples and advanced image acquisition and analysis techniques. Many of these studies have supported the PFIT and confirmed the distributed nature of brain areas related to intelligence, even in children (Kim et al., 2016) and in patients with brain lesions (Barbey, Colom, & Grafman, 2013; Barbey, Colom, Paul, & Grafman, 2014; Glascher et al., 2009, 2010). Most of the newer studies show that parietal and frontal areas have the strongest associations with intelligence. A group of German researchers replicated and extended our 2007 PFIT paper (Basten, Hilger, & Fiebach, 2015). Whereas the PFIT was based on a qualitative analysis, they used a quantitative method (voxel-based morphometry) to find common brain areas where structural and functional measures were related to individual differences in intelligence across 28 studies with a combined sample of over 1,000 participants. The findings supported a parieto-frontal network of areas related to intelligence and added some other areas that might be involved, pending replication.

Studies with newer imaging methods show that the integrity of white matter fibers throughout the brain is related to IQ (Penke et al., 2012a, 2012b). As noted, other studies that include imaging and genetic analyses show that there are likely genes in common for brain structures and intelligence. For example, a study of Australian twins reported that white matter integrity is most hereditable in frontal and parietal areas, and that IQ scores were correlated to some of these tracts (Chiang et al., 2009, 2011, 2012; Kohannim et al., 2012).

CAN INTELLIGENCE BE ASSESSED WITH NEUROIMAGING?

It had been apparent from the earliest studies that neuroimaging might provide a new means of assessing intelligence based on brain function and structure, in place of psychometric measures. The PFIT ushered in a basis for such attempts by identifying specific brain areas and networks to target. Despite all the studies showing such brain correlates of intelligence, however, the goal of predicting IQ from neuroimages has been elusive. Wide individual differences both in intelligence and in brain characteristics make prediction difficult. In fact, a number of attempts using multivariate statistical approaches have claimed some success, but until recently all have failed to be cross-validated in independent samples. A major problem is that not all brains work the same way, so a subset of areas important for intelligence in one person may not be the same as the subset for another person.

A powerful tool for brain imaging is graph analysis, which has opened the door to predicting intelligence from brain images (Stam & Reijneveld, 2007; van den Heuvel & Sporns, 2011; van den Heuvel, Stam, Kahn, & Pol, 2009). *Graph analysis* is a statistical method that establishes how any designated brain area (called a *node*) is connected to all other brain areas (analogous to how flights from your local airport are connected to all other airports around the world). The method can be applied to structural or function data (say, gray matter volume or white matter integrity or blood flow during a cognitive task). The essence of the method is to find correlations among many spots in the brain, to identify what areas systematically vary in their structural or functional connectivity with other areas. Networks of brain areas can be identified by using this method, and the number and strength of connections among areas can be quantified. Areas with many connections to other areas are called *hubs*. Some connections reach far across the brain, whereas other connections are more local. For example, an early study using graph analysis on 207 individuals found that IQ scores were related to connections among many areas, including ones identified in the PFIT (Santarnecchi, Galli, Polizzotto, Rossi, & Rossi, 2014).

There are now many studies that use graph analyses of one kind or another to investigate structural and functional connectivity across the entire brain and to examine how differences in connectivity among people relate to differences in cognitive abilities. A major study like this was published in 2015 as part of the Human Connectome Project, a multinational collaboration (Finn et al., 2015). In our view, it is the first study including an independent sample for cross-validation to predict IQ from functional magnetic resonance imaging (fMRI). This research is a remarkable achievement. Here is what the researchers did (as described by Haier, 2017, p. 182):

> They started with fMRI data from 126 people collected during six sessions, including four task and

> two resting conditions. The typical analysis would have compared the average connectivity for the entire group among the task and rest conditions. These researchers, however, focused on individual differences. The simple question was whether connectivity patterns were stable within a person. To address this question, functional connectivity patterns among 268 brain nodes (making up 10 networks) were calculated for each person separately for each session. Not only was the connectivity pattern stable within a person when the two resting conditions were compared, it was also stable across the four different tasks. In addition, each person's pattern was unique enough that it could be used to identify the person. Because these remarkable results combined stability and uniqueness, the connectivity pattern was characterized as a brain fingerprint. Of particular interest to us, individual brain fingerprints predicted individual differences in fluid intelligence. It gets even better. The strongest correlations with fluid intelligence were in frontoparietal networks. And, best of all, cross-validation was included in the report. The authors note, "These results underscore the potential to discover fMRI-based connectivity 'neuromarkers' of present or future behavior that may eventually be used to personalize educational and clinical practices and improve outcomes." They conclude, "Together, these findings suggest that analysis of individual fMRI data is possible and indeed desirable. Given this foundation, human neuroimaging studies have an opportunity to move beyond population-level inferences, in which general networks are derived from the whole sample, to inferences about single subjects, examining how individuals' networks are functionally organized in unique ways and relating this functional organization to behavioral phenotypes in both health and disease."

Haier (2017, pp. 118–119) speculates on a possible use of these findings:

> Imagine if colleges and universities gave applicants for admission a choice between submitting either SAT scores or a brain image . . . SAT scores are a ood estimate of general intelligence and that is an important reason they are good predictors of academic success. Can a better estimate of intelligence or predictor of academic success be extracted from a brain image? This is an empirical question, and a positive answer is probably far less scary than you might think. In fact, brain images are likely to be more objective, especially structural images, and not sensitive to a host of factors that potentially can influence psychometric test scores, like motivation or anxiety. Whether you are a good test-taker or not is irrelevant for getting a scan. Brain images generally are less expensive than SAT preparation courses or formal IQ testing and getting a brain image is far less time consuming. There is no preparation, you spend about 20 minutes in the scanner, and you can have a nap during structural image acquisition.

Of course, more research is required to confirm the brain fingerprint story and its relationship to intelligence. In addition to predicting intelligence from brain images, it might be possible to develop a new definition of intelligence based on quantified brain connectivity variables. Measurement of these brain characteristics might replace traditional intelligence tests if neuroimaging-based assessment is found to predict academic and life outcomes better than IQ tests currently do.

Predicting and defining intelligence from neuroimages are key steps toward understanding how the brain and intelligence are related. Understanding the influence of specific genes and networks of genes on intelligence may be potentially more important. Inevitably, advances in this understanding can lead to ways of influencing brain mechanisms that underlie intelligence. It might even be possible to dramatically increase intelligence through methods based on neuroscientific advances, but developing this technology would be a long and difficult endeavor. Understanding the complex cascade of genetic and neurobiological events that underlie the *g* factor and other aspects of intelligence is a formidable task. But the breathtaking sweep of advances in intelligence research over the last four decades suggests that this task is not impossible.

CONCLUSION

Remember that these exciting advances increase our understanding of the single most predictive variable for education and life success. Perhaps increasing intelligence is the key to eliminating achievement gaps. Perhaps neuroimaging assessments will help realize the goal of individualized education based on a student's cognitive strengths and weaknesses. We don't know, but the future of intelligence assessment is already moving beyond psychometrics and deep into the brain itself. And that will change everything.

REFERENCES

Asbury, K., & Plomin, R. (2014). *G is for genes: The impact of genetics on education and achievement*, Chichester, UK: Wiley-Blackwell.

Barbey, A. K., Colom, R., & Grafman, J. (2013). Architecture of cognitive flexibility revealed by lesion mapping. *NeuroImage, 82*, 547–554.

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure and Function, 219*, 485–494.

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence, 51,* 10–27.

Chiang, M. C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Johnson, K., Montgomery, G. W., et al. (2012). Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *Journal of Neuroscience, 32,* 8732–8745.

Chiang, M. C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., et al. (2009). Genetics of brain fiber architecture and intellectual performance. *Journal of Neuroscience, 29,* 2212–2224.

Chiang, M. C., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Hickie, I., Toga, A. W., et al. (2011). Genetics of white matter development: A DTI study of 705 twins and their siblings aged 12 to 29. *NeuroImage, 54,* 2308–2317.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience, 12,* 489–501.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35,* 13–21.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience, 18,* 1664–1671.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or *g*?: The relationship between the Scholastic Assessment Test and general cognitive ability (Vol. 15, pg. 373, 2004). *Psychological Science, 15,* 641.

Glascher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., et al. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences USA, 107,* 4705–4709.

Glascher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, A., et al. (2009). Lesion mapping of cognitive abilities linked to intelligence. *Neuron, 61,* 681–691.

Haier, R. J. (2009). Neuro-intelligence, neuro-metrics and the next phase of brain imaging studies. *Intelligence, 37,* 121–123.

Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in Systems Neuroscience, 8,* 34.

Haier, R. J. (2017). *The neuroscience of intelligence.* New York: Cambridge University Press.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., et al. (1988). Cortical glucose metabolic-rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12,* 199–217.

Hunt, E. (2007). P-FIT: A major contribution to theories of intelligence. *Behavioral and Brain Sciences, 30,* 158–159.

Hunt, E. B. (2011). *Human intelligence.* New York: Cambridge University Press.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence, 32,* 95–107.

Jung, R. E., & Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30,* 135–154.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*?: An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123–138.

Kim, D.-J., Davis, E. P., Sandman, C. A., Sporns, O., O'Donnell, B. F., Buss, C., et al. (2016). Children's intellectual ability is associated with structural network integrity. *NeuroImage, 124*(Pt. A), 550–556.

Kohannim, O., Jahanshad, N., Braskie, M. N., Stein, J. L., Chiang, M.-C., Reese, A. H., et al. (2012). Predicting white matter integrity from multiple common genetic variants. *Neuropsychopharmacology, 37,* 2012–2019.

Mackintosh, N. J. (2011). *IQ and human intelligence.* New York: Oxford University Press.

Penke, L., Maniega, S. M., Bastin, M. E., Hernandez, M. C. V., Murray, C., Royle, N. A., et al. (2012a). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry, 17,* 1026–1030.

Penke, L., Maniega, S. M., Bastin, M. E., Hernandez, M. C. V., Murray, C., Royle, N. A., et al. (2012b). Brain-wide white matter tract integrity is associated with information processing speed and general intelligence. *Molecular Psychiatry, 17,* 955.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135,* 1–22.

Santarnecchi, E., Galli, G., Polizzotto, N. R., Rossi, A., & Rossi, S. (2014). Efficiency of weak brain connections support general cognitive functioning. *Human Brain Mapping, 35,* 4566–4582.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics, 1,* 3.

van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience, 31,* 15775–15786.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience, 29,* 7619–7624.

CHAPTER 8

Intelligence-as-Process, Personality, Interests, and Intelligence-as-Knowledge

A Framework for Adult Intellectual Development

Phillip L. Ackerman

Historically, intellectual assessment of adults has presented several critical challenges to the scientific community. These challenges were first encountered when the scales originally created by Binet and Simon (1905/1916) for children were adapted for testing of adolescents and adults. The results of these assessments yielded confusion and controversy, but eventually were interpreted to mean that middle-aged and older adults, on average, have lower levels of intelligence than the average 18- to 21-year-old. However, it later became clear that the concept of *mental age* is not readily applicable to adults. More importantly, some researchers realized by the middle of the 20th century that the intelligence of children and intelligence of adults differ in important ways, which render traditional IQ-type scores somewhat ambiguous in terms of interpreting an adult's intellect. These ambiguities compromise the validity of predicting an adult's ability to perform intellectually demanding tasks. The *intelligence-as-process, personality, interests, and intelligence-as-knowledge* (PPIK) framework expands on the notions introduced by Cattell (1971, 1987) in his *investment hypothesis*. This framework enables us to understand adult intellect in general, and to consider how the interactions among process-type intellectual abilities; non-ability traits, such as affective (personality) traits and conative (interests, motivation) traits; and an individual's investment of time and effort lead to the development and maintenance of domain knowledge in adulthood. Evidence from this line of inquiry suggests that adults are, on average, far more intelligent than traditional IQ-based assessments suggest. Moreover, the development of intelligence-as-knowledge in late adolescence and adulthood appears to be related to an individual's standing on a relatively small number of constellations of non-ability traits, called *trait complexes*.

BACKGROUND

Binet and Simon

The seminal work by Binet and his colleagues in the late 1800s and early 1900s, which ultimately resulted in the Binet and Simon (1905/1961, 1908/1961) scales, provided a formalized system for the psychological assessment of intelligence that, with relatively few significant changes over the subsequent decades, is still the dominant approach for assessing child intellect.

Binet's major insight was that intelligence *develops* during childhood, such that older children, on average, could be expected to be more intelligent than younger children. This insight was, of course, a substantial departure from the notion of "fixed" intelligence put forth by Spearman (e.g., see

Spearman, 1904) and other psychologists. Binet used this insight to design assessment scales that were age-norm-referenced, so that a child's intelligence was reported specifically with reference to other children of the same age or to those younger or older than the child being assessed. Two other aspects of the Binet–Simon intelligence assessments are also important to note here, as they affected how later developments were applied to the assessment of adult intelligence. First, Binet and Simon (1905/1961) differentiated what they called their "psychological method" of assessing intelligence from two others: the "medical method," which attempts to determine a child's intelligence through examination of the individual's physical capabilities or appearance; and the "pedagogical method," which determines a child's intelligence though assessment of what the child knows (i.e., the assessment of knowledge).

The reason Binet and Simon rejected the medical method is somewhat obviously a question of invalidity: Simply looking at the slope of a child's forehead, or at how closely set his or her eyes are, provides no useful prediction of the child's academic success. But the reason Binet and Simon rejected the pedagogical method of assessing intelligence was one primarily aimed toward socioeconomic equity. They believed that assessing a child's ability to read and write, for example, would be influenced by the class of the child's family: Children in affluent families would be more likely to have access to books, or to have instruction or intellectual experiences outside the classroom that would lead to greater knowledge, when compared to children from less affluent families. So, for example, the items that appear on the Binet–Simon scales (and indeed on subsequent tests inspired by the Binet–Simon scales) do not require that a child know how either to read or write.

When knowledge was assessed in the Binet–Simon scales, the content was limited to those domains representing information that would be widely available to children across a large cross-section of all the socioeconomic strata in a particular culture (defining common words, recognizing common objects from pictures, etc.). The central theme, then, for the Binet–Simon "psychological method" of intelligence assessment, in contrast to the pedagogical method, was to limit knowledge assessment to information that one could expect most (if not all) children to have had substantial exposure to and experience with.

The other aspect of the Binet–Simon framework for intelligence assessment was the testing procedure's requirement that the child put forth his or her *maximal* effort during the test. Binet and Simon recommended individualized approaches to elicit such effort, such as supportive encouragement or strict authority, depending on what the examiner determined was necessary to maximize the child's motivation for performance on the test. Thus the general assumption was that a child's intelligence test performance represented what the child was *capable* of doing intellectually, not necessarily what the child was *likely* to do in the classroom.

The key aspects of the Binet-Simon approach are as follows:

- Preference for the "psychological method" of assessing intelligence of children over the "pedagogical method."
- Specification that intelligence increases with age.
- Assessment of maximal performance rather than typical effort.

Early Studies of Adult Intelligence

The first sign of a conceptual problem in extending the testing inspired by Binet and Simon was essentially a controversy that played out in the pages of *The New Republic* after World War I. Ironically, if Terman (1916) hadn't modified the Binet–Simon procedure of simply reporting that a child was "advanced" or "retarded" in intelligence, in comparison to his or her age peers, by adopting Stern's (1914) *intelligence quotient,* the controversy might have been much less severe than it turned out to be for the community of psychologists involved in intelligence testing. Basically, Binet's approach did not imply that a child's relative standing with respect to similar-age peers was going to be consistent from one year to the next. (In fact, there is ample reason to believe from Binet and Simon's writings that they did not believe in fixed relative standing on intelligence; see Fancher, 1985.) Stern, and later Terman, believed that a child's *relative* standing was indeed constant, such that from a single testing occasion, one could predict the child's relative standing years later. The initial formulation of the IQ (mental age divided by chronological age) by Stern, later modified by Terman by multiplying this value by 100, inevitably yielded scores for adults that made little sense, given that it was not reasonable to expect that the development of intelligence from childhood to adulthood would follow a linear path. Thus it made no sense to talk of a chronological 40-year-old with the mental age of a 20-year-old, or a 20-year-old with the mental age of a 40-year-old. But when the Stanford–Binet was extended to adolescents and adults (Terman, 1916), and administered to a small sample of U.S. Army conscripts in World War I (with comparison to the Army Alpha tests; see Yerkes, 1918), the conclusion was that the average conscript had a mental age of 13 (e.g., see Brigham, 1922). Walter Lippmann (1922) made light of this particular result, and psychologists like Terman (1922) and Boring (1923) attempted to defend the result *as if* such results were meaningful in and of themselves (though see Freeman, 1922, for a more coherent discussion).

Fortunately, some investigators, such as Conrad, Jones, and Hsiao (e.g., see Conrad, 1930; Hsiao, 1927; Jones & Conrad, 1933), decided to focus not on adult IQ per se, but on raw scores from the individual component scales that made up the overall Army Alpha test. What they found, in cross-sectional studies of individuals between 10 and 60 years of age, was that although overall scores were highest among 20-year-olds, two scales of the test showed a strikingly different pattern of results. These scales were Synonym/Antonym and general Information, two scales tapping an individual's vocabulary and general knowledge. For these scales, there was no pattern of lower scores among older individuals. In hindsight, possibilities of *cohort differences*—that is, the fact that the 60-year-olds tested by Conrad and colleagues (e.g., Jones & Conrad, 1933) had been born in an earlier time, with different exposures to language, education, news media, and so on—would have made it impossible to say that by the time the 20-year-olds in the sample reached the age of 60, they would perform as well as or better than the 60-year-olds in the sample (for a discussion of this issue, see Schaie & Strother, 1968).[1]

Nonetheless, there is a fundamental question about how to interpret these particular results, in comparison to the overall pattern of lower average IQ scores for adults beyond the age of 20. Jones and Conrad (1933), for example, concluded that intelligence did decline rapidly after age 20, but the Synonym/Antonym and Information scales "present an unfair advantage to those in the upper age brackets" (p. 271). There are, in retrospect, quite different alternative interpretations. One interpretation might be that what it means for adults to be "intelligent" is different from what it means for children to be "intelligent," and that the scales inspired by Binet and Simon's work were not assessing adult intelligence. After all, the Binet–Simon scales were designed specifically to predict academic success in early childhood years, and the day-to-day intellectual demands on typical 40-year-olds may diverge substantially from those of the elementary school classroom. Another interpretation might be that vocabulary and general knowledge are indicative of adult intelligence, while the other scales are not germane to the adult intellectual experience—with the implication being that middle-aged adults are, on average, just as intelligent as adolescents. Much later, Demming and Pressey (1957) examined adults on a variety of different tasks (e.g., using a telephone directory, getting professional assistance, etc.), and found that middle-aged adults performed much better than young adults or adolescents. Their conclusion was that measures of intelligence that were designed to assess the kinds of tasks more central to an adult's life—many of them centered on an individual's knowledge—were more appropriate for assessing adult intelligence.

Thus the key issues from early studies of adult intelligence are as follows:

- Overall, the intelligence assessments based on or extended from the Binet–Simon scales indicated that intelligence peaks at about age 18.
- The concept of mental age cannot be readily applied to adult intelligence, partly because of the pattern of slowing development as adolescents make the transition to adulthood.
- At the individual scale level (especially in the domains of verbal ability and general knowledge), middle-aged adults may perform as well as or better than adolescents and young adults.

Hebb and Cattell

Donald Hebb (1939, 1942), a neuropsychologist, was the first researcher to come to the conclusion that there are two fundamentally different aspects of adult intelligence. He based his reasoning on clinical examinations of patients who had experienced removal of brain tumors or other surgical losses of brain tissue. Hebb also noted that such brain traumas have different effects on intelligence, depending on whether the trauma is experienced relatively early in life or later in adulthood. Neurological trauma during infancy or childhood, when intelligence ordinarily develops at a rapid rate, has widespread negative effects on intellectual functioning. But for adults, for whom intellectual growth has largely reached plateau levels, some components of intellectual functioning may be substantially compromised by significant loss of neural tissue, while other components appear to be largely preserved. Hebb named these two components *Intelligence* A and *Intelligence B*. Intelligence A was identified as "direct intellectual power" (Hebb, 1942)—and it was hypothesized to represent the aspect of intelligence most involved in new learning, abstract reasoning, and similar aspects that make up the "process" aspect of general intellectual ability. In contrast, Intelligence B was thought to involve "the establishment of routine modes of response to common problems" (p. 289). For Hebb, Intelligence B was more than just an individual's repertoire of knowledge; he viewed it as including vocabulary and verbal comprehension abilities that are essential in an individual's day-to-day occupational functioning, but that also represent skills and procedures acquired through long-term exposure and experience.

Shortly after Hebb's introduction of the Intelligence A and Intelligence B constructs, Cattell (1943), drawing on the psychometric literature, introduced his constructs of fluid intelligence (g_f, as he abbreviated it, or Gf as it is now known) and crystallized intelligence (g_c or Gc). As Cattell noted, Hebb's framework accounted for "two-thirds of [Cattell's] present theory" (p. 179). Cattell's initial theory added to Hebb's framework by integrating research from studies of normal aging, and by hypothesizing a developmental trajectory related explicitly to the growth of Gc out of the investment of Gf during childhood and adolescence (Cattell, 1971, 1987).

Both Hebb (1942) and Cattell (1943) rejected the conclusions of psychologists such as Jones and Conrad, who, as discussed earlier, argued that if intelligence scales showed little or no decline in adulthood, then such tests were not adequate measures of adult intelligence. Instead, Hebb and Cattell argued that it was critical to consider that adult intelligence is composed of both "process"-type abilities that evidence declines in adulthood, and "content"-type abilities that show stable or somewhat increasing scores throughout much of adulthood. Cattell's framework also provided an explicit explanation for why large increases in Gc do not ordinarily take place in adulthood, in that Gf peaks in late adolescence, and that Gf is important for development of further gains in Gc. However, Cattell did not take full account of the importance of *transfer* in the development of new knowledge—an important contribution made later by Ferguson (1954, 1956). That is, Ferguson noted that learning, in the absence of transfer of training, only takes place with infants; by the time an individual has reached adulthood, acquisition of new knowledge and skills is critically dependent on the transfer of knowledge and skills that have already been acquired (Gc). Thus, especially for older individuals, it is most likely that the efficacy of new learning is determined both by individual differences in Gf and by individual differences in the breadth and depth of relevant knowledge and skills, which may serve as a foundation for transfer to the new knowledge and skills.

Of fundamental importance for the current discussion, both Hebb and Cattell proposed that adult intelligence, far from being represented by inevitable decline after about age 20, is associated with two components, of which only one component (Intelligence A/Gf) shows declines after adolescence. The other component (Intelligence B/Gc) does not show such declines, and this component is also of critical importance to the success of adults in intellectually demanding day-to-day activities.

From the work of Hebb and Cattell, the central points are as follows:

- Adult intelligence can be meaningfully divided into two components. One (Intelligence A/Gf) is composed of abstract reasoning, short-term memory, and other process abilities; the other (Intelligence B/Gc) is composed of experiential and educational knowledge and skills.
- Intelligence A/Gf shows rapid development during childhood and plateaus during late adolescence; Intelligence B/Gc develops well into adulthood, and is much better preserved in normal individuals as well as in individuals with various types of neurological traumas.

Cattell's Investment Theory

Although Hebb did not further explore Intelligence A/Intelligence B in later work (though see Hebb, 1949), Cattell individually, and with Horn's contributions, expanded both the theoretical and empirical foundations for his theory of intelligence. Cattell originally proposed that Gc grows out of Gf in childhood, but in a later discussion, Cattell (1971, 1987) proposed that Gc develops specifically as a result of the *investment* of Gf during this period. The details of how this investment takes place were not spelled out or investigated; for example, Cattell noted that development of Gc was "a function both of g_f and a bunch of opportunity, motivation and memory factors" (1987,

p. 152). However, he clearly implied that he considered differentiated Gc knowledge and skills within and between individuals to be the results of the time and intellectual (Gf) effort allocated to particular domains of inquiry. Although Cattell later noted relations between personality traits and measures of academic achievement, these did not play an explicit role in his investment theory.

The other major consideration that Cattell offered in this framework was a distinction between "historical" Gc and "present" Gc, especially with respect to assessing adult intelligence. The problem actually is partly a consequence of the Binet–Simon paradigm for assessing intelligence of children. As discussed earlier, Binet and Simon decided against using the "pedagogical method" of assessing intelligence—that is, assessing what the individual "knows"—because they thought that such measures would be overly influenced by socioeconomic status. So, to the degree that knowledge and skills are assessed by Binet-inspired tests, the content of tests is limited to that which is common to the wider culture. Practically speaking, knowledge and skills assessed on IQ-type tests are those that one can expect all or nearly all children and adolescents to have had ample exposure to. For elementary school children, vocabulary knowledge and verbal comprehension represent domains of common experiences or "core" curricula in the verbal domain; simple math operations (addition, subtraction, multiplication, and division) represent core curricula in the math domain. But as children reach secondary school and beyond, or as they move from school to work, their intellectual experiences become quite varied. For example, high school students often have opportunities for multiple elective courses, whether in different foreign languages, various arts and humanities fields, sciences, and mathematics, leaving a smaller number of common experiences or core courses. Cattell (1987) recognized the inherent problem associated with assessing Gc in adults:

> . . . we begin to ask what happens to crystallized general intelligence, and the traditional intelligence tests that measure it, *after* school. The crystallized intelligence then goes awry both conceptually and in regard to the practical predictions to be made from traditional intelligence tests. In the twenty years following school, the judgmental skills that one should properly be measuring as the expression of learning by fluid ability must become different from different people. If these are sufficiently varied and lack any common core, the very concept of general intelligence begins to disappear.
>
> . . . [The psychologist's] alternatives are then: (a) to sample behavior still more widely than in the traditional test, using a formula expressing the role of fluid intelligence in learning in each of many different fields (an approach which, in practice, might amount to producing as many different tests as there are occupations, etc.); (b) to change completely to fluid intelligence measures . . . ; or (c) to continue to measure by the "school version" of crystallized ability essentially learning on what the individual's intelligence was at the time of leaving school. (Cattell, 1987, pp. 143–144)

That is, to assess current Gc in adults, one would need to assess knowledge and skills as varied as various educational specialties (physical and natural sciences, social sciences, technology, math, engineering, humanities, business, law, medicine), but also nonacademic domains, such as carpentry, plumbing, gardening, popular music, current events, and so on. Otherwise, it would be impossible to determine, for example, whether a carpenter has higher or lower levels of current Gc than a physicist, a concert violinist, or a nurse.

When it comes to assessing Gc in late adolescents and adults, nearly all intelligence researchers have adopted the third strategy described by Cattell—that is, to measure historical Gc, rather than attempt to provide a comprehensive assessment of current Gc. Aptitude test designers take a similar approach: The math content of the SAT is limited to algebra and geometry (which are required in nearly all secondary school programs). No trigonometry or calculus problems appear on these tests, mainly because such courses are often electives that are not completed by significant proportions of students. Even the general tests of the Graduate Record Examinations (GRE), which are completed by students seeking admission to graduate study at the end of their college/university programs, do not test advanced mathematics; they too are limited to the same common core curriculum (algebra and geometry) that high school students complete. In this sense, such tests do not directly assess "current" Gc, but rather content that individuals have acquired in the past. In the case of the GRE, the content may have been acquired 6 or 7 years previous to the administration of the test, making the test much more of an assessment of historical Gc than of current Gc. Nontraditional students, who may be 30 or 40 years old, are thus at a considerable disadvantage in comparison to younger students when it comes to completing tests like the SAT or GRE—given that 10 or 20 (or more) years may have elapsed between the time they acquired algebra and geometry knowledge and skills and the time when such a test occurs, and given that such skills may have been rarely if ever used by these adults in their occupational or avocational activities.

The central issues, then, from Cattell's framework for intellectual development are as follows:

- Investment of Gf capabilities, through some unknown combination of factors, leads to development of Gc.
- For assessment of Gc in adults, either one must develop a wide array of tests that adequately sample the "current" levels of common and a variety of noncommon domains of knowledge and expertise of adults, or one must limit the assessment to "historical" Gc—that is, the knowledge and skills to which nearly all members of the population were exposed as adolescents or children.

Mid-20th-Century Status of Intelligence Assessment of Adults

The introduction of an omnibus test of intelligence specifically designed for adults (Wechsler, 1939) removed the problems caused by the concept of mental age and by reliance on tests that had been extended from assessment of children. Wechsler's test also included consideration of diagnostic medical criteria (e.g., Korsakoff syndrome, organic brain disease) rather than school achievement. Assessment of adult IQ could therefore be accomplished without many of the limitations posed by the earlier extensions of the Binet–Simon scales. Although Wechsler's scale represented a substantial improvement over existing measures, cross-sectional norms still indicated that average peak IQ was obtained by young adults, with declining levels of intelligence as age progressed into middle age and beyond (see Wechsler, 1944). Later studies of longitudinal design (e.g., Owens, 1953) and lagged cross-sectional (also called cross-sequential) design (e.g., see Schaie, 1996) suggested that cohort differences resulted in somewhat exaggerated declines in intellectual abilities with increasing age in adulthood, but with the exception of some Gc (historical) abilities, most other intellectual abilities showed declines with age across most of the adult lifespan. Even considering that Gc (historical) abilities might increase during adulthood (Horn, 1965), the relative gradient of improvement through middle age on Gc was expected to be much smaller than the gradient of decline on Gf abilities. Thus an average intelligence score that equally weighted Gf and Gc abilities was expected to peak no later than the late 20s.

Development of an Alternative Theory

The PPIK theory (Ackerman, 1996) was born from a conclusion that the Binet–Simon paradigm for intelligence assessment of children is inadequate for assessing adults, even with the kinds of innovations provided by the Wechsler Adult Intelligence Scale. To review, the Binet–Simon paradigm consists of four major components, as follows:

1. Intelligence assessment is mainly concerned with individual differences in mental *processes* (e.g., abstract verbal, spatial, numerical reasoning, short-term memory, mathematical computation).
2. Intelligence assessments eliminate (as far as possible) consideration of knowledge acquired through *specific* educational or experiential sources. Knowledge assessed in standard intelligence measures is limited to content that is common to a dominant culture or included in a core educational curriculum.
3. Elicitation of *maximal* effort from the examinee is expected.
4. School achievement is the fundamental criterion for external validation (except for Wechsler's scales, which include medical criteria for validation).

Each of these components seems incomplete for representing the intelligence of adults. Moreover, these conditions appear to reduce or eliminate the possibility of understanding how non-ability traits might influence the development and maintenance of intelligence, from the time that individuals stop sharing most of their educational and vocational experiences in adolescence and beyond. From these initial considerations, I developed four alternative propositions:

1. Although process aspects of intelligence are important to the acquisition of new knowledge and skills (e.g., see Ackerman, 1988), these aspects of intelligence have less importance in adult intelligence, which depends to a substantial degree on prior knowledge and transfer of training (Ferguson, 1956). Thus, in agreement with Cattell's broad theory, Gf (process) abilities are most important during early development, and are less important in adulthood.

2. For adults, Binet's "psychological method" of intelligence assessment is inadequate. It must be supplemented by the pedagogical method—that is, assessment of what adults actually *can do* (i.e., knowledge and skills that are not necessarily common to a core educational curriculum). The pedagogical method requires consideration of *current* Gc, Accordingly, assessments of intellectual knowledge and skills must be both broad and deep, ranging across a wide array of different domains that represent the diversity of expertise in the adult population. Thus intelligence-as-knowledge represents an important component of adult intelligence.

3. In the school system—where educational assessments often involve students' "cramming" for end-of-term examinations, or when the goal is to estimate what a student might be capable of, rather than what the student is likely to do—it makes sense to attempt to assess an individual's maximal performance. However, for an adult, one is arguably more interested in how the individual typically performs, so considerations of likely performance seem to be more relevant to assessing adult intelligence than assessments of maximal performance.

4. One major shortcoming of both the Binet–Simon scales and the Wechsler scales is that while they correlate well with academic performance, these tests correlate significantly, but much less highly, with measures of occupational performance (e.g., see Schmidt & Hunter, 2004). But relatively few adults are engaged as students in educational contexts. A more useful criterion for adults is occupational performance, or performance on relevant intellectual tasks outside the classroom or work environment. In general, tests of general intelligence correlate only moderately at best with adult criteria such as job performance. In hindsight, this result should not be all that surprising because intelligence tests were not originally designed with job performance as the criterion on which the tests were developed or validated.

In addition, through an examination of correlations among personality traits, interest and motivational traits, and intellectual abilities (e.g., see Ackerman, 1997; Ackerman & Heggestad, 1997), it became clear that these non-ability traits play an important role in adult intellectual development. From a theoretical perspective, the justification for finding these relations to be more pronounced in adulthood, compared with children, is that post-secondary educational and occupational environments provide much lower situational "press" than the mandatory secondary school environment. This means that adults are much less constrained in how they engage the intellectual environment, especially once they leave secondary school. With lower situational press, it is generally believed that non-ability traits play an increasingly important role in determining the direction and intensity of an individual's effort (for a review, see Tett & Burnett, 2003).

THE PPIK THEORY OF ADULT INTELLECTUAL DEVELOPMENT

Components of the Theory

There are four components to the PPIK theory: intelligence-as-Process, Personality, Interest, and intelligence-as-Knowledge. These components are defined in turn below.

Intelligence-as-Process

As Cattell (1943) and others have pointed out, performance on any particular test of intelligence is almost certainly determined by both Gf and Gc, although to different degrees, depending on (1) the content of the test (e.g., familiar words or numbers vs. artificial stimuli or novel figures/words); (2) the characteristics of the examinee sample (in terms of their age, prior experiences, education, etc.); and (3) most likely a variety of other test characteristics (e.g., speededness, need for precision in attending or responding, complexity of the items). It is, however, possible to specify that underlying cognitive processes such as Spearman's "eduction of relations" (i.e., abstract reasoning), short-term memory, working memory, and similar operations represent the "process" aspects of intellectual abilities. (As an aside, it is useful to note that because Gc is involved in all intellectually demanding tests, there really is no test that can be reasonably classified as truly "culture-free" or "culture-fair," given that the makeup of Gc is highly determined by the cultural milieu of the individual examinees. Instead, many investigators quite reasonably refer to tests with mostly Gf content as being "culture-reduced.") Intelligence-as-process, whether it is referred to as Cattell's Gf or Hebb's Intelligence A, is envisioned to be (1) dominant early in childhood, reaching a plateau in adolescence; (2) important as a determinant of the level of intellect that can be invested in development of intelligence-as-knowledge; and (3) relevant, but not univocally so, in the acquisition of new knowledge throughout adulthood.

Personality

A meta-analysis we conducted on personality and ability correlations makes it possible to articulate some general trends in the associations between these two families of traits (Ackerman & Heggestad, 1997). Although in general, personality traits do not share a large degree of common variance with measures of intellectual ability, there are some important exceptions. First, the traits of neuroticism, anxiety, and psychoticism all appear to have pervasive significant negative correlations across a wide variety of intellectual abilities, even though the correlations are of modest magnitude. Second, there is a group of personality traits that appear to have generally positive relations with intellectual abilities, especially for the verbal/Gc component of intelligence. These traits have a variety of names—openness to experience, intellectance, culture, need for cognition, and typical intellectual engagement—but in general they refer to an individual's general orientation toward thinking, problem solving, reading, and seeking intellectual and cultural experiences. Scales that represent these traits are often highly correlated, so it is generally concluded that they represent one overarching general personality construct of intellectual orientation (e.g., see von Stumm & Ackerman, 2013).

Other personality traits appear to have more complex associations with intellectual abilities, such as masculinity–femininity, which is generally found to be associated with math and spatial abilities, consistent with gender differences on these types of ability tests. Mixed results have been found for associations between introversion–extroversion and intellectual ability. Although Cattell (1945) speculated that introverts would prefer reading books to going to a social event, and as such should be expected to have higher abilities (especially in the Gc domain), only some studies have supported this conjecture; others have failed to find any significant associations. Undoubtedly, part of the difficulty with this particular trait is the fact that extremes on either end of the introversion–extroversion continuum are considered to be associated with poor adaptation, so that individuals closer to the center of the distribution would be expected to have better overall adaptation to society. If this is correct, the relationship between introversion–extroversion and intellectual abilities should have an inverted-U shape, depending on the particular characteristics of the test sample and the particular assessment scale. This is something that cannot readily be resolved with a linear correlational analysis.

Interests

Examining vocational interests and intellectual abilities provides a much more varied set of associations than for personality traits, partly for reasons that are specific to the nature of interest theories and to the specific design of interest assessments. Historically, vocational interest measures were not designed to provide trait-like scores; the original Strong Vocational Interest Blank (Strong, 1943) provided estimates of similarity between the patterns of likes and dislikes between various job types and the individual completing the questionnaire. So one could inquire, for example, whether those individuals with interests similar to those of engineers or doctors had higher or lower IQ scores than those with interests similar to those of plumbers or carpenters, but it was otherwise impossible to determine associations between intellectual abilities on the one hand, and degree of interest in particular occupations on the other hand (Darley & Hagenah, 1955). In addition, many vocational interest measures were designed in a way to reduce or eliminate associations with intellectual abilities. Typically, individuals are asked to consider which jobs they might like or dislike, independently of whether they think they have the ability to perform the jobs.

By the 1950s, however, vocational psychologists had started to converge on a consensus list of vocational interest traits (e.g., see Guilford, Christensen, Bond, & Sutton, 1954; Roe, 1956), which allowed for correlational assessments of interest–intelligence associations. The current dominant approach to assessing vocational interests, developed by Holland (1959, 1973), has particular importance for interest–intelligence relations, but it has one substantial shortcoming. The shortcoming is that Holland's model is essentially a typology, meaning that it does not readily lend itself to correlational analyses; each individual is described in terms of his or her top interest themes, rather than receiving a score on each of the interest theme traits. Holland's original theory of the development and expression of vocational interests, however, does make explicit connections to intellectual abilities. First, interests are hypothesized to develop as a function of early experiences and challenges; children who are initially successful with math problems or verbal problems, for example, are expected to develop higher levels of self-concept for these domains, in contrast to those children who struggle with the same materials. Thus interests and intellectual ability are expected to be associated: Higher levels of intellectual abilities should lead to greater successes, which in turn should lead to higher self-concept and then to higher levels of interest in the domain, whereas lower levels of intellectual abilities should lead to fewer successes, which in turn should lead to lower self-concept and thus lower levels of interest in the domain. By the time children reach adolescence, there should be an alignment between their patterns of ability strengths and weaknesses and their high and low levels of vocational interests.

As for particular interest themes, Holland originally specified two themes as particularly associated with higher intellectual abilities: *intellectual* interests and *esthetic* interests. In addition, *motoric* interests were identified as being associated with hands-on kinds of skills (one could imagine mechanical reasoning and knowledge as key abilities associated with these interests). *Conforming* interests were thought to be mainly associated with lower-order abilities, such as math computation and perceptual speed abilities. Other interest themes (*supporting* and *persuasive*) were hypothesized to be either uncorrelated with abilities, or negatively associated with intelligence. Because most modern interest assessments do not provide scores for each individual on each of these six interest themes, there have been relatively few sources of evidence to assess these relations. However, based on the existing corpus of data, it appears that interests in two themes (intellectual, which was retitled as *investigative,* and esthetic, which was retitled as *artistic*), are substantially positively correlated with intellectual abilities. Investigative interests are associated with a wide array of intellectual abilities, while artistic interests are mostly associated with Gc abilities. Conforming interests (retitled as *conventional*) are positively correlated with measures of math computation and perceptual speed abilities. The remaining two interest themes (*supporting,* retitled as *social,* and *persuasive,* retitled as *enterprising*) tend to have negligible or negative associations with traditional measures of intellectual abilities (for a review, see Ackerman & Heggestad, 1997).

Intelligence-as-Knowledge

The fourth component of the PPIK theory is the construct of intelligence-as-knowledge. The construct has much in common with Hebb's Intelligence B and Cattell's Gc, but the main differences are these: (1) Within the PPIK framework, current Gc is the central focus, in contrast to historical Gc; and (2) a much broader range of knowledge is included, such as cognitive and perceptual skills, cultural knowledge (e.g., knowledge of current events), occupational knowledge, and avocational knowledge (cooking, gardening, technology, hobbies of various sorts, etc.). Fundamentally, an adult's entire declarative (factual), procedural (skill), and tacit (implicit) knowledge and skill intellectual repertoire is considered to represent his or her intelligence-as-knowledge. A full accounting of an individual's intelligence-as-knowledge would be contained in the answers to the two questions "What do you know?" and "What can you do?"—at least within domains that require cognitive effort to acquire. Of course, not all knowledge and skills are equally indicative of intelligence-as-knowledge. Acquiring knowledge of how to perform neurosurgery or design a nuclear power plant obviously requires greater investment of cognitive effort over a longer period of time than learning all of the different configurations for ordering an automobile from a dealer. One could argue whether learning trivia such as sports team statistics is "intellectual" to a greater or lesser degree than memorizing the scores for all of Beethoven's piano sonatas, but clearly individuals differ in terms of such diverse knowledge in measurable ways. Ultimately, an individual's intelligence-as-knowledge is represented by both the depth and breadth of the person's knowledge across the various topical domains that make up the wider civilization.

Investment

At its core, the PPIK theory considers adult intelligence to be represented by what an individual *can do*—that is, what the individual is capable of accomplishing. This in turn is determined by the individual's relevant repertoire of knowledge and skills (i.e., current Gc). The development of Gc in turn is determined by the individual's investment of both process-type intelligence early in life, and increasingly by the individual's use of transfer of prior knowledge and skills for the development of refined (and in some cases automatized) procedures for accomplishing intellectual tasks. Tasks that are exceptionally novel require greater investments in intelligence-as-process; tasks that are more familiar, or those that represent building on existing knowledge structures, require less investment of intelligence-as-process and depend more on the foundation of existing intelligence-as-knowledge. A few examples might illuminate this particular proposition. Extensive research has suggested that general intelligence, made up of both Gf and Gc components, is importantly related to individual differences in task performance when the task is novel. But with extensive practice and skill development, individual differences in performance on such tasks are less related to general intelligence, and more related to the component skills (e.g., perceptual speed and psychomotor abilities for tasks that require rapid perception, stimulus recognition, and responses). See Ackerman (1988) for an extensive review.

The process of learning to drive a car, or even learning to fly an airplane, for example, has these kinds of characteristics. When a new driver or pilot first confronts the task, it requires a great amount of dedicated attention and cognitive effort to remember all the component tasks that are required for the safe and efficient operation of the vehicle. The use of secondary task probes provides a good indication of the cognitive/intellectual demands of the task. If the operator cannot successfully perform the target task and carry out a secondary task (such as carrying on a coherent conversation), then the target task is presumed to be consuming all of the individual's intellectual effort resources. New drivers or new pilots are ordinarily seriously impaired in attempting to carry out a conversation (or even to sing along to a familiar tune on the radio) when they are first learning the task. But with extensive practice, many aspects of the skills required to drive a car or fly an airplane become part of the intelligence-as-knowledge repertoire, and as such require little involvement from intelligence-as-process resources. Under normal conditions, most adult drivers (and experienced pilots) show little or no significant impairments in task performance while carrying out a conversation at the same time.

Similarly, the effects of declines in Gf/intelligence-as-process with advancing age are often not inherently problematic for task accomplishment in a variety of domains. Michael DeBakey was successfully performing heart bypass surgeries well into his 80s, partly because by the time he reached that age, he had performed over 60,000 such procedures (DeBakey, 1999). There should be little doubt that a 30-year-old MD from a prestigious medical school would have higher Gf abilities than the 80+-year-old Dr. DeBakey, but it is much easier and more effective, in general, to recall the correct answer to a problem that has been previously encountered many times before than it is to derive an answer depending only or mostly on intelligence-as-process resources. There are numerous other examples of middle-aged and older individuals who perform their specialized skills or employ their specialized knowledge at levels exceeding those of much younger professionals, who are expected to have higher Gf abilities but less acquired/Gc skills and knowledge.

Ultimately, though, there is a point at which the demands for new learning cannot be fully compensated by preexisting intelligence-as-knowledge. For example, older musicians may struggle with— new music scores; older dentists may struggle with using 3D computer-aided design/manufacturing systems for creating porcelain crowns; and older behavioral psychologists may struggle with integrating their knowledge with recent advances in studies with functional magnetic resonance imaging systems. In such cases, declines in intelligence-as-process with increasing age may limit or slow the acquisition of relatively novel knowledge and skills. It then becomes a question of whether further investment of intelligence-as-process resources is worth the potential return in terms of new knowledge and skills (e.g., see Posner, 1995).

Actions and Influences in PPIK Theory

According to the original PPIK theory, intelligence-as-knowledge is initially developed, especially during adolescence and through early adulthood, through the investment of Gf/intelligence-as-process resources into Gc as traditionally assessed (such as verbal abilities and general core knowledge). These abilities then jointly influence the development of specialized intelligence-as-knowledge, with key personality traits (e.g., openness to experience, typical intellectual engagement) and interest themes (realistic, artistic, investigative) influencing the direction and intensity of effort toward acquisition and maintenance of intelligence-as-knowledge. Knowledge in the physical sciences and mathematics domains, for example, is more highly influenced by Gf/intelligence-as-process abilities, and by the realistic and investigative interest themes; knowledge in the arts, literature, and social sciences is more highly influenced by Gc (historical), the personality traits of openness to experience and typical intellectual engagement, and both the artistic and investigative interest themes.

TRAIT COMPLEXES AND REVISIONS TO THE PPIK THEORY

In the original specification of the PPIK theory, as noted above, the roles of personality traits and interest themes were considered at the level of individual constructs. Although these non-ability traits were viewed as related to one another (e.g., artistic interests and openness to experience), and as such were thought to have possible influences on one another during development, their actions were considered separately. Subsequent to the initial specification of the theory, a review of the literature and meta-analysis of interest–personality–ability relations, along with a series of empirical studies, together suggested that something more fundamental could be revealed. This led to refinements to the theory.

The first indication of commonality among these traits came from an analysis of interest–personality communality. In contrast to personality–intelligence relations, the associations between personality traits and interest themes are much more substantial, suggesting that there are relatively close associations between particular affective and conative/motivational traits. Enterprising and social interest themes, in particular are associated with personality traits of extroversion and positive affect/well-being. Conventional interests are associated with conscientiousness and traditionalism; investigative and artistic interests are associated with absorption, and a wide array of personality traits is associated with an intellectual orientation (as previously discussed, including openness to experience, typical intellectual engagement, etc.)

Additional communalities were found for traits in the domain of academic self-concept and self-estimates of abilities (e.g., see Ackerman, 1997; Ackerman & Wolman, 2007). It is ultimately useful to consider whether an individual's investment of intelligence-as-process and Gc-type abilities is more influenced by the individual's self-assessments of his or her abilities than by the person's objectively determined abilities.

Trait Complexes

Finding common variance among personality traits and interest themes, along with some intellectual ability traits, led to a suggestion that constellations of groups of traits across trait families are more or less likely to cluster together. From reviewing the literature on cross-trait associations, we (Ackerman & Heggestad, 1997) suggested four constellations of traits: *social, clerical/conventional, science/math,* and *intellectual/cultural,* as shown in Figure 8.1. These are groups of traits that tend to be moderately to highly correlated with one another. For example, this means that, everything else being equal, individuals with high levels of traditionally measured Gc also have higher level of investigative and artistic interests, and higher levels of the personality traits of absorption, openness to experience, and typical intellectual engagement.

There are two noteworthy features of these trait complexes, and one overarching issue not depicted in Figure 8.1. First, these trait complexes were established on the basis of existing data, which means that trait relations that had not been investigated in prior studies would not appear in the representation. Second, of the four identified trait complexes, the science/math complex had no personality traits with substantial associations; this *might* mean that there are no specific personality traits common to this complex of ability and interest themes, but it might also indicate that there are not-yet-identified personality traits that would be associated with these other traits (e.g., see Toker & Ackerman, 2012). Also, the social trait complex contains no abilities. A similar interpretation of this finding might be offered; that is, abilities are not associated with this constellation of personality traits and interest themes. However, it is also possible that what is missing for this trait complex is a reliable and valid indicator of so-called "social intelligence," a construct that has proved elusive to intelligence researchers over the modern history of intelligence assessment (e.g., see Jones & Day, 1997; Thorndike & Stein, 1937). The last point, not directly captured in the figure, is that these are relatively independent or uncorrelated constellations of traits. This means that although an individual's position on one trait complex provides little information about his or her standing on the others, it is quite possible that an individual could have high or low standing on multiple trait complexes.

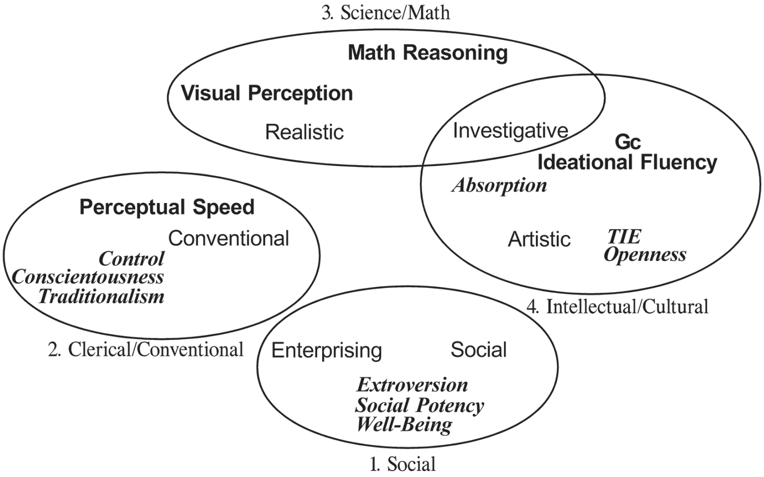


FIGURE 8.1. Trait complexes, including abilities (bold), interests (regular), and personality (italic) traits, showing positive communalities. Number categories are trait complexes. Gc, crystallized intelligence; TIE, typical intellectual engagement. From Ackerman and Heggestad (1997, Figure 7). Copyright © American Psychological Association. Reprinted by permission.

Because of the patterns of ability associations with the personality and interest constructs, we conjectured that these constellations of traits might have facilitative or impeding influences in the development and maintenance of intelligence-as-knowledge, in much the same way that Snow's (1963) concept of *aptitude complexes* might be supportive or not supportive of the acquisition of knowledge and skills in educational contexts. In particular, the intellectual/cultural and science/math trait complexes were hypothesized to be facilitative of new learning and existing knowledge structures in adults, especially in the domains of physical sciences/math and arts/humanities/social sciences, respectively. Higher levels of the social trait complex were expected to have impeding influences on new learning and existing knowledge structures for most academic and related knowledge domains, while the clerical/conventional trait complex was hypothesized to be relatively unrelated to knowledge across a wide range of domains (see Figure 8.2). Furthermore, we hypothesized that with increasing age and development during adulthood, individuals might develop more "coherent" trait complexes; that is, the traits within a complex should become more highly correlated with one another.

Empirical Evidence

There are several implications of the PPIK theory, and some of them have been subjected to empirical testing. The first implication is that if the definition of adult intelligence is broadened from what is assessed by Binet-inspired IQ-type tests to give credit for the breadth and depth of intelligence-as-knowledge, then peak intelligence is less likely to be identified with adolescents and young adults. Or, to put it in Cattell's terms, including Gc (current), either in addition to or instead of Gc (historical), in an evaluation of intelligence will indicate that middle-aged adults fare better than they do on traditional intelligence tests. As Cattell pointed out and as mentioned earlier, to comprehensively assess Gc (current) in adults, one would need to create as many different domain knowledge tests as there are differentiable domains of knowledge—a task that is not realistically feasible.

As a modest attempt to sample a wide range of domains, my students and colleagues and I initially assembled a set of 20 different domain knowledge tests (ranging from physical and social sciences to literature, history, law, business, etc.). We then administered the tests to samples of young adults and middle-aged adults (e.g., Ackerman, 2000; Ackerman, Bowen, Beier, & Kanfer, 2001; Ackerman & Rolfhus, 1999), along with traditional measures of Gf and historical Gc, and a wide array of personality, interest, and self-concept assessments. The pattern of results from these investigations was largely supportive of the PPIK theory. First, with the exception of some science domains (e.g., chemistry and physics), Gc (historical) was a better predictor than Gf of domain knowledge. Also, although middle-aged participants performed more poorly than young adults on Gf tests, and somewhat better on Gc (historical) tests, the domain knowledge scores of middle-aged adults were equal to or better than those of young adults in nearly every domain, with the exception of those domains also most highly associated with Gf abilities. In a sample where ages ranged from 21 to 62 years (Ackerman, 2000), Gf correlated negatively with age (*r* = –.39), Gc (historical) correlated positively with age (*r* = .14), and a general knowledge composite correlated positively with age (*r* = .21). An equally weighted general intelligence composite of Gf and Gc (historical) yielded a negative correlation with age (*r* = –.14). But, a composite that equally weighted Gf, Gc, and a general knowledge composite yielded an essentially zero correlation with age (*r* = –.02), meaning that with this weighting, middle-aged adults were as intelligent as young adults. Combinations that provided greater weight to intelligence-as-knowledge and less proportional weight to Gf would result in positive correlations of intelligence and age across a wide range of adult ages. Similar results were found for knowledge domains that are not primarily academic, such as knowledge of current events (Beier & Ackerman, 2001) and of health (Beier & Ackerman, 2003). Also, for middle-aged adults, individual differences in the acquisition of new knowledge (e.g., technology and health domains) were more highly related to measures of Gc than they were to measures of Gf, partly attributable to transfer of knowledge from existing domain knowledge to novel information (e.g., see Ackerman & Beier, 2006).

Finally, personality traits (such as typical intellectual engagement) and interest themes (such as investigative interests) were also positively associated with individual differences in domain

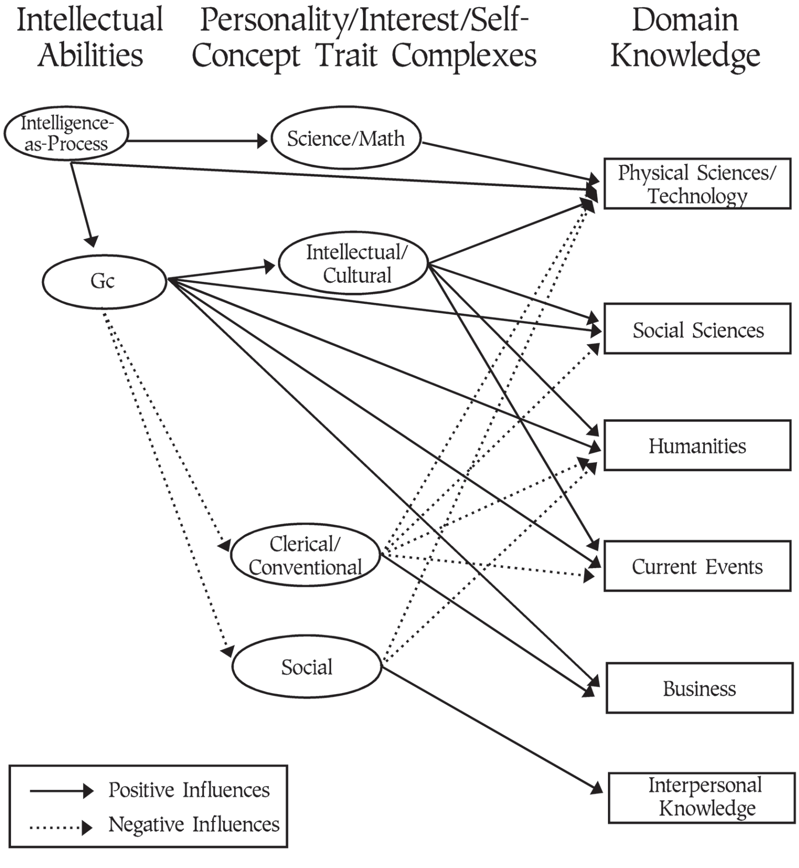


FIGURE 8.2. An illustration of the revised PPIK theory. Shown are positive (facilitative) and negative (impeding) relations between abilities, trait complexes, and individual differences in domain knowledge of adults. Gc, crystallized intelligence.

knowledge, as specified by the PPIK theory. Subsequent studies yielded results consistent with these findings, but also extended the examination of non-ability constructs from individual scales to assessments of trait complexes. In particular, for both young adults and middle-aged adults, scores on the science/math and intellectual/cultural trait complexes were positively associated with domain knowledge in academic domains, and the social trait complex was negatively correlated with knowledge across nearly all domains.

Other Evidence

Although there have been no direct investigations of the PPIK theory in a longitudinal study of adults, various other sources of evidence are consistent with the general tenets of the theory. Studies of the importance of job knowledge (including studies indicating that knowledge is a more potent predictor of job performance than general intelligence) support the emphasis on intelligence-as-knowledge as a key indicator of adult intellect (e.g., see Hunter, 1983; Schmidt, Hunter, & Outerbridge, 1986). Additional studies have supported the PPIK framework in demonstrating the importance of domain-specific knowledge in academic success at the postsecondary level (e.g., Ackerman, Kanfer, & Calderwood, 2013).

Other studies have provided convergent evidence regarding the trait complexes identified in the Ackerman and Heggestad (1997) meta-analysis (e.g., see Armstrong, Day, McVay, & Rounds, 2008; Staggs, Larson, & Borgen, 2007; Sullivan & Hansen, 2004). Finally, several recent studies have shown personality–ability relations during later adulthood, in patterns that are largely supportive of the hypothesized role of non-ability traits in adult intellectual development (e.g., see Ziegler, Cengia, Mussel, & Gertsorf, 2015).

CHALLENGES AND CONTROVERSIES

The PPIK theory provides just one perspective on adult intelligence and adult intellectual development among many competing theories. The most prominent competitor approach is the one that defines intelligence within Spearman's *g* or working memory (for a review, see Ackerman, Beier, & Boyle, 2005). Over the past century, there have been many attempts to localize general intelligence as marked by specific processes or abilities, including Spearman's attempt to identify general intelligence with the Raven's Progressive Matrices test; various attempts in the 1970s and 1980s involving the search for fundamental information-processing building blocks for intelligence; or more recent efforts to identify working memory ability as the central construct for general intelligence or fluid intelligence. None of these efforts have been successful in terms of providing more valid predictors for the academic success of children and adolescents, and to the limited degree that they have been explored, none have improved upon either traditional Binet-inspired tests or knowledge tests for predicting occupational performance for adults.

Yet one should not overlook the evidence that older adults do perform worse than younger adults, on average, on measures of speeded processing or highly Gf-loaded ability tests (Tucker-Drob & Salthouse, 2011), or other evidence that older adults perform more poorly than young adults in new learning, occupational training, and skill acquisition (e.g., Kubeck, Delp, Haslett, & McDaniel, 1996). The PPIK theory does not readily address these issues, except indirectly, to the degree that it provides a framework for understanding under what circumstances a middle-aged adult is likely to have more or less difficulty in acquiring new knowledge and skills (based on the degree of transfer from the individual's repertoire of existing knowledge and skills to the new tasks).

Ultimately, the question remains: What is adult intelligence? Numerous definitions of intelligence have been provided from academics and assessment professionals over the past 100 years. As Ackerman (2017) recently noted "Perhaps it is worthwhile to consider E. L. Thorndike's (1921) proposition: "Realizing that definitions and distinctions are pragmatic, we may then define intellect in general as *the power of good responses from the point of view of truth or fact*" (1921, p. 124; original emphasis). From this perspective, it seems somewhat quaint to think that many of the items on existing intelligence tests are representative, to the degree that they demand processes resembling how adults actually reach "good responses" in the current environment. Questions of general knowledge (common cultural information) can be readily answered by most adults in the space of a few seconds to search the Internet (e.g., "Who was Neil Armstrong?" or "Who wrote the play *Waiting for Godot*?") Similarly, although adults of a "certain age" can easily add, subtract, multiply, or divide a string of numbers in their heads or with paper and pencil, less error-prone (and typically) faster answers can be derived with a calculator, a computer, or the ubiquitous smartphone application. So what does it mean to be "intelligent" in this context? Do we give credit only to those who can retrieve correct information from long-term memory, or is the person who has achieved the requisite skills at rapid information search and retrieval from the Internet or other sources equal in ability? How would different adults perform on an "open-book" test of intelligence, where they can attempt to answer the test questions by using their own memory and mental skills, but they also have access to the kinds of tools they typically use when confronted with similar tasks on a day-to-day basis?

Some items on current intelligence tests will be resistant to online solutions. Items of abstract reasoning, spatial ability, and comprehension are not readily answered with these kinds of external aids, but the nature of some of these items (e.g., abstract reasoning) essentially begs the question of their relevance to the expression of adult intelligence. This question, then, leads me to the main issue: Namely, what activities do adults typically confront that require intelligence? If the task is to determine how many tiles necessary to cover a shower stall, or how much paint is needed to cover a room with two coats, many adults would refer to Internet sources rather than attempting to work out the answers in their heads or with an electronic calculator. Other tasks might have rough analogues to intelligence test items. For example, consider the task of installing a graphics board in a computer, or fixing a lawn mower. An adult unfamiliar with either of these tasks is perhaps more likely to search YouTube for an audio and video demonstration of how to accomplish the task. Under such circumstances, the intellectual task the adult faces is to take note of the critical steps in the demonstration video, and work that into an action plan for accomplishing the goal. One might conjecture that higher-intelligence adults will require fewer viewings or review occasions with the YouTube video, compared to lower-intelligence adults" (Ackerman, 2017, p. 994). But these are open questions that will (let us hope) be answered by future research.

In the final analysis, the PPIK theory attempts to shift the orientation away from a g/Gf/intelligence-as-process orientation, in that it maintains that the vast majority of "intellectually demanding" activities performed by adults are well learned and as such, demand relatively little involvement of intelligence-as-process capabilities. How much weight should be accorded to intelligence-as-knowledge for evaluating adult intelligence? The question cannot be answered by only appealing to correlations between measures, factor structures, age changes, or other aspects of construct validity. The answer to that question will depend on the utility of the assessments for predicting occupational and other kinds of performance in the real world, as a true extension of Binet's approach to predicting academic achievement of children in school.

NOTE

1. Much later, Owens (1953) found that in a sample of 127 men who had originally been tested on the Army Alpha test in 1919 (at about age 18) and then retested in 1950 (at about age 49), there were significant gains in scores on the verbal/information subtests, and mostly stable scores on the other tests.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22,* 229–259.

Ackerman, P. L. (1997). Personality, self-concept, interests, and intelligence: Which construct doesn't fit? *Journal of Personality, 65*(2), 171–204.

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *Journal of Gerontology Series B: Psychological Sciences, 55B*(2), 69–84.

Ackerman, P. L. (2017). Adult intelligence: The construct and the criterion problem. *Perspectives in Psychological Science, 12*(6), 987–998.

Ackerman, P. L., & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology, 98,* 366–381.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Ackerman, P. L., Bowen, K. R., Beier, M. B., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93,* 797–825.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121,* 219–245.

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2013).

High school Advanced Placement and student performance in college: STEM majors, non-STEM majors, and gender differences. *Teachers College Record, 115*(10), 1–43.

Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and non-ability traits. *Psychology and Aging, 14,* 314–330.

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied, 13,* 57–78.

Armstrong, P. I., Day, S. X., McVay, J. P., & Rounds, J. (2008). Holland's RIASEC model as an integrative framework for individual differences. *Journal of Counseling Psychology, 55,* 1–18.

Beier, M. E., & Ackerman, P. L. (2001). Current events knowledge in adults: An investigation of age, intelligence and non-ability determinants. *Psychology and Aging, 16,* 615–628.

Beier, M. E., & Ackerman, P. L. (2003). Determinants of health knowledge: An investigation of age, gender, abilities, personality, and interests. *Journal of Personality and Social Psychology, 84*(2), 439–448.

Binet, A., & Simon, T. (1916). Application of the new methods to the diagnosis of the intellectual level among normal and subnormal children in institutions and in the primary schools. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.) (pp. 91–181). Baltimore: Williams & Wilkins. (Original work published 1905)

Binet, A., & Simon, T. (1961). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux (*Année Psychologique, 11,* 191–336); and Le développment de l'intelligence chez les enfants (*Année Psychologique, 14,* 1–94). In J. J. Jenkins & D. G. Paterson (Eds., & Trans.), *Studies in individual differences: The search for intelligence.* New York: Appleton-Century-Crofts. (Original works published 1905 and 1908)

Boring, E. G. (1923, June 6). Intelligence as the tests measure it. *The New Republic, 35,* 35–37.

Brigham, C. C. (1922). *A study of American intelligence.* Princeton, NJ: Princeton University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cattell, R. B. (1945). Personality traits associated with abilities: II. With verbal and mathematical abilities. *Journal of Educational Psychology, 36,* 475–486.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* New York: Elsevier.

Conrad, H. S. (1930). General-information, intelligence, and the decline of intelligence. *Journal of Applied Psychology, 14,* 592–599.

Darley, J. G., & Hagenah, T. (1955). *Vocational interest measurement: Theory and practice.* Minneapolis: University of Minnesota Press.

DeBakey, M. E. (1999). *Timeline and accomplishments.* Houston, TX: Author.

Demming, J. A., & Pressey, S. L. (1957). Tests "indigenous" to the adult and older years. *Journal of Counseling Psychology, 4,* 144–148.

Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy.* New York: Norton.

Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8,* 95–112.

Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology, 10,* 121–131.

Freeman, F. N. F. (1922). The mental age of adults. *Journal of Educational Research,* 6(5), 441–444.

Guilford, J. P., Christensen, P. R., Bond, N. A., Jr., & Sutton, M. A. (1954). A factor analysis study of human interests. *Psychological Monographs: General and Applied,* 68(4, No. 375).

Hebb, D. O. (1939). Intelligence in man after large removals of cerebral tissue: Report of four frontal lobe cases. *Journal of General Psychology, 21,* 73–87.

Hebb, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society, 85,* 275–292.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology,* 6(1), 35–45.

Holland, J. L. (1973). *Making vocational choices: A theory of careers.* Englewood Cliffs, NJ: Prentice-Hall.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign. (UMI No. AAT6507113)

Hsiao, H. H. (1927). *The performance of the Army Alpha as a function of age.* Unpublished master's thesis, Columbia University, New York.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Erlbaum.

Jones, H. E., & Conrad, H. S. (1933). The growth and decline of intelligence. *Genetic Psychology Monographs, 13,* 223–298.

Jones, K., & Day, J. D. (1997). Discrimination of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology,* 89(3), 486–497.

Kubeck, J. E., Delp, N. D., Haslett, T. K., & McDaniel, M. A. (1996). Does job-related training performance decline with age? *Psychology and Aging, 11*(1), 92–107.

Lippmann, W. (1922). The mental age of Americans. *The New Republic, 32,* 213–215.

Owens, W. A., Jr. (1953). Age and mental abilities: A longitudinal study. *Genetic Psychology Monographs, 48,* 3–54.

Posner, R. A. (1995). *Aging and old age.* Chicago: University of Chicago Press.

Roe, A. (1956). *The psychology of occupations.* New York: Wiley.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study.* New York: Cambridge University Press.

Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin, 70,* 671–680.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162–173.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432–439.

Snow, R. E. (1963). *Effects of learner characteristics in learning from instructional films.* Unpublished doctoral dissertation, Purdue University, Lafayette, IN. (UMI No. 6404928)

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Staggs, G. D., Larson, L. M., & Borgen, F. H. (2007). Convergence of personality and interests: Meta-analysis of the Multidimensional Personality Questionnaire and the Strong Interest Inventory. *Journal of Career Assessment, 15,* 423–445.

Stern, W. (1914). *The psychological methods of testing intelligence* (G. M. Whipple, Trans.). Baltimore: Warwick & York.

Strong, E. K., Jr. (1943). *Vocational interests of men and women.* Stanford, CA: Stanford University Press.

Sullivan, B. A., & Hansen, J.-I. C. (2004). Mapping associations between interests and personality: Toward a conceptual understanding of individual differences in vocational behavior. *Journal of Counseling Psychology, 51,* 287–298.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet–Simon Intelligence Scale.* Boston: Houghton Mifflin.

Terman, L. M. (1922). The great conspiracy: The impulse imperious of intelligence testers psychoanalyzed by Mr. Lippmann. *The New Republic, 33,* 116–117.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*(3), 500–517.

Thorndike, E. L. (1921). Intelligence and its measurement: A symposium—I. *Journal of Educational Psychology, 12*(3), 124–127.

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin, 34*(5), 275–285.

Toker, Y., & Ackerman, P. L. (2012). Utilizing occupational complexity levels in vocational interest assessments: Assessing interests for STEM areas. *Journal of Vocational Behavior, 80,* 524–544.

Tucker-Drob, E. M., & Salthouse, T. A. (2011). Individual differences in cognitive aging. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 242–267). Chichester, UK: Wiley-Blackwell.

von Stumm, S., & Ackerman, P. L. (2013). Investment and intelligence: A review and meta-analysis. *Psychological Bulletin, 139,* 841–869.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

Yerkes, R. M. (1918). Measuring the mental strength of an army. *Proceedings of the National Academy of Sciences USA, 4,* 295–297.

Ziegler, M., Cengia, A., Mussel, P., & Gerstorf, D. (2015). Openness as a buffer against cognitive decline: The openness–fluid–crystallized intelligence (OFCI) model applied to late adulthood. *Psychology and Aging, 30(3),* 573–588.

PART III

Contemporary Intelligence, Cognitive, and Neuropsychological Batteries and Associated Achievement Tests

CHAPTER 9

The Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition, Wechsler Intelligence Scale for Children—Fifth Edition, and Wechsler Individual Achievement Test—Third Edition

Dustin Wahlstrom
Susan Engi Raiford
Kristina C. Breaux
Jianjun Zhu
Lawrence G. Weiss

The Wechsler intelligence scales are a suite of individually administered instruments for assessing the intellectual functioning of children and adults across the lifespan. Of these tests, two are designed specifically for the assessment of children: the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV; Wechsler, 2012a) for children ages 2 years, 6 months through 7 years, 7 months (i.e., 2:6 through 7:7), and the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014a) for children ages 6:0 through 16:11. Since the original publication of the Wechsler–Bellevue Intelligence Scale in 1939, the Wechsler scales have had a tremendous influence on the field of psychological assessment. They are the most widely used and researched intelligence tests in the field today, due in part to their comprehensiveness, clinical utility, ecological validity, and theoretical soundness (Kamphaus, 1993; Kaplan & Saccuzzo, 2005; A. S. Kaufman, Raiford, & Coalson, 2016; Sattler, Dumont, & Coalson, 2016; Wasserman & Maccubbin, 2002).

The Wechsler family of tests has expanded over the years to include measures of other constructs related to intelligence. One of the more widely used is the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), which was developed to address the growing demand for standardized achievement testing in special education and clinical settings. Intelligence and achievement tests are frequently used in conjunction with each other, whether to identify learning disabilities or to assess the impact of clinical syndromes, acquired injury, or delays in other cognitive domains on learning and academic potential. The Wechsler tests facilitate this process by providing statistical linkage between the intelligence and achievement measures, which supplies practitioners with important information to support their clinical interpretations and conclusions. The current chapter provides an overview of the WPPSI-IV, WISC-V, and WIAT-III, beginning with the intelligence tests, and concluding with a joint interpretive section. An appendix provides a clinical case example to illustrate how results from intelligence and achievement measures can be integrated to provide richer clinical information than either type of test provides alone.

HISTORY AND THEORY BEHIND THE WPPSI-IV AND THE WISC-V

For overviews of the theory underlying the Wechsler scales and the history of their development, see Tulsky and colleagues (2003) and A. S. Kaufman and colleagues (2016). David Wechsler's reputation as an astute clinician, and his inclusion of existing measures with proven clinical and practical utility, did not preclude him from considering the most modern theories of the day when constructing his measures. He was inspired by Charles Spearman's and Edward Thorndike's theories of intelligence at the time. Each successive revision of the scales has been based on the latest contemporary knowledge available.

The modern Wechsler theoretical framework of intelligence uses dialectical reasoning to simultaneously consider structural models of intellect (e.g., Cattell–Horn–Carroll [CHC] theory, *g* theory), evidence of clinical utility (e.g., clinical sensitivity, predictive validity, neuroscience), functional models of cognition (e.g., neuropsychological processing theory), and specific ability models (e.g., working memory). The Wechsler theoretical framework continually informs development to ensure that each revision of the scales is innovative, is contemporary, and draws on the best and most current information these different lines of inquiry have to offer.

Modern editions of Wechsler's tests have changed dramatically compared to their original versions. This is evidenced by the fact that 62% of WISC-V subtests have been developed since the publication of the original WISC in 1949, and all the original items on the retained subtests have been replaced. Similarly, 60% of WPPSI-IV subtests have been developed since the publication of the first WPPSI in 1967. Of the WPPSI subtests that have been retained, all of the original item content has been replaced. Reflected in these changes is the updated theoretical structure of the Wechsler scales that results from added subtests of visual–spatial processing (Visual Puzzles) and fluid reasoning (Matrix Reasoning, Figure Weights, and Picture Concepts), new working memory subtests for young children (Picture Memory and Zoo Locations), and a new visual working memory task for school-age children (Picture Span), as well as improved and more developmentally appropriate measures of processing speed for young children (Bug Search, Cancellation, and Animal Coding). These constructs are increasingly recognized as critical contributors to general intellectual ability, as well as to the understanding of intelligence as a dynamic construct that involves the interaction of working memory, processing speed, and fluid reasoning with other cognitive variables (Dang, Braeken, Colom, Ferrer, & Lui, 2014; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Pineda-Pardo, Martínez, Román, & Colom, 2016; Salthouse, 1996; Weiss, Saklofske, Holdnack, & Prifitera, 2016). The importance of working memory and processing speed are especially relevant to the childhood populations served by the WPPSI-IV and WISC-V. The development of processing speed (Nettelbeck & Burns, 2010) and working memory skills (Tourva, Spanoudis, & Demetriou, 2016) appears to underlie the development of fluid reasoning, with processing speed mediating developmental gains in working memory as well (Fry & Hale, 1996; Nettelbeck & Burns, 2010).

Furthermore, the Working Memory Index (WMI) is strongly related to reading, writing, and math as measured by standardized achievement tests (Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; Konold, 1999; Rowe, Miller, Ebenstein, & Thompson, 2012; Wechsler, 2014b), and working memory is related to academic skills such as spelling, reading decoding, reading comprehension, reading fluency, note taking, and following instructions (Bergman-Nutly & Klingberg, 2014; De Clercq-Quaegebeur et al., 2010; Giofrè, Stoppa, Ferioli, Pezzuti, & Cornoldi, 2016; Jacobson et al., 2011; Malstädt, Hasselhorn, & Lehmann, 2012; Wechsler, 2014b).

Although many of these advances have been made by a team of dedicated clinical researchers since Wechsler's death in 1981, he recognized the importance of the processes that underlie working memory, processing speed, and fluid reasoning (even before they were named as such), and utilized subtests measuring these processes in his early intelligence batteries. For example, Digit Span and Arithmetic, which were initially part of Wechsler's Verbal IQ (VIQ), are now primarily considered tests of working memory and fluid reasoning, respectively. Similarly, Digit–Symbol Coding originally contributed to the Performance IQ (PIQ), but is currently understood as a measure of processing speed; and Wechsler also developed the Symbol Search subtest, even though it didn't appear in the earliest editions of this test.

The result of Wechsler's foresight and the changes to the tests since his passing is a suite of tests that, despite being criticized as atheoretical, yield constructs that are aligned very closely with recent models of intelligence. Independent examinations of recent editions of the WISC indicate that it measures several constructs central to CHC-based models of intelligence, including crystallized ability, visual processing, fluid reasoning, working memory, and processing speed (Keith, Fine, Taub, Reynolds, & Kranzler, 2006). These findings have been supported in studies using normative samples from other countries (Chen, Keith, Chen, & Chang, 2009; Georgas, Van de Vijver, Weiss, & Saklofske, 2003; Reverte, Golay, Favez, Rossier, & Lecerf, 2015). Finally, the WISC-V and WPPSI-IV correlate highly with other intelligence tests explicitly based on neurocognitive theories, such as the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004) and the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007).

WISC-V AND WPPSI-IV TEST FRAMEWORKS AND SUBTESTS

The WISC-V and WPPSI-IV yield subtest scaled scores, index scores, and a general intellectual ability score (the Full Scale IQ [FSIQ]). The subtests that contribute to the composite scores vary between the two instruments, but their test frameworks and many of the subtests for these two measures are similar.

Figures 9.1, 9.2, and 9.3 show the WISC-V and WPPSI-IV composite scores and the subtests that contribute to each score. A great deal of cognitive development takes place during early childhood, so the WPPSI-IV is divided into two batteries for children ages 2:6–3:11 and 4:0–7:7 years to account for these changes. The battery for younger children consists of 5 primary subtests and 2 secondary subtests, whereas the one for older children is composed of 10 primary subtests and 5 secondary subtests. Children in the younger age band take fewer tests with expressive language demands, thus reducing these confounds for a developmental period during which the acquisition of these skills varies widely. The WPPSI-IV younger battery yields an FSIQ, three primary index scores, and three ancillary index scores. The primary index scores are the Verbal Comprehension Index (VCI), the Visual Spatial Index (VSI), and the Working Memory Index (WMI). The ancillary index scores consist of the Vocabulary Acquisition Index (VAI), the Nonverbal Index (NVI), and the General Ability Index (GAI). The older battery provides the same composite scores as the younger battery, but adds two more primary index scores for a total of five, and one more ancillary index score for a total of four. The additional primary index scores include the Fluid Reasoning Index (FRI) and the Processing Speed Index (PSI), and the additional ancillary index score is the Cognitive Proficiency Index (CPI). All primary and ancillary index scores are available in the published test materials.

The WISC-V comprises 10 primary subtests, 6 secondary subtests, and 5 complementary subtests, all of which can be administered to the full age range of 6:0–16:11. The WISC-V yields a Full Scale IQ, five primary index scores, seven ancillary index scores, and three complementary index scores. The primary index scores are the VCI, the VSI, the FRI, the WMI, and the PSI. The ancillary index scores consist of the Verbal (Expanded Crystallized) Index (VECI), the Expanded Fluid Index (EFI), the Quantitative Reasoning Index (QRI), the Auditory Working Memory Index (AWMI), the NVI, the GAI, and the CPI. The VECI and the EFI can be obtained through using tables in a technical report (Raiford, Drozdick, Zhang, & Zhou, 2015) or the scoring software on Q-global or Q-interactive. The remainder of the ancillary index scores are available in the published test materials. The complementary index scores are the Naming Speed Index (NSI), the Symbol Translation Index (STI), and the Storage and Retrieval Index (SRI). All are available in the published test materials.

The WPPSI-IV includes five new subtests: Picture Memory, Zoo Locations, Bug Search, Cancellation, and Animal Coding. Eight new subtests are included in the WISC-V: Visual Puzzles, Figure Weights, Picture Span, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation.

Descriptions of Cognitive Domains, Composite Scores, and Subtests

Full Scale

The Full scale includes subtests that are routinely used to derive the FSIQ, as well as subtests that may be substituted for one invalid subtest for this

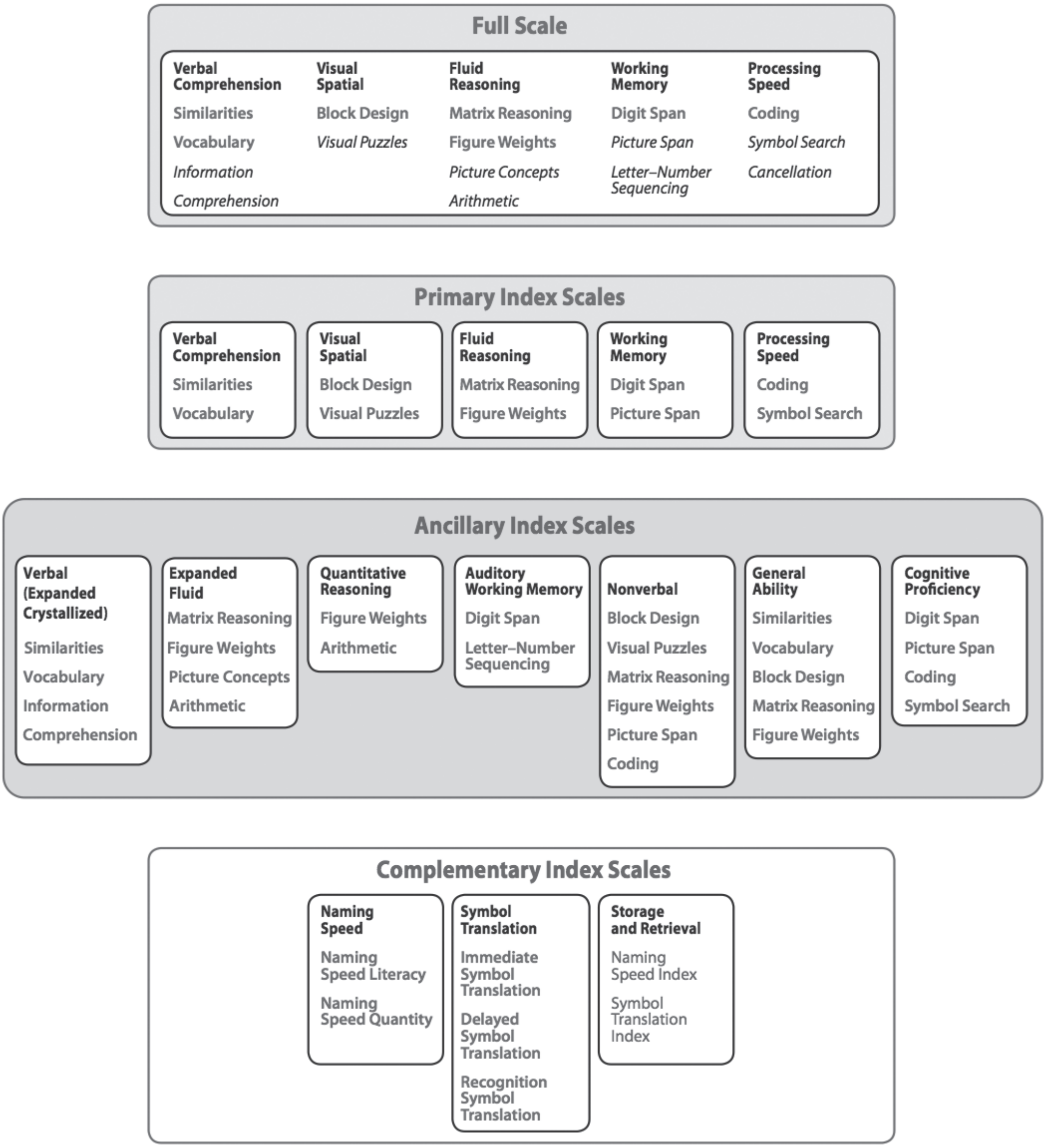


FIGURE 9.1. WISC-V index scores and subtests. Figures found in the manual for the *Wechsler Intelligence Scale for Children®, Fifth Edition (WISC®-V)*. Copyright © 2014 NCS Pearson, Inc. Reproduced with permission. All rights reserved.

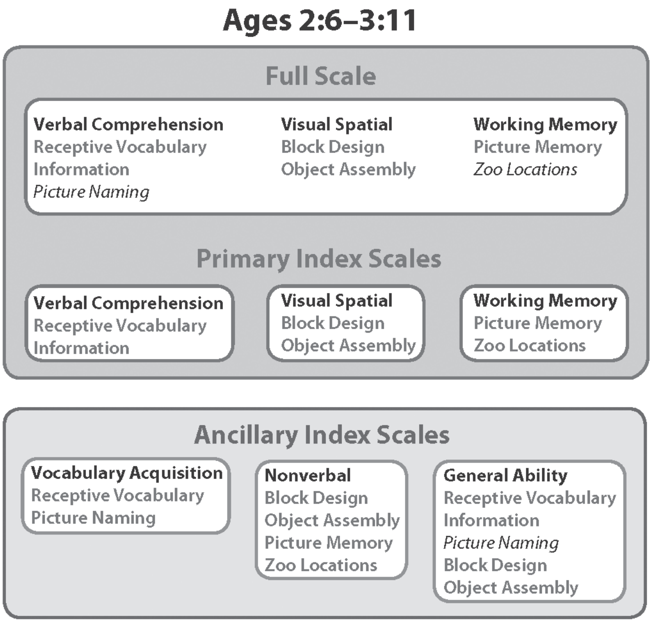


FIGURE 9.2. WPPSI-IV index scores and subtests for ages 2:6–3:11. Figures found in the Manual for the *Wechsler Preschool and Primary Scale of Intelligence®, Fourth Edition (WPPSI®–IV)*. Copyright © 2012 NCS Pearson, Inc. Reproduced with permission. All rights reserved.

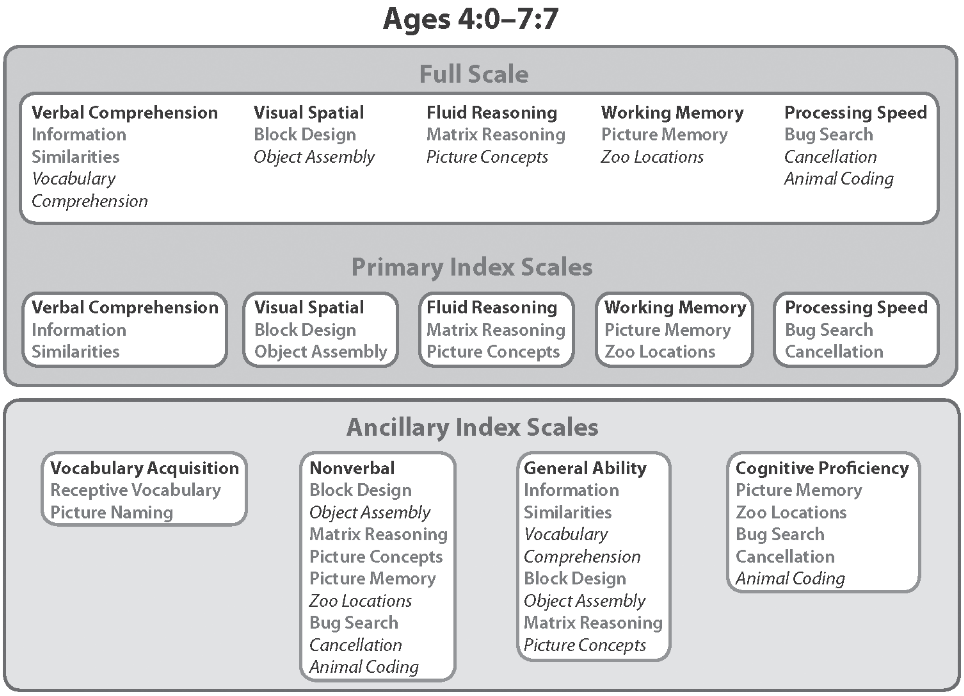


FIGURE 9.3. WPPSI-IV index scores and subtests for ages 4:0–7:7. Figures found in the Manual for the *Wechsler Preschool and Primary Scale of Intelligence®, Fourth Edition (WPPSI®–IV)*. Copyright © 2012 NCS Pearson, Inc. Reproduced with permission. All rights reserved.

purpose. FSIQ is a measure of general intellectual ability that is derived from a subset of subtests on the Full scale. The subtests vary across the WISC-V and WPPSI-IV.

The WISC-V FSIQ is derived from the seven FSIQ subtests. Two of the subtests are drawn from the Verbal Comprehension domain, and two are from the Fluid Reasoning domain. One each are from the Visual Spatial, Working Memory, and Processing Speed domains. Relative to the WISC-IV, the WISC-V FSIQ places greater emphasis on fluid reasoning.

The WPPSI-IV FSIQ subtests vary across the two age bands. The FSIQ for the younger age band is derived from five subtests. Two subtests are drawn from the Verbal Comprehension domain, two from the Visual Spatial domain, and one from the Working Memory domain. For the older WPPSI-IV age band, the FSIQ is derived from six subtests. Two are drawn from the Verbal Comprehension domain, and one each from the Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed domains. The WPPSI-IV FSIQ for the younger age band places greater emphasis on working memory relative to its predecessor. The FSIQ for the older age band emphasizes working memory more heavily than did the WPPSI–III.

The FSIQ is the most reliable score on the WISC-V and the WPPSI-IV, and is considered to be the best predictor of general intellectual functioning (*g*). It is usually the first score to be considered in profile interpretation, but should be interpreted within the context of the subtests that contribute to it and the primary index scores (A. S. Kaufman et al., 2016; Raiford & Coalson, 2014; Weiss et al., 2016).

Verbal Comprehension Domain

The Verbal Comprehension domain contains a variety of subtests that measure verbal concept formation, verbal reasoning, verbal comprehension and expression, acquired knowledge, and practical knowledge and judgment. The VCI is a composite score that is derived from two subtests in the Verbal Comprehension domain, but the contributing subtests vary across instruments (and across WPPSI-IV age bands). The WISC-V VCI measures the ability to acquire, access, and apply semantic knowledge and verbal reasoning. The WPPSI-IV VCI also involves acquisition, access, and application of semantic knowledge, with an added emphasis on acquired knowledge and long-term retrieval.

Because Verbal Comprehension subtests measure concepts such as semantic or factual knowledge, they require a certain level of exposure to cultural and educational opportunities. However, it would be incorrect to assume that the VCI is simply a reflection of semantic knowledge, learned facts, or education, as individual differences in knowledge acquisition depend not only on opportunity, but also on the application of other cognitive abilities during those experiences to take advantage of opportunity (Weiss et al., 2016) and on intellectual curiosity. As such, Verbal Comprehension subtests also reflect the utilization of experience to learn verbal concepts and the ability to apply them in novel situations.

On the WISC-V, all Verbal Comprehension subtests require some form of verbal response from the examinee. On the WPPSI-IV, some of the floor items allow the child to respond nonverbally by pointing, but all stimuli are presented verbally by the examiner. The younger age band further reduces expressive requirements because receptive vocabulary is included.

Similarities

For each Similarities item, the examinee is presented with two words that represent common concepts and asked to describe how they are alike (e.g., "In what way are fish and birds alike?"). On the WPPSI-IV, there are also picture items requiring the examinee to point to the object that is from the same category as two other objects. Similarities measures verbal concept formation and abstract reasoning. It also involves crystallized intelligence, word knowledge, memory, distinctions between nonessential and essential features, and verbal expression (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Vocabulary

Vocabulary consists of both picture-naming items and word definition items. For picture-naming items, the examinee names the object presented visually. For word definition items, the examinee defines words that are read aloud by the examiner and presented visually in the stimulus book (e.g., "What does *conceive* mean?"). Vocabulary measures an individual's verbal concept formation and word knowledge. It also measures crystallized intelligence, fund of knowledge, verbal conceptualization, learning ability, long-term memory, and degree of language development (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Information

Information requires the examinee to answer questions that address a broad range of general knowledge topics (e.g., "Name the country that launched the first man-made satellite"). The WPPSI-IV also contains picture items in which the examinee selects the response option that corresponds to an answer for a general knowledge question. It measures an individual's ability to acquire, retain, and retrieve general factual knowledge. It involves crystallized intelligence, long-term memory, and verbal expression, and the ability to acquire knowledge from formal instruction or other environments (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Comprehension

Comprehension requires the examinee to use his or her understanding of general principles and social situations to answer questions (e.g., "What is the advantage of keeping money in a bank?"). On the WPPSI-IV, there are also picture questions in which the examinee selects the best response option for a general principle or social situation. The subtest measures practical reasoning and judgment. It also involves verbal comprehension and expression, crystallized knowledge, common sense, and verbal reasoning (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Receptive Vocabulary

Receptive Vocabulary is available only on the WPPSI-IV. For Receptive Vocabulary, the child looks at a group of four pictures and points to the one the examiner names aloud. It is a measure of receptive language ability and vocabulary development. It also measures lexical knowledge, fund of information, and perception of visual stimuli (Lichtenberger & Kaufman, 2003; Sattler et al., 2016).

Picture Naming

Picture Naming is available only on the WPPSI-IV. For each Picture Naming item, the child names a picture that is displayed in the stimulus book. It measures expressive language ability, as well as acquired knowledge, fund of information, long-term memory, and perception of visual stimuli (Lichtenberger & Kaufman, 2003; Sattler et al., 2016).

Visual Spatial Domain

The Visual Spatial domain contains subtests that measure the ability to evaluate visual details, understand spatial relations among objects, and construct geometric designs using a model. This requires visual spatial processing, synthesis of part–whole relations, attention to visual detail, and (for two of the subtests) visual–motor integration.

The VSI is a composite score that is derived from two subtests in the Visual Spatial domain, but the contributing subtests vary across instruments. The WISC-V VSI includes one task, Block Design, that involves visual–motor construction and uses the same stimuli throughout the subtest (i.e., two-color blocks). Another task, Visual Puzzles, involves mental rotation and varies stimuli throughout the subtest; the puzzles are abstract shapes (e.g., colored squares, circles, triangles). The WPPSI-IV VSI also includes Block Design, but floor items involve use of different (single-colored) blocks. Object Assembly, in which children use physical pieces to assemble puzzles of meaningful stimuli, is the other WPPSI-IV subtest used to obtain the VSI.

Block Design

Block Design requires the examinee to view a constructed model or a picture and to use one-color or two-color blocks to recreate the design within a specified time limit. It measures the ability to analyze and synthesize abstract visual stimuli. It also involves nonverbal concept formation, visual perception and organization, broad visual intelligence, visual–motor coordination, learning, and the ability to separate figure and ground in visual stimuli (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Visual Puzzles

Visual Puzzles is a new subtest for the WISC-V, adapted from the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). On Visual Puzzles, examinees reproduce a geometric image by choosing three response options that can be combined to form the image from six available options within a predetermined time limit. It was designed to measure mental and nonmotor construction skills, visual–spatial processing and reasoning, and mental rotation. It is also thought to involve visual working memory, broad visual intelligence, fluid reasoning, and simultaneous processing (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Object Assembly

Object Assembly is available only on the WPPSI-IV. For Object Assembly, the examinee fits the pieces of a puzzle together to form an identifiable image within specified time limits. It is designed to assess visual–spatial processing and visual–motor integration. It also involves perceptual integration and synthesis of part–whole relationships, nonverbal reasoning, trial-and-error learning, spatial ability, cognitive flexibility, and persistence (Groth-Marnat, 2009; Lichtenberger & Kaufman, 2003; Sattler et al., 2016).

Fluid Reasoning Domain

The Fluid Reasoning domain contains subtests that measure the ability to use reasoning to identify and apply solutions to problems. The FRI is a composite score that is derived from two subtests in the Fluid Reasoning domain. As with the VCI and the VSI, the contributing subtests vary, depending on the instrument. The WPPSI-IV younger age battery does not include the Fluid Reasoning domain because many younger children cannot comprehend the tasks involved.

The WISC-V FRI is derived from two tasks that involve detecting conceptual relations among abstract visual objects and then applying that relation to identify a solution (i.e., Matrix Reasoning and Figure Weights). The WPPSI-IV FRI, available only for ages 4:0–7:7, includes one task that utilizes both realistic pictures and abstract stimuli and requires selection of a single response (i.e., Matrix Reasoning), and another involving a set of realistic pictures that are scanned to determine multiple options which comprise a unified set (i.e., Picture Concepts).

Matrix Reasoning

For each Matrix Reasoning item, the examinee looks at an incomplete matrix or series and selects the missing portion from five response options. This subtest is designed to measure fluid intelligence, classification and spatial ability, simultaneous processing, attention to visual detail, and working memory (Flanagan & Alfonso, 2017; Groth-Marnat, 2009; Sattler et al., 2016).

Figure Weights

Figure Weights is another new subtest for the WISC-V that is adapted from the WAIS-IV. For each item, individuals must balance a scale by identifying the correct response option within a specified time limit. In order to determine the correct response, the examinee must figure out the relationships between shapes that balanced a previous scale and apply these relationships to the incomplete scale. The task measures quantitative and inductive reasoning. It also requires mental flexibility and set shifting (Flanagan & Alfonso, 2017; Flanagan, Alfonso, & Ortiz, 2012; Flanagan & Kaufman, 2009; Sattler et al., 2016).

Picture Concepts

On Picture Concepts, the examinee is presented with two or three rows of pictures and chooses one picture from each row to form a group with a common characteristic. It measures fluid and inductive reasoning, visual-perceptual recognition and processing, conceptual thinking, visual scanning, working memory, and conceptual thinking (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Sattler et al., 2016).

Arithmetic

Arithmetic is available only on the WISC-V. In Arithmetic, the examinee mentally solves a series of orally presented arithmetic problems within a specified time limit (e.g., "Jim buys five stamps, and his mother gives him two more. He then uses one to mail a letter. How many stamps does he have left?"). Arithmetic involves mental manipulation, concentration, and working memory; short- and long-term memory; fluid, quantitative, and logical reasoning; sequential processing; and quantitative knowledge (Benson, Hulac, & Kranzler, 2010; Flanagan & Alfonso, 2017; Sattler et al., 2016).

Arithmetic has been moved from the Working Memory domain of the WISC-IV to the Fluid Reasoning domain of the WISC-V. Including the new Visual Puzzles, Figure Weights, and Picture Span subtests in the WISC-V permitted a split of the former Perceptual Reasoning factor into Visual Spatial and Fluid Reasoning factors, which increased the number of factors in the test model from four to five. This is not surprising because the results of factor analyses can vary, depending on the subtests in the set being tested. In addition, studies of the WISC-IV (Keith et al., 2006; Weiss, Keith, Zhu, & Chen, 2013b) and the WAIS-IV (Benson et al., 2010; Weiss, Keith, Zhu, & Chen, 2013a) had previously indicated that Arithmetic loads on the Working Memory factor in four-factor models, but on Fluid Reasoning in five-factor models. Therefore, the WISC-V confirmatory factor-analytic studies compared several models with Arithmetic on the Fluid Reasoning or the Working Memory factor. These models showed the best fit with Arithmetic on the Fluid Reasoning factor, with a cross-loading on the Working Memory factor. Mental arithmetic based on word problems is a complex mental task involving successful integration of several cognitive abilities, and is therefore highly *g*-loaded.

Working Memory Domain

The Working Memory domain contains subtests that measure the ability to consciously register, maintain, and manipulate auditory and visual information. This requires paying attention and focusing, keeping the information in conscious awareness, mentally processing the information in a manner that conforms to the task demands, and then providing a result.

All of the WISC-V and WPPSI-IV Working Memory subtests involve use of proactive interference. Proactive interference involves repeated exposure to stimuli across items, such that prior exposure interferes with memory for the present item. A number of studies support the effectiveness of this method (Lipinski, Simmering, Johnson, & Spencer, 2010; Makovski & Jiang, 2008; Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). Its first use with infants and toddlers was described by Piaget (1952).

Proactive interference is particularly important within the WPPSI-IV subtests, which rely on proactive interference to introduce a competing cognitive processing demand that tax working memory. Because a number of challenges related to cognitive development (e.g., limited working memory capacity, distractibility, inability to use rehearsal strategies, lack of comprehension of complex instructions) render complex tasks difficult to teach and perform.

Both of the WPPSI-IV Working Memory subtests involve visual stimuli rather than auditory. At first blush, this may seem to limit the comprehensiveness of construct measurement because there are not auditory working memory subtests on the WPPSI-IV. Whereas the domain-specific storage components of working memory appear to be distinct in young children in the WPPSI-IV age range, and research with young children supports Baddeley's (2012) multicomponent model (Alloway, Gathercole, & Pickering, 2006; Hornung, Brunner, Reuter, & Martin, 2011), memory storage in the visual domain is strongly linked with cognitive processing, but auditory memory storage is not (Alloway et al., 2006; Hornung et al., 2011). This suggests that simple visual memory tasks involving storage alone are preferable to an analogous auditory task when attempting to measure working memory for young children.

The WMI is a composite score that is derived from two subtests in the Working Memory domain. As with other primary index scores, the contributing subtests vary, depending on the instrument. The WISC-V WMI is based on the sum of scaled scores for Digit Span and Picture Span. Both subtests require resequencing of information. However, Digit Span involves numerical auditory stimuli and oral responses, whereas Picture Span involves pictorial stimuli that can feasibly (for many items) be verbally mediated and rehearsed, and either gestured or oral responses. The WPPSI-IV WMI involves Picture Memory, which is similar to Picture Span but does not require resequencing, and Zoo Locations, which involves visual–spatial information and a performance response.

Digit Span

Digit Span is available only on the WISC-V. It has traditionally been composed of two parts: Digit Span Forward and Digit Span Backward. Digit Span Forward requires the examinee to repeat numbers in the same order as read aloud by the examiner, and Digit Span Backward requires the examinee to repeat the numbers in the reverse order of that presented by the examiner. The WISC-V (and the WAIS-IV) have added Digit Span Sequencing, which requires examinees to sequentially order the numbers presented by the examiner. This subtest measures working memory, in addition to auditory short-term memory, sequential processing, and mental manipulation (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Groth-Marnat, 2009; Sattler et al., 2016).

Picture Memory/Picture Span

Picture Span is a new Working Memory subtest for the WISC-V, and Picture Memory is a new Working Memory subtest for the WPPSI-IV. The tasks are very similar; in each, the examinee views a stimulus page with one or more objects for a predefined period and is then asked to select those objects among a larger group of options on a response page. On Picture Span, the examinee is asked only to select the objects seen on the stimulus page in order. Additional credit is awarded if the examinee can select the objects in the order they were presented on the stimulus page. For Picture Memory, the examinee only selects the objects seen on the stimulus page, regardless of order.

These subtests measure visual working memory and working memory capacity; they are similar to other tasks that are known to measure attention, visual processing, and response inhibition (Flanagan & Alfonso, 2017; Flanagan et al., 2012; Flanagan, Alfonso, Ortiz, & Dynda, 2010; Miller, 2010, 2013; Sattler et al., 2016).

Letter–Number Sequencing

Letter–Number Sequencing is only available for the WISC-V. For this task, the examiner reads a sequence of numbers and letters to the examinee, who recalls the numbers in ascending order and the letters in alphabetical order. Similar to Digit Span, it measures sequencing and working memory, as well as auditory sequential processing, immediate auditory memory, attention, numerical ability, auditory working memory, visual–spatial imaging, and processing speed (Flanagan & Alfonso, 2017; Groth-Marnat, 2009; Sattler et al., 2016).

Zoo Locations

Zoo Locations is a visual working memory subtest on the WPPSI-IV. It requires the examinee to view one or more animal cards on a zoo layout for a specified period of time and then re-place them in the correct locations. It is designed to measure visual–spatial working memory (Sattler et al., 2016), similar to established spatial working memory tasks (Blalock & McCabe, 2011; Lipinski et al., 2010). Proactive interference is utilized to increase the working memory load across items.

Processing Speed Domain

The Processing Speed domain contains subtests that measure the ability to use reasoning to identify and apply solutions to problems. The PSI is a composite score that is derived from two subtests in the Processing Speed domain. As with the VCI and the VSI, the contributing subtests vary, depending on the instrument. The WPPSI-IV younger age battery does not include the Processing Speed domain because many younger children have difficulty with the concept of working quickly.

The WISC-V PSI is derived from two tasks that involve simple perceptual speed. One of these tasks, Coding, places demands on associative memory, and the other, Symbol Search, requires visual scanning and discrimination. The WPPSI-IV PSI, available only for ages 4:0–7:7, includes two tasks that involve simple perceptual speed. One of these tasks, Bug Search, is an adaptation that serves as a downward extension of Symbol Search, and the other, Cancellation, involves categorical reasoning as well as visual scanning and discrimination.

Coding

Coding is a WISC-V subtest that requires the examinee to associate symbols paired with numbers, but the exact task demands differ slightly, depending on whether the paper or digital format of the task is being administered. For the paper format, using a key, the examinee draws a symbol in each numbered box within a specified time limit. For the digital format, the examinee taps the symbol associated with the correct number. Both the paper and digital formats of Coding measure processing speed, as well as short-term memory, learning ability, visual acuity, sequential processing, and attention to visual stimuli (Flanagan & Alfonso, 2017; Groth-Marnat, 2009; Sattler et al., 2016). The paper format also measures graphomotor speed due to the writing demands—a construct not measured by the digital format because the correct answers are tapped rather than written. This reduces the unwanted influence of fine motor skills in the measurement of visual processing speed. See “Administration Options,” below, for a more thorough discussion of the digital adaptation of Coding and Symbol Search.

Symbol Search

Similar to Coding, Symbol Search is a WISC-V subtest that can be administered with a paper response booklet or in digital format without a response booklet. However, the two formats are very similar in their response demands. In each, the examinee is required to scan a group of symbols and indicate whether the target symbols match any of the symbols in the group within a specified time limit. In paper, the examinee indicates the correct response by drawing a line through it. For the digital format, the correct answer is tapped. Symbol Search measures visual–motor processing speed, as well as short-term visual memory, perceptual organization, learning ability, perceptual and psychomotor speed, visual–motor coordination, visual discrimination, and attention to visual stimuli (Groth-Marnat, 2009; Sattler et al., 2016).

Bug Search

Bug Search is a WPPSI-IV subtest that is an adaptation and downward extension of Symbol Search for younger children. It requires the examinee to mark a bug in the search group that matches the target. Several modifications make it more age-appropriate than Symbol Search. For example, the stimuli are larger and color-coded to make them easier to discriminate, and the examinee indicates the correct response by using a dauber rather than a pencil. It is intended to measure many of the same constructs as Symbol Search. It measures processing speed, attention, and concentration (Sattler et al., 2016).

Cancellation

On Cancellation, the examinee is required to identify and mark target objects interspersed among distractors. The stimuli are presented in a random array on the first trial and organized into rows on the second trial. Cancellation measures rate of test taking, perceptual speed, visual discrimination, visual scanning, and visual-perceptual processing (Flanagan & Alfonso, 2017; Sattler et al., 2016).

Animal Coding

Animal Coding is a WPPSI-IV subtest that is intended to be an age-appropriate version of Coding. Using a key, the examinee marks shapes that correspond to symbols within a specified time limit. It differs from Coding in that the child marks the correct response by using a dauber rather than a pencil, as well as the fact that there are fewer shape–animal relationships to scan than number–symbol relationships. It is intended to be a measure of visual–motor processing speed that measures constructs similar to those involved in Coding. It measures processing speed (Sattler et al., 2016).

Quantitative Reasoning Index

The QRI is a WISC-V ancillary index score derived from Figure Weights and Arithmetic that is intended to measure the examinee's quantitative reasoning skills. Quantitative reasoning is closely related to *g* (Weiss et al., 2013a, 2013b) and predictive of reading and math achievement, creativity, and success in giftedness programs (Lakin & Lohman, 2011; Robertson, Smeets, Lubinski, & Benbow, 2010). The QRI may be of special interest in cases where a specific learning disability in mathematics is suspected. Low quantitative reasoning may be due to difficulties with mental math manipulation, poor understanding of quantitative relationships, or low working memory, and thus may constitute a specific area of focus for intervention (Wechsler, 2014b; Zheng, Flynn, & Swanson, 2013).

Auditory Working Memory Index

The addition of Picture Span to the WISC-V WMI results in that index's measuring a combination of auditory and visual working memory; this is a departure from previous editions of the WISC, in which the WMI measured auditory working memory only. The AWMI consists of Digit Span and Letter–Number Sequencing, the same subtests used to derive the WISC-IV WMI (though there is an increased emphasis on sequencing, due to the new Digit Span Sequencing condition). In this respect, the AWMI may be useful when an examinee's score is being compared with scores on previous evaluations that included the WISC-IV. Furthermore, contemporary working memory research suggests that domain-specific neuropsychological systems support auditory working memory versus visual–spatial working memory: These are the *phonological loop* and *visuospatial sketchpad*, respectively (Baddeley, 2012). Within this context, the AWMI is a purer measure of auditory working memory processes supported by the phonological loop.

Nonverbal Index

The NVI is derived from subtests that do not require an expressive response. Because it consists of subtests from all available cognitive domains excluding Verbal Comprehension, the NVI can be considered a good measure of general intellectual functioning that minimizes expressive language demands. It may be useful in testing children who are English-language learners (Raiford & Coalson, 2014) or who have language-based clinical disorders, such as autism spectrum disorder with language impairment. However, the NVI requires items to be presented verbally; as such, it should not be considered a "language-free" assessment, nor should it be substituted for measures like the Wechsler Nonverbal Scale of Ability (WNV).

The WISC-V NVI is derived from Block Design, Visual Puzzles, Matrix Reasoning, Figure Weights, Picture Span, and Coding. The WPPSI-IV NVI subtests vary across the two age bands. The NVI for the younger age band is derived from four subtests (i.e., Block Design, Object Assembly, Picture Memory, and Zoo Locations). Two subtests are drawn from the Visual Spatial domain, and two from the Working Memory domain. For the older WPPSI-IV age band, the NVI is derived from five subtests (i.e., Block Design, Matrix Reasoning, Picture Concepts, Picture Memory, and Bug Search). One is drawn from the Visual Spatial domain, two from the Fluid Reasoning domain, and one each from the Working Memory and Processing Speed domains.

General Ability Index and Cognitive Proficiency Index

The GAI and CPI are provided in the published test manuals as ancillary index scores. The GAI is derived from all the subtests that contribute to the FSIQ except for the Working Memory and Processing Speed subtests. The CPI is derived from all four subtests that contribute to the WMI and the PSI. Thus the GAI measures a subset of intellectual functioning with reduced influences of working memory and processing speed, and the CPI represents an index of cognitive processing proficiency that reduces crystallized knowledge, verbal reasoning, and fluid reasoning demands. These index scores may be especially useful in the context of evaluations for specific learning disorders.

Because working memory and processing speed subtests contribute to the FSIQ, lower FSIQ scores may occur in the presence of neurodevelopmental disorders known to be associated with difficulties in working memory and processing speed. These include specific learning disorders, attention-deficit/hyperactivity disorder (ADHD), language disorders, and autism spectrum disorder (Akbar, Loomis, & Paul, 2013; Horowitz-Kraus, 2014; Kasper, Alderson, & Hudec, 2012; Niileksela & Reynolds, 2014; Pimperton & Nation, 2014; Thaler, Bello, & Etcoff, 2012; Travers et al., 2014; Vugs, Cuperus, Hendriks, & Verhoeven, 2013). The lower FSIQ scores may in turn obscure meaningful differences between intellectual ability (represented by the FSIQ) and other cognitive functions (e.g., achievement and memory).

Some have advocated use of the GAI in placement decisions for gifted and talented or similar programs (Rimm, Gilman, & Silverman, 2008). There are indications that some children in gifted education programs tend to obtain higher scores on the GAI than on the CPI or FSIQ. For example, 34% of children in the WISC-V intellectually gifted special group study showed GAI > FSIQ discrepancies. Although a lower percentage, this is similar to findings for the WISC-IV (Raiford, Rolfhus, Weiss, & Coalson, 2005), the WPPSI-IV (Wechsler, 2012a), and the WAIS-IV (Wechsler, 2008). Furthermore, 72% of children in the intellectually gifted special group study obtained GAI > CPI discrepancies.

Discrepancies in which the GAI is higher than the CPI are also implicated in traumatic brain injury and autism spectrum disorder, and the GAI is discrepant from the FSIQ in a multitude of other clinical populations, including children with intellectual disability, traumatic brain injury, ADHD, and autism and Asperger's disorder as previously defined by DMS-IV (Saklofske, Weiss, Raiford, & Prifitera, 2006; Strauss, Sherman, & Spreen, 2006). However, the GAI should not be used as a substitute for FSIQ solely because the WMI or the PSI is low, as working memory and processing speed represent important aspects of general cognitive ability.

Vocabulary Acquisition Index

The VAI is a WPPSI-IV ancillary index score derived from the Receptive Vocabulary and Picture Naming subtests. It is intended to provide a measure of vocabulary development. When combined with other sources of information such as parent and teacher report and behavioral observations, low scores on the VAI may warrant further evaluation with a speech–language pathologist.

WISC-V Complementary Scales

Naming Speed Literacy

Naming Speed Literacy requires the examinee to name objects, letters, or numbers as quickly as possible within a specified time limit. There are three different conditions: (1) Color–Object Naming, in which the examinee is required to say the name and color of an object (e.g., "red cat"); (2) Size–Color–Object Naming, in which the size of the object must also be named (e.g., "big green cat"), and (3) Letter–Number Naming, which requires the examinee to name letters and numbers. Naming Speed Literacy is a rapid naming task intended to measure storage and retrieval fluency and naming facility (Flanagan & Alfonso, 2017). It is not intended as a measure of intelligence, but rather a complementary subtest that can assist in the identification of cognitive weaknesses associated with academic learning. As such, similar tasks have been related to the development of reading and spelling skills, as well as learning disabilities, ADHD, and autism spectrum disorder (Crews & D'Amato, 2009; Korkman, Kirk, & Kemp, 2007; Powell, Stainthorp, Stuart, Garwood, & Quinlan, 2007; Weiss et al., 2016; Willburger, Fussenegger, Moll, Wood, & Landerl, 2008).

Naming Speed Quantity

Naming Speed Quantity requires the examinee to name the number of squares inside a larger box (one item varies the number of boxes from one to four, and another from one to five). The task measures storage and retrieval fluency and naming facility (Flanagan, Alfonso, & Ortiz, 2012). It is intended to complement Naming Speed Literacy in instances where math weaknesses are suspected, as tasks similar to Naming Speed Quantity have been shown to be more closely related to mathematics skills and specific learning disorders in mathematics than have letter–number or object-naming tasks (Pauly et al., 2011; van der Sluis, de Jong, & van der Leij, 2004; Weiss et al., 2016; Willburger et al., 2008).

Immediate Symbol Translation

Immediate Symbol Translation is intended to assist in the identification of cognitive weaknesses that may be having an impact on academic learning. For this subtest, the examinee is shown a series of symbols and taught told that each symbol is associated with a word, after which he or she is asked to "read" sentences made out of the symbols. New symbol–word associations are taught as the task progresses, and the sentences increase in length and complexity. Immediate Symbol Translation measures verbal–visual associative memory, storage and retrieval fluency and accuracy, and immediate recall (Flanagan & Alfonso, 2017; Weiss et al., 2016). These tasks are related to decoding skills, word-reading accuracy and fluency, and reading comprehension, as well as math calculation and reasoning skills (Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007; Hulme, Goetz, Gooch, Adams, & Snowling, 2007; Litt, de Jong, van Bergen, & Nation, 2013). As might be expected, they are sensitive to dyslexia as long as the examinee is required to respond verbally (Gang & Siegel, 2002; Litt & Nation, 2014).

Delayed Symbol Translation

Delayed Symbol Translation is administered 20–30 minutes after Immediate Symbol Translation and requires the examinee to translate the symbols into sentences from memory. No more symbol–word associations are taught as part of the delayed subtest. Tasks similar to Delayed Symbol Translation measure verbal–visual associative memory, storage and retrieval fluency and accuracy, and delayed recall (Flanagan & Alfonso, 2017).

Recognition Symbol Translation

On Recognition Symbol Translation, the examinee views one of the previously learned symbols and selects the word that corresponds to it from a list read aloud by the examiner. Performance on Recognition Symbol Translation can be compared to Delayed Symbol Translation to compare the ability to encode the symbol–word associations with the ability to retrieve them (e.g., intact recognition with impaired delayed recall suggests a weakness in retrieval). This subtest measures verbal–visual associative memory, storage and retrieval fluency and accuracy, and delayed recognition (Flanagan & Alfonso, 2017; Weiss et al., 2016).

Naming Speed Index

The NSI is derived from Naming Speed Literacy and Naming Speed Quantity. At a high level, it is intended to reflect the automaticity of naming as measured by a wide range of tasks (Wechsler, 2014b). In addition, it measures the ability to register visual stimuli and to retrieve verbal labels from long-term memory, working memory, visual processing speed, and oral–motor sequencing (Flanagan & Alfonso, 2017). The NSI is not a measure of intelligence, but rather an optional score that can enhance the assessment of examinees suspected of specific learning disabilities or other neurodevelopmental disorders.

Symbol Translation Index

The STI comprises all three Symbol Translation subtests and is a summary measure of visual–verbal associative memory across the different types of conditions each subtest represents (e.g., encoding, recall, recognition). It also measures auditory and visual processing, auditory discrimination, and attention to visual and auditory processing (Flanagan & Alfonso, 2017; A. S. Kaufman et al., 2016). Like the NSI, it is not intended to be a measure of intelligence, but rather is intended to enhance the assessment of examinees with specific learning disabilities, memory impairment, or other neurodevelopmental conditions.

Storage and Retrieval Index

The SRI is derived from the sum of standard scores of the NSI and STI. As such, it reflects the ability to accurately and efficiently store and retrieve auditory and visual information from long-term memory. High performance on this index may reflect well-developed capacity for learning and easy access to stored verbal knowledge. On the other hand, poor performance may reflect a number of things, including deficits in the encoding or retrieval of long-term memory stores, inattention, or visual or verbal processing weaknesses.

PSYCHOMETRIC PROPERTIES OF THE WISC-V AND WPPSI-IV

According to the criteria proposed by Flanagan and Alfonso (1995) and Bracken (1987), the WISC-V and the WPPSI-IV have outstanding psychometric properties.

Normative Samples

The WISC-V and the WPPSI-IV have excellent normative samples (Dumont & Willis, 2014; A. S. Kaufman et al., 2016). The sizes of the normative samples are 2,200 for the WISC-V and 1,700 for the WPPSI-IV, respectively. The sample size for most norming age groups is 200 cases. The normative samples are stratified to closely match contemporary U.S. census data for race/ethnicity, parent education level, and geographic region. Each age group contains 200 children, except for the oldest group at the upper end of the WPPSI-IV (i.e., 7:0–7:7), which is composed of 100 children.

Reliability

The WPPSI-IV and WISC-V composite scores have strong reliabilities (Wechsler, 2012b, 2014b). The overall internal-consistency reliability coefficients for the normative sample are in the .90s for all IQ and index scores except the WPPSI-IV VSI, which is .89. The PSI composites for both the WPPSI-IV and WISC-V are slightly below .90, but were calculated using test–retest methods because they are speeded tests. Test–retest typically yields lower reliability coefficients compared to internal-consistency methods. At the subtest level, the overall internal-consistency reliability coefficients of the normative sample are in the .80s or .90s for all of the WPPSI-IV subtests with the exceptions of Cancellation and Animal Coding (with test–retest reliability coefficients of .76 and .75, respectively. On the WISC-V, all subtests have internal-consistency reliability of above .80.

Overall, the reliability of the WPPSI-IV and the WISC-V for special group samples is consistent with reliability estimates for the normative samples (Wechsler, 2012b, 2014b). The test-retest stability coefficients of the WPPSI-IV and WISC-V FSIQ scores are .93 and .92, respectively. The test–retest coefficients of the index scores range from .82 to .94, with the exception of the WISC-V FRI, which has a test–retest reliability of .75. The subtest coefficients range from .71 to .90. Finally, the interscorer agreements of the WPPSI-IV and WISC-V subtests are all .96, even for the verbal subtests that require more subjective scoring techniques.

Validity

There is ample evidence to support the validity of the WISC-V and the WPPSI-IV. One of the largest changes introduced in these revisions was the five-factor index score test structure, consisting of Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed domains. Evidence supporting this model comes from the factor analyses of the WPPSI-IV and WISC-V normative samples contained in the test manuals (Wechsler, 2012b, 2014b), as well as studies demonstrating that the five-factor model had support similar to that for the four-factor model when applied to the WISC-IV and WAIS-IV (Benson et al., 2010; Keith et al., 2006; Weiss et al., 2013a, 2013b).

Taken together, results from these studies indicate that the latent traits measured by the Wechsler scales appear consistently across different ages, ethnicities, cultures, specific clinical populations, and sexes. Furthermore, they support the updated theoretical foundations of the Wechsler scales, as converging evidence suggests that the tests measure working memory, processing speed, and fluid reasoning, among other cognitive abilities recently identified as critical to the construct of intelligence.

In addition to the factor-analytic work supporting the construct validity of the WPPSI-IV and WISC-V, a multitude of evidence supports their utility as clinical tools. This is consistent with David Wechsler's original intent of developing an intelligence battery that was, above all else, a clinically useful instrument. Data collected during the standardization phase for both tests with a number of special groups (children with ADHD, autism spectrum disorder, learning disabilities, intellectual disability, etc.) demonstrate unique patterns of index score strengths and weaknesses relative to the normative samples (Wechsler, 2012b, 2014b). These results are corroborated by a number of independent samples that show similar patterns of performance (Calhoun & Mayes, 2005; Ghaziuddin & Mountain-Kimchi, 2004; Harrison, DeLisle, & Parker, 2008; Mayes & Calhoun, 2006; Rackley, Allen, Fuhrman, & Mayfield, 2012; Sweetland, Reina, & Tatti, 2006), though much of this evidence stems from previous versions of the tests, given their recent revisions. In addition, the Wechsler scales show significant utility for identification of specific learning disabilities (Flanagan, Ortiz, & Alfonso, 2013; Hale et al., 2008).

Finally, the Wechsler scales also correlate highly with other measures of intelligence. For example, the correlation between the WISC-V FSIQ and KABC-II Fluid–Crystallized Index (FCI) and Mental Processing Index (MPI) are .81 and .77, respectively (Wechsler, 2014b). Similarly, the correlation between the WPPSI-IV FSIQ and the DAS-II General Conceptual Ability (GCA) and Special Nonverbal Composite (SNC) scores are .81 and .75, respectively (Wechsler, 2012a). The magnitude of these correlations is high and provides evidence of concurrent validity, as the Wechsler scales appear to measure very similar constructs relative to other modern intelligence batteries. Similarly, the WPPSI-IV and WISC-V correlate highly with academic achievement as measured by the WIAT-III (discussed later in this chapter; Wechsler, 2012a, 2014a). Correlations between the WPPSI-IV FSIQ and WIAT-III index scores range from .50 to .75, and correlations between the WISC-V FSIQ and WIAT-III composites range from .58 to .81.

WIAT-III BACKGROUND AND HISTORY

The WIAT-III is a comprehensive, individually administered achievement test of listening, speaking, reading, writing, and mathematics skills. Both grade and age norms are provided for testing individuals who are in prekindergarten (PK) through grade 12 or ages 4:0 through 50:11 years. Separate norms are provided for fall, winter, and spring for grades PK–12. Administration time varies according to a number of factors, such as the grade and skill level of the examinee and the number of subtests administered; however, subtest administration time is typically between 1 and 15 minutes, depending on the subtest.

The original WIAT (Psychological Corporation, 1992) was designed to measure the academic achievement of students in kindergarten through high school, ages 5:0–19:11. The WIAT provided eight subtests to correspond to each of the areas of learning disability identification specified by the Education for All Handicapped Children Act of 1975 (P.L. 94-142) and was the first test of its kind to be linked with the Wechsler ability scales for conducting ability–achievement discrepancy analyses. Nine years later, the WIAT-II (Psychological Corporation, 2001) was published, with one new subtest (Pseudoword Decoding) and significant revisions to the existing subtests. Subsequently, updated scoring and normative materials were released in 2002 and 2005. The WIAT-II was designed for children, adolescents, college students, and adults ages 4:0 through 85:11 years. The WIAT-III retains updated versions of the subtests included in the previous editions and adds five new subtests: Early Reading Skills, Oral Reading Fluency, Math Fluency—Addition, Math Fluency—Subtraction, and Math Fluency—Multiplication. In addition, the former Written Expression subtest was split into three distinct subtests: Alphabet Writing Fluency, Sentence Composition, and Essay Composition. It is normed for individuals between the ages of 4:0 and 50:11 years.

WIAT-III SUBTESTS AND COMPOSITES

As shown in Table 9.1, the WIAT-III covers all 8 areas specified by the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), with 16 subtests and 8 composite scores (7 achievement area composites and 1 Total Achieve-

TABLE 9.1. Alignment of WIAT-III with IDEA 2004

IDEA 2004 areas of achievement	WIAT-III subtests	WIAT-III composites	
Oral expression	Oral Expression	Oral Language	
Listening comprehension	Listening Comprehension		
Written expression	Alphabet Writing Fluency Sentence Composition Essay Composition Spelling	Written Expression	
Basic reading skills	Early Reading Skills[a] Word Reading Pseudoword Decoding	Basic Reading	Total Reading
Reading fluency skills	Oral Reading Fluency	Reading Comprehension and Fluency	
Reading comprehension	Reading Comprehension		
Mathematics calculation	Numerical Operations	Mathematics	
Mathematics problem solving	Math Problem Solving		
	Math Fluency—Addition Math Fluency—Subtraction Math Fluency—Multiplication	Math Fluency	

[a] Early Reading Skills does not contribute to either the Basic Reading or Total Reading composite.

ment composite). With the exception of the three Math Fluency subtests, each subtest contributes to the Total Achievement composite to provide an estimate of overall academic achievement. An examiner may choose to administer as many or as few subtests as he or she deems appropriate for the purpose of the evaluation, the types of scores required, and the student's grade level. Figure 9.4 shows all WIAT-III composites and subtests.

Subtest Descriptions

Listening Comprehension, administered to examinees ages 4–50 (grades PK–12+), contains two components, Receptive Vocabulary and Oral Discourse Comprehension. For Receptive Vocabulary, the examinee points to the picture that best illustrates the meaning of each word he or she hears. For Oral Discourse Comprehension, the examinee listens to sentences and passages, and orally responds to comprehension questions.

Early Reading Skills, administered to examinees ages 4–9 (grades PK–3), requires the examinee to name letters, identify and generate rhyming words, identify words with the same beginning and ending sounds, blend sounds, match sounds with letters and letter blends, and match written words with pictures that illustrate their meaning.

Reading Comprehension, administered to examinees ages 6–50 (grades 1–12+), requires the examinee to read narrative and expository passages (either aloud or silently), and then orally respond to literal and inferential comprehension questions read aloud by the examiner.

Math Problem Solving, administered to examinees ages 4–50 (grades PK–12+), measures untimed math problem-solving skills in the areas of basic concepts, everyday applications, geometry, and algebra. The items require oral or pointing responses.

Alphabet Writing Fluency, administered to examinees ages 4–9 (grades PK–3), measures the ability to write letters of the alphabet within a 30-second time limit. The examinee may write letters in any order, in cursive or print, in uppercase or lowercase.

Sentence Composition, administered to examinees ages 6–50 (grades 1–12+), contains two components: Sentence Combining and Sentence Building. Sentence Combining requires the examinee to combine two or three sentences into one sentence that preserves the meaning of the original sentences. Sentence Building requires the examinee to write sentences that include a target word.

Word Reading, administered to examinees ages 6–50 (grades 1–12+), measures speed and accuracy

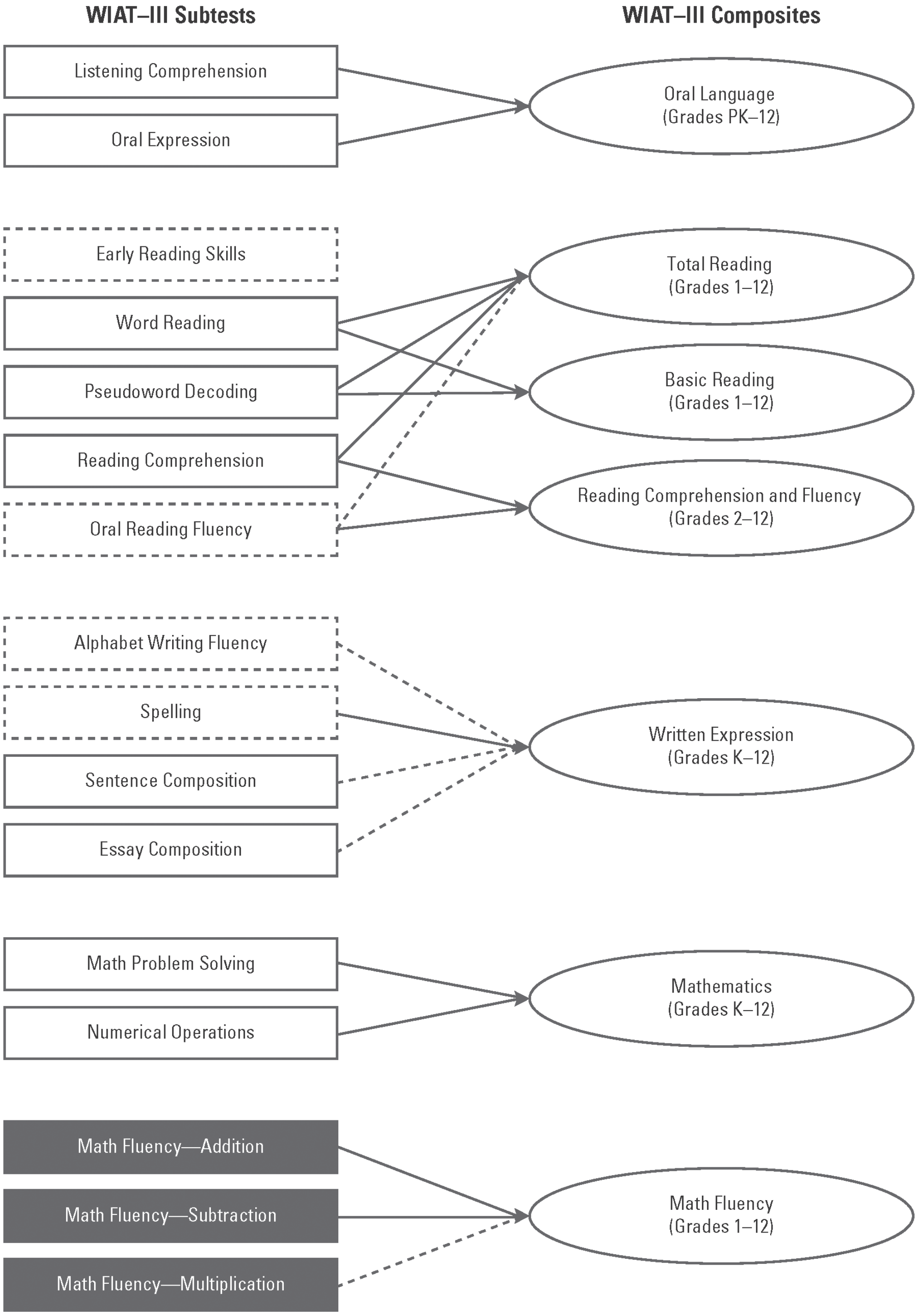


FIGURE 9.4. WIAT-III index scores and subtests. *Note.* A solid arrow indicates that the subtest contributes to the composite for all composite grade levels; a dashed arrow indicates that the subtest contributes to the composite for some, but not all, possible composite grade levels. The Early Reading Skills subtest *only* contributes to the Total Achievement composite. Boxes with dotted lines indicate that the subtest is included in the Total Achievement composite for some, but not all, possible grades. Subtests in shaded boxes do *not* contribute to the Total Achievement composite. Figures found in the Manual for the *Wechsler Individual Achievement Test®, Third Edition (WIAT®-III)*. Copyright © 2009 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

of oral single-word reading. The examiner records the item reached at 30 seconds to obtain a supplemental measure of reading speed; however, the examinee continues reading the list of words without a time limit until the discontinue rule is met or the last item is read.

Essay Composition, administered to examinees ages 8–50 (grades 3–12+), measures spontaneous, compositional writing skills. The examinee is given 10 minutes to respond to one essay prompt.

Pseudoword Decoding, administered to examinees ages 6–50 (grades 1–12+), requires the examinee to read aloud from a list of pseudowords. As in the Word Reading subtest, the examiner records the item reached at 30 seconds to obtain a supplemental measure of reading speed; however, the examinee continues reading the list of words without a time limit until the discontinue rule is met or the last item is read.

Numerical Operations, administered to examinees ages 5–50 (grades K–12+), measures untimed, written math calculation skills in the areas of basic skills, basic operations with integers, geometry, algebra, and calculus.

Oral Expression, administered to examinees ages 4–50 (grades PK–12+), contains three components: Expressive Vocabulary, Oral Word Fluency, and Sentence Repetition. Expressive Vocabulary measures speaking vocabulary and word retrieval ability by requiring the examinee to say the word that best corresponds to a given picture and definition. Oral Word Fluency measures how quickly and easily the examinee can name things belonging to a given category within 60 seconds. Sentence Repetition measures oral syntactic knowledge and short-term memory by asking the examinee to repeat sentences verbatim.

Oral Reading Fluency, administered to examinees ages 6–50 (grades 1–12+), measures oral reading speed, accuracy, fluency, and prosody by requiring the examinee to read passages aloud. Fluency is calculated as the average number of words read correctly per minute. A qualitative scale is used to assess reading prosody.

Spelling, administered to examinees ages 5–50 (grades K–12+), requires the examinee to spell (write) letter sounds or words, depending on grade level. The examinee hears each letter sound within the context of a word, and each word within the context of a sentence.

Math Fluency—Addition, administered to examinees ages 6–50 (grades 1–12+), requires the examinee to solve written addition problems within a 60-second time limit. Math Fluency—Subtraction, administered to examinees ages 6–50 (grades 1–12+), requires the examinee to solve written subtraction problems within a 60-second time limit. Math Fluency—Multiplication, administered to examinees ages 8–50 (grades 3–12+), requires the examinee to solve written multiplication problems within a 60-second time limit.

PSYCHOMETRIC PROPERTIES OF THE WIAT-III

The normative sample of the WIAT-III meets or exceeds current practice recommendations, and the psychometric properties are generally strong (see Dumont & Willis, 2010).

Normative Samples

The WIAT-III was standardized on nationally stratified samples of 2,775 students in the grade-norm sample (grades PK–12), 1,826 students in the age-norm sample (ages 4:0–19:11), and 225 individuals in the adult age-norm sample (ages 20:0–50:11 years). The stratification of the normative samples matches recent U.S. census data closely on the following key demographic variables: grade/age, sex, race/ethnicity, parental educational level, and geographic region. Approximately 8% of the school-age normative samples included individuals with diagnosed clinical disorders, and approximately 2% were identified as academically gifted.

Reliability

The reliability coefficients for the school-age and adult samples are generally consistent. The average internal-consistency (split-half) reliability coefficients for the composite scores and for Math Problem Solving, Word Reading, Pseudoword Decoding, Numerical Operations, Spelling, Math Fluency—Subtraction, and Math Fluency—Multiplication (for the adult sample only) are in the .90s. The average split-half reliability coefficients for Listening Comprehension, Early Reading Skills (school-age sample only), Reading Comprehension, Sentence Composition, Essay Composition, Essay Composition: Grammar and Mechanics, and Oral Expression are predominantly in the .80s and .90s.

The WIAT-III composites and subtests also demonstrate strong test–retest stability coefficients. For the school-age sample, the average corrected stability coefficients are excellent (.87–.96) for the composite scores; excellent (.90–.94) for Reading Comprehension, Word Reading, Pseudoword Decoding, Oral Reading Fluency, Oral Reading Rate, and Spelling; good (.82–.89) for Early Reading Skills, Math Problem Solving, Essay Composition, Essay Composition: Grammar and Mechanics, Numerical Operations, Oral Expression, Oral Reading Accuracy, Math Fluency—Addition, Math Fluency—Subtraction, and Math Fluency—Multiplication; and adequate (.75, .79) for Listening Comprehension and Sentence Composition. Alphabet Writing Fluency is a speeded subtest with a restricted raw score range, so a lower average stability coefficient (.69) is expected. For the adult sample, the average corrected stability coefficients are excellent (.90–.97) for the composite scores; excellent (.90–.97) for Math Problem Solving, Word Reading, Pseudoword Decoding, Numerical Operations, Oral Expression, Oral Reading Fluency, Oral Reading Rate, Spelling, Math Fluency—Subtraction, and Math Fluency—Multiplication; good (.81–.87) for Listening Comprehension, Reading Comprehension, Sentence Composition, Essay Composition: Grammar and Mechanics, and Math Fluency—Addition; and adequate (.78, .74, respectively) for Essay Composition and Oral Reading Accuracy.

Validity

The validity of the WIAT-III has been demonstrated by using intercorrelation data, correlations with other measures, and clinical studies. The subtests that make up each composite are generally moderately correlated with one another and show expected relationships (e.g., strong correlations between Math Problem Solving and Numerical Operations, and between Word Reading and Pseudoword Decoding) and discriminant evidence of validity (e.g., the mathematics subtests correlate more highly with each other than with other subtests). Correlations among the composite scores range from .45 to .93, with stronger correlations among the reading composites, and weaker correlations between the Math Fluency composite and other composites. Construct validity was established by correlating the WIAT-III with other tests. Correlations with the WIAT-II indicate that the two tests measure similar constructs. The corrected correlations between the composite scores for the two tests ranged from .76 (Oral Language) to .93 (Total Achievement), and correlations between the common subtests ranged from .62 (Oral Expression) to .86 (Spelling). Consistent with expectations, the corrected correlations were high for subtests that are highly similar in content and structure, and relatively low for subtests in which content and structure were modified considerably for the WIAT-III.

Correlations between the WIAT-III and the Wechsler intelligence scales (WPPSI-IV, WISC-V, WAIS-IV, and WNV), and between the WIAT-III and the DAS-II, are consistent with expectations regarding typical correlations between cognitive ability and achievement measures. The correlations between the WIAT-III Total Achievement score and the overall cognitive ability scores for these measures range from .60 to .82. These correlations provide divergent evidence of validity, suggesting that the WIAT-III and the cognitive ability tests are measuring different constructs with varying degrees of overlap in the cognitive skills required.

To establish the validity and clinical utility of the WIAT-III, clinical studies were conducted with students identified as academically gifted and talented (GT), students with mild intellectual disability (ID), students with expressive language disorder (ELD), and students with specific learning disabilities in the areas of reading (SLD-R), written expression (SLD-W), and mathematics (SLD-M). Results showed expected patterns of performance in each study. Students identified as GT scored consistently higher than a matched control group across all subtests and composites except Early Reading Skills and Alphabet Writing Fluency. Students with mild ID scored significantly lower than a matched control group across all subtests and composites. Results from the ELD study confirmed that the oral language subtests and several other subtests and composites requiring expressive language and related skills reliably differentiated between students with ELD and their age-matched peers. Students with SLD-R (approximately 10% of whom were diagnosed with comorbid SLD-W) performed significantly lower than a matched control group on all reading-related subtests and composites, in addition to the Total Achievement composite, the Spelling subtest, and the Written Expression composites. Students with SLD-W (approximately 18% of whom were diagnosed with comorbid SLD-R) performed significantly lower than a matched control group on all writing-related subtests and composites (except the Alphabet Writing Fluency subtest) in addition to the Total Achievement composite and some reading-related subtests and composites. Students with LD-M performed significantly lower than a matched control group on all math-related subtests and composites. These results provide evidence that the WIAT-III reliably differentiates between students in these special groups and their age-matched peers.

ADMINISTRATION OPTIONS

The WISC-V, WPPSI-IV, and WIAT-III can be administered in paper-and-pencil format with traditional test kits, or digitally via the Q-interactive™ system. Q-interactive allows tests to be administered by using two tablets connected with Bluetooth. One tablet is used by the examiner to read instructions, record and score examinee responses, and send visual stimuli to the examinee's tablet. The examinee's tablet allows the test taker to view and respond to visual stimuli. The Q-interactive system differs from the well-known form of computer-administered testing found in education and the workplace, whereby the examinee self-guides the on-screen administration of items and responds by using a mouse or keyboard. Instead, it is specifically intended to enhance the types of testing in which the examinee interacts with a trained examiner who presents the test items, records and scores responses, and follows up as necessary. This facilitates the clinician's ability to establish rapport, observe qualitative aspects of performance such as problem solving style, and ensure that the examinee is demonstrating his or her best performance. An in depth review of how to use Q-interactive to administer the WISC-V, WPPSI-IV, and WIAT-III is outside the scope of this chapter, but refer to Wahlstrom, Daniel, Weiss, and Prifitera (2016) and Cayton, Wahlstrom, and Daniel (2012) for more detailed descriptions of the system. In addition, for an in-depth review of how to use Q-interactive to administer the WIAT-III, refer to Witholt, Breaux, and Lichtenberger (2016).

To date, all tests on Q-interactive are scored by using the original normative data collected via traditional paper-based administration methods. In place of renorming the Q-interactive versions of tests, studies have been conducted to establish the raw score equivalence of scores generated through paper- and digital-based administrations. The study methodologies vary, depending on test characteristics, but in all cases equivalence is defined a priori as a paper–digital format effect size less than or equal to 0.20, which is a little more than 0.5 scaled score point or 3 standard score points (Daniel, 2012a).

The goals of the Q-interactive equivalence program have shifted over time. Originally, all subtests placed on Q-interactive underwent an equivalence study. This included subtests of the WAIS-IV (Daniel, 2012a), the WISC-IV (Daniel, 2012b), and the Wechsler Memory Scale—Fourth Edition (WMS-IV) (Daniel, 2013a), among others. As these studies were completed, a collective knowledge of how test designs were affected by digital adaptation was acquired, and from then on the team only tested new designs that could potentially introduce an equivalence threat. For example, with the WIAT-III, the raw score equivalence of Oral Reading Fluency was assessed because it was the first time an examinee read paragraphs on the iPad screen. However, this was not done for Oral Word Fluency because it was identical to Delis–Kaplan Executive Function System (D-KEFS) Verbal Fluency—which was found equivalent in previous studies—in terms of how the items were administered and scored (Daniel, 2013b). All WISC-V subtests underwent an equivalence study in order to replicate the WISC-IV findings, with the exception of Coding, Symbol Search, and Cancellation (which at the time were completed in paper response booklets and thus were very unlikely to reveal format effects). In addition, WPPSI-IV subtests that required use of the client tablet were assessed for format effects because it was unknown whether the findings for school-age children could be applied to preschoolers on these types of tasks.

Overall, the equivalence study data indicate that there are few, if any, differences between subtests administered in paper versus digital formats. The WISC-IV study revealed that all subtests had a format effect size of 0.20 or lower (the selected threshold), with the exception of Matrix Reasoning and Picture Concepts, which had effect sizes of 0.27 and 0.21, respectively (Daniel, 2012b). When the WISC-IV study was replicated for the WISC-V, the effect sizes for Matrix Reasoning and Picture Concepts met the equivalence threshold (0.17 and 0.02, respectively), along with all of the other subtests assessed (Daniel, Wahlstrom, & Zhang, 2014). For the WPPSI-IV, it was found that the correspondence between the paper and digital formats of the Picture Memory subtest did not support equivalence (Drozdick, Getz, Raiford, & Zhang, 2016). As a result, the subtest is still administered by using a paper stimulus book, and the examiner uses the practitioner's device to record and score the examinee's responses. All other subtests met the equivalence threshold. On the WIAT-III, only Word Reading, Pseudoword Decoding, Sentence Repetition, and Oral Reading Fluency were assessed because they are the only subtests with unique response demands relative to subtests that had been previously found equivalent. The study did not reveal any effect of format (Daniel, 2013b).

These findings are not surprising, as equivalence is considered during all phases of subtest development on Q-interactive and is a major driver of design decisions on the system. For example, one question that clinicians often ask is this: Why, when a child selects the correct answer on WISC-V Matrix Reasoning or Visual Puzzles, does the selection flash rather than stay active on the screen? The reason is that some subtests, such as Visual Puzzles, require the examinee to select multiple answers that go together. In these subtests, there is a working memory component, whereby the examinee must hold stimuli in mind while considering whether they fit together. If responses could be selected and remain highlighted, would that reduce the working memory requirements and change slightly the difficulty of the task? The answer is unknown, but decisions were made to keep the experience in digital format as similar as possible to that in the paper format, in order to avoid potential equivalence threats such as this. For the same reason, physical manipulatives such as the Block Design blocks or WPPSI-IV Object Assembly puzzle pieces are used with Q-interactive, despite the fact that they could be redesigned as completely digital experiences.

There is one notable exception to the equivalence approach, which is the digital adaptation of WISC-V Coding and Symbol Search. These have been developed as completely digital subtests, in which the examinee responds by tapping the screen rather than writing in a response booklet. Figures 9.5 and 9.6 depict samples of the tasks. For Coding, the examinee views a box that contains a missing symbol, and uses a key at the top of the screen to select the correct symbol from among five responses at the bottom of the screen. For Symbol Search, the task's demands are similar to those in the paper format, but the child taps the correct response rather than crossing it out with a pencil. Because the changes to task demands were more drastic than those in other subtests adapted for Q-interactive, there was no assumption that the digital Coding and Symbol Search would be equivalent to the paper versions. Rather, the goal was to equate them to the paper norms, since the paper and digital versions are assumed to be measuring the same core constructs, as evidenced by a high correlation between the different formats. The first attempt at digital adaptation of Coding required the examinee to use a stylus to write the correct response in the box, which more closely represents the paper response processes. However, pilot data suggested that the correlations with the paper subtest were insufficiently low, and after consultation with child design experts the task was redesigned in its current form. The equating data collected on the revised digital formats of Coding and Symbol Search suggest impressive similarities to their paper counterparts, as the raw score correlations with the paper tasks are .88 for Coding and .85 for Symbol Search (Raiford et al., 2016). In addition, confirmatory factor analyses indicated that the goodness-of-fit statistics were similar, regardless of whether the digital or paper formats of Coding and Symbol Search were used in the model. Moreover, special group studies indi-

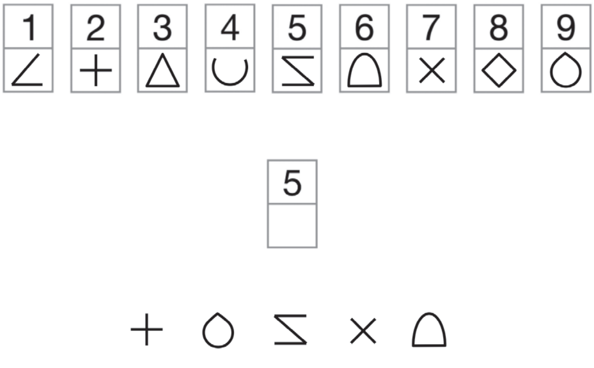


FIGURE 9.5. Example of an item from WISC-V Coding on Q-interactive.

FIGURE 9.6. Example of an item from WISC-V Symbol Search on Q-interactive.

cated similar patterns of performance across paper and digital subtests in individuals with intellectual disability, intellectual giftedness, specific learning disorders in math and in reading, ADHD, autism spectrum disorder with accompanying language impairment, and motor impairment. The finding that children with motor impairment performed similarly across paper and digital formats of Coding is especially important, given the reduction of graphomotor requirements in the digital format of the task. These results converge to indicate that the cognitive processes measured by traditional and digital formats of Coding and Symbol Search are very similar, and that administering the tasks in either format produces highly consistent results.

What are some considerations when practitioners are deciding whether to administer digital formats of the WISC-V, WPPSI-IV, or WIAT-III? One is how the testing experience will change for the test taker. Although the results of the equivalence studies suggest that raw scores do not change when a test is administered using a tablet device, many practitioners report that students are more engaged when interacting with technology (Daniel, 2013c). For the most part, this engagement is positive, as it suggests that scores are more likely to be valid because they aren't being negatively affected by boredom or fatigue. However, it may lead to new types of distractions for children who are inattentive and would rather do other things on the tablet than take a test. As always, the clinical skills used to manage behavior in traditional testing formats are important to keep an examinee motivated and on task.

Another consideration is that the test-taking experience with Q-interactive also changes significantly for the practitioner, and thus it is important to have appropriate training in the technology before using it for clinical cases. Basic familiarity with mobile technology is a must in order to feel comfortable during the testing session and provide the examinee with full attention. Finally, technology such as Q-interactive is accompanied by major changes in how test data are accessed, secured, and stored. Before using a system like Q-interactive, practitioners should understand their data privacy responsibilities as defined by the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH), as well as how the technology they use helps them comply with these requirements.

INTERPRETATION

The WPPSI-IV, WISC-V, and WIAT-III technical manuals provide basic instruction regarding their interpretation (Pearson, 2009; Wechsler, 2012a, 2014a). Leading clinicians and researchers have developed more in-depth and alternative interpretive strategies for these tests, all of which represent valid and useful approaches to test interpretation (Breaux & Lichtenberger, 2016; Flanagan & Kaufman, 2009; A. S. Kaufman et al., 2016; Raiford & Coalson, 2014; Sattler et al., 2016; Weiss et al., 2016). A detailed guide to interpretation is outside the scope of this chapter, and we refer the reader to these other outstanding volumes, as well as the tests' technical and interpretive manuals, for further information. Instead, we provide basic information illustrating how the WPPSI-IV, WISC-V, and WIAT-III can be utilized by practitioners. In addition, we include guidance for how these measures can be integrated within the context of learning disability evaluations. When reading this section, the reader is encouraged to keep in mind that psychological assessment is a complex problem-solving process (Prifitera, Weiss, & Saklofske, 1998), and test results should always be interpreted in conjunction with a thorough history and careful clinical observations of the individual.

Common Wechsler Scores

The WPPSI-IV and the WISC-V utilize two types of age-adjusted scores: scaled scores and standard scores (for all composite scores and for the WISC-V complementary subtests and process scores). For subtest scores, an examinee's raw scores are compared to others within his or her age group. The WIAT-III provides both age-based and grade-based subtest and composite standard scores, which are computed by comparing an examinee's raw scores to others within his or her age or grade group. These scores permit comparison of an individual's performance to that of others in his or her peer group; because they are based on a similar metric, they also allow for comparisons within and across Wechsler scales, as well as between Wechsler tests and other commonly used assessment batteries.

For those without formal training in psychometrics (examinees, parents, educators, etc.), scaled and standard scores are often difficult to understand. Thus the Wechsler scales provide a number of other metrics to help communicate test results, which include percentile ranks and age equivalents (and grade equivalents for the WIAT-III); both of these should be provided when appropriate on the WPPSI-IV, WISC-V, and WIAT-III. A *percentile rank* is the percentage of individuals within the examinee's age range obtaining scores below or equivalent to that of the examinee, whereas an *age equivalent* is the age of individuals who, on average, obtain the examinee's raw score. Composite scores always should be accompanied by a *confidence interval,* which is based on the scale's reliability and expresses the likelihood that an examinee's true score (i.e., the score he or she would obtain if there was no error in the test) falls within a specified range around the actual score that he or she obtained. The confidence interval is important because it communicates the error that is inherent in all psychological tests.

Levels of Performance

Wechsler standard scores can also be communicated by describing a range of performance within which each score falls, which correspond to the relationship of the examinee's score to that of his or her normative group. These qualitative descriptors are particularly useful when describing test results to individuals unfamiliar with quantitative test data. The traditional labels used to describe the different qualitative categories on Wechsler tests were revised with the WISC-V. The reasons are that (1) some of the traditional labels are confusing (e.g., *borderline*) or value-laden (e.g., *superior*), and (2) there is little consistency across various tests used by clinicians (Breaux & Lichtenberger, 2016; Raiford & Coalson, 2014).

Table 9.2 presents the qualitative descriptors used for the WISC-V, WPPSI-IV, and WIAT-III, which highlight the recent shift in terminology. These terms are not empirically based, which means that alternative labels can be used, depending on the reporting needs for an examinee. For example, as seen in Table 9.2, the WIAT-III descriptive categories are based on a 15-point scale, whereas the WISC-V and WPPSI-IV categories are based on a 10-point scale. Recently, an option to use the 10-point scale has been added to the WIAT-III scoring software on Q-global™, Pearson's scoring and reporting platform, and to Q-interactive.

The 15-point scale allows for finer differentiation at the high and low ends of the distribution, whereas the 10-point scale allows for finer differentiation in the middle of the distribution. When reporting on WIAT-III scores that were obtained as part of a broader evaluation including the WISC-V or WPPSI-IV, it may be helpful to use the 10-point scale for consistency, unless the examinee's achievement scores are at the tails of the distribution and would thus benefit from the greater differentiation yielded by the 15-point scale. Regardless of the scale used, it is recommended that *extremely high* and *very high* be used in place of *very superior* and *superior,* respectively.

Basic Interpretation of the Intelligence Scales (WISC-V and WPPSI-IV)

Reporting and Describing Index Scores

The FSIQ and index scores are much more reliable measures than the subtest scaled scores. In general, they are the first scores examined by practitioners.

The FSIQ is often used as the primary level of interpretation. Some controversy exists in the literature regarding the interpretability and clinical utility of the FSIQ in children, especially in clinical samples characterized by significant variation

TABLE 9.2. Qualitative Descriptions for the WISC-IV, WPPSI-IV, and WIAT-III

10-point scale (WISC-V and WPPSI-IV)			15-point scale (WIAT-III)	
Score range	WISC-V descriptive classification	WPPSI-IV descriptive classification	Score range	WIAT-III descriptive classification
130 and above	Extremely high	Very superior	146 and above	Very superior
120–129	Very high	Superior	131–146	Superior
110–119	High average	High average	116–130	Above average
90–109	Average	Average	85–115	Average
80–89	Low average	Low average	70–84	Below average
70–79	Very low	Very low	55–69	Low
69 and below	Extremely low	Extremely low	40–54	Very low

in index scores (e.g., see the 2007 special issue of *Applied Neuropsychology*, Volume 14, Issue 1). However, the FSIQ is the best single-point predictor of school achievement, work performance, and many other important life outcomes known to modern science (Gottfredson, 2008), and evidence suggests that FSIQ remains valid and predictive even in cases characterized by extreme index score discrepancies (Daniel, 2007; Watkins, Lei, & Canivez, 2007).

However, some previous interpretive approaches posit that the FSIQ is valid and interpretable only if no 23-point or greater discrepancy exists between the highest and lowest primary index scores. There is no evidence that there is a discrepancy or index score scatter (or subtest score scatter) beyond which the FSIQ becomes invalid and uninterpretable (see Flanagan & Alfonso, 2017, for a discussion). When great variability or discrepancy characterizes the primary index scores or the subtest scaled scores, the FSIQ alone is insufficient to describe a child's intellectual abilities. However, reliance on any single score is never recommended for describing a child's intellectual abilities or for identifying his or her strengths and needs. Also, not all subtests that contribute to the primary index scores are used to create the FSIQ; the FSIQ is derived from a subset of those primary subtests on the WISC-V and the WPPSI-IV.

Research indicates that the construct and predictive validity of the FSIQ are not affected by an index score discrepancy (Daniel, 2007). It is quite typical to have a discrepancy of greater than 23 points between two primary index scores: For the WISC-V, 56.6% of the normative sample (i.e., 1,246 of the 2,200 children) had such a discrepancy, and 52.5% of the special group study cases (261 of the 497 children from these groups) had such a discrepancy (Kaufman et al., 2016). Similar frequencies were found among the WPPSI-IV normative and special group study samples (Raiford & Coalson, 2014). Given the vast evidence in support of the predictive validity of *g* and the FSIQ (Daniel, 2007; Deary & Johnson, 2010; Deary, Strand, Smith, & Fernandes, 2007; Johnson, Deary, & Iacono, 2009; S. B. Kaufman et al., 2012), it is counterintuitive to assume that for almost 60% of children, the FSIQ is not valid due to such discrepancies. Identifying index score discrepancies can reveal rich clinical information about a child, which is why comprehensive assessment of these diverse cognitive domains is important. Furthermore, more specific domains of intellectual ability do not show the same broad degree of predictive validity as does *g* (Hartmann, Larsen, & Nyborg, 2009; Kotz, Watkins, & McDermott, 2008; Reeve & Charles, 2008). The FSIQ consistently provides essential, clinically rich information when practitioners are attempting to understand the expression of intelligent behavior in real-world settings (Jacobson, Delis, Hamilton, Bondi, & Salmon, 2004).

While interpretation should include consideration of the FSIQ, it should be augmented by examination of index-level results. This is important because index-level strengths and weaknesses can provide important insight into a child's functioning that cannot be obtained from the FSIQ alone. For example, many clinical disorders are characterized by unique patterns of cognitive strengths and weaknesses, though it should be noted that there is enough overlap of profile patterns across disorders to suggest caution against the use of profile patterns for diagnostic purposes in the absence of additional information.

Assessing Index-Level Strengths and Weaknesses

As just noted, analyzing index-level strengths and weaknesses is important because they often yield important clinical information and augment the interpretation of the FSIQ. A significant difference between a primary index score and an overall performance indicator (i.e., the mean of the primary index scores [MIS] or the FSIQ) suggests a strength or weakness in that specific cognitive area. Therefore, the next step is to compare the primary index scores to the overall performance indicator, which reduces the error rate resulting from multiple comparisons (e.g., doing pairwise comparisons between all primary index scores). This approach to interpretation is advocated by multiple researchers and within various interpretive approaches (Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Longman, 2005; Naglieri & Paolitto, 2005; Raiford & Coalson, 2014) and represents a significant advancement in the interpretive approach advocated for these measures. The preferred overall performance indicator is the MIS because it is based on a larger sample of behavior and is not subject to scaling, but is truly a mean indicator. However, there may be occasions where the MIS is unavailable (e.g., invalid subtests, choosing not to assess all five cognitive domains, choosing to administer only the FSIQ subtests and derive only the FSIQ, VCI, and FRI), so the FSIQ can be used in its place.

Determining whether the difference between a primary index score and the overall performance indicator is statistically significant is the first step in this level of interpretation; however, not all statistically significant differences are clinically meaningful. Some degree of cognitive variability is typical. Thus the significance and frequency of a discrepancy have different implications for test interpretation (Flanagan & Alfonso, 2017; A. S. Kaufman et al., 2016; Matarazzo & Herman, 1985; Payne & Jones, 1957; Raiford & Coalson, 2014; Silverstein, 1981; Weiss et al., 2016). The Wechsler scales provide base rate data for discrepancies to assist practitioners in determining the rarity of significant differences. Because the values differ by ability level, the base rate data are also provided by five ability levels. The base rates are provided by the direction of the discrepancy because the frequencies of score differences vary with the direction of the difference (Sattler et al., 2016). For instance, the percentage of the WISC-V normative sample with VCI scores greater than the MIS is not the same percentage of the normative sample with an MIS greater than the VCI. Clinically speaking, the direction of discrepancy is related to different patterns of cognitive strengths and weaknesses that may be unique to particular clinical disorders. As discussed, some statistically significant discrepancies are commonly found among individuals in the normative sample, whereas others are relatively rare.

As with the index-level strengths and weaknesses comparisons, the WISC-V provides critical value and base rate information relevant to index-level pairwise comparisons. Pairwise comparisons are best used to evaluate specific hypotheses in interpretation. Each primary index score is a summary of performance on two subtests. It is therefore important to evaluate the variability between the two subtest scaled scores that are used to derive each of the primary index scores being compared, to ensure clear interpretation of discrepancies between the index scores.

Examining Subtest-Level Variability

Composite scores are estimates of overall functioning in their respective areas and should always be evaluated within the context of the subtests that contribute to them (Flanagan & Kaufman, 2009; A. S. Kaufman et al., 2016; Raiford & Coalson, 2014). It is good practice to evaluate the level of score consistency or variability among subtest scaled scores before interpreting composite scores. In cases characterized by extreme subtest score variability, consideration of that variability augments interpretation of the composite scores and may yield a more complete picture of the child's performance. Subtest-level variability is best examined to inform composite score interpretation using (1) subtest-level strengths and weaknesses, and (2) pairwise subtest comparisons that are relevant to interpretation of index scores derived from only two subtests.

It is a good practice to conduct an analysis of subtest strengths and weaknesses when the subtest scaled scores appear variable. Such an analysis is usually hypothesis-driven and focused on certain cognitive domains associated with the referral question (Flanagan & Alfonso, 2017; A. S. Kaufman et al., 2016; Raiford & Coalson, 2014; Weiss et al., 2016). As at the index level, the subtest-level strengths-and-weaknesses analysis determines whether or not the difference between each primary subtest scaled score and the overall indicator of subtest performance is statistically significant and unusual. As with index-level strengths and weaknesses, there are two indicators of overall subtest performance: the mean of the primary subtest scaled scores (MSS-P) and the mean of the FSIQ subtest scaled scores (MSS-F).

For interpretation of the primary index scores, each of which consists of only two subtests, pairwise subtest-level comparisons are informative. For example, comparing subtest-level performance on Digit Span and Picture Span can provide information relevant to interpretation of the WMI, as well as potentially relevant to auditory and visual working memory performance. Such a comparison may also suggest that further subtest administration would be helpful. For example, if the difference between Digit Span and Picture Span is significant, administering Letter–Number Sequencing yields an Auditory Working Memory Index, and administering Spatial Span from the WISC-V Integrated provides a Visual Working Memory Index. (See Raiford, Chapter 11 this volume, for a complete overview of the use and interpretation of the WISC-V Integrated, which enhances interpretation by extending the information available about WISC-V performance.) As with other types of comparisons, the WISC-V provides critical value and base rate information relevant to subtest-level pairwise comparisons.

Interpretation of the WIAT-III

The interpretation of the WIAT-III follows the same basic outline as that of the WPPSI-IV and WISC-V. The practitioner should begin by interpreting the composite scores to determine if the scores are consistent with the Total Achievement composite and with the other composite scores. Next, he or she should interpret the individual subtest standard scores for more detailed information about the examinee's performance. Next, the practitioner should identify academic strengths and weaknesses by evaluating the pattern of composite and subtest scores. As with the Wechsler intelligence tests, both the statistical significance and base rate of score discrepancies should be considered to determine if differences are clinically meaningful.

Growth scale values (GSVs) and subtest skills analysis are two aspects of interpretation included in the WIAT-III but not the Wechsler intelligence scales. GSVs are provided at the subtest level and are intended for comparing performance across test sessions and measuring change over time. Unlike standard scores and percentile ranks, GSVs do not compare performance to a normative sample, but rather describe performance in absolute terms. Comparing GSVs over time indicates whether an examinee's skill level has changed relative to his or her own previous skill level, and this growth can be compared to the typical growth rate in the normative sample. The key advantage of GSVs is that they are sensitive to small changes over time; however, the practitioner must be careful not to overinterpret progress, as GSVs only represent growth in the limited range of academic skills measured by a given subtest. It is often useful to interpret GSVs within the context of the examinee's standard scores. For example, increased GSVs associated with increased standard scores indicate that the examinee is progressing at a faster rate than his or her peers are. Increased GSVs with consistent standard scores indicate that the examinee is developing additional academic skills but at a rate consistent with peers, whereas increased GSVs with lower standard scores indicate that academic skills are increasing but at a slower rate than peers' skills.

Subtest skills analysis data provide more in-depth information about the examinee's strengths and weaknesses, and assist in the formulation of intervention plans. There are two types of skills analysis within the WIAT-III: item-level and within-item-level. For item-level skills analysis, an item is assigned to a particular skill category; if that item is scored as incorrect, an error is marked for that particular category. Within-item-level skills analysis is used when a particular item may reflect multiple error categories, depending on the type of incorrect response that the examinee provides. In this case, the examiner is required to interpret the error and assign it to the relevant error category. Skills analysis is particularly useful in identifying nuanced skill weaknesses for students who exhibit average or below-average performance on a subtest. Importantly, an examinee's pattern of errors can be used to generate goal statements for intervention planning.

Integration of Cognitive Ability and Achievement Results for Learning Disability Evaluations

IDEA 2004 increases flexibility in determining the method to be used for learning disability identification. Analyses to support two of the more commonly used methods in the field are available for use with the WIAT-III and the WISC-V or WPPSI-IV: the ability–achievement discrepancy

(AAD) analysis and the pattern of strengths and weaknesses (PSW) discrepancy analysis. Both of these analyses require the integration of standardized achievement and ability measures, and the statistical information supporting them can be found in the WISC-V and WPPSI-IV technical and interpretive manuals and scoring software.

It should be noted at the outset that significant discrepancies as described by either the AAD or PSW approach are insufficient to diagnose a learning disorder, but they are often used as administrative criteria to determine eligibility for special education services in public school systems. Although they are important empirical sources of information, they must be interpreted as part of a comprehensive evaluation that includes a variety of other information (e.g., educational records, medical history, family history, social and emotional functioning).

AAD Analysis

An AAD is characterized by performance on an academic achievement test that is significantly lower than what would be expected, given an examinee's performance on a cognitive ability measure. While discrepancies of this sort sometimes indicate the presence of a learning disability, many other factors may cause academic achievement to be discordant with cognitive ability. For example, children who struggle with extreme inattention or anxiety may be at a disadvantage relative to peers with respect to knowledge acquisition. Similarly, bright and motivated children may compensate for their learning difficulties through the use of executive compensatory strategies, hard work, or external support. In other words, although the presence of an AAD indicates the child is not achieving at his or her potential, the reason for underachievement requires further investigation and cannot be inferred from the presence of a discrepancy alone. For this reason, AAD analyses should be accompanied by other sources of information (Berninger & O'Donnell, 2005; Shinn, 2007; Siegel, 2003) in the process of identifying a learning disability.

There are two methods for calculating AAD with the WIAT-III: the simple-difference method and the predicted-difference method. (Formulas to calculate simple-difference critical values and base rates, as well as those for the predicted-achievement method, can be found in Rust & Golombok, 1999.) The predicted-achievement method is typically preferred because it accounts for regression to the mean, and the correlations between the two tests used in the analysis. WIAT-III standard scores can be predicted from a number of cognitive batteries, including the WISC-V and WPPSI-IV. The FSIQ and a number of other index scores may be used to predict achievement scores.

PSW Analysis

IDEA 2004 allows clinicians to consider a child's "pattern of strengths and weaknesses" in determining eligibility for learning disability services. A PSW analysis may be conducted utilizing scores from the WIAT-III and the WISC-V or WPPSI-IV, based on the concordance–discordance model developed by Hale and Fiorello (2004). The PSW model is often preferred because it requires the identification of a relative processing weakness, which helps to clarify the factors underlying academic difficulties (e.g., a cognitive processing deficit as opposed to emotional factors or poor instruction). Doing so can facilitate differential diagnosis and treatment planning, which is often difficult, given that students with learning problems are a heterogeneous group (Hale et al., 2008).

PSW analyses can be completed digitally only, within Q-interactive (for digital administration) or the Q-global scoring system (for paper administration). Both systems conduct two score comparisons, each of which must be significantly different to meet the model's criteria for learning disability identification: (1) processing strength versus achievement weakness, and (2) processing strength versus processing weakness. In order to calculate the PSW analysis with the WIAT-III and either the WISC-V or WPPSI-IV, the following steps are necessary. First, the WIAT-III subtest or composite score that reflects the student's achievement weakness should be identified; the practitioner should check first to make sure that if a composite score is chosen, it is interpretable (i.e., subtest scores do not show significant scatter). Scores below 85 are typically chosen, though scores above 85 may be acceptable in some cases (e.g., gifted students). Second, the cognitive weakness is identified, once the practitioner has again made sure that the index is interpretable and theoretically related to the achievement weakness (see Hale et al., 2001; Hale, Fiorello, Bertin, & Sherman, 2003). Finally, a cognitive strength is identified that is interpretable and not related to the academic weakness identified in the first step. The manuals advise against the use of some index scores (e.g., WMI, PSI, AWMI, NSI) to identify cognitive strengths because they have lower *g* loadings and are therefore less consistent with the conceptualization of specific learning disability as representing unexpected underachievement in comparison to intellectual ability (Pearson, 2009; Wechsler, 2012a, 2014a).

Appendix 9.1 presents a case study that demonstrates how the WISC-V and WIAT-III can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

APPENDIX 9.1

Illustrative Case Study

Report date: 11/18/2017	Grade: 6
Examinee: Andy Ford	Ethnicity: Mixed
Age: 12 years, 6 months	Sex: Male
Date of birth: 5/13/05	Tests administered: WISC-V and WIAT-III (11/13/17)

REASON FOR REFERRAL

Andy is a 12-year-old male referred for evaluation at the recommendation of his pediatrician. He has a history of school underachievement, and the purpose of the current evaluation is to clarify the factors underlying his academic difficulties.

FAMILY HISTORY

Andy lives with his parents and two older sisters, with whom he has lived since birth. Andy comes from a culturally diverse family background; his father is European American, and his mother is Hispanic. His maternal grandmother is a first-generation Mexican immigrant. Andy's mother obtained her master's degree and is employed as a nurse. His father has a bachelor's degree and works in sales. No significant stressors were reported at home.

LANGUAGE DEVELOPMENT

Andy's dominant language is English, and his parents and siblings speak English at home. Andy's maternal grandmother, who lives nearby and sees Andy several times a month, speaks to him in Spanish. Andy's mother reported that Andy understands Spanish and speaks some Spanish to his grandmother, but prefers speaking in English in all settings. No history of speech or language delays was reported, although he spoke his first words and used short sentences slightly later than expected.

MEDICAL AND DEVELOPMENTAL HISTORY

Mrs. Ford reported that Andy was the product of a full-term gestation free of complications. With the exception of his language development, Andy met all major developmental milestones within normal limits. Medical history is significant for frequent ear infections at the age of 2 years, which were treated with the placement of pressure equalizer tubes. No other hospitalizations, surgeries, or chronic illnesses were reported. A recent physical exam revealed normal hearing and vision. Andy is not currently on medication.

EDUCATIONAL HISTORY

Andy is in the sixth grade, his first year in middle school. He attends a mainstream classroom and has never been evaluated for special education services. However, Mrs. Ford noted that he has generally struggled with reading and writing tasks, but formerly compensated for these difficulties by spending more time studying with his parents. As the coursework became more complex in middle school, Andy struggled to keep up in class and earned below-average grades in English language arts his first semester. He was assigned a small-group tutoring service and support from a classroom aide, neither of which was effective in improving his test scores.

Mrs. Ford reported that Andy is motivated to do well in school, but frustrated by his lack of success. His attendance record is excellent, and he has no reported disciplinary problems. Andy spends approximately 1 hour per night completing his homework. Mrs. Ford reported that Andy rarely forgets to bring assignments home or turn in completed work.

BEHAVIORAL OBSERVATIONS

Andy appeared alert and oriented. He seemed shy at the beginning of testing, but rapport was established quickly and maintained throughout testing. His speech was goal-directed and fluent, without errors of articulation, and his receptive language appeared adequate for testing purposes. Andy appeared cooperative and motivated on all tasks.

INTERPRETATION OF WISC-V RESULTS

Andy was administered all of the WISC-V subtests (see Tables 9.A.1 and 9.A.2 for his composite and subtest scores, respectively). Andy's general cognitive ability is within the average range of intellectual functioning, as measured by the Full Scale IQ (FSIQ). His overall thinking and reasoning abilities exceed those of approximately 27% of children his age (FSIQ = 91; 95% confidence interval = 86–97), indicating well-developed overall cognitive functioning. Another overall measure of cognitive ability, the Nonverbal Index (NVI), was examined due to Andy's lower performance on Verbal Comprehension subtests, because such a pattern may be related to academic achievement issues. The NVI was closer to the population average of 100 (NVI = 98, 45th percentile; 95% confidence interval = 92–104). A comparison of higher-order reasoning abilities and efficiency of cognitive processing indicated that these abilities are comparable (General Ability Index [GAI] = 91, 27th percentile; Cognitive Processing Index [CPI] = 94, 54th percentile).

Andy's verbal reasoning abilities, as measured by the Verbal Comprehension Index (VCI), are in the low average range and above those of approximately 14% of his peers (VCI = 84; 95% confidence interval = 78–93). The VCI is designed to measure verbal reasoning and concept formation. Andy performed comparably on the subtests contributing to the VCI, suggesting that these verbal cognitive abilities are similarly developed. An expanded measure of verbal reasoning, the Verbal (Expanded Crystallized) Index (VECI), measures broad verbal reasoning and includes knowledge

TABLE 9.A.1. Andy's WISC-V Composite Scores

Composite	Sum of scaled scores	Composite score	Percentile rank	95% confidence interval	Qualitative description
		Primary			
Verbal Comprehension Index (VCI)	14	84	14	78–93	Low average
Visual Spatial Index (VSI)	18	94	34	87–102	Average
Fluid Reasoning Index (FRI)	21	103	58	96–110	Average
Working Memory Index (WMI)	20	100	50	92–108	Average
Processing Speed Index (PSI)	17	92	30	84–102	Average
Full Scale IQ (FSIQ)	61	91	27	86–97	Average
		Ancillary			
Verbal (Expanded Crystallized) Index (VECI)	28	82	12	77–89	Low average
Expanded Fluid Index (EFI)	40	100	50	94–106	Average
Quantitative Reasoning Index (QRI)	23	109	73	102–115	Average
Auditory Working Memory Index (AWMI)	16	89	23	83–97	Low average
Nonverbal Index (NVI)	59	98	45	92–104	Average
General Ability Index (GAI)	43	91	27	86–97	Average
Cognitive Proficiency Index (CPI)	37	94	34	87–102	Average
		Complementary			
Naming Speed Index (NSI)	158	78	7	72–89	Very low
Symbol Translation Index (STI)	251	82	12	76–90	Low average
Storage and Retrieval Index (SRI)	160	76	5	71–84	Very low

TABLE 9.A.2. Andy's WISC-V Subtest Scores

Subtest name	Total raw score	Scaled score	Percentile rank	Age equivalent
	Verbal Comprehension			
Similarities[a]	24	7	16	9:6
Vocabulary[a]	23	7	16	9:2
Information[b]	15	6	9	8:10
Comprehension[b]	19	8	25	10:2
	Visual Spatial			
Block Design[a]	28	8	25	10:6
Visual Puzzles	18	10	50	13:6
	Fluid Reasoning			
Matrix Reasoning[a]	19	9	37	10:6
Figure Weights[a]	26	12	75	16:2
Picture Concepts[b]	13	8	25	9:2
Arithmetic[b]	23	11	63	13:6
	Working Memory			
Digit Span[a]	25	9	37	10:6
Picture Span	34	11	63	15:10
Letter–Number Sequencing[b]	14	7	16	8:2
	Processing Speed			
Coding[a]	48	9	37	11:2
Symbol Search	23	8	25	10:2
Cancellation[b]	70	10	50	12:2
	Naming Speed			
Naming Speed Literacy	63	70	2	<9:2
Naming Speed Quantity	26	88	21	10:2
	Symbol Translation			
Immediate Symbol Translation	55	82	12	6:2
Delayed Symbol Translation	40	84	14	7:2
Recognition Symbol Translation	24	85	16	7:2

[a] Subtests that contribute to the Full Scale IQ.

[b] Subtests that do not contribute to the primary index score but can be substituted into the Full Scale IQ for that index.

about words, the world at large, and social situations. His performance on the VECI was also in the low average range relative to other children his age.

Andy's low average performance on the VCI was significantly lower than the mean of his primary index scores (MIS). This indicates that his skills in the areas of nonverbal reasoning, sustained attention, and effortful mental control are better developed than are his verbal reasoning skills, and these discrepancies suggest a weakness in complex verbal information processing. Fewer than 15% of children with overall cognitive skills similar to Andy's have a higher MIS relative to the VCI, indicating a relatively rare discrepancy that may be of clinical import.

Andy's visual–spatial skills, as measured by the Visual Spatial Index (VSI), are in the average range and above those of approximately 34% of his peers (VSI = 94; 95% confidence interval = 87–102). The VSI is designed to measure the ability to evaluate visual details and understand visual–spatial relationships. He performed similarly on subtests contributing to the VSI, suggesting that his skills in this domain are developing at a comparable rate.

Andy's nonverbal reasoning abilities, as measured by the Fluid Reasoning Index (FRI), are in the average range and above those of approximately 58% of his peers (FRI = 103; 95% confidence interval = 96–110). The FRI is designed to measure the ability to identify conceptual relationships between visual objects and use reasoning skills to apply rules to them. An expanded measure of reasoning, the Expanded Fluid Index (EFI,) provides a broader measure of his ability to detect underlying conceptual relations, extract important information, and use reasoning to identify and apply rules. His performance on the EFI is also in the average range.

Andy's average performance on the FRI was significantly higher than the mean of his primary index scores (MIS), suggesting a strength in nonverbal reasoning relative to verbal reasoning, visual–spatial skills, sustained attention and effortful mental control, and processing speed. Andy also displayed variable performance on subtests contributing to the FRI, as his score on Figure Weights was higher than his mean score on the 10 primary index scores (scaled score = 12). Figure Weights measures quantitative and inductive reasoning, as well as mental flexibility and set shifting. His performance indicates a strength in this area that was demonstrated by less than 10% of the normative sample, suggesting a clinically important result.

Andy's abilities to sustain attention, concentrate, and exert mental control, as measured by the Working Memory Index (WMI), are in the average range. He performed better than approximately 50% of his same-age peers in this area (WMI = 100; 95% confidence interval = 92–108). These results suggest that relative to other children his age, Andy demonstrates the mental control and attentional skills necessary to facilitate learning and support more complex information processing. Andy performed comparably on the Working Memory subtests that contribute to the WMI. His Auditory Working Memory Index (AWMI), which provides a purer measure of auditory working memory, is in the low average range.

Andy's abilities to process simple or routine visual material without making errors, as measured by the Processing Speed Index (PSI), are in the average range compared to those of his peers. He performed better than approximately 30% of children his age in this area (PSI = 92; 95% confidence interval = 84–102). Learning often involves both routine and complex information processing, and Andy's performance on the PSI suggests that he demonstrates the basic processing abilities necessary to acquire new information.

Rapid automatic naming involves recognizing and recalling overlearned information, like letters and numbers or quantities, as efficiently as possible. Overall, Andy's rapid automatic naming speed is in the very low range relative to other children his age (Naming Speed Index [NSI] = 78, 7th percentile; 95% confidence interval = 72–89). He recognizes, recalls, and recites letters and numbers far more slowly than other young people his age. This skill is related to and predictive of reading ability. His ability to rapidly recognize and name quantities is more efficient and close to average. Rapid quantity recognition is related to math skills, because working with these basic building blocks of overlearned information is involved in solving math problems.

Andy's visual–verbal associative memory, or the ability to form new associations between symbols and meanings, is in the low average range relative to other young people his age (Symbol Translation Index [STI] = 82, 12th percentile; 95% confidence interval = 76–90). This type of memory is especially relevant to learning to read, write, and do math, because all of these skills involve learning systems of symbols that are assigned meanings. His performance suggests uniform abilities when he is recalling associations learned within the past few seconds or minutes, or when trying to recall them or recognize them half an hour later.

In summary, Andy demonstrates overall cognitive skills in the average range. His scores on nonverbal reasoning, sustained attention, mental control, and ability to process simple or routine information were average. His verbal reasoning skills were low average. He exhibited a relative strength in his nonverbal reasoning skills and a relative weakness in his verbal reasoning skills relative to visual–spatial skills, sustained attention and mental control, and processing speed. He has particular weaknesses in rapid automatized naming and visual–verbal associative memory, which are commonly observed in conjunction with specific learning disabilities in the areas of reading and written expression.

INTERPRETATION OF WIAT-III RESULTS

Given Andy's history of difficulties in reading and writing, he was administered subtests of word recognition and decoding, reading comprehension, and written expression. To evaluate a possible area of strength, a subtest of math problem solving was administered. Results were interpreted by using age-based norms.

Andy's WIAT-III results are presented in Table 9.A.3. Andy demonstrated overall reading skills in the below-average range, performing better than approximately 9% of his peers (Basic Reading = 80; 95% confidence interval = 76–84). His ability to read single words aloud was below average relative to his peers. He correctly read 33 words, which is better than approximately 6% of children his age (Word Reading = 77; 95% confidence interval = 72–82). His pseudoword decoding skills (i.e., the ability to sound out made up words) were below average (Pseudoword Decoding = 82; 95% confidence interval = 77–87). He was able to sound out 20 pseudowords correctly. An evaluation of the skills analysis data for Word Reading and Pseudoword Decoding revealed specific weaknesses in recognizing more advanced vowel types (e.g., vowel digraphs and diphthongs), consonant blends, and common suffixes. Finally, Andy's reading comprehension was average (Reading Comprehension = 100; 95% confidence interval = 88–112). His ability to read passages and provide oral responses to open-ended comprehension questions was better than 47% of his peers.

Andy was administered two subtests of written expression. His ability to spell single words read aloud by the examiner was at the low end of average and better than 23% of his peers (Spelling = 89; 95% confidence interval = 83–95). A skills analysis revealed particular difficulty in spelling irregular vowels and certain suffixes (e.g., *-ous, -ious*), suggesting lower skills in orthography relative to phonology. His written productivity, theme development, and text organization were also at the low end of average and better than approximately 19% of children his age (Essay Composition = 87; 95% confidence interval 77–97). His one-paragraph essay included a simple introduction, two transition words, and two clearly stated reasons; however, he did not include elaborations or a conclusion.

Finally, Andy's math problem-solving skills were average (Math Problem Solving = 108; 95% confidence interval = 100–116). His ability to complete math calculations quickly and accurately, as well as solve word problems, was better than 70% of his peers.

Overall, the results of Andy's academic testing reveal relative weaknesses in the areas of single-word recognition and decoding. Despite these weaknesses, Andy demonstrates average reading comprehension skills. Consistent with his average overall cognitive functioning and strength in fluid reasoning, he seems to utilize compensatory strategies and context cues effectively to circumvent his word recognition difficulties while reading in context. Andy's written expression skills were at the low end of average. His highest score was in the area of math problem solving, which represents a relative strength for Andy.

TABLE 9.A.3. Andy's WIAT-III Composite and Subtest Scores

Subtest/composite	Raw score	Standard score	95% confidence interval	Percentile rank	Grade equivalent	Age equivalent	Qualitative description
Reading Comprehension	33	100	88–112	50	6.7	12:4	
Math Problem Solving	56	108	100–116	70	10.4	15:0	
Essay Composition	[a]	87	77–97	19	4.0	9:0	
Spelling	27	89	83–95	23	5.2	10:4	
Basic Reading composite		80	76–84	9			Below average
Word Reading	33	77	72–82	6	2.9	8:4	
Pseudoword Decoding	20	82	77–87	12	2.5	7:8	

[a] This subtest has multiple raw scores.

ABILITY–ACHIEVEMENT COMPARISONS

Ability–Achievement Discrepancy

The predicted-difference method was utilized to assess for discrepancies between his predicted achievement (based on his WISC-V FSIQ) and actual achievement on the Basic Reading composite, as well as the Spelling and Essay Composition subtests. This analysis revealed that Andy's performance on the Basic Reading composite was significantly below what would be predicted from his FSIQ ($p < .01$). This level of discrepancy was evident in 10% of Andy's peers, a relatively rare occurrence. No discrepancies were revealed between Andy's cognitive functioning and written expression skills. This analysis suggests that relative to his overall cognitive functioning, Andy demonstrates significant underachievement in the area of basic reading skills, a necessary but insufficient condition for identifying a specific learning disability. To further evaluate the factors underlying his reading difficulties (e.g., a possible learning disability), an analysis of his pattern of strengths and weaknesses (PSW) was conducted.

PSW Analysis

Andy's score on the Basic Reading composite was used as the academic weakness in the PSW analysis. His NSI on the WISC-V was identified as his cognitive weakness, as naming speed is related to basic reading skills. Andy's FRI score was identified as his cognitive strength, as his skills in this area are average and significantly higher than the mean of his index scores, and fluid reasoning is not thought to be theoretically related to basic reading skills. This analysis revealed that Andy's academic and cognitive weaknesses were significantly lower than his cognitive strength ($p < .01$ for both the Basic Reading composite and the NSI compared to the FRI). Like the discrepancy analysis, the PSW results support the diagnosis of a specific learning disorder in reading with impairment in word recognition skills.

CONCLUSIONS AND RECOMMENDATIONS

Andy is a 12-year-old male with a history of difficulties in reading and spelling. The results of the current evaluation reveal overall cognitive skills in the average range, with his verbal comprehension, naming speed, and visual–verbal associative memory abilities underdeveloped relative to his fluid reasoning, visual–spatial, working memory, and processing speed abilities. He exhibits a relative strength in fluid reasoning. Andy's reading is below average and a relative weakness for him; his written expression skills are at the low end of average; and his math problem solving skills are average. Analyses reveal a pattern of strengths and weaknesses consistent with a specific learning disorder in the area of basic reading: Andy's academic weakness (Basic Reading) and cognitive weakness (NSI) scores were significantly lower than his cognitive strength (FRI) score. These results are consistent with parent and teacher reports. Based on these findings, Andy meets criteria for special education services for a specific learning disability in the area of reading with impairment in word recognition skills.

Andy's present level of reading achievement, as well as the interventions already implemented, suggest that he would be likely to benefit from individualized instruction. Reading instruction should integrate reading and spelling to reinforce what is taught, and should incorporate single-word reading of both real words and nonwords. He may benefit most from instruction that is explicit and systematic, offering repetition and opportunities to practice and reinforce skills in a variety of contexts to develop automaticity. Specifically, Andy needs work in the areas of phoneme–grapheme relationships, sight words, and structural analysis. Given Andy's strength in fluid reasoning and math, he may enjoy opportunities to track and chart his progress over time. Decisions regarding classroom accommodations and modifications will be made in collaboration with Andy's parents and teachers.

REFERENCES

Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with ASD. *Research in Autism Spectrum Disorders, 7,* 494–501.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77,* 1698–1716.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22*(1), 121–130.

Bergman-Nutley, S., & Klingberg, T. (2014). Effect of working memory training on working memory, arithmetic and following instructions. *Psychological Research, 78*(6), 869–877.

Berninger, V. W., & O'Donnell, L. (2005). Research-supported differential diagnosis of specific learning disabilities. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 189–233). Boston: Elsevier Academic Press.

Blalock, L. D., & McCabe, D. P. (2011). Proactive interference and practice effects in visuospatial working memory span task performance. *Memory, 19*(1), 83–91.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal level of technical adequacy. *Journal of Psychoeducational Assessment, 4,* 313–326.

Breaux, K. C., & Lichtenberger, E. O. (2016). *Essentials of KTEA-3 and WIAT-III assessment.* Hoboken, NJ: Wiley.

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools, 42*(4), 333–343.

Cayton, T. G., Wahlstrom, D., & Daniel, M. H. (2012). The initial digital adaptation of the WAIS-IV. In E. O. Lichtenberger & A. S. Kaufman, *Essentials of WAIS-IV assessment* (2nd ed., pp. 389–426). Hoboken, NJ: Wiley.

Chen, H. Y., Keith, T. Z., Chen, Y. H., & Chang, B. S. (2009). What does the WISC-IV measure?: Validation of the scoring and CHC-based interpretative approaches. *Journal of Research in Education Studies, 54*(3), 85–108.

Crews, K. J., & D'Amato, R. C. (2009). Subtyping children's reading disabilities using a comprehensive neuropsychological measure. *International Journal of Neuroscience, 119,* 1615–1639.

Dang, C., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2014). Why is working memory related to intelligence?: Different contributions from storage and processing. *Memory, 22*(4), 426–441.

Daniel, M. H. (2007). Scatter and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology, 14*(4), 291–295.

Daniel, M. H. (2012a). *Equivalence of Q-interactive-administered cognitive tasks: WAIS-IV* (Q-interactive Technical Report No. 1). Bloomington, MN: Pearson.

Daniel, M. H. (2012b). *Equivalence of Q-interactive administered cognitive tasks: WISC-IV* (Q-interactive Technical Report No. 2). Bloomington, MN: Pearson.

Daniel, M. H. (2013a). *Equivalence of Q-interactive and paper administration of WMS-IV cognitive tasks* (Q-interactive Technical Report No. 6). Bloomington, MN: Pearson.

Daniel, M. H. (2013b). *Equivalence of Q-interactive and paper scoring of academic tasks: Selected WIAT-III subtests* (Q-interactive Technical Report No. 5). Bloomington, MN: Pearson.

Daniel, M. H. (2013c). *User survey on Q-interactive examinee behavior.* Bloomington, MN: Pearson. Retrieved November 13, 2016, from *www.helloq.com/content/dam/ped/ani/us/helloq/media/User_Survey_on_Effect_of_Q-interactive_Examinee_Behavior_final.pdf.*

Daniel, M. H., Wahlstrom, D., & Zhang, O. (2014). *Equivalence of Q-interactive and paper administrations of cognitive tasks: WISC-V* (Q-interactive Technical Report No. 8). Bloomington, MN: Pearson.

De Clercq-Quaegebeur, M., Casalis, S., Lemaitre, M., Bourgois, B., Getto, M., & Vallée, L. (2010). Neuropsychological profile on the WISC-IV of French children with dyslexia. *Journal of Learning Disabilities, 43*(6), 563–574.

Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology, 39*(5), 1362–1369.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21.

Drozdick, L. W., Getz, K., Raiford, S. E., & Zhang, O. (2016). *WPPSI-IV: Equivalence of Q-interactive and paper formats* (Q-interactive Technical Report No. 14). Bloomington, MN: Pearson.

Dumont, R., & Willis, J. O. (2010). Strengths and weaknesses of the WIAT-III and KTEA-II. In E. O. Lichtenberger & K. C. Breaux, *Essentials of WIAT-III and KTEA-II assessment* (pp. 239–269). Hoboken, NJ: Wiley.

Dumont, R., & Willis, J. O. (2014). Strengths and weaknesses of the WPPSI-IV. In D. Coalson & S. E. Raiford, *Essentials of WPPSI-IV assessment* (pp. 189–213). Hoboken, NJ: Wiley.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Flanagan, D. P., & Alfonso, V. C. (1995). A critical review of the technical characteristics of new and recently revised intelligence test for preschool children. *Journal of Psychoeducational Assessment, 13,* 66–90.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment.* Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459–483). New York: Guilford Press.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155–171.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*(2), 200–233.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237–241.

Gang, M., & Siegel, L. S. (2002). Sound–symbol learning in children with dyslexia. *Journal of Learning Disabilities, 35*(2), 137–157.

Georgas, J., Van de Vijver, F., Weiss, L., & Saklofske, D. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, L. Weiss, F. Van de Vijver, & D. Saklofske (Eds.), *Cross-cultural analysis of the WISC-III: Cultural considerations in assessing intelligence* (pp. 277–313). San Diego, CA: Academic Press.

Ghaziuddin, M., & Mountain-Kimchi, K. (2004). Defining the intellectual profile of Asperger syndrome: Comparison with high-functioning autism. *Journal of Autism and Developmental Disorders, 34*(3), 279–284.

Giofrè, D., Stoppa, E., Ferioli, P., Pezzuti, L., & Cornoldi, C. (2016). Forward and backward digit span difficulties in children with specific learning disorder. *Journal of Clinical and Experimental Neuropsychology, 38*(4), 478–486.

Gottfredson, L. S. (2008). Of what value is intelligence? In A. Prifitera, D. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 545–564). San Diego, CA: Academic Press.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hale, J. B., Fiorello, C. A., Bertin, M., & Sherman, R. (2003). Predicting math achievement through neuropsychological interpretation of WISC-III variance components. *Journal of Psychoeducational Assessment, 21,* 358–380.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16*(1), 31–55.

Hale, J. B., Fiorello, C. A., Miller, J. A., Wenrich, K., Teodori, A., & Henzel, J. N. (2008). WISC-IV interpretation for specific learning disabilities identification and intervention: A cognitive hypothesis testing approach. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 109–171). San Diego, CA: Academic Press.

Harrison, A. G., DeLisle, M. M., & Parker, K. C. H. (2008). An investigation of the General Abilities Index in a group of diagnostically mixed patients. *Journal of Psychoeducational Assessment, 26*(3), 247–259.

Hartmann, P., Larsen, L., & Nyborg, H. (2009). Personality as a predictor of achievement. *Journal of Individual Differences, 30,* 65–74.

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence, 39,* 210–221.

Horowitz-Kraus, T. (2014). Pinpointing the deficit in executive functions in adolescents with dyslexia performing the Wisconsin Card Sorting Test: An ERP study. *Journal of Learning Disabilities, 47*(3), 208–223.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96,* 150–166.

Jacobson, L. A., Ryan, M., Martin, R. B., Ewen, J., Mostofsky, S. H., Denckla, M. B., et al. (2011). Working memory influences processing speed and reading fluency in ADHD. *Child Neuropsychology, 17*(3), 209–224.

Jacobson, M. W., Delis, D. C., Hamilton, J. M., Bondi, M. W., & Salmon, D. P. (2004). How do neuropsychologists define cognitive constructs?: Further thoughts on limitations of factor analysis used with normal or mixed clinical populations. *Journal of the International Neuropsychological Society, 10*(7), 1020–1021.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences USA, 105*(19), 6829–6833.

Johnson, W., Deary, I. J., & Iacono, W. G. (2009). Genetic and environmental transactions underlying educational attainment. *Intelligence, 37*(5), 466–478.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence.* Needham Heights, MA: Allyn & Bacon.

Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues.* Belmont, CA: Thomson Wadsworth.

Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical Psychology Review, 32,* 605–617.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V.* Hoboken, NJ: Wiley.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic *g* one and the same *g*?: An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123–138.

Keith, T. Z., Fine, J., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 35*(1), 108–127.

Konold, T. R. (1999). Evaluating discrepancy analysis with the WISC-III and WIAT. *Journal of Psychoeducational Assessment, 17,* 24–35.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II.* San Antonio, TX: Harcourt Assessment.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability Score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review, 37*(2), 261–278.

Lakin, J. M., & Lohman, D. F. (2011). The predictive accuracy of verbal, quantitative, and nonverbal reasoning tests: Consequences for talent identification and program diversity. *Journal for the Education of the Gifted, 34*(4), 595–623.

Lichtenberger, E. O., & Kaufman, A. S. (2003). *Essentials of WPPSI-III assessment.* Hoboken, NJ: Wiley.

Lipinski, J., Simmering, V. R., Johnson, J. S., & Spencer, J. P. (2010). The role of experience in location estimation: Target distributions shift location memory biases. *Cognition, 115*(1), 147–153.

Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL–reading relationship? *Journal of Experimental Child Psychology, 115,* 137–149.

Litt, R. A., & Nation, K. (2014). The nature and specificity of paired associate learning deficits in children with dyslexia. *Journal of Memory and Language, 71,* 71–88.

Longman, S. R. (2005). Ipsative comparisons of index scores for the Canadian WAIS-III. *Canadian Journal of Behavioural Science, 37*(2), 155–158.

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory and Cognition, 36*(1), 43–52.

Malstädt, N., Hasselhorn, M., & Lehmann, M. (2012). Free recall behaviour in children with and without spelling impairment: The impact of working memory subcapacities. *Dyslexia, 18*(4), 187–198.

Matarazzo, J. D., & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardization sample. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 899–932). New York: Wiley.

Mayes, S. D., & Calhoun, S. L. (2006). WISC-IV and WISC-III profiles in children with ADHD. *Journal of Attention Disorders,* 9(3), 486–493.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A., & Paolitto, A. W. (2005). Ipsative comparisons of WISC-IV index scores. *Applied Neuropsychology, 12*(4), 208–211.

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences, 48*(4), 379–384.

Niileksela, C. R., & Reynolds, M. R. (2014). Global, broad, or specific cognitive differences?: Using a MIMIC model to examine differences in CHC abilities in children with learning disabilities. *Journal of Learning Disabilities, 47*(3), 224–236.

Pauly, H., Linkersdörfer, J., Lindberg, S., Woerner, W., Hasselhorn, M., & Lonnemann, J. (2011). Domain-specific rapid automatized naming deficits in children at risk for learning disabilities. *Journal of Neurolinguistics, 24,* 602–610.

Payne, R. W., & Jones, H. G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13,* 115–121.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition.* San Antonio, TX: Author.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Pimperton, H., & Nation, K. (2014). Poor comprehenders in the classroom: Teacher ratings of behavior in children with poor reading comprehension and its relationship with individual differences in working memory. *Journal of Learning Disabilities, 47*(3), 199–207.

Pineda-Pardo, J. A., Martínez, K., Román, F. J., & Colom, R. (2016). Structural efficiency within a parieto-frontal network and cognitive differences. *Intelligence, 54,* 105–116.

Powell, D., Stainthorp, R., Stuart, M., Garwood, H., & Quinlan, P. (2007). An experimental comparison between rival theories of rapid automatized naming performance and its relationship to reading. *Journal of Experimental Child Psychology,* 98(1), 46–68.

Prifitera, A., Weiss, L. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 1–38). New York: Academic Press.

Psychological Corporation. (1992). *Wechsler Individual Achievement Test.* New York: Author.

Psychological Corporation. (2001). *Wechsler Individual Achievement Test—Second Edition.* San Antonio, TX: Author.

Rackley, C., Allen, D. N., Fuhrman, L. J., & Mayfield, J.

(2012). Generalizability of WISC-IV index and subtest score profiles in children with traumatic brain injury. *Child Neuropsychology, 18*(5), 512–519.

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV assessment*. Hoboken, NJ: Wiley.

Raiford, S. E., Drozdick, L., Zhang, O., & Zhou, X. (2015). *Expanded index scores* (WISC-V Technical Report No. 1). Bloomington, MN: Pearson.

Raiford, S. E., Rolfhus, E., Weiss, L. G., & Coalson, D. (2005). *General Ability Index* (WISC-IV Technical Report No. 4). San Antonio, TX: Psychological Corporation.

Raiford, S. E., Zhang, O., Drozdick, L. W., Getz, K., Wahlstrom, D., Gabel, A., et al. (2016). *WISC-V Coding and Symbol Search in digital format: Reliability, validity, special group studies, and interpretation* (Q-interactive Technical Report No. 12). Bloomington, MN: Pearson.

Reeve, C. L., & Charles, J. E. (2008). Survey of opinions on the primacy of *g* and social consequences of ability testing: A comparison of expert and non-expert views. *Intelligence, 36,* 681–688.

Reverte, I., Golay, P., Favez, N., Rossier, J., & Lecerf, T. (2015). Testing for multigroup invariance of the WISC-IV structure across France and Switzerland: Standard and CHC models. *Learning and Individual Differences, 40,* 127–133.

Rimm, S., Gilman, B. J., & Silverman, L. K. (2008). Non-traditional applications of traditional testing. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 175–202). Waco, TX: Prufrock Press.

Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science, 19*(6), 346–351.

Rowe, E. W., Miller, C., Ebenstein, L. A., & Thompson, D. F. (2012). Cognitive predictors of reading and math achievement among gifted referrals. *School Psychology Quarterly, 27*(3), 144–153.

Rust, J., & Golombok, S. (1999). *Modern psychometrics: The science of psychological assessment* (2nd ed.). New York: Routledge.

Saklofske, D. H., Weiss, L. G., Raiford, S. E., & Prifitera, A. (2006). Advanced clinical interpretation of WISC-IV index scores. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 139–179). Amsterdam: Elsevier.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Sattler, J. M., Dumont, R., & Coalson, D. L. (2016). *Assessment of children: WISC-V and WPPSI-IV.* San Diego, CA: Jerome M. Sattler.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*(4), 601–617.

Siegel, L. S. (2003). IQ-discrepancy definitions and the diagnosis of LD: Introduction to the special issue. *Journal of Learning Disabilities, 36,* 2–3.

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37*(2), 392–394.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Sweetland, J. D., Reina, J. M., & Tatti, A. F. (2006). WISC-III verbal/performance discrepancies among a sample of gifted children. *Gifted Child Quarterly, 50*(1), 7–10.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 137–151.

Thaler, N. S., Bello, D. T., & Etcoff, L. M. (2012). WISC-IV profiles are associated with differences in symptomatology and outcome in children with ADHD. *Journal of Attention Disorders, 17*(4), 291–301.

Tourva, A., Spanoudis, G., & Demetriou, A. (2016). Cognitive correlates of developing intelligence: The contribution of working memory, processing speed and attention. *Intelligence, 54,* 136–146.

Travers, B. G., Bigler, E. D., Tromp, D. P. M., Adluru, N., Froehlich, A. L., Ennis, C., et al. (2014). Longitudinal processing speed impairments in males with autism and the effects of white matter microstructure. *Neuropsychologia, 53,* 137–145.

Tulsky, D. S., Saklofsky, D. H., Chelune, G. J., Heaton, R. K., Ivnik, R. J., Bornstein, R., et al. (Eds.). (2003). *Clinical interpretation of the WAIS-III and WMS-III.* San Diego, CA: Academic Press.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology, 87*(3), 239–266.

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities, 34,* 2586–2597.

Wahlstrom, D., Daniel, M., Weiss, L. G., & Prifitera, A. (2016). Digital assessment with Q-interactive. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (pp. 347–372). Boston: Elsevier Academic Press.

Wasserman, J. D., & Maccubbin, E. M. (2002). *Wechsler at The Psychological Corporation: Chorus girls and taxi drivers.* Paper presented at the 110th annual convention of the American Psychological Association, Chicago, IL.

Watkins, M. W., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence, 35,* 59–68.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition.* San Antonio, TX: Pearson.

Wechsler, D. (2012a). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition.* Bloomington, MN: Pearson.

Wechsler, D. (2012b). *WPPSI-IV technical and interpretive manual.* Bloomington, MN: Pearson.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D. (2014b). *WISC-V technical and interpretive manual.* Bloomington, MN: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31*(2), 114–131.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013b). WISC-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31*(2), 94–113.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (Eds.). (2016). *WISC-V assessment and interpretation: Scientist-practitioner perspectives.* Boston: Elsevier Academic Press.

Willburger, E., Fussenegger, B., Moll, K., Wood, G., & Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learning and Individual Differences, 18,* 224–236.

Witholt, T., Breaux, K., & Lichtenberger, E. O. (2016). Q-interactive for KTEA-3 and WIAT-III. In K. Breaux & L. O. Lichtenberger (Eds.), *Essentials of KTEA-3 and WIAT-III assessment* (pp. 379–393). Hoboken, NJ: Wiley.

Zheng, X., Flynn, L. J., & Swanson, H. L. (2013). Experimental intervention studies on word problem solving and math disabilities: A selective analysis of the literature. *Learning Disability Quarterly, 36*(2), 97–111.

CHAPTER 10

The Woodcock–Johnson IV Tests of Early Cognitive and Academic Development

Fredrick A. Schrank
Barbara J. Wendling

The Woodcock–Johnson IV (WJ IV) Tests of Early Cognitive and Academic Development (ECAD; Schrank, McGrew, & Mather, 2015) is a special-purpose early childhood battery of cognitive and academic skills tests that is co-normed with, and conceptually related to, the Woodcock–Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014a). The WJ IV includes the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b), the Woodcock–Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b), and the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a). We discuss these three batteries in Chapter 14 of this volume.

The ECAD measures emergent cognitive abilities and academic skills. It is designed primarily for use with children from 30 months through 7 years of age, although the battery is also appropriate for use with children ages 8 through 9 years who have cognitive-developmental delays or intellectual disability. In contrast, the WJ IV COG and WJ IV ACH tests are designed primarily for use with individuals of school age through adulthood, although there is some overlap in age ranges of the WJ IV batteries; for example, many of the WJ IV OL tests can also be used with young children. The ECAD stimulus materials are designed to be engaging and attractive to young children, and administration of the entire battery can be completed in less than an hour by an examiner using only one test easel and only one test record. Some of the tests are unique to the ECAD, while others are adapted and alternate forms of tests included in other parts of the WJ IV that have been widely used at the prekindergarten (PreK) through grade 1 levels.

The ECAD was developed in response to a need for a dedicated WJ IV battery of cognitive and academic skills tests to measure both the emergence and development of the different broad Cattell–Horn–Carroll (CHC) abilities at the early childhood level. CHC theory is especially helpful for determining if a child's cognitive abilities and academic skills are developing as expected: Identification of specific cognitive or academic delays at a young age can suggest a need for early intervention. The ECAD combines an individualized, age-appropriate, norm-referenced assessment with a criterion-referenced (authentic) interpretive model to target early learning needs and to suggest related instructional practices or intensive interventions.

ORGANIZATION OF THE ECAD

Figure 10.1 presents the Selective Testing Table for the ECAD. Tests 1 through 7 are the cognitive tests that make up the General Intellectual Ability—Early Development (GIA-EDev) cluster;

		General Intellectual Ability–Early Development	Early Academic Skills	Expressive Language
ECAD 1	Memory for Names	■		
ECAD 2	Sound Blending	■		
ECAD 3	Picture Vocabulary	■		■
ECAD 4	Verbal Analogies	■		
ECAD 5	Visual Closure	■		
ECAD 6	Sentence Repetition	■		■
ECAD 7	Rapid Picture Naming	■		
ECAD 8	Letter-Word Identification		■	
ECAD 9	Number Sense		■	
ECAD 10	Writing		■	

■ Tests required to create the cluster listed.

FIGURE 10.1. Selective Testing Table (tests and clusters) for the Woodcock–Johnson IV Tests of Early Cognitive and Academic Development (ECAD). From *Woodcock–Johnson IV*™ (*WJ IV*™). Copyright © The Riverside Publishing Company. All rights reserved. Used by permission of the publisher.

Tests 8 through 10 are the early academic tests that make up the Early Academic Skills cluster. In addition, Test 3: Picture Vocabulary and Test 6: Sentence Repetition constitute the Expressive Language cluster.

Each of the seven cognitive tests represents one of seven broad CHC abilities: long-term storage and retrieval (Glr), auditory processing (Ga), comprehension–knowledge (Gc), fluid reasoning (Gf), visual processing (Gv), short-term working memory (Gwm), and cognitive processing speed (Gs). The broad CHC ability of reading–writing (Grw) is represented by one test of reading and one test of writing. The broad CHC ability of quantitative knowledge (Gq) is represented by one early math knowledge test.

HOW TO ADMINISTER THE ECAD TESTS

Administration and scoring of the ECAD tests require knowledge of the exact administration and scoring procedures, as well as an understanding of the importance of adhering to standardized procedures. The ECAD comprehensive manual (Wendling, Mather, LaForte, McGrew, & Schrank, 2015) provides extensive information on administration and scoring. Additionally, the test book contains an abbreviated set of instructions for administering and scoring items on each test; these brief instructions are found on the introductory page of the test (the first printed page after the tabbed page). Additional instructions may appear on the test pages, such as in boxes with special instructions.

For most purposes, examiners should begin testing with Test 1: Memory for Names. The format of this test provides the opportunity for a child to become engaged with attractive, storybook-like art, and the unusual, game-like names provide a verbal focus of attention. This test also provides a useful segue into the testing phase of an evaluation, as it requires only a pointing response from the child. Although tests can be administered in any order, most examiners administer the additional tests in sequential order, and discontinue testing at a point that provides the desired or required level of information. The cognitive tests (Tests 1–7) are administered to obtain an evaluation of any intraindividual differences among emergent cognitive abilities. An identified strength or weakness among the seven cognitive tests can provide insights into the relative development of different cognitive abilities.

Tests 8 through 10 are administered to determine a child's levels of early academic skills in reading, mathematics, and writing. An intra-achievement analysis can yield information about whether reading, math, or writing skills are better developed than the other academic areas. Many examiners routinely administer all 10 tests because of the depth of interpretive information obtained, including an overall comparison of early intellectual development to early academic skill acquisition.

Examiners will need to learn how to establish a basal and a ceiling level for several tests. Basal and ceiling criteria are included in the test book for each test requiring them. If an examinee fails to meet the basal criterion for any test, the examiner is directed to test backward until the examinee has met the basal criterion or until Item 1 has been administered. When stimuli are visible to the examinee, the examiner is required to test by complete pages. The complete page rule may impact the process of establishing a basal or a ceiling. For example, if the basal is not established on the first page administered, the examiner tests backward full page by full page until the basal is established or item 1 is administered. Conversely, if the ceiling appears to be reached on a page, the examiner must complete that page. If, in the process of completing the page, the examinee responds correctly to an item, testing must continue until the ceiling rule is met and the page is completed. Note that the basal rule is based on the lowest-numbered correct responses specified, and that the ceiling is based on the highest-numbered incorrect responses specified. For some tests, examinees begin with item 1 and test until they reach their ceiling level or time limit; these tests do not require a basal level. Test 1: Memory for Names has items arranged within sets; for this test, specific cutoff criteria are outlined in the test record, test book, and comprehensive manual.

Individual test items are scored during test administration. All tests use a 1 (correct) or 0 (incorrect) scoring rule for determining raw scores. Examiners will need to learn how to score items and calculate the raw score for each test. Generally, raw scores are determined by adding the number of correctly completed items to the number of unadministered test items below the basal. Scores for sample or practice items should not be included in calculating raw scores. The correct and incorrect keys in the test books are intended to be guides to demonstrate how certain responses are scored; not all possible or acceptable responses are included in the keys. Completion of the scoring procedure requires using the Woodcock–Johnson Online Scoring and Reporting System (Schrank & Dailey, 2014). Additionally, examiners may purchase and use the WJ IV Interpretation and Instructional Interventions Program (WIIIP; Schrank & Wendling, 2015b), an online option for generating narrative interpretive text and related interventions. The WIIIP simplifies the process of linking assessment to instruction by using ECAD test results to generate evidence-based interventions. These scoring and interpretive programs provide a number of different interpretive report options, profiles, and a full array of interpretive scores—including age- and grade-equivalent scores, relative proficiency indexes, and peer comparison scores (e.g., standard scores and percentile ranks).

Administration Time

As a general rule, experienced examiners require about 35 minutes to administer the seven cognitive tests and an additional 15 minutes to administer the three academic skills tests. Although there can be wide variation in the amount of time it takes to administer standardized tests to a young child, typically the entire ECAD can be administered in less than 1 hour, including the time needed for setting up the testing materials, establishing rapport, completing the identifying information on the test record, and calculating raw scores.

Testing Materials

Testing materials required for administration of the ECAD include the test book, test record, and response worksheet; the audio CD and appropriate audio equipment; pencils; and a stopwatch.

Summary of Key Administration and Scoring Points

Most tests use suggested starting points, basal and ceiling rules, and item response scores of 1 or 0. Table 10.1 summarizes the key administration and scoring points for each test.

SPECIAL ADMINISTRATION AND SCORING CONSIDERATIONS

Although the ECAD comprehensive manual and test book provide detailed rules for test administration, this section presents important reminders about tests that have special administration or scoring rules.

Test 1: Memory for Names

In Test 1: Memory for Names, when a child makes an error, the examiner must score the item 0 and follow the error correction procedure outlined in the test book. The final number correct is based on the number of correct responses the child obtained on the items administered before the cutoff criterion was met. At each potential cutoff point, the examiner tallies the number of cumulative correct responses to determine whether to continue or discontinue testing. In the score entry section of the test record, the examiner enters the total number correct and a letter corresponding to the set of items administered. To obtain estimated age and grade equivalents, the examiner locates the number of points in the appropriate column corresponding to the set of items administered.

Audio Tests

Two tests, Test 2: Sound Blending and Test 6: Sentence Repetition, require the use of an audio

TABLE 10.1. Summary of ECAD Test Administration and Scoring Rules

Test name	Item scoring rule	Sample items?	Basal/ceiling rules?[a]	Administration notes	Additional materials required for administration
Test 1: Memory for Names	1, 0	No	All begin with Introduction 1; ceiling determined by cutoffs	Give immediate corrective feedback when child makes an error. Scoring is based on items administered.	None
Test 2: Sound Blending	1, 0	Yes	All begin with item 1; ceiling is 6 highest incorrect		Audio
Test 3: Picture Vocabulary	1, 0	Yes	Yes; 6/6	Suggested starting points are available.	
Test 4: Verbal Analogies	1, 0	Yes	Yes; 6/6	Suggested starting points are available.	None
Test 5: Visual Closure	1, 0	Yes	All begin with item 1; ceiling is 6 highest incorrect		None
Test 6: Sentence Repetition	1, 0	Yes	Yes; 5/5	Suggested starting points are available.	Audio
Test 7: Rapid Picture Naming	1, 0	Yes	No; 2-minute time limit	All begin with sample items, then proceed to Item 1.	Stopwatch: Timed test
Test 8: Letter–Word Identification	1, 0	No	Yes; 5/5	Suggested starting points are available; word must be pronounced smoothly to receive credit.	None
Test 9: Number Sense	1, 0	No	Yes; 5/5	Suggested starting points are available.	None
Test 10: Writing	1, 0	No	Yes; 6/6	Suggested starting points are available.	Response worksheet

[a] May be modified by complete-page rule.

recording. The audio recording is used to ensure standardized presentation of these tasks. The audio equipment used must have a good speaker, be in good working order, and produce a faithful, clear reproduction of the test items. Using headphones is recommended, as they were used in the standardization. An examiner can wear a monaural earphone or wear only one headphone over one ear to monitor the audio recording while an examinee is also listening through his or her headphones. The examiner should not repeat or replay any item unless the directions permit it (e.g., on sample items). In some cases, a young child may need to be administered this test via live voice. If presenting the items orally, the examiner should say the words or sentences in an even voice, exactly as presented on the audio recording.

Timed Test

Test 7: Rapid Picture Naming has a 2-minute time limit. Administering this test requires using a stopwatch or a watch with a second hand. Because this is a test of cognitive processing speed, if a child uses a synonym or a word that is similar in meaning for the pictured item, the item should be scored as correct.

Response Worksheet

Administration of Test 10: Writing requires the examiner to use the response worksheet. Directions for using the response worksheet are provided in the test book.

Test Session Observations Checklist

To help document general testing observations, each test record includes a brief, seven-category behavior rating scale on the first page. The categories include levels of conversational proficiency, cooperation, level of activity, attention and concentration, self-confidence, care in responding, and response to difficult tasks. Each category has a range of possible responses, in order to help identify whether the behavior is typical or atypical for the age of the child being assessed. This checklist should be completed at the end of the testing session.

BASIC TECHNICAL INFORMATION

Following is a summary of reliability and validity information for the ECAD. For more complete information, readers should consult the ECAD's comprehensive manual (Wendling et al., 2015).

Median reliability coefficients for the ECAD tests are reported in Table 10.2. With the exception of the one timed test, coefficients of stability are reported as median internal-consistency (r_{11}) and median standard error of measurement (*SEM*) values. For the timed test, Test 7: Rapid Picture Naming, reliabilities were calculated using the split-half procedure. The median reliability for 7 of the 10 tests met or exceeded .80, with the other 3 exceeding .90. The 3 tests with median reliabilities below .80 demonstrated higher reliabilities at specific ages. For example, Test 3: Picture Vocabulary's median reliability is .74, but it has reliabilities of .94 at ages 2 through 4.

TABLE 10.2. ECAD Median Test Reliability Statistics

Test	Median r_{11}	Median *SEM* (SS)
Test 1: Memory for Names	.97	2.60
Test 2: Sound Blending	.84	4.74
Test 3: Picture Vocabulary	.74	7.79
Test 4: Verbal Analogies	.80	6.18
Test 5: Visual Closure	.77	7.19
Test 6: Sentence Repetition	.86	5.61
Test 7: Rapid Picture Naming	.86	5.81
Test 8: Letter–Word Identification	.96	2.60
Test 9: Number Sense	.79	6.87
Test 10: Writing	.91	4.50

Note. r_{11}, internal-consistency reliability. Split-half procedure used on Test 7.

Table 10.3 contains the median cluster reliability coefficients (r_{cc}) and median standard errors of measurement for obtained cluster standard scores, or *SEM* (SS), as calculated via Mosier's (1943) formula. Of the three ECAD clusters, the GIA-EDev cluster and the Early Academic Skills cluster exceed the commonly referenced professional standard of .90 for decision-making purposes. The exception is the Expressive Language cluster, with a median reliability of .89. It is of interest to note that the reliability of this cluster exceeds .90 at ages 2 through 4. Although the test reliabilities support use and interpretation of the test-level scores as single indicators of an ability, the ECAD clusters are preferred for most decision-making purposes because of their higher reliabilities.

Table 10.4 contains a summary of correlations for the ECAD clusters with other measures of cognitive abilities, including the Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III; Wechsler, 2002), the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV; Wechsler, 2012), and the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007). The correlations in Table 10.4 provide validity evidence for the ECAD GIA-EDev cluster as a measure of general intelligence and as an indicator of verbal abilities. There is also evidence, based on the high correlation (.91) with the DAS-II School Readiness cluster, that the ECAD

TABLE 10.3. ECAD Median Cluster Reliability Statistics

Cluster	Median r_{cc}	Median *SEM* (SS)
General Intellectual Ability—Early Development (GIA-EDev)	.95	3.35
Early Academic Skills	.96	2.60
Expressive Language	.89	4.97

TABLE 10.4. Correlations for ECAD Clusters and Other Measures of Cognitive Abilities

Other measure	GIA-EDev	Expressive Language	Early Academic Skills
Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III)			
Full Scale IQ	.75	.72	—
Verbal IQ	.82	.81	—
General Learning Quotient	.80	.81	—
Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV)			
Full Scale IQ	.78	.62	—
Verbal Comprehension Index	.77	.72	—
Visual Spatial Index	.77	.51	—
Differential Ability Scales—Second Edition (DAS-II)			
General Conceptual Ability	.87	.71	.90
Verbal Ability	.79	.73	.81
Nonverbal Reasoning Ability	.70	.56	.74
Spatial Ability	.69	.43	.79
School Readiness	.82	.71	.91
Working Memory	.79	.72	.84
Processing Speed	.68	.58	.66

Early Academic Skills cluster is a strong indicator of general school readiness.

Table 10.5 contains a summary of correlations for selected ECAD tests and clusters with other measures of oral language, including the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003), the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007), the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999), the Oral and Written Language Scales: Listening/Comprehension/Oral Expression (OWLS; Carrow-Woolfolk, 1995), and the WJ IV OL (Schrank, Mather, & McGrew, 2014b). The ECAD Expressive Language cluster shows moderate to strong correlations with all the CELF-4 composites and the total PPVT-4 score, and moderate correlations with the CASL Core Composite and the OWLS Oral composite. Correlations with the WJ IV OL clusters may be high because the tests are parallel forms. Overall, these results support the validity of the ECAD Expressive Language cluster and several of the tests as valid measures of general oral language abilities.

Table 10.6 contains a summary of correlations for selected ECAD tests and clusters with measures from the WJ IV ACH (Schrank, Mather, & McGrew, 2014a) for ages 6 through 8. Other age groups are reported in the ECAD comprehensive manual. Again, some of the correlations may be high because some tests are parallel forms. For example, Letter–Word Identification is in both tests and is in the ECAD's Early Academic Skills cluster, as well as the WJ IV ACH's Brief Achievement and Reading clusters. Although the item content is not identical, the tasks are essentially the same. The ECAD tests shown in Table 10.6 correlate more strongly with their respective domain-specific WJ IV ACH clusters. For example, Test 9: Number Sense correlates more highly with Mathematics than with any of the other WJ IV ACH clusters reported. In addition, Number Sense does not share any content with the WJ IV ACH Mathematics cluster. These correlations support the use of the ECAD tests and Early Academic Skills cluster as valid measures of early achievement in reading, math, and writing.

A summary of correlations between ECAD and other early childhood developmental measures are reported in Table 10.7. These measures include the Battelle Developmental Inventory, Second Edition (BDI-2; Newborg, 2005) and the Riverside Early Assessments of Learning IDEA Observational Version (REAL IDEA-OV; Bracken, 2013). The relationship between the ECAD cognitive tests (Tests 1–7) and two clusters (GIA-EDev and Expressive Language) and the BDI-2 Communication and Cognitive domains are reported. Not surprisingly, the GIA-EDev cluster and the BDI-2

TABLE 10.5. Correlations for Select ECAD Measures and Other Measures of Oral Language Abilities

Other measure	Expressive Language cluster	Test 2: Sound Blending	Test 3: Picture Vocabulary	Test 6: Sentence Repetition
Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4)				
Core Language	.82	.58	.70	.74
Receptive Language	.74	.40	.70	.61
Expressive Language	.81	.54	.69	.73
Language Content	.76	.47	.78	.55
Language Structure	.81	.53	.71	.71
Working Memory	.68	.50	.54	.64
Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)	.79	.39	.75	.65
Comprehensive Assessment of Spoken Language (CASL): Core Composite	.52	—	.58	.53
Oral and Written Language Scales: Listening/Comprehension/Oral Expression (OWLS)				
Oral Composite	.50	—	.57	.52
Woodcock–Johnson Tests of Oral Language (WJ IV OL)				
Broad Oral Language	.79	.41	.85	.55
Listening Comprehension	.70	.41	.64	.55
Phonetic Coding	.36	.80	.29	.32

Cognitive domain had the highest correlation (.79), but there was also a strong correlation between GIA-EDev and the BDI-2 Communication domain (.75). This finding provides evidence to support the use of the ECAD GIA-EDev cluster as a valid measure of cognitive ability in developmental assessments with young children. Although the REAL IDEA-OV is an observational assessment scored as examiner ratings, the correlations with the ECAD were moderate to strong. In particular, the highest correlations were between GIA-EDev and the Cognitive domain (.80), the Communication domain (.82), and the Academic domain (.83). The ECAD's Expressive Language cluster had the strongest correlations with those same three domains: Cognitive (.78), Communication (.80), and Academic (.80). The pattern of correlations in Table 10.7 suggests that the ECAD is a valid measure of early cognitive abilities, expressive language skills, and preacademic skills.

TABLE 10.6. Correlations for Select ECAD Measures and WJ IV ACH Measures, Ages 6–8

	ECAD clusters and tests			
WJ IV ACH clusters	Early Academic Skills cluster	Test 8: Letter–Word Identification	Test 9: Number Sense	Test 10: Writing
Brief Achievement	.98	.94	.69	.93
Reading	.93	.96	.60	.83
Written Language	.91	.87	.60	.89
Mathematics	.81	.70	.79	.71
Phoneme–Grapheme Knowledge	.82	.81	.57	.75
Academic Knowledge	.59	.51	.61	.50

The information presented in Tables 10.2 through 10.7 provides a summary of the evidence supporting the ECAD's use as a measure of cognitive abilities, expressive language, and early academic skills during early childhood. The ECAD is a reliable and valid measure for evaluating children's abilities (Carney, 2017; Wright, 2017) .

INTERPRETATION

The ECAD can be interpreted at the test and cluster levels, depending on the purposes of an evaluation. The broad and narrow abilities described in contemporary CHC theory provide the basic interpretive architecture and the validity of the tests and clusters. Interpretation is based on extensive psychometric data and analyses provided by McGrew, LaForte and Schrank (2014), as well as related research outside the WJ IV literature that provides support for our interpretation of the emergent cognitive processes and developmental skills required for performance on each test. In addition to CHC theory, an analysis of task demands helps provide links to interventions or accommodations that may be helpful in many situations for developing a targeted educational plan, based in part on the trajectory of a child's early cognitive and academic development.

The ECAD tests and clusters are described in this section. Table 10.8 identifies the broad and narrow cognitive abilities measured by each test in the ECAD; it also includes brief test descriptions. The ECAD tests are organized into clusters for interpretive purposes; these clusters are outlined in Table 10.9. The ECAD's methodology for determining developmental levels is described. The remainder of the chapter describes the variation and discrepancy procedures that can be used to compare scores, determine a developmental profile of abilities, and assist in diagnostic decision making.

General Intellectual Ability—Early Development

The GIA-EDev score is a special-purpose measure of overall cognitive development that is based on seven tests, each representing a broad CHC factor. When combined and differentially weighted, the overall score measures psychometric *g* or general intellectual ability. In addition, each of the seven tests measures the emergence and developmental progression of a specific broad cognitive ability. Tests were selected for inclusion in the GIA-EDev score to best measure each of the seven broad CHC factors in the most developmentally appropriate format. The seven tests represent the following CHC constructs: long-term storage and

TABLE 10.7. Correlations for ECAD Tests/Clusters and Other Measures of Early Childhood Development

	Battelle (BDI-2)		REAL IDEA-OV		
ECAD tests and clusters	Communication domain	Cognitive domain	Cognitive domain	Communication domain	Academic domain
Tests					
Test 1: Memory for Names	.40	.39	.44	.39	.48
Test 2: Sound Blending	.53	.61	.68	.68	.75
Test 3: Picture Vocabulary	.67	.64	.71	.77	.78
Test 4: Verbal Analogies	.58	.66	.67	.69	.69
Test 5: Visual Closure	.68	.71	.78	.76	.79
Test 6: Sentence Repetition	.70	.71	.74	.75	.74
Test 7: Rapid Picture Naming	.44	.47	.63	.70	.68
Test 8: Letter–Word Identification	—	—	.82	.82	.89
Test 9: Number Sense	—	—	.80	.82	.85
Test 10: Writing	—	—	.41	.44	.46
Clusters					
GIA-EDev	.75	.79	.80	.82	.83
Expressive Language	.74	.72	.78	.80	.80
Early Academic Skills	—	—	.39	.42	.43

TABLE 10.8. ECAD Tests, CHC Broad and Narrow Abilities Measured, and Brief Test Descriptions

Test name	Primary broad CHC ability *Narrow ability*	Brief test description
Test 1: Memory for Names	Long-term storage and retrieval (Glr) *Associative memory* (MA)	Measures auditory–visual paired associate encoding in the learning phase, visual identification in the response phase.
Test 2: Sound Blending	Auditory processing (Ga) *Phonetic coding* (PC)	Assesses the ability to synthesize speech sounds into a whole word.
Test 3: Picture Vocabulary	Comprehension–knowledge (Gc) *Language development* (LD)	Evaluates general vocabulary knowledge through object recognition.
Test 4: Verbal Analogies	Fluid reasoning (Gf) *General sequential reasoning* (RG) *Induction* (I) Comprehension–knowledge (Gc) *Language development* (LD) *General (verbal) information* (KO)	Measures the ability to reason by using relationships between words.
Test 5: Visual Closure	Visual processing (Gv) *Closure speed* (CZ)	Assesses the ability to identify an object from an incomplete or masked visual representation.
Test 6: Sentence Repetition	Short-term working memory (Gwm) *Memory span* (MS) Comprehension–knowledge (Gc) *Listening ability* (LS)	Evaluates ability to listen to and repeat sentences presented orally.
Test 7: Rapid Picture Naming	Cognitive processing speed (Gs) *Naming facility* (NA) *Speed of lexical access* (LA)	Measures the speed/fluency of retrieval and oral production of recognized objects.
Test 8: Letter–Word Identification	Reading–writing ability (Grw) *Reading decoding* (RD)	Assesses skill in identifying printed letters and words.
Test 9: Number Sense	Quantitative knowledge (Gq) *Mathematics achievement* (A3) Fluid reasoning (Gf) *Quantitative reasoning* (RQ)	Evaluates understanding of numbers, number vocabulary, concepts, and relationships required to compare, judge, estimate size, quantity, position, or volume.
Test 10: Writing	Reading–writing ability (Grw) *Spelling ability* (SG)	Measures prewriting skills and details of letter and word forms.

TABLE 10.9. ECAD Clusters and Brief Cluster Descriptions

Cluster	Brief cluster description
General Intellectual Ability—Early Development (GIA-EDev)	Measures psychometric *g* as distilled from a broad spectrum of emergent cognitive abilities and functions, including auditory–visual paired associate encoding; phonological coding; knowledge of object names; analogic verbal reasoning; ability to identify and name an object when presented with only limited visual cues; listening and oral expression abilities; speed and fluency of retrieval and oral production of object names.
Expressive Language	Assesses single-word and sentence-level oral language skills.
Early Academic Skills	Evaluates emergent reading, mathematics, and writing skills.

retrieval (Glr), auditory processing (Ga), comprehension–knowledge (Gc), fluid reasoning (Gf), visual processing (Gv), short-term working memory (Gwm), and cognitive processing speed (Gs).

In the ECAD, CHC theory provides a foundation for interpretation of test results and is especially useful for determining whether different cognitive abilities and academic skills are developing as expected. Related research (much of it from outside the WJ IV literature) provides support to interpretation of the CHC constructs that the GIA-EDev cluster measures. Finally, all tests selected for inclusion in the ECAD have relevance for educational programming and intensive intervention. Table 10.10 includes a summary of suggested educational enrichment goals or individually targeted interventions related to developmental delays identified on the ECAD tests. As noted earlier, the GIA-EDev cluster includes the following: Test 1: Memory for Names; Test 2: Sound Blending; Test 3: Picture Vocabulary; Test 4: Verbal Analogies; Test 5: Visual Closure; Test 6: Sentence Repetition; and Test 7: Rapid Picture Naming.

TABLE 10.10. Examples of Educational Enrichment Goals or Targeted Interventions Related to Limitations in Performance on ECAD Tests

Test	Brief intervention, strategy, or accommodation
Test 1: Memory for Names	Active learning environments; picture matching games with oral elaboration; verbal associations; overlearning; mnemonics.
Test 2: Sound Blending	Language-rich environments; increased exposure to words and word sounds; early exposure to sounds, music, and rhythms; reading aloud to a child; opportunities to explore and manipulate word sounds; nursery rhymes and songs with rhyming words; explicit, systematic instruction in phonics; practice with blending sounds into words.
Test 3: Picture Vocabulary	Language- and experience-rich environments; vocabulary-intensive curriculum; explicit teaching of key vocabulary words; frequent exposure and practice with words; reading aloud to a child; text talks; picture books.
Test 4: Verbal Analogies	Vocabulary-rich learning activities; planned activities that promote thinking; use of words that describe relationships; similarities and differences; reading aloud to a child; dialogic and interactive book-reading techniques; inclusion in adult conversations.
Test 5: Visual Closure	Visual discrimination games; use of words describing the visual–spatial characteristics of objects; building blocks, log and block building sets, puzzles; dot-to-dot drawings; pegboards; visual matching games.
Test 6: Sentence Repetition	Songs, games, nursery rhymes, and musical instruments to develp auditory memory span; modeling of correct language use; use of increasingly complex syntactic forms in everyday conversations; oral rehearsal of information to be remembered.
Test 7: Rapid Picture Naming	Picture-naming activities; repetitive practice; serial naming; repeated reading.
Test 8: Letter–Word Identification	Reading to a young child; oral production of sound representations of words; matching games with letter-like symbols; direct and fully guided instruction; explicit, systematic, synthetic phonics programs; integrated word phonology, orthography, and morphology approaches; increased exposure to printed words; word recognition strategies (word walls, flow lists); systematic literacy programs.
Test 9: Number Sense	Number awareness; early number concepts; number vocabulary; manipulatives and board games; use of number words; counting skills; adding and subtracting with number line; direct instruction; guided practice.
Test 10: Writing	Prewriting play activities; tracing letters and saying letter names; integrated word phonology, orthography, and morphology programs of instruction; frequent writing practice; teaching spelling of common irregular words; saying letter sounds when spelling; explicit instruction in six syllable types.

Test 1: Memory for Names

In CHC theory, Memory for Names measures the narrow ability of associative memory, an aspect of long-term storage and retrieval (Glr). This test measures a developmentally important way in which semantic memories (Tulving, 1972, 1985) are created: the association of words with pictured objects. The controlled memory consolidation format utilized in task performance represents an important link between working memory and long-term memory. This link is critically important for education; it is the gateway between information to be learned and learned information (Schrank, Decker, & Garruto, 2016).

Memory for Names measures the ability to associate unfamiliar auditory and visual stimuli, encode the association as a semantic memory, and subsequently reidentify the paired visual association upon presentation of a verbal cue. Colorful, storybook-like art enhances the attractiveness of this learning task for young children. The stimulus presentation phase of this task employs a directed, spotlight-attention paired-associate encoding procedure that prepares the child to encode a stimulus (Brefczynski & DeYoe, 1999; Gazzaniga, Irvy, & Mangun, 1998; Klingberg, 2009; Sengpiel & Hubener, 1999). The response phase of this task measures whether consolidation of semantic memory has occurred in the learning phase. The test's error correction procedure utilizes a controlled-learning format that corrects any faulty hypothesis and provides the opportunity for additional spotlight-attention encoding, from which the examiner may glean qualitative information about the level of repetition required for paired-associate consolidation of semantic (meaning-based) representations into long-term memory.

The controlled-learning procedure employed in Memory for Names provides a cue to the intervention known as *active learning* (Marzano, Pickering, & Pollock, 2001). Active learning is especially important in early education because it facilitates the creation of meaning-based codes that can be used to relate new information or task requirements to previously acquired knowledge. Varying the learning tasks, incorporating emotions and novelty, and fostering creativity are ways to promote active learning in early education. See Table 10.10.

Test 2: Sound Blending

Sound Blending measures the ability to blend sounds into words, an important prerequisite to reading competence (Schrank & Wendling, 2015a). Blending sounds to create words is a key aspect of phonological coding, a narrow auditory processing (Ga) ability. A young child may have low scores on this test for several reasons, such as poor phonological awareness, English as a second language, articulation difficulties, weak memory, or inadequate instruction. However, with appropriate instruction, blending skills can be developed.

For preschool children with limitations or delays in sound-blending ability, an environment rich in words and the sounds that make up words is particularly important for language development. Increased exposure to words may help develop a greater sensitivity to the sound patterns within words (Biemiller & Boote, 2006; Carlisle & Fleming, 2003; Dickinson & Tabors, 2001; McDowell, Lonigan, & Goldstein, 2007; Missall, Reschly, & Betts, 2007; Poe, Burchinal, & Roberts, 2004; Roth, Speece, & Cooper, 2002). For children in the early primary grades, interventions include explicit, systematic instruction in phonics (Ehri, 1998; Jenkins & O'Connor, 2002; National Reading Panel, 2000; Snow, Burns, & Griffin, 1998; Stanovich, 1994; Torgesen, 1997; Torgesen et al., 1999), including practice blending sounds into words (Adams, 1990). See Table 10.10.

Test 3: Picture Vocabulary

Picture Vocabulary measures vocabulary, verbal ability, and background knowledge, all aspects of the narrow ability of language development (LD) and the broad ability of comprehension–knowledge (Gc). Picture Vocabulary is primarily a single-word expressive vocabulary task that requires the cognitive processes of object recognition, lexical access, and lexical retrieval. A child may have difficulty with this test due to limited knowledge of word meanings, word retrieval difficulties, English as a second language, limited experiences and opportunities, or cultural differences (Mather & Wendling, 2015). Observing performance and analyzing errors may help determine if poor performance is a result of retrieval difficulties or limited vocabulary. When an error is related to the correct response, it may indicate a retrieval or word-finding problem. Although scored as incorrect, this type of error indicates some level of object knowledge. For example, an individual may not be able to provide the exact name of an object, but can describe an appropriate function or attribute. An error that is not directly associated with the correct response may indicate a weakness in vocabulary knowledge.

Early interventions for the development of vocabulary knowledge include creating a language- and experience-rich environment (Gunn, Simmons, & Kame'enui, 1995; Hart & Risley, 2003), explicit teaching of key vocabulary words (Beck & McKeown, 2001; Graves, Juel, & Graves, 2004; Nagy, 2005; National Reading Panel, 2000), frequent exposure and practice with words (Gunn et al., 1995; Hart & Risley, 2003), reading aloud to the child (Adams, 1990), and text talks (structured dialogue about a text that has been read) (Beck & McKeown, 2001). See Table 10.10.

Test 4: Verbal Analogies

Verbal Analogies measures the ability to reason by using relationships between words, an aspect of fluid reasoning (Gf). As an early development task, this test taps the emergence of reasoning ability as the child applies his or her knowledge of the relationship between the words in the first part of each item to induce and say a fourth word that completes the analogy (an aspect of general sequential reasoning [RG]). Each item requires induction (I) of the structure for the first part of each analogy and then mapping (or projecting) that structure onto the second part (Gentner & Markham, 1997). Forming verbal analogies involves determining general principles from specific examples and establishing a connection between previously unrelated pieces of information (Bunge, Mackey, & Whitaker, 2009). As a young child matures and cognition progresses, language development (LD) increases, and comprehension–knowledge (Gc; particularly general verbal knowledge [KO]) facilitates verbal analogy solutions. In reciprocal fashion, increased word, object, and concept knowledge facilitates increased analogical reasoning ability (Goswami, 1989).

Analogical thought is an important cognitive process through which learning occurs (Goswami, 1989). Children as young as 36 months are able to utilize verbal analogical reasoning if they are familiar with the words and objects included in the analogy (Goswami, 1989). Analogies are often used to teach and learn new words; concepts are developed by association (analogous relationships) with prior knowledge (Gentner, 1983). Vocabulary-rich learning activities can increase a child's level of word knowledge (Beck & McKeown, 2001; Graves et al., 2004; National Reading Panel, 2000), which is required for reasoning with words. In the course of everyday conversations with a child, parents, teachers, and caregivers should be encouraged to increase their use of words that describe abstract relationships, such as similarities and differences (Engle et al., 2011; Goldin-Meadow et al., 2014; Marulis & Neuman, 2010; Roberts & Kaiser, 2011). Parents or caregivers should be encouraged to include the child in adult conversations, using high-level vocabulary words (Nisbett, 2009).

Test 5: Visual Closure

Visual Closure measures the ability to identify an object from an incomplete or masked visual representation. Recognition occurs when the incomplete picture matches a representation in semantic memory. In CHC theory, this is called closure ability (CZ), a narrow aspect of visual processing (Gv)—specifically, the ability to identify an object by name when provided only with a limited portion of a pen-and-ink drawing of the object. The attention to detail that is required to identify objects from limited, but key, features is an important developmental task that precedes the ability to decode and assign meaning to printed text.

Young children who have difficulty with the Visual Closure test may benefit from interventions that are designed to develop the ability to discriminate visual features and/or match visual information with object names (Greenleaf & Wells-Papanek, 2005). Parents, caregivers, and teachers of a young child with limitations in closure ability should be encouraged to increase use of words that describe the visual–spatial features of objects as part of everyday conversations with the child, and to spend more time with the child doing hands-on activities and playing games that facilitate conversations about visual–spatial features of objects. Building blocks, log and block building sets, and puzzles are useful for this purpose (Engle et al., 2011; Goldin-Meadow et al., 2014; Marulis & Neuman, 2010; Roberts & Kaiser, 2011).

Test 6: Sentence Repetition

Sentence Repetition measures auditory memory span (MS), a narrow ability of short-term working memory (Gwm). An expressive language task that requires listening ability (LS), this test requires the ability to encode and maintain verbal information in primary memory and then accurately reproduce it in sequence. Test performance is aided by semantic knowledge (word meaning), an aspect of comprehension–knowledge (Gc). A child may obtain low scores on this test for several reasons, such as limited attention, poor memory, or limited oral language (Mather & Wendling, 2015). Evaluators should encourage teachers, parents, and caregivers to model correct language use, monitor the child's understanding of sentence structure, and use increasingly complex syntactic forms in everyday conversations with the child (Engle et al., 2011; Goldin-Meadow et al., 2014; Marulis & Neuman, 2010; Roberts & Kaiser, 2011).

Test 7: Rapid Picture Naming

Rapid Picture Naming measures the speed or fluency of recognition, retrieval, and oral production of names of pictured objects (naming facility; NA). In CHC theory, this test is primarily a measure of cognitive processing speed (Gs) at the preschool level (McGrew et al., 2014), but speed of lexical access (LA) or retrieval ability (Gr) in the school years.

Evaluators should compare performance on Rapid Picture Naming to performance on Test 3: Picture Vocabulary. If limited proficiency on this test is a function of lack of knowledge of object names (i.e., if a child also does poorly on Picture Vocabulary), then a foundational and primary instructional goal would be to increase use of object names in conversations with the child (Engle et al., 2011; Goldin-Meadow et al., 2014; Marulis & Neuman, 2010; Roberts & Kaiser, 2011). Pictures can be used to help stimulate production and fluency with names of objects (Nickels & Best, 1996). The most widely cited intervention for limited proficiency in speeded object naming is serial naming training, or game-like activities that encourage naming speed (Bowers, Golden, Kennedy, & Young, 1994; Bowers & Wolf, 1993; Conrad & Levy, 2007; Hayward, Das, & Janzen, 2007; Kirby, Georgiou, Martinussen, & Parrila, 2010; Logan, Schatschneider, & Wagner, 2009; Nelson, Benner, & Gonzalez, 2005; Nelson, Stage, Epstein, & Pierce, 2005; Sternberg & Wagner, 1982).

Early Academic Skills

The Early Academic Skills cluster is an aggregate measure of emergent or developing reading, math, and writing skills. The Early Academic Skills cluster can also be useful as a screening measure for any delay in the development of early academic skills. This cluster includes Test 8: Letter–Word Identification, Test 9: Number Sense; and Test 10: Writing.

Test 8: Letter–Word Identification

Letter–Word Identification measures reading readiness skills and the development of reading decoding ability (RD), a narrow aspect of reading–writing ability (Grw) in CHC theory. This test requires a child to identify or read isolated letters and words orally. The child does not need to know the meaning of the words. In Letter–Word Identification, recognized letters and words are accessed from the child's mental lexicon (i.e., his or her store of word knowledge) and recoded phonologically as the sounds the letters or letter combinations make.

Although there are many intervention approaches for limited reading readiness or reading ability, delays in development of letter and word identification skills can be addressed through a highly structured and intensive program of direct and fully guided instruction (Chambers, Cheung, Slavin, Smith, & Laurenzano, 2010; Englemann, 1968; Miller & Dyer, 1975; Salaway, 2008), including explicit, systematic, synthetic phonics programs (Jenkins & O'Connor, 2002; National Reading Panel, 2000; Stanovich, 1994; Torgesen, 1997). Integrated word phonology, orthography, and morphology approaches are frequently recommended (Carlisle & Stone, 2005; Moats, 2005; Venezky, 1970, 1999).

Test 9: Number Sense

Number Sense measures a child's understanding of numbers and their relationships to other numbers, as well as the vocabulary and concepts necessary to compare, judge, and estimate size, quantity, position, or volume. This test measures a critical developmental familiarity with numbers and how to think with numbers; it employs a broad sampling of number development skills, such as number recognition, counting, sequencing, and understanding of magnitude and quantity estimation. In CHC theory, these skills are part of the broad ability of quantitative knowledge (Gq) and the narrow ability of mathematics achievement (A3). The test also taps fluid reasoning (Gf), specifically quantitative reasoning (RQ).

Young children who perform poorly on Number Sense will benefit from well-designed activities to help develop an awareness of numbers, knowledge of the meaning of numbers, early number concepts, and related vocabulary (Baroody, Eiland, & Thompson, 2009; Chambers et al., 2010; Gelman & Gallistel, 1978; Gersten, Jordan, & Flojo, 2005; Griffin, 2002, 2004; Malofeeva, Day, Saco, Young, & Ciancio, 2004; National Research Council, 2009). Number sense skills can be developed with manipulatives and board games that require counting and which help develop a sense of numeric quantity, patterns, core mathematical concepts, and the relationships among numbers (Gersten et al., 2008; National Council of Teachers of Mathematics, 2000; Ramani & Siegler, 2005). Basic counting skills should be explicitly taught (Gersten et al., 2008). High-quality instruction, with guided practice, helps develop the ability to solve basic mathematics facts and executive mathematics procedures quickly and efficiently (Gersten et al., 2008).

Test 10: Writing

The Writing test measures prewriting skills and knowledge of the details of letter and word forms, and focuses on early pencil-and-paper writing skills. In CHC theory, this test assesses part of the broad reading–writing ability (Grw), with later items requiring spelling ability (SG). When a child progresses to writing words, the task involves mapping phonology onto orthographic representations of words. A careful analysis of errors can often result in specific instructional recommendations.

Remediation of difficulties in early writing skill development may require use of multisensory techniques, such as repeatedly tracing and saying letter names (Carreker, 2005). Multisensory techniques involving repeatedly tracing and saying the letters and words can also be used to build word-spelling skills, and may be especially helpful when irregular words are introduced (Fernald, 1943). Integrated word phonology, orthography, and morphology programs of instruction are often effective when the child is ready to learn to spell words (Carlisle & Stone, 2005; Moats, 2005; Templeton & Bear, 1992; Venezky, 1970, 1999).

Expressive Language

The Expressive Language cluster is a combined measure of single-word and sentence-level oral language skills. This cluster is intended as a screening measure for any delay in the development of expressive language abilities. This cluster includes Test 3: Picture Vocabulary and Test 6: Sentence Repetition.

SCORES, DEVELOPMENTAL LEVEL INDICATORS, AND SCORE COMPARISON PROCEDURES

The Woodcock–Johnson Online Scoring and Reporting Program (Schrank & Dailey, 2014, 2015) calculates all derived scores based on age norms. As in other WJ IV batteries, age equivalents, grade equivalents (if appropriate), *W* scores (a Rasch-based metric; Wendling et al., 2015), relative proficiency indexes (RPIs), standard scores, and percentile ranks may be elected for inclusion in a score report. The scores selected for inclusion in a report should be determined by the purposes of an evaluation. In addition, the ECAD includes descriptive terms to help determine levels of cognitive and academic development, including the presence and severity of any developmental delay.

Levels of Cognitive and Academic Development

The ECAD introduces a set of terms to describe a child's levels of cognitive and academic development. The cutoff scores for the terms are derived from the *W* difference (*W* Diff) values obtained by the child (defined as the difference between a child's *W* score and the median *W* score for the reference group in the norming sample; Wendling et al., 2015).

The ECAD's underlying scaling metric (the *W* scale) is a Rasch-based metric (Bond & Fox, 2007; Rasch, 1960/1980; Wright & Stone, 1979; Woodcock & Dahl, 1971) that provides the basis for a set of criterion-referenced labels tied directly to proficiency with the measured tasks (LaForte, McGrew, & Schrank, 2015). The ECAD levels of development can be compared and contrasted to other methods sometimes used for determining developmental delay: months delay, percentage delay, and standard deviation (*SD*) delay—each of which has measurement limitations. Because a calculation of months delay, percentage delay, or *SD* delay is not comparable at all ages during early childhood (in terms of ability difference required to determine a delay), the ECAD's levels can draw attention to developmental needs in some cases or ages when the other score metrics are insensitive to identification of a delay. The ECAD method may be particularly accurate for identifying the presence and severity of cognitive or academic delay during a critical window of time when early identification and targeted intervention can reduce or even eliminate a developmental delay (Baroody et al., 2009; Biemiller & Boote, 2006; Case & Griffin, 1990; Chambers et al., 2010; Dickinson & Tabors, 2001; Engle et al., 2011; Gersten et al., 2005; Goldin-Meadow et al., 2014; Griffin, 2002; Marulis & Neuman, 2010; Miller & Dyer, 1975; National Research Council, 2009; Poe et al., 2004; Roberts & Kaiser, 2011; Roth et al., 2002; Salaway, 2008; Wasik, Bond, & Hindman, 2006; Whitehurst et al., 1994).

Table 10.11 illustrates the six basic terms for levels of development, as well as two important blended-category terms (*age-appropriate to advanced* and *mildly delayed to age-appropriate*), which draw attention to task performance at or near a critical change in interpretation. The RPIs that are reported on the table of scores are also based on *W* Diff values and can be used to provide meaningful interpretations regarding a child's proficiency with cognitive and academic tasks.

Score Comparison Procedures

Scores from the ECAD can be compared with intracognitive and intra-achievement variations, as well as with an ability–achievement comparison procedure. These procedures can help determine the presence and severity of any intraindividual strengths and weaknesses or performance discrepancies.

Intracognitive Variation Procedure

The intracognitive variation allows examiners to analyze a child's cognitive test scores to explore relative strengths and weaknesses among developing cognitive abilities. For example, a child may have a strength in learning ability, as shown in Test 1: Memory for Names, but a weakness in background knowledge, as shown in Test 3: Picture Vocabulary. Any variations within cognitive development can help determine a child's present educational needs. Examiners can calculate intracognitive variations if Tests 1 through 7 have been administered.

Intra-Achievement Variation Procedure

The intra-achievement variation allows examiners to analyze a child's academic test scores and explore relative strengths and weaknesses among developing academic skills. For example, a child may have a strength in number awareness, but a weakness in early reading skills. Any variations within achievement can help determine a child's present educational needs. Examiners can calculate intra-achievement variations if Tests 8 through 10 have been administered.

Ability–Achievement Comparison Procedure

Because the GIA-EDev score is a broad-based measure of intellectual ability, it serves as a good predictor of academic development. This procedure allows an examiner to determine if a child's overall achievement level is commensurate with a child's overall cognitive ability. The GIA-EDev score is used to calculate predicted scores for expected achievement. The child's predicted achievement is then compared to his or her actual achievement. Children with predicted scores significantly higher than their actual achievement scores exhibit an

TABLE 10.11. ECAD Developmental Levels, *W* Diff Values, and Reported RPIs

Level of development	*W* Diff values	Reported RPIs
Very advanced	+31 and above	100/90
Advanced	+14 to +30	98/90 to 100/90
Age-appropriate to advanced	+7 to +13	95/90 to 98/90
Age-appropriate	–6 to +6	82/90 to 95/90
Mildly delayed to age-appropriate	–13 to –7	67/90 to 82/90
Mildly delayed	–30 to –14	24/90 to 67/90
Moderately delayed	–50 to –31	3/90 to 24/90
Extremely delayed	–51 and below	0/90 to 3/90

ability–achievement discrepancy in achievement, which can suggest a need for academic intervention.

SUMMARY AND IMPLICATIONS

The ECAD (Schrank et al., 2015) is a special-purpose WJ IV battery of 10 early cognitive and academic skills tests, contained within a single test easel, that measure the emergence and development of CHC abilities at the early childhood level (ages 3 through 9). Four of the ten tests are unique to the ECAD, while the other six are adapted and alternate forms of tests included in other parts of the WJ IV that are useful for assessment in preschool, kindergarten, and the early primary grades. The ECAD forms of the tests have greater item density, to help capture changes in growth and development at the preschool and early elementary ages.

In addition to providing a highly reliable CHC-theory-based evaluation of cognitive abilities, the ECAD can be used to identify any delay in the development of expressive language or early academic skills. Test-level information can provide early indicators of specific cognitive and early academic skills that might be targets for intervention, if a need is identified. Early intervention that specifically targets cognitive deficits or delays in emerging academic skills can foster learning readiness and enhance the probability of success for a child entering kindergarten or progressing to the early grades.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development, 20*(1), 49.

Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *The Reading Teacher, 55,* 10–20.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology,* 98(1), 44–62.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Bowers, P. G., Golden, J., Kennedy, A., & Young, A. (1994). Limits upon orthographic knowledge due to processes indexed by naming speed. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge 1: Theoretical and developmental issues* (pp. 173–218). Dordrecht, The Netherlands: Kluwer Academic.

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing: An Interdisciplinary Journal, 5*(1), 69–85.

Bracken, B. A. (2013). *Riverside Early Assessments of Learning IDEA Observational Version*. Rolling Meadows, IL: Riverside.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience, 2,* 370–374.

Bunge, S. A., Mackey, A. P., & Whitaker, K. J. (2009). Brain changes underlying the development of cognitive control and reasoning. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 73–85). Cambridge, MA: MIT Press.

Carlisle, J. F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading, 7,* 239–253.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*(4), 428–449.

Carney, R. N. (2017). [Test review of Woodcock–Johnson IV Tests of Early Cognitive and Academic Development]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurement yearbook*. Lincoln, NE: Buros Center for Testing.

Carreker, S. (2005). Teaching spelling. In J. Birsh (Ed.), *Multisensory teaching of basic language skills* (2nd ed., pp. 257–295). Baltimore: Brookes.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales*. Circle Pines, MN: American Guidance Service.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: American Guidance Service.

Case, R., & Griffin, S. (1990). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. In E. A. Hauert (Ed.), *Developmental psychology: Cognitive, perceptual–motor, and neurological perspectives* (pp. 193–230). New York: Elsevier.

Chambers, B., Cheung, A., Slavin, R. E., Smith, D., & Laurenzano, M. (2010). *Effective early childhood education programs: A systematic review.* Baltimore: Johns Hopkins University, Center for Research and Reform in Education.

Conrad, N. J., & Levy, B. A. (2007). Training letter and orthographic pattern recognition in children with slow naming speed. *Reading and Writing: An Interdisciplinary Journal, 20,* 201–223.

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school.* Baltimore: Brookes.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Minneapolis, MN: Pearson Assessments.

Ehri, L. C. (1998). Learning to read and learning to spell are one and the same, almost. In C. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages* (pp. 237–269). Mahwah, NJ: Erlbaum.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition*. San Antonio, TX: Harcourt Assessment.

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafazi, A., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet, 378,* 1339–1353.

Englemann, S. (1968). The effectiveness of direction instruction on IQ performance and achievement in reading and arithmetic. In J. Hellmuth (Ed.), *Disadvantaged child* (Vol. 3). New York: Brunner/Mazel.

Fernald, G. M. (1943). *Remedial techniques in basic school subjects*. New York: McGraw-Hill.

Gazzaniga, M. S., Irvy, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind.* New York: Norton.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56.

Gersten, R., Ferrini-Mundy, J., Benbow, D., Clements, D. H., Loveless, T., Williams, V., et al. (2008). Report of the task group on instructional practices. In *National Mathematics Advisory Panel: Reports of the task groups and subcommittees* (pp. 45–53). Washington, DC: U.S. Department of Education.

Gersten, R., Jordon, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304.

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study. *American Psychologist, 69*(6), 588–599.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development, 4,* 251–268.

Graves, M. F., Juel, C., & Graves, B. B. (2004). *Teaching reading in the 21st century* (3rd ed.). Boston: Allyn & Bacon.

Greenleaf, R. K., & Wells-Papanek, D. (2005). *Memory, recall, the brain, and learning.* Newfield, ME: Greenleaf & Papanek.

Griffin, S. (2002). The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies. In J. M. Roher (Ed.), *Mathematical cognition* (pp. 1–32). Greenwich, CT: Information Age.

Griffin, S. (2004). Teaching number sense. *Educational Leadership, 61*(5), 39–42.

Gunn, B. K., Simmons, D. C., & Kame'enui, E. J. (1995). *Emergent literacy: A synthesis of the research.* Eugene, OR: National Center to Improve the Tools of Educators.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1), 4–9.

Hayward, D., Das, J. P., & Janzen. T. (2007). Innovative programs for improvement in reading through cognitive enhancements: A remediation study of Canadian First Nations children. *Journal of Learning Disabilities, 40*(5), 443–457.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning problems. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–149). Mahwah, NJ: Erlbaum.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*(3), 341–362.

Klingberg, T. (2009). *The overflowing brain: Information overload and the limits of working memory* (N. Betteridge, Trans.). New York: Oxford University Press.

LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2015). *Woodcock–Johnson IV Tests of Early Cognitive and Academic Development: Overview and Technical Abstract* (Woodcock–Johnson IV Assessment Service Bulletin No. 4). Rolling Meadows, IL: Riverside.

Logan, J., Schatschneider, C., & Wager, R. (2009). Rapid serial naming and reading ability: The role of lexical access. *Reading and Writing: An Interdisciplinary Journal, 24*(1), 1–25.

Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with Head Start children. *Journal of Educational Psychology,* 96(4), 648–659.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research,* 80, 300–335.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mather, N., & Wendling, B. J. (2015). *Essentials of WJ IV Tests of Achievement.* Hoboken, NJ: Wiley.

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations among socioeconomic status, age, and predictors of phonological awareness. *Journal of Speech, Language, and Hearing Research, 50,* 1079–1092.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Miller, L. B., & Dyer, J. L. (1975). Four preschool programs: Their dimensions and effects. *Monographs of* *the Society for Research in Child Development, 40*(5–6), Serial No. 162.

Missall, K., Reschly, A., & Betts, J. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36,* 433–452.

Moats, L. C. (2005). How spelling supports reading. *American Educator, 6,* 12–22, 42–43.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8,* 161–168.

Nagy, W. E. (2005). Why vocabulary instructions needs to be long-term and comprehensive. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 27–44). Mahwah, NJ: Erlbaum.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity.* Washington, DC: National Academies Press.

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2005). An investigation of the effects of a prereading intervention on the early literacy skills of children at risk of emotional disturbance and reading problems. *Journal of Emotional and Behavioral Disorders, 13*(1), 3–12.

Nelson, J. R., Stage, S. A., Epstein, M. H., & Pierce, C. D. (2005). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children, 72*(1), 29–45.

Newborg, J. (2005). *Battelle Developmental Inventory—Second Edition.* Itasca, IL: Riverside.

Nickels, L., & Best, W. (1996). Therapy for naming disorders (Part I): Principles, puzzles and progress. *Aphasiology, 10*(1), 21–47.

Nisbett, R. (2009). *Intelligence and how to get it: Why schools and cultures count.* New York: Norton.

Poe, M. D., Burchinal, M. R., & Roberts, J. E. (2004). Early language and the development of children's reading skills. *Journal of School Psychology, 42,* 315–332.

Ramani, G. R., & Siegler, R. S. (2005). *It's more than just a game: Effects of children's board games play on the development of numerical estimation.* Poster presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.

Rasch, G. (1980). *Probabalistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press. (Original work published 1960)

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech–Language Pathology, 20,* 180–199.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95,* 259–272.

Salaway, J. L. (2008). *Efficacy of a direct instruction approach to promote early learning.* Unpublished doctoral dissertation, Duquesne University, Pittsburgh, PA.

Schrank, F. A., & Dailey, D. (2014, 2015). *Woodcock–Johnson online scoring and reporting system* [Online format]. Rolling Meadows, IL: Riverside.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV cognitive abilities assessment.* Hoboken, NJ: Wiley.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock–Johnson IV Tests of Oral Language.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2015). *Woodcock–Johnson IV Tests of Early Cognitive and Academic Development.* Rolling Meadows, IL: Riverside.

Schrank, F. A., & Wendling, B. J. (2015a). *Manual and checklists: WJ IV interpretation and instructional interventions program.* Rolling Meadows, IL: Riverside.

Schrank, F. A., & Wendling, B. J. (2015b). *WJ IV interpretation and instructional interventions program.* Rolling Meadows, IL: Riverside.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Sengpiel, F., & Hubener, M. (1999). Visual attention: Spotlight on the primary visual cortex. *Current Biology, 9,* 318–321.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stanovich, K. E. (1994). Romance and reality. *The Reading Teacher, 4,* 280–290.

Sternberg, R. J., & Wagner, R. K. (1982). Automatization failure in learning disabilities. *Topics in Learning Disabilities, 2,* 1–11.

Templeton, S., & Bear, D. (Eds.). (1992). *Development of orthographic knowledge and the foundations of literacy.* Hillsdale, NJ: Erlbaum.

Torgesen, J. K. (1997). The prevention and remediation of reading disabilities: Evaluating what we know from research. *Journal of Academic Language Therapy, 1,* 11–47.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing difficulties: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. Oxford, UK: Oxford University Press.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385–398.

Venezky, R. (1970). *The structure of English orthography*. The Hague: De Gruyter.

Venezky, R. (1999). *The American way of spelling*. New York: Guilford Press.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*(1), 63–74.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence—Third Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition*. Bloomington, MN: Pearson.

Wendling, B. J., Mather, N., LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2015). *Comprehensive manual: Woodcock–Johnson IV Tests of Early Cognitive and Academic Development*. Rolling Meadows, IL: Riverside.

Whitehurst, G. J., Epstein, J. N., Angell, A. C., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86*, 542–555.

Woodcock, R. W., & Dahl, M. N. (1971). *A common scale for the measurement of person ability and test item difficulty* (AGS Paper No. 10). Circle Pines, MN: American Guidance Service.

Wright, B. D., & Stone, M. S. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press.

Wright, C. R. (2017). [Test review of Woodcock–Johnson IV Tests of Early Cognitive and Academic Development]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurement yearbook*. Lincoln, NE: Buros Center for Testing.

CHAPTER 11

The Wechsler Intelligence Scale for Children—Fifth Edition Integrated

Susan Engi Raiford

The Wechsler Intelligence Scale for Children—Fifth Edition Integrated (WISC-V Integrated; Wechsler & Kaplan, 2015) is a companion measure to the most widely used intelligence test for children in the world, the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014). It allows practitioners to learn more about the cognitive processes and test-taking behaviors that affect performance on the WISC-V and that also have an impact on school performance.

Because the WISC-V represents a substantial revision of this popular test, the WISC-V Integrated also reflects a good deal of change since the previous edition. Major modifications have been made to both the WISC-V content and structure to stay current with advances in the field of intellectual assessment. With updated content and structure, the WISC-V Integrated can be used to understand WISC-V results in greater detail at the item, subtest, and composite level. The test allows practitioners to weigh evidence that supports or refutes hypotheses about WISC-V results, and it informs accommodation, modification, and intervention recommendations.

The current chapter provides an overview of the WISC-V Integrated. It concludes with a section on interpretation. An appendix provides a clinical case example to illustrate how results can be used to understand WISC-V results; to understand problem-solving and learning issues; and to facilitate accommodation, modification, and intervention planning.

HISTORY AND THEORY

The *process approach* to cognitive assessment has its roots in ideas put forth by Heinz Werner (1937). Werner proposed that cognitive assessment involving careful and systematic observation of problem-solving strategies yields richer and more useful information than correct–incorrect scoring does. He suggested that cognitive task performance involves multiple, dynamic processes. Any of these processes may contribute individually or in combination with others to exert an influence on responses and performance. For example, a child may have difficulty with Block Design due to low motor skills, but another child's difficulties on the same task may be related to impaired visual–spatial processing. While both children may score similarly on Block Design, the poor performance can be traced back to two different origins.

Process approaches to assessment recognize that in completing any cognitive task, regardless of its complexity, a number of mental processes are involved. These processes can range from simple (e.g., sensing, perceiving, motor functioning) to

---

Portions of this chapter are adapted from *Essentials of WISC-V Integrated Assessment* (Raiford, 2017). Copyright © John Wiley & Sons, Inc. I express my thanks to John Wiley & Sons for granting permission to adapt this material.

more complex (e.g., reasoning, decision making; Kaufman, Raiford, & Coalson, 2016). A problem with any of these processes can result in difficulties with completing the overall task. The process approach aims to understand the reasons for low scores through identifying the cognitive processes that operate jointly on performance, breaking them down to more specific components in a stepwise fashion or altering the task in some manner, and testing hypotheses through employing the tasks that involve fewer or different cognitive processes. For example, the hypothesis that low motor skills result in a child's lower Block Design score could be more closely examined by presenting the completed design and multiple pictured sets of blocks in a stimulus book and asking the examinee to select the set of blocks that, when assembled, make the design.

There are many approaches to understanding the cognitive processes involved in performance on the Wechsler scales, beginning with an approach used by David Wechsler himself. He stated that "individuals attaining identical scores on intelligence tests cannot always be classified in the same way" (Wechsler, 1944, p. 12). His approach was somewhat psychoanalytic in nature, and described how information at the task and item levels could provide clues as to the nature of impairment.

Kaufman (1979, 1994) and Kaufman and Lichtenberger (1999, 2000, 2002) encouraged clinicians to adopt a process approach to interpretation of test results by analyzing input and output modalities (e.g., verbal input, visual output) as a means of better understanding the underlying causes of a child's pattern of strengths and weaknesses. This evolved into the *shared-abilities analysis* approach to interpretation of Wechsler intelligence scale performance. Ultimately, the approach was thought to be not specific enough and too loosely integrated with factor-analytic research to be clinically useful (Floyd & Kranzler, 2012).

Edith Kaplan and others at the Boston Veterans Administration Medical Center operationalized and refined Heinz Werner's original work; they termed their version the *Boston process approach*. They provided supporting evidence that this approach improved diagnostic and clinical utility in neuropsychological assessment, and that problem solving varied across various neurological conditions, even in the presence of similar scores (Kaplan, 1988).

The Wechsler Intelligence Scale for Children—Third Edition (WISC-III) as a Process Instrument (Kaplan, Fein, Kramer, Delis, & Morris, 1999) and the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) Integrated (Wechsler et al., 2004) were the predecessors of the WISC-V Integrated. Therefore, the WISC-V is the third edition of the WISC to feature a process approach measure that complements interpretation of results.

Psychologists utilize the process approach to investigate the reasons for low scores, learn about cognitive strengths and weaknesses, and develop recommendations for learning accommodations. Studies suggest that the process approach is clinically useful for assessing a wide variety of neuropsychological and neurodevelopmental problems, such as autism spectrum disorder, attention-deficit/hyperactivity disorder (ADHD), traumatic brain injury, and specific learning disorders (Boxer, Jackson, & Kohlman, 2014; Halleland, Sorensen, Posserud, Haavik, & Lundervold, 2015; Hoffmann, Donders, & Thompson, 2000; Kramer, Knee, & Delis, 2000; Mayfield, Reyes, Mayfield, & Allen, 2014; McLean, Johnson, Zimak, Joseph, & Morrow, 2014).

THE WISC-V INTEGRATED TEST FRAMEWORK

The WISC-V Integrated yields subtest-level scaled scores and index scores. Figure 11.1 shows the cognitive domains represented by the WISC-V Integrated and their corresponding WISC-V subtests, the WISC-V Integrated index scores published in the test manual, and the subtests that contribute to each index score. Descriptions of the cognitive domains have been provided earlier in this volume (see Wahlstrom, Raiford, Breaux, Zhu, & Weiss, Chapter 9).

The WISC-V Integrated battery consists of 14 subtests. Children are administered only the subtests necessary to clarify and extend WISC-V results.

An *adaptation* subtest includes the same item content as the corresponding WISC-V subtest, but the mode of presentation, response method, or item administration has been modified. Eight subtests (i.e., Similarities Multiple Choice, Vocabulary Multiple Choice, Picture Vocabulary Multiple Choice, Information Multiple Choice, Comprehension Multiple Choice, Figure Weights Process Approach, Arithmetic Process Approach, and Written Arithmetic) are adaptations of WISC-V subtests. A *variation* subtest has item content that does not correspond to that on a WISC-V subtest. and the mode of presentation, response method, and/or administration procedures have been modified. Two subtests (i.e., Block Design Multiple Choice and Cancellation Abstract) are variations of the WISC-V subtests. Two subtests (i.e., Coding Recall and Coding Copy) are designed to clarify WISC-V Coding performance. Two subtests (i.e., Spatial Span and Sentence Recall) are included because they improve the breadth of working memory construct coverage in the WISC-V/WISC-V Integrated system.

Each composite score in the published test is derived from two subtests. The Visual Working Memory Index can be derived after summing the WISC-V Picture Span scaled score and the Spatial Span scaled score. No substitution is permitted in deriving these scores.

The cognitive domains corresponding to the subtests appear in the first column of Figure 11.1. Because several of the WISC-V Integrated subtests are adaptations or variations, some of the WISC-V subtests appear in the second column. The third column lists the WISC-V Integrated subtests, in most cases next to the corresponding WISC-V subtest for which it is an adaptation or variation. The subtests used to derive the two index scores appear in the shaded boxes in the second and third columns, and an arrow designates the index score to which they contribute.

SUBTESTS AND INDEX SCORES

Table 11.1 provides a description of all subtests and information on the constructs and abilities thought to be involved with each subtest.

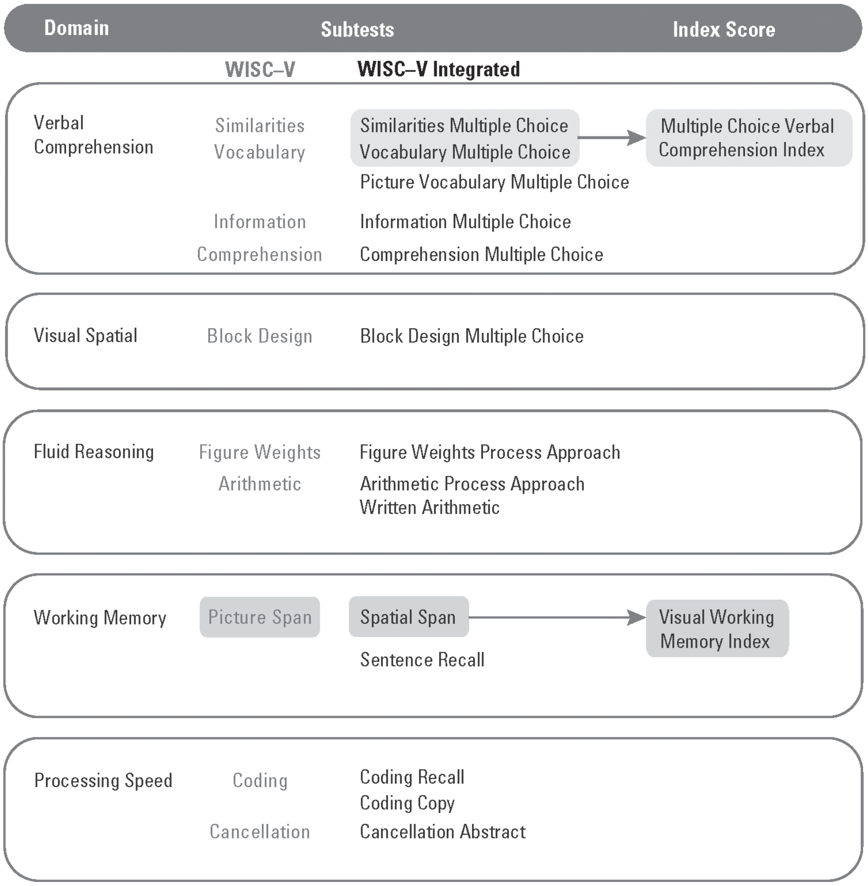


FIGURE 11.1. WISC-V Integrated test framework. Figures found in the manual for the *Wechsler Intelligence Scale for Children®, Fifth Edition Integrated (WISC®-V Integrated)*. Copyright © 2015 NCS Pearson, Inc. Reproduced with permission. All rights reserved.

TABLE 11.1. Subtest Descriptions, Constructs, and Abilities

Subtest	Description	Constructs and abilities
Similarities Multiple Choice	Multiple-choice adaptation (same items) of the WISC-V Similarities subtest. Each item and its response options are presented visually and read aloud. The child selects the response option that best represents how the common objects or concepts are similar.	Designed to measure: Verbal reasoning and concept formation. Decreased demands relative to WISC-V Similarities: Verbal expression and memory retrieval. Possibly increased demands relative to WISC-V Similarities: Receptive language skills, reading skills, decision-making skills, and working memory. Related to: Crystallized ability, associative and categorical thinking, Gf-I (induction), concept recognition and generation. May also involve: Auditory perception.
Vocabulary Multiple Choice	Multiple-choice adaptation (same items) of the WISC-V Vocabulary subtest. For picture items, the child views pictures and selects the best response from options read aloud. For verbal items, each item and its response options are presented visually and read aloud. The child selects the response option that best represents the definition of the word.	Designed to measure: Word knowledge, verbal concept formation, semantic memory. Decreased demands relative to WISC-V Vocabulary: Verbal expression and memory retrieval. Possibly increased demands relative to WISC-V Vocabulary: Receptive language skills, reading skills, decision-making skills, and working memory. Related to: Crystallized ability, Gc-VL (lexical knowledge), fund of knowledge, learning, verbal expression, long-term semantic memory, vocabulary development. May also involve: Auditory perception, auditory comprehension, abstract thinking, receptive vocabulary.
Picture Vocabulary Multiple Choice	Pictorial multiple-choice adaptation (same items) of the WISC-V Vocabulary subtest. The child views four pictures and selects the picture that best depicts the definition of the word that is read aloud.	Designed to measure: Word knowledge, verbal concept formation, receptive vocabulary. Decreased demands relative to WISC-V Vocabulary: Verbal expression, memory retrieval, receptive language. Possibly increased demands relative to WISC-V Vocabulary: Visual perception, decision-making skills, working memory. Related to: Crystallized ability, Gc-VL (lexical knowledge), fund of knowledge, learning, long-term memory, visual comprehension, visual–verbal association formation. May also involve: Visual perception, auditory comprehension.
Information Multiple Choice	Multiple-choice adaptation (same items) of the WISC-V Information subtest. Each item and its response options are presented visually and read aloud. The child selects the response option that best represents an understanding of the general knowledge topic.	Designed to measure: Acquisition, retention, and retrieval of general facts and knowledge. Decreased demands relative to WISC-V Information: Verbal expression and memory retrieval. Possibly increased demands relative to WISC-V Information: Receptive language skills, reading skills, decision-making skills, and working memory. Related to: Crystallized ability, Gc-K0 (general information), Glr (retention and retrieval of learned information). May also involve: Auditory perception, verbal expression.

(continued)

TABLE 11.1. *(continued)*

Subtest	Description	Constructs and abilities
Comprehension Multiple Choice	Multiple-choice adaptation (same items) of the WISC-V Comprehension subtest. Each item and its response options are presented visually and read aloud. The child selects the response option that best represents an understanding of the general principle or social situation.	Designed to measure: Verbal reasoning, verbal conceptualization, verbal comprehension, verbal expression, practical knowledge, judgment. Decreased demands relative to WISC-V Comprehension: Verbal expression and memory retrieval. Possibly increased demands relative to WISC-V Comprehension: Receptive language skills, reading skills, decision-making skills, and working memory. Related to: Crystallized ability (Gc), understanding of societal standards and conventional behavior, social judgment, Glr, common sense. May also involve: Auditory perception.
Block Design Multiple Choice	Multiple-choice variation (different items) of the WISC-V Block Design subtest. The child views a picture of a constructed block design and selects the pictured block set that produces a matching composition, within a specified time limit.	Designed to measure: Visual–spatial processing, analysis and synthesis of abstract visual stimuli, mental imaging. Decreased demands relative to WISC-V Block Design: Relaying response to motor channels, motor skills. Possibly increased demands relative to WISC-V Block Design: Decision-making skills, working memory. Related to: Gv-SR (spatial relations), Gv-Vz (visualization), Gv-CS (closure speed), mental rotation, nonverbal reasoning, visual perception, simultaneous processing, problem solving, cognitive flexibility, planning.
Figure Weights Process Approach[a]	Adaptation (same items) of the WISC-V Figure Weights subtest. The child is given additional time to respond. Within an extended time limit, the child is readministered Figure Weights items previously scored 0 points.	Designed to measure: Quantitative fluid reasoning/intelligence, inductive reasoning. Decreased demands relative to WISC-V Figure Weights: Speeded performance. Related to: Gf-I, Gf-RQ, simultaneous and successive processing, problem solving, cognitive flexibility. May also involve: Working memory, math problem solving, math computation.
Arithmetic Process Approach	Adaptation (same items) of the WISC-V Arithmetic subtest. Items 6–34 are presented in multiple modalities for the child to solve within a specified time limit. For Part A, Arithmetic items on which the child scored 0 points are presented visually and simultaneously read aloud. For Part B, the child is provided with pencil and paper, and is readministered the items scored 0 points in Part A.	Designed to measure: Quantitative, fluid, and logical reasoning, mental manipulation. Decreased demands relative to WISC-V Arithmetic: Attention, auditory working memory, short-term memory, auditory discrimination, auditory comprehension. Possibly increased demands relative to WISC-V Arithmetic: Reading, graphomotor skills. Related to: Gf-RQ, sequential processing, working memory, quantitative knowledge, applied computation, logical reasoning. May also involve: Auditory discrimination.
Written Arithmetic	Adaptation (same items) of the WISC-V Arithmetic subtest. The child is presented with the mathematical computations for Arithmetic items and uses paper and pencil to complete them.	Designed to measure: Numerical reasoning ability, acquired knowledge of mathematical calculations, math computation. Decreased demands relative to WISC-V Arithmetic: Attention, mental efficiency, verbal aspects of cognitive arithmetic, math problem solving.

(continued)

TABLE 11.1. *(continued)*

Subtest	Description	Constructs and abilities
Written Arithmetic *(continued)*		Possibly increased demands relative to WISC-V Arithmetic: Reading, graphomotor skills. Related to: Gf-RQ, sequential processing, working memory, quantitative knowledge, applied computation, logical reasoning, calculation skills, counting skills, math facts retrieval. May also involve: Working memory, knowledge of mathematical symbols and syntax, order-of-operations knowledge.
Spatial Span	Expands construct coverage of working memory. Spatial Span is composed of two tasks: Forward and Backward. For Spatial Span Forward, the child reproduces a sequence of tapped blocks. For Spatial Span Backward, the child reproduces a sequence of tapped blocks, in reverse order.	Designed to measure: Visual–spatial working memory. Related to: Gsm-MW (working memory capacity), Gsm-MS (memory span), Gv-MV (visual memory), attention and attentional capacity, simultaneous and successive processing, planning and metacognition, visual immediate memory, spatial locations, response inhibition. May also involve: Motor integration and programming, motor and self-regulation, cognitive flexibility, mental alertness, primacy effects, recency effects.
Sentence Recall[a]	Expands construct coverage of working memory. Sentence Recall items are composed of two tasks: a question task and a recall task. For the question task, the child responds either "Yes" or "No" to one or more simple questions. For the recall task, the child recalls the last word of each question, in the order presented.	Designed to measure: Auditory working memory with cognitive processing, working memory capacity. Related to: Storage during cognitive processing, reactivation of attention.
Coding Recall	Provides more information about performance on the WISC-V Coding subtest. Working within a specified time limit and without a key, the child attempts to remember the corresponding pairs from Coding in three formats: cued recall, free recall, and pairing.	Designed to measure: Incidental learning, associative memory. Decreased demands relative to WISC-V Coding: Graphomotor speed, timed performance. Possibly increased demands relative to WISC-V Coding: Associative memory. Related to: Short-term visual recall and recognition memory and learning ability. May also involve: Visual–motor skills, procedural learning.
Coding Copy	Provides more information about performance on the WISC-V Coding subtest. The child copies symbols within a specified time limit.	Designed to measure: Speed, fluency, and efficiency of processing; performance fluency, graphomotor speed, perceptual speed, visual–motor integration. Decreased demands relative to WISC-V Coding: Incidental learning, associative memory. Possibly increased demands relative to WISC-V Coding: Graphomotor speed. Related to: Selective and sustained attention, visual scanning and tracking, response inhibition. May also involve: Visual–motor skills.

(continued)

TABLE 11.1. *(continued)*

Subtest	Description	Constructs and abilities
Cancellation Abstract[a]	A variation of the WISC-V Cancellation subtest. Working within a specified time limit, the child scans two arrangements of shapes (one random, one structured) and marks target shapes.	Designed to measure: Processing speed; speed, fluency, and efficiency of processing; performance fluency. Decreased demands relative to WISC-V Cancellation: categorical knowledge, visual immediate memory. Possibly increased demands relative to WISC-V Coding: Response inhibition. Related to: Gs-P (perceptual speed), speed and efficiency, Gs-P (perceptual speed), simultaneous processing, planning and metacognition, speed and efficiency, selective and sustained attention, visual scanning and tracking, visual immediate memory, response inhibition. May also involve: Visual–motor skills.

Note. References: Cardoso, Branco, Cotrena, and Fonseca (2015); Carroll (1993); Demakis, Sawyer, Fritz, and Sweet (2001); Flanagan and Alfonso (2017); Flanagan, Alfonso, and Ortiz (2012); Flanagan, Alfonso, Ortiz, and Dynda (2010); Gagnon and Belleville (2011); Goldstein and Green (1995); Groeger, Field, and Hammond (1999); Groth-Marnat (2009); Joy, Fein, Kaplan, and Freedman (1999); Joy, Kaplan, and Fein (2003); Kreiner and Ryan (2001); Lezak, Howieson, Bigler, and Tranel (2012); Lichtenberger and Kaufman (2013); Mainela-Arnold, Misra, Miller, Poll, and Park (2012); McCloskey (2009); McCloskey and Maerlender (2005); Milberg, Hebben, and Kaplan (1986); Miller (2010, 2013); Miller and Jones (2016); Sattler (2008); Sattler, Dumont, and Coalson (2016); Schneider and McGrew (2012); Schroeder (2014); Service and Maury (2015); Smyth and Scholey (1992).

Adapted from Table 1.1 of the *WISC-V Integrated Administration and Scoring Manual* (Wechsler & Kaplan, 2015), part of the *Wechsler Intelligence Scale for Children®, Fifth Edition Integrated (WISC®-V Integrated)*. Copyright © 2015 NCS Pearson, Inc. Adapted with permission. All rights reserved. Also adapted from Rapid References 1.4 and 1.5 of *Essentials of WISC-V Integrated Assessment* (Raiford, 2017). Copyright © John Wiley & Sons, Inc. Adapted by permission.

[a] New subtest.

Multiple Choice Verbal Comprehension Index

The Multiple Choice Verbal Comprehension Index (MCVCI) is a new composite score that is designed to measure the same cognitive processes as the WISC-V Verbal Comprehension Index (VCI), but in a multiple-choice format to reduce expressive language requirements. It is derived from Similarities Multiple Choice and Vocabulary Multiple Choice, which are the multiple-choice adaptations of the subtests that contribute to the WISC-V VCI (i.e., Similarities and Vocabulary). The multiple-choice format reduces the demands on retrieval and expressive ability because correct responses are merely recognized to demonstrate acquired semantic knowledge.

Visual Working Memory Index

Picture Span from the WISC-V and Spatial Span are summed to derive the Visual Working Memory Index (VWMI). Contemporary working memory research indicates that domain-specific neuropsychological systems support auditory working memory versus visual–spatial working memory (Baddeley, 2012). The VWMI facilitates an expanded examination of a child's working memory abilities by providing a purer measure of visual working memory processes than the Working Memory Index (WMI), and by allowing examination of relative strength and weakness across the working memory system when compared with the Auditory Working Memory Index (AWMI).

PSYCHOMETRIC PROPERTIES

Normative Sample

The WISC-V Integrated normative sample consisted of 550 children, with 50 children (equal numbers male and female) in each of 11 age groups corresponding to those of the WISC-V normative sample. The normative sample was stratified to closely match 2014 U.S. census data for parental education level, race/ethnicity, and geographic region.

Reliability

The WISC-V Integrated composite scores have strong reliabilities (Wechsler & Kaplan, 2015). The overall internal-consistency reliability coefficients for the normative sample are .87 for the MCVCI and .90 for the VWMI. At the subtest level, the overall internal-consistency reliability coefficients of the normative sample are in the .80s or .90s, with the exceptions of Similarities Multiple Choice and Comprehension Multiple Choice (with coefficients of .79 and .77, respectively).

Overall, the reliability for special-group samples is consistent with reliability estimates for the normative sample (Wechsler & Kaplan, 2015). The test–retest stability coefficient of the MCVCI is .72, and that of the VWMI is .91. The subtest-level coefficients range from .69 to .86. Finally, the subtest-level interscorer agreement range is .98–.99.

Validity

The evidence to support the validity of the WISC-V Integrated is derived from its relations with the WISC-V and its intercorrelations. The subtests and index scores show moderate to high correlations with the WISC-V subtests and index scores from the same domain (Wechsler & Kaplan, 2015), the highest of which are with the WISC-V subtests and index scores from the same cognitive domain. These results indicate that the WISC-V Integrated subtests and index scores measure cognitive processes and abilities similar to those measured by the WISC-V, and that these cognitive processes and abilities are highly relevant to WISC-V performance. The subtest-level scores from each cognitive domain are generally moderately to highly correlated with other subtest-level scores in that domain. They show evidence of discriminant validity with subtests from other cognitive domains as well.

A great deal of evidence also supports the utility of the WISC-V Integrated as a clinical tool for understanding children's WISC-V performance, cognitive abilities, and problem solving. Studies conducted during the standardization stage with a number of special groups (e.g., children with intellectual disability, intellectual giftedness, ADHD, autism spectrum disorders, learning disabilities, traumatic brain injury, language disorders) demonstrate unique patterns of performance relative to the normative samples (Wechsler & Kaplan, 2015). These results are corroborated by a number of independent samples that show similar patterns of performance on the WISC-V and other measures of cognitive ability (see Wahlstrom et al., Chapter 9, this volume).

Moreover, the WISC-V Integrated correlates highly with academic achievement (Wechsler & Kaplan, 2015) as measured by the Wechsler Individual Achievement Test—Third Edition (WIAT-III) and the Kaufman Test of Educational Achievement, Third Edition (KTEA-3). Correlations of the index scores with achievement composites are generally in the .40s–.70s, and a great number of the WISC-V Integrated subtest-level scores have similar relationships with the achievement subtests and composites. These results indicate that the cognitive processes measured by the WISC-V Integrated are important to academic success. They also provide evidence of divergent validity (i.e., they suggest that the WISC-V Integrated and achievement tests measure different constructs, with varying degrees of overlap in the cognitive skills required).

INTERPRETATION

The WISC-V Integrated technical and interpretive manual (Wechsler & Kaplan, 2015) provides basic instruction regarding its interpretation. Clinicians and researchers have developed more in-depth and alternative interpretive strategies for the WISC-V Integrated (Raiford, 2017) and its predecessor, the WISC-IV Integrated (McCloskey, 2009; McCloskey & Maerlender, 2005). A detailed guide to interpretation is outside the scope of this chapter, and the reader is referred to these other works, as well as the technical and interpretive manual, for further information. Instead, basic information is provided to illustrate how the WISC-V Integrated can be utilized with the WISC-V to better understand low performance. The reader is cautioned that this is not a comprehensive review of interpretive strategies, and that test results should always be interpreted in conjunction with history, background, and careful clinical observations of the individual and corroborated with school-related information.

Scores and Levels of Performance

There are several types of scores available on the WISC-V Integrated, including age-adjusted scores, total raw scores, base rates for errors or other aspects of performance or observations, and percentile norms.

Age-Adjusted Scores

As the WISC-V does, the WISC-V Integrated utilizes two types of age-adjusted scores: scaled scores for subtest-level information, and standard scores for all composite scores. These are described in greater detail by Wahlstrom and colleagues in Chapter 9.

These scores permit comparison of the child's performance to his or her same-age peers and allow comparisons within the WISC-V Integrated and with the WISC-V and measures of achievement, memory, and other cognitive processes. To facilitate communication of results with parents and teachers, test results are communicated with percentile ranks (i.e., the percentage of children in the age group that obtain scores below or consistent with the child's performance). The index scores, like those on the WISC-V, should always be accompanied by a confidence interval. The index scores also utilize the same qualitative descriptors established for the WISC-V, as shown in Table 9.2 of this volume.

For most scores, comparisons can only take place at the scaled score level because interpretation must take into account performance relative to same-age peers. However, it is possible to conduct meaningful comparisons of WISC-V subtests with their corresponding adaptation subtests, as well as some of the WISC-V Integrated subtest-level scores, using both the scaled score and the raw score metrics.

Raw Scores

Subtest-Level Total Raw Scores

Total raw scores may be used to ascertain the effect of a modified presentation and/or response format on a given child's intraindividual performance. That is, comparisons at the raw score level answer the following question: How does changing the time given to respond, presentation method, or response format of a task affect the child's expression of a given ability? This information can be used to ascertain the benefit of those accommodations and modifications, and can inform recommendations for the same in the classroom. The scaled score metric comparison answers a different question. That is, what is the impact of the accommodation or modification on the child's performance relative to the impact on same-age peers' performance? This information provides a different take on the child's ability in that area, and also can clarify the need for accommodations or modifications in specific scenarios.

Base Rates and Percentile Norms

The WISC-V and the WISC-V Integrated contain a variety of other raw scores that are converted to base rates or percentile norms. They include longest span and sequence scores (e.g., the number of blocks recalled on the last Spatial Span Forward trial scored 1 point), error scores (e.g., the number of yes–no responses to Sentence Recall questions that were incorrect), and process observations (e.g., the number of times the examinee indicated not knowing the answer to an Information Multiple Choice item). It also provides percentile norms for the number of correct responses on the various tasks for Coding Recall.

Basic Interpretation

WISC-V interpretation generally proceeds from the global (e.g., Full Scale IQ) to the specific (e.g., index scores, subtest-level scores, item-level performance). The WISC-V Integrated can be used to clarify the child's WISC-V results at any point in the process.

Reporting and Describing Global Performance

All of the WISC-V interpretive systems recommend examining global performance first. Three global composite scores are available in the published WISC-V: the Full Scale IQ, the Nonverbal Index (NVI), and the General Ability Index (GAI). Each may be appropriate to describe overall intellectual ability, depending on the clinical situation and the purpose of the evaluation.

In some specific clinical situations, however, access to an alternate global score is useful. This is particularly important when issues related to language or motor impairment make it impossible to calculate a Full Scale IQ, NVI, or GAI. This occurs more frequently with the WISC-V than with its predecessors because all but one of the WISC-V composite scores do not allow substitution or proration. The sole exception is the Full Scale IQ, which allows only one substitution or proration—not enough to exclude all subtests that require verbal expression or all subtests that have motor requirements from the calculation.

Global Scores for Expressive Language Difficulties and/or Motor Impairment

Essentials of WISC-V Integrated Assessment (Raiford, 2017) provides several additional global composite scores that use combinations of WISC-V and WISC-V Integrated subtests to accommodate situations when expressive language, motor development, or both are concerns: the Nonexpressive Full Scale Score (NEFSS), the Nonmotor Full Scale Score (NMFSS); the Nonmotor Nonverbal Index (NMNVI), and the Nonmotor General Ability Index (NMGAI). The global composite scores are therefore expanded to seven options, which are listed in Table 11.2. Each score is listed with its abbreviation, subtest composition, and an example of potentially appropriate use. Interpretive details and the composite norms tables for these new scores are available in Raiford (2017).

For the NMFSS, it is notable that the complementary WISC-V Naming Speed Quantity subtest replaces Coding. Both Coding and Naming Speed Quantity involve cognitive speed and visual scanning. However, Coding involves encoding and rapidly using newly encoded associations and motor production, but Naming Speed Quantity involves rapid retrieval of frequently used long-term associative memories and expressive responses. To obtain the Naming Speed Quantity scaled score, the standard score must be converted to a scaled score from a standard score. The conversion can be completed automatically with the software that is included with the Raiford (2017) book.

Naming Speed Quantity was selected to contribute to the NMFSS for a number of reasons. First, if choosing between the two WISC-V Naming Speed subtests, Naming Speed Quantity has reduced language requirements relative to Naming Speed Literacy. Second, a review of the WISC-V technical and interpretive manual reveals that its clinical sensitivity is similar to that of Coding for many neurodevelopmental conditions commonly assessed with the WISC-V (i.e., intellectual disability, specific learning disorders, ADHD, autism spectrum disorder with language impairment, autism spectrum disorder without language impairment, and traumatic brain injury). Third, it has a superior *g* loading relative to Naming Speed Literacy (Raiford, 2017).

TABLE 11.2. Global Composite Scores in the WISC-V and WISC-V Integrated System and Examples of Potentially Appropriate Uses

Global composite score	Source	Subtest composition	Example of potentially appropriate use
Full Scale IQ	Wechsler (2014)	SI, VC, BD, MR, FW, DS, CD	Default global score
Nonexpressive Full Scale Score (NEFSS)	Raiford (2017)	SIMC,[a] VCMC,[a] BD, MR, FW, PS, CD	Expressive language issues, but ideal for crystallized ability to still contribute to the global score
Nonmotor Full Scale Score (NMFSS)	Raiford (2017)	SI, VC, BDMC,[a] MR, FW, DS, NSQ	Motor impairment
Nonverbal Index (NVI)	Wechsler (2014)	BD, VP, MR, FW, PS, CD	Language issues (expressive and mild receptive)
Nonmotor Nonverbal Index (NMNVI)	Raiford (2017)	BDMC,[a] VP, MR, FW, PS	Expressive issues accompanied by motor impairment
General Ability Index (GAI)	Wechsler (2014)	SI, VC, BD, MR, FW	Neurodevelopmental issue affecting working memory and/or processing speed performance
Nonmotor General Ability Index (NMGAI)	Raiford (2017)	SI, VC, BDMC,[a] MR, FW	Neurodevelopmental issue affecting working memory and/or processing speed performance, also accompanied by motor impairment

Note. Subtest abbreviations: SI, Similarities; VC, Vocabulary; BD, Block Design; MR, Matrix Reasoning; FW, Figure Weights; DS, Digit Span; CD, Coding; SIMC, Similarities Multiple Choice; VCMC, Vocabulary Multiple Choice; PS, Picture Span; BDMC, Block Design Multiple Choice; NSQ, Naming Speed Quantity.

From Raiford (2017). Copyright © John Wiley & Sons, Inc. Adapted by permission.

[a] WISC-V Integrated subtest.

Reporting and Describing Index Scores

The published WISC-V Integrated index scores are useful when a more in-depth explanation of low scores is necessary. For example, if a child unexpectedly receives a low score on the WISC-V VCI, the MCVCI can be used to clarify whether the child merely performs more poorly than peers due to expressive issues or whether other weaknesses are implicated. As another instance, the VWMI can be used together with the WISC-V WMI and AWMI to provide a broader assessment of working memory. Table 11.3 lists the alternate index scores that correspond to primary index scores, and examples of their appropriate use.

Expanded Index Scores

The WISC-V and WISC-V Integrated system's expanded index scores may be used in various situations where a broader measure of the ability of interest would be helpful. For example, practitioners who assess intellectually gifted children, who perform comparisons with achievement scores in some states, or who complete evaluations for private school admissions have expressed interest in the expanded scores for their purposes. As another example, some practitioners utilize expanded index scores when there is a significant discrepancy between the two subtest scaled scores that contribute to the corresponding primary index score.

The names, sources, and examples of appropriate uses of each expanded index score are listed in Table 11.4. The appropriate uses listed are only examples; other clinical situations may also call for their use.

Index-Level Strengths and Weaknesses

It is crucial to ground interpretation of the index-level strengths and weaknesses in the context of index score interpretation, describing them in relation to the normative sample as well as to intrapersonal global performance. Kaufman and colleagues (2016) recommend using the Full Scale IQ for strengths-and-weaknesses comparisons even if another global composite score (i.e., the NVI or the GAI) is selected to represent overall ability. Even if another global score is selected, the authors suggest substituting the selected composite score for the Full Scale IQ score and performing the comparisons anyway because the correla-

TABLE 11.3. Sources and Examples of Potentially Appropriate Uses for the Alternate Index Scores Corresponding to Primary Index Scores

Alternate index score	Alternate for:	Issue	Source	Examples of potentially appropriate uses
Multiple Choice Verbal Comprehension Index	Verbal Comprehension Index	Expressive	Wechsler and Kaplan (2015)	Measuring verbal comprehension in the presence of expressive language difficulties. Investigating the impact of reducing expressive and retrieval demands on the measure of verbal comprehension.
Nonmotor Visual Spatial Index	Visual Spatial Index	Motor	Raiford (2017)	Measuring visual–spatial processing in the presence of motor impairment. Investigate the impact of reducing motor demands and trial-and-error problem solving on the measure of visual–spatial ability.
Auditory Working Memory Index	Working Memory Index	Vision	Wechsler (2014)	Measuring working memory in the presence of visual impairment. Investigating the impact of eliminating visual requirements on the measure of working memory.
Visual Working Memory Index	Working Memory Index	Expressive	Wechsler and Kaplan (2015)	Measuring working memory in the presence of expressive language difficulties. Investigating the impact of eliminating expressive language demands on the measure of working memory.

TABLE 11.4. WISC-V and WISC-V Integrated Expanded Index Scores

Expanded index score	Source	Examples of appropriate use
Verbal (Expanded Crystallized) Index	Kaufman, Raiford, and Coalson (2016); Raiford, Drozdick, Zhang, and Zhou (2015)	• An expanded measure of crystallized ability is needed. • Intellectual giftedness. • Comparisons with achievement scores. • Evaluations for private school admissions. • Significant discrepancy between Similarities and Vocabulary (WISC-V VCI).
Nonexpressive (Expanded Crystallized) Index	Raiford (2017)	• An expanded measure of crystallized ability is needed in the presence of expressive language issues. • Intellectual giftedness in the presence of expressive language issues. • Comparisons with achievement scores in the presence of expressive language issues. • Evaluations for private school admissions in the presence of expressive language issues. • Significant discrepancy between Similarities Multiple Choice and Vocabulary Multiple Choice (WISC-V Integrated MCVCI).
Expanded Visual Spatial Index	Raiford (2017)	• An expanded measure of visual–spatial ability is needed. • Intellectual giftedness. • Comparisons with achievement scores. • Evaluations for private school admissions. • Significant discrepancy between Block Design and Visual Puzzles (WISC-V VSI).
Expanded Fluid Index	Kaufman, Raiford, and Coalson (2016); Raiford, Drozdick, Zhang, and Zhou (2015)	• An expanded measure of fluid reasoning ability is needed. • Intellectual giftedness. • Comparisons with achievement scores. • Evaluations for private school admissions. • Significant discrepancy between Matrix Reasoning and Figure Weights (WISC-V FRI).
Expanded Working Memory Index	Raiford (2017)	• An expanded measure of working memory ability is needed. • Strengths-and-weaknesses analysis for specific learning disability evaluation. • Significant discrepancy between Digit Span and Picture Span (WISC-V WMI). • Significant discrepancy between Digit Span and Letter–Number Sequencing (WISC-V AWMI). • Significant discrepancy between Picture Span and Spatial Span (WISC-V Integrated VWMI).
Expanded Auditory Working Memory Index	Raiford (2017)	• An expanded measure of auditory working memory ability is needed. • Strengths-and-weaknesses analysis for specific learning disability evaluation. • Significant discrepancy between Digit Span and Letter–Number Sequencing (WISC-V AWMI).
Expanded Processing Speed Index	Raiford (2017)	• An expanded measure of processing speed ability is needed. • Strengths-and-weaknesses analysis for specific learning disability evaluation. • Significant discrepancy between Coding and Symbol Search (WISC-V PSI).

tions of the Full Scale IQ with these other global scores is better than the Full Scale IQ retest stability value and only differ slightly from a prorated Full Scale IQ in terms of content. The Raiford (2017) book extends this approach to include the global scores created for children with expressive or motor difficulties provided in that book (e.g., NEFSS, NMFSS). There, I demonstrate that the correlations of those scores with the Full Scale IQ are comparable to those of the NVI and the GAI, and that their reliabilities and standard errors of measurement are comparable to those of the Full Scale IQ. I conclude that substituting any of those scores for the Full Scale IQ in strengths-and-weaknesses comparisons is likely to provide a good approximation.

It is typical to have some areas of strength and weakness across cognitive ability domains. A significant strength or weakness relative to an overall indicator of intellectual ability is a normal occurrence and should not be taken as a sign of pathology or abnormality. The Raiford (2017) book presents data demonstrating that such discrepancies are extremely common and occur in close to 90% of results, whether in clinical or nonclinical populations.

Composite-Level Pairwise Comparisons

A number of pairwise composite-level comparisons can clarify interpretation as framed within the process approach. These comparisons can elucidate the examinee's cognitive strengths and needs and reasons for poor WISC-V performance, as well as the potential effect of accommodations and modifications. The WISC-V Integrated provides critical values, base rates, and interpretive hypotheses relevant to three composite-level pairwise comparisons (i.e., VCI–MCVCI, WMI–VWMI, and AWMI–VWMI). The Raiford (2017) book provides additional composite score comparisons and similar information; the additional comparisons are listed in Table 11.5, along with the sources for each composite score. In that book, I also provide interpretive hypotheses for these comparisons, which can shed light on the impact

TABLE 11.5. WISC-V and WISC-V Integrated Composite Score Pairwise Comparisons

Score 1	Source	Score 2	Source
Full Scale IQ	Wechsler (2014)	Nonexpressive Full Scale Score	Raiford (2017)
Full Scale IQ	Wechsler (2014)	Nonmotor Full Scale Score	Raiford (2017)
General Ability Index	Wechsler (2014)	Nonmotor General Ability Index	Raiford (2017)
Nonmotor General Ability Index	Raiford (2017)	Nonmotor Full Scale Score	Raiford (2017)
Nonverbal Index	Wechsler (2014)	Nonmotor Nonverbal Index	Raiford (2017)
Verbal (Expanded Crystallized) Index	Kaufman, Raiford, and Coalson (2016); Raiford, Drozdick, Zhang, and Zhou (2015)	Nonverbal Index	Wechsler (2014)
Nonexpressive Expanded Crystallized Index	Raiford (2017)	Nonverbal Index	Wechsler (2014)
Verbal (Expanded Crystallized) Index	Kaufman, Raiford, and Coalson (2016); Raiford, Drozdick, Zhang, and Zhou (2015)	Nonexpressive Expanded Crystallized Index	Raiford (2017)
General Verbal Information Index	Kaufman, Raiford, and Coalson (2016)		
Nonexpressive General Verbal Information Index	Raiford (2017)		

Note. From Raiford (2017). Copyright © John Wiley & Sons, Inc. Adapted by permission.

of expressive or motor difficulties on WISC-V composite scores.

Examining Subtest-Level Performance

Pairwise subtest-level comparisons can inform interpretation of performance, particularly within the process approach. Such comparisons shed light on the reasons for poor WISC-V performance, as well as the potential effect of accommodations and modifications. The WISC-V Integrated provides critical value and base rate information relevant to subtest-level pairwise comparisons.

The pairs of subtest-level scores with meaningful comparisons are listed in Table 11.6. Each comparison is designated as meaningful, using a scaled score metric or both a scaled and raw score metric.

It is worth noting that comparisons of Picture Vocabulary Multiple Choice with Vocabulary and Vocabulary Multiple Choice should only be made at the scaled score or item level because Picture Vocabulary Multiple Choice is scored 0 or 1 at the

TABLE 11.6. Pairs of WISC-V and WISC-V Integrated Subtest-Level Scores Producing Meaningful Comparisons Using Scaled and/or Raw Score Metrics

	Meaningful comparison score metric	
Comparison	Scaled	Raw
Similarities and Similarities Multiple Choice	✓	✓
Vocabulary and Vocabulary Multiple Choice	✓	✓
Vocabulary and Picture Vocabulary Multiple Choice	✓	
Vocabulary Multiple Choice and Picture Vocabulary Multiple Choice	✓	
Information and Information Multiple Choice	✓	✓
Comprehension and Comprehension Multiple Choice	✓	✓
Block Design and Block Design Multiple Choice	✓	
Block Design No Time Bonus and Block Design Multiple Choice	✓	
Figure Weights and Figure Weights Process Approach	✓	✓
Arithmetic[a] and Arithmetic Process Approach Part A	✓	✓
Arithmetic[a] and Arithmetic Process Approach Part B	✓	✓
Arithmetic Process Approach Part A and Arithmetic Process Approach Part B	✓	✓
Arithmetic Process Approach Part B and Written Arithmetic	✓	✓
Arithmetic[a] and Written Arithmetic	✓	✓
Digit Span and Sentence Recall	✓	
Letter–Number Sequencing and Sentence Recall	✓	
Picture Span and Spatial Span	✓	
Spatial Span Forward and Spatial Span Backward	✓	
Digit Span Forward and Spatial Span Forward	✓	
Digit Span Backward and Spatial Span Backward	✓	
Coding and Coding Copy	✓	✓
Cancellation and Cancellation Abstract	✓	✓
Cancellation Random and Cancellation Abstract Random	✓	✓
Cancellation Structured and Cancellation Abstract Structured	✓	✓

Note. From Raiford (2017). Copyright © by John Wiley & Sons, Inc. Adapted by permission.

[a] Ensure that the five picture items are subtracted from the Arithmetic total raw score before comparing this score to Arithmetic Process Approach or Written Arithmetic total raw scores. However, *do not* attempt to adjust any scaled scores for these picture items.

item level, whereas the others are scored 0, 1, or 2 at the item level. Total raw score comparisons may be made between Arithmetic and its adaptations, but Arithmetic has five picture items that do not appear on the corresponding adaptation scores. These must be taken into account when comparisons are made at the total raw score level. Detailed information relevant to interpreting subtest-level differences is offered in the Raiford (2017) book. A few illustrative examples extracted from that book appear in Table 11.7.

Examining Qualitative Aspects of Performance

Qualitative evaluation of WISC-V Integrated performance varies by cognitive domain and can take place at the scale, task, or item level. For example, such an evaluation can involve examining patterns of responses and performance across and within tasks, among items grouped by content, or across items with similar stimulus characteristics or response requirements.

TABLE 11.7. Sample Interpretive Hypotheses for Pairwise Comparisons of Various WISC-V and WISC-V Integrated Subtest Scores

Comparison	Interpretive hypotheses
Vocabulary and Vocabulary Multiple Choice	Vocabulary > Vocabulary Multiple Choice: • Expression of verbal concept formation and semantic knowledge not significantly influenced by expressive problems or retrieval deficits. • Difficulty inhibiting urge to respond to a partially correct response that is of lower quality than another option. • Working memory difficulties create issues with expression of verbal concept formation and semantic knowledge. Vocabulary < Vocabulary Multiple Choice: • Difficulty accessing or retrieving semantic knowledge when no external prompts or cues present. • Decision-making skills bolster expression of verbal concept formation and semantic knowledge. • Expressive language difficulties hamper expression of verbal concept formation and semantic knowledge.
Arithmetic and Arithmetic Process Approach Part A	Arithmetic > Arithmetic Process Approach Part A: • Working memory demands do not substantially reduce expression of quantitative reasoning skills. • Increasing reading demands hamper the expression of quantitative reasoning skills. • Listening comprehension demands do not constrain the expression of quantitative reasoning skills. Arithmetic < Arithmetic Process Approach Part A: • Reducing auditory working memory demands bolsters expression of quantitative reasoning skills. • The child was relatively more fatigued during Arithmetic, due to the additional items administered. • Simplifying sensory processing demands improves expression of quantitative reasoning skills. • Listening comprehension demands constrain the expression of quantitative reasoning skills.
Coding and Coding Copy	Coding > Coding Copy: • Graphomotor skills hamper the expression of processing speed ability. • Associative memory skills are relatively stronger than graphomotor skills (confirm by examining relative performance on Coding Recall). Coding < Coding Copy: • Associative memory skills reduce the expression of processing speed ability. • Graphomotor skills are relatively more well developed than associative memory skills (confirm by examining relative performance on Coding Recall).

Note. From Raiford (2017). Copyright © John Wiley & Sons, Inc. Adapted by permission.

Variability of performance may be due to issues with fatigue, distraction, inattention, persistence, motivation, receptive language, listening comprehension, reading, working memory, or decision-making skills. It could also be due to difficulties with a particular subject matter or area of knowledge. Common performance across subtests with similar types of content, stimulus, or response demands may indicate an area of weakness.

Each domain of the WISC-V and the WISC-V Integrated system provides unique opportunities to discover areas for remediation or growth. For example, the Raiford (2017) book provides grouping systems for error analysis of Verbal Comprehension subtest items that may be examined together to screen for weak content areas (e.g., human health, science, calendar information), as well as sets of Arithmetic, Arithmetic Process Approach, Written Arithmetic, Figure Weights, and Figure Weights Process Approach items that may be related to emerging math score weaknesses.

Appendix 11.1 presents a case study that demonstrates how the WISC-V, WISC-V Integrated, KTEA-3, NEPSY-II, California Verbal Learning Test for Children (CVLT-C), and Behavior Assessment System for Children, Third Edition (BASC-3) can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

APPENDIX 11.1

Illustrative Case Study

Name: Bill D.

School: B Elementary School

Date of birth: 8/25/2006

Age: 10 years, 10 months

Race/ethnicity: European American

Parents: Dennis and Ann D.

Date of testing: 7/1/2017

Date of report: 7/5/2017

REASON FOR REFERRAL

Mrs. D. requested an evaluation of Bill's cognitive strengths and weaknesses, as well as his emotional and behavioral functioning. Mrs. D. reported that Bill shows many signs of attention-deficit/hyperactivity disorder (ADHD). Bill says he has a hard time staying in his seat and paying attention. Specific questions to be answered:

- (Parents:) Does Bill have ADHD?
- (Parents:) What can we do to help Bill stay organized and on track at school?
- (School:) Are there issues that may require school accommodations?

ANSWERS TO REFERRAL QUESTIONS

Does Bill Have ADHD?

Yes. I have diagnosed Bill with ADHD, combined presentation. The diagnosis of this subtype means that he exhibits symptoms of both inattention and hyperactivity/impulsivity. Various educators, Mrs. D., and Bill report these symptoms, which are also observable in objective test results.

What Can We Do to Help Bill Stay Organized and on Track at School?

Pharmacotherapy (medication), working memory training, vigorous exercise, limiting screen time, executive function training, and parent support are all possible courses of action that are recommended to assist in improving school functioning; many of these will need to be implemented outside school. Numerous possible interventions are listed in the "Recommendations" section of this report. Because Bill attends a private parochial school, not all accommodations listed in the "Recommendations" section may be available or feasible.

Are There Issues That May Require School Accommodations?

Bill's diagnosis qualifies him to be recognized as having a disability in one of the categories designated under the Individuals with Disabilities Education Improvement Act of 2004, which governs free public school education: *other health impairment (ADHD)*.

Accommodations for these conditions, if available within the private school environment, are appropriate. Refer to the accommodations listed in the "Recommendations" section. Ultimately, the accommodation plan should be discussed and agreed upon in a follow-up meeting with parents, educators, and Bill himself.

EVALUATION METHODS AND PROCEDURES

Evaluation methods and procedures included the following:

- History and background review
- Parent interview
- Teacher interview
- Child interview
- Behavioral observations
- Review of group achievement testing results
- Psychological testing

Psychological tests administered were as follows (see "Complete Test Data" at the end of this report for full results):

- Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V)
- Wechsler Intelligence Scale for Children—Fifth Edition Integrated (WISC-V Integrated), selected subtests
- Kaufman Test of Educational Achievement, Third Edition (KTEA-3), selected subtests
- NEPSY: A Developmental Neuropsychological Assessment—Second Edition (NEPSY-II), selected subtests
- California Verbal Learning Test—Children's Version (CVLT-C)
- Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)
- Behavior Assessment System for Children, Third Edition (BASC-3):
    - Self-Report Form
    - Parent Rating Scales
    - Teacher Rating Scales

CURRENT SYMPTOMS

Bill's mother and teacher have observed that he shows many present signs of inattention. Both report consistent observations, as his mother helps him with schoolwork at home. He is off task when he is supposed to do homework or chores at home, has trouble completing such tasks, and appears disorganized. He avoids tasks that are effortful. He acts childish, loses track of what he is doing (e.g., because of noises), forgets what he is supposed to do, and lets his mind wander if an adult does not keep him on task. He reportedly has a short attention span, acts without thinking, listens to directions sporadically, and often does not listen carefully or pay attention when being spoken to. He has great trouble concentrating almost all the time and acts confused. He says he often gets in trouble for not paying attention. He is described by his mother and teacher as highly disorganized, with weak planning skills.

Bill has many behaviors consistent with hyperactivity. His teacher and mother report the following: He often interrupts, acts without thinking, is overly active, and fiddles with things. He disrupts other children's activities and schoolwork, has poor self-control, and is unable to slow down. He has difficulty waiting his turn, has trouble keeping his hands to himself, speaks out of turn, and has trouble staying seated during class. He is described as almost always being in motion. He acts out of control at times as well.

Bill shows some signs of social difficulties. He has some difficulty understanding certain social cues. For example, he sometimes becomes angry because he misinterprets lighthearted joking as a personal affront. However, he generally interacts well with the other boys in class and with adults. His classmates find him to be funny, and he uses humor as a method to make friends.

HISTORY OF PRESENTING PROBLEMS

Bill has never been formally evaluated for ADHD. His teacher says that Bill is easily sidetracked and does not focus on class activities. His mother works with him on homework and observes that Bill tries to avoid schoolwork in the evenings by diverting attention to martial arts or piano practice. He has attended his current private school since he was in prekindergarten at age 3. His teachers have consistently reported hyperactivity, but his parents have hoped that the issues would decrease with maturity. His mother reported that he is well liked at the school, despite some misunderstandings with others. His friends enjoy him for his sense of humor and imagination on the playground, although his teacher reports that sometimes their play degenerates into yelling, and that Bill cannot stand to lose at kickball and is sometimes a sore loser.

FAMILY HISTORY

Mrs. D. is a stay-at-home mother. Mr. D. is a sales manager. Mr. D. has a bachelor's degree in business, and his mother finished 3 years of college before becoming pregnant; she plans to finish her degree when Bill is able to drive. Bill has one brother, age 6. His 6-year-old brother also attends school at B Elementary and is in the first grade. Bill says he has a good relationship with his mother and father. His mother reports that apart from the difficulties with homework, getting Bill to finish chores can be challenging, but he is a sweet, loving boy.

MEDICAL/DEVELOPMENTAL HISTORY

Mrs. D. reported that her pregnancy with Bill was normal, but that he was born at 34½ weeks because her water broke early. The labor and delivery were unremarkable, but Bill spent 6 days in the NICU due to electrolyte imbalances and low bilirubin levels. He was released afterward and followed by his pediatrician for a short time, during which the issues resolved. No other health or medical concerns occurred after the birth.

Her mother reported that Bill always has been large and tall for his age. He likes sports and athletics, but she describes him as clumsy and easily frustrated (e.g., with waiting his turn or losing). Bill started walking late (at 19 months), but was saying simple words at 10 months and has normal language development. Bill was not toilet-trained until sometime after age 3.

Bill was previously diagnosed with developmental coordination disorder and fine and gross motor delays by the occupational therapist who assessed him 2 years ago. He underwent treatment for these issues and for a reflex (asymmetric tonic neck reflex) that was not integrated. He continues to attend occupational therapy once a week for fine and gross motor delays, and to work on his handwriting due to motor dysgraphia.

Bill is not presently prescribed any medications. His parents indicated that they are open to the possibility of medication to treat him if necessary. He has had recent vision and hearing screenings, which showed no abnormalities.

ACADEMIC HISTORY AND STATUS

When Bill was age 1½, he began to recognize numbers, letters, colors, and shapes. He began reading well at age 3.

Bill has just completed the fifth grade. Mrs. D. stated that his lowest grade this past year was a B– (80) in math, and that he reads far above grade level. His handwriting is somewhat messy. He has an excellent vocabulary and his spelling and grammar skills are above average. These reports were confirmed with review of school records. His group achievement scores indicate that he is functioning at or above grade level in all subjects.

Bill wants to be a school teacher and a martial arts instructor in the future. His parents anticipate that he will attend college after high school.

PATIENT'S STRENGTHS, COPING MECHANISMS, AND AVAILABLE SUPPORT SYSTEMS

Mrs. D. describes Bill as funny, smart, creative, caring, and kind toward others. His teacher describes him as a bright child with a great sense of humor who enjoys playing with the other children. She indicated that Bill likes to be with others and focuses well when he is in a one-on-one situation. She describes Bill as polite to adults and says he wants to do what is right.

MENTAL STATUS AND BEHAVIORAL OBSERVATIONS

Bill arrived on time and was oriented to person, place, time, and purpose of the testing. He was accompanied by Mrs. D. He was groomed neatly and dressed appropriately, and he was taller and heavier than his same-age peers. Activity level was higher than typical of other children his age, and eye contact was inconsistent. He turned away from testing on a number of occasions, and he sometimes attempted to change the subject by discussing comic books and sports and recounting movies he had recently seen. His speech revealed very strong vocabulary. His reported mood was "good," and his affect was generally mood-congruent. However, he became impatient and irritable after sitting for approximately 30 minutes; therefore, he was given multiple breaks.

Bill showed reactions to perceived success and failure, seeming happy and proud when he thought his performance was good, and irritated and impatient when he thought he was doing poorly. He asked if his answers were correct on multiple occasions. He verbally coached himself through several tasks involving visual stimuli. He required some corrective feedback and prompting, but quickly learned the rules of tasks and how they were performed when he was attentive. Given the effort and level of engagement Bill displayed in the session, the results of the assessment are likely to reflect accurate estimates of his cognitive functioning.

TEST RESULTS AND INTERPRETATION

Cognitive Functioning

Intellectual Ability

Bill's performance on cognitive measures suggests that his overall intellectual ability is in the very high range compared with that of other children his age. His Full Scale IQ is high average, but the Nonmotor General Ability Index, an indicator of ability that reduces the contribution of motor skills and of working memory and processing speed, indicates that Bill is functioning in the very high range. Given his known issues with motor functioning and attentional difficulties, this finding is not surprising. Because the Nonmotor General Ability Index eliminates motor production demands and minimizes the contribution of working memory and processing speed, the best estimate of his global ability is the Nonmotor General Ability Index. Results suggest relatively weaker working memory and motor production skills relative to higher-order reasoning. In the classroom, these weaknesses might be manifested as appearing to forget or failing to track complex information that is being presented; being slow to respond; having difficulty with tasks that require use of hands to manipulate objects; and messy handwriting.

In large studies, out of all the scores on cognitive tests, overall intellectual ability does the best job at predicting children's school success. However, it doesn't always tell us everything we need to know about a child's individual strengths and needs. An individual's cognitive picture is usually better and more deeply understood when narrower areas of ability are also considered.

Language

Speech

Bill responded readily to items that required expressive responses. He was articulate and highly verbal. He elaborated sufficiently.

Verbal Comprehension and General Verbal Knowledge

Bill's ability to access, apply, and express knowledge he has gained about words and their meanings and to reason with verbal material is in the extremely high range relative to that of other children his age, and is a personal strength relative to his other cognitive abilities. When acquired knowledge and social/practical knowledge are taken into account as well, Bill's verbal crystallized abilities are slightly lower (but still in the very high range). These results are not surprising, given reports of inattention in class and some social skill difficulties. Providing multiple choices for the social/practical knowledge questions, so that he was merely required to recognize the correct responses rather than to recall and express them, did not result in substantial improvement; he was able to respond slightly more accurately to just two more questions under these conditions.

Because he has very strong verbal skills and a well-developed vocabulary, teachers may expect him to perform other tasks more easily or grasp information more readily and quickly than he does in the classroom because of his inattention and hyperactivity.

Listening Comprehension

The kind of listening comprehension that students must do in school involves understanding relatively formal speech, rather than casual speech. Bill's listening comprehension abilities are average compared with those of other children his age. This finding suggests that Bill grasps the meaning of what is said in the classroom at a level that is typical of his peers. However, this skill is lower than expected relative to his intellectual abilities.

Associative Verbal Fluency

Bill's ability to call to mind and verbalize categorically associated information (associative verbal fluency) is high average relative to that of his same-age peers. However, when semantic constraints are applied (naming objects that begin with a specific letter rather than objects in a category), his verbal fluency is in the average range relative to other children his age. This decrease relative to his initial performance probably reflects the increased demands on executive functions and strategy inherent in the initial-letter condition.

Visual–Spatial Processing

Visual–spatial processing is the ability to evaluate visual details, to understand relationships of visual parts in space, and to use such information to assemble a geometric design that matches a model (either pictured or real). Sometimes the expression of this ability is tested through manipulation of real objects. Bill's visual–spatial processing ability is in the low average range compared with that of other children his age and is a personal weakness in relation to his other cognitive abilities. Reducing the influence of speedy performance on the task did not result in substantive improvement.

Bill frequently stopped attempting to solve problems requiring visual–spatial abilities, such as constructing a design with blocks or selecting pieces corresponding to a completed puzzle. However, his difficulty in this area was greater on a task involving physical (motor) manipulation of components of a construction, rather than mental imagery. This observation is not surprising because of his known motor skill delay. He is better able to quickly recognize correct solutions among multiple choices than to quickly construct them. His estimated visual–spatial ability is in the average range when motor production requirements are eliminated.

Reasoning and Problem Solving

Reasoning involves detecting and applying the underlying rules or relationships that define how objects or ideas are understood as a group or work together. Bill's visual reasoning ability is in the high average range compared with that of other children his age. He is better able to engage in proper reasoning and problem solving if tasks involve relationships among meaningful objects or verbal material rather than abstract shapes, and he has slightly lower performance on tasks that involve quantitative reasoning. It was noted that he had greater difficulty on problems requiring multiple operational steps or multiplication than on items merely requiring simple addition or subtraction.

Giving more time and allowing him to use pencil and paper, especially when math is involved in the problem solving, should help Bill to better express verbal quantitative reasoning relative to his own prior performance as well as in relation to his same-age peers. If the problem-solving load is reduced to requiring only simple responses to written math problems, his performance improves further.

Learning and Memory

Learning and memory were assessed to examine their role in, and implications for, Bill's future academic success. The two are closely related to each other and are very important to school achievement. Bill's performance on a broad measure of long-term storage and retrieval, which involves tasks requiring accurate and fluent retrieval of overlearned associations and accurate retrieval of new associations, is in the high average range.

Working Memory

Working memory is the ability to take in, keep, and manipulate information in one's awareness to get some type of output that can be expressed. Bill's working memory ability is in the low average range relative to that of other children his age, and is an intrapersonal weakness relative to his overall cognitive abilities. His auditory working memory (i.e., verbal stimuli and responses) is in the average range and is better than his visual working memory (i.e., visual stimuli and spatial locations), which is in the very low range. His working memory performance was consistent across items rather than variable, which suggests that his attention was consistent during these particular tasks.

Associative Memory

Bill's visual–verbal associative memory, or the ability to form new associations between symbols and meanings, is in the high average range relative to that of other young people Bill's age. His performance suggests uniform ability when he is recalling associations learned within the past few seconds or minutes, or when trying to recall or recognize them half an hour later.

Short-Term and Long-Term Auditory Memory

Relative to that of his peers, Bill's ability to encode and recall details of organized verbal material and lists of items from a few categories (short- and long-term auditory/verbal memory) is very high.

Short-Term and Long-Term Visual Memory

Bill's short- and long-term visual memory for visual abstract content and spatial locations is in the low average range.

Rapid Automatic Naming

Rapid automatic naming involves recognizing and recalling overlearned information, like letters and numbers or quantities, as efficiently as possible. Bill's ability to recognize, recall, and recite letters and numbers quickly is in the high average range relative to that of other children his age. His ability to rapidly recognize and name quantities is slightly less efficient, and is in the average range compared with that of others his age. Overall, his rapid automatic naming is in the high average range.

Cognitive Speed

Processing Speed

Processing speed is the ability to make speedy and accurate judgments about visual information and act on those judgments. Slow processing speed can lower academic performance, because information tends to be lost if it is not processed quickly. Bill's processing speed is in the high average range relative to other children his age. His score on a subtest that required him to quickly associate and write codes paired with numbers in a key was lower than his score on a task that required him to scan rows for symbols matching one of two target symbols. Results suggest that these difficulties are due to low graphomotor speed (the quickness with which one can reproduce symbols in written format) rather than weak associative memory for visual material.

Rapid Automatic Naming

Rapid automatic naming is also sometimes thought of as an aspect of cognitive speed and has been discussed above under "Learning and Memory."

Attention and Executive Functioning

Attention and executive functioning were evaluated to provide objective results to examine, along with the subjective reports (from Bill, his mother, and his educators) of Bill's issues with hyperactive and inattentive behavior. Objective findings show low cognitive flexibility and self-monitoring, low selective and sustained attention, and slowed responding when cognitive flexibility is required. These results are consistent with observations in the testing session and by educators and Bill himself. In other words, he has difficulty adapting to changing expectations.

Because Bill's attention and other executive functions are impaired, performance in other areas may improve after these issues are addressed.

Emotional and Behavioral Functioning

Bill has clinically significant attention and hyperactive/impulsive symptoms, according to subjective informants and objective test findings. These symptoms were also observed in the testing session. Bill's symptoms and test results are consistent with a diagnosis of ADHD, combined presentation (both inattentive and hyperactive/impulsive symptoms).

Social Perception and Skills

Because children with attention problems sometimes have deficits in social skills and perception, and informants indicate that Bill may be struggling in this area, this area was assessed. Objective testing results suggest that he has some difficulties reading others well and identifying emotional expressions in others (particularly misinterpreting facial expressions corresponding with neutral emotion, angry feelings, and disgust). These problems could lead to difficulties in the classroom or with peers.

For example, Bill may not distinguish nonverbal cues and facial expressions of teachers as readily as do his peers; if an instructor or his mother appears annoyed, he may not readily recognize that cue. He may also interpret neutral faces as conveying negative feelings or disapproval. In addition, he is likely to misinterpret nonverbal cues that most children his age comprehend; such cues are very valuable to provide ongoing feedback about performance and behaviors.

Bill's pragmatic language, or use of language for social purposes, was assessed through a measure completed by his mother as informant. He has some social communication deficits in the areas of awareness of voice pitch, loudness, and rhythm and their influence on meaning; and of reading and using body language. These difficulties are likely to interfere with his ability to respond appropriately on a consistent basis with peers and teachers.

SUMMARY AND DIAGNOSTIC IMPRESSIONS

Bill qualifies for a diagnosis of ADHD, combined presentation (both inattentive and hyperactive/impulsive type), moderate. Bill himself reports these symptoms, as do his educators and mother. These symptoms are sometimes observed in a one-on-one situation and are very apparent in the classroom. He has several of the symptoms of inattention required for this diagnosis. These problems cause him difficulty in the classroom and impair his academic and social functioning. Given his level of intelligence, he is likely to be capable of greater academic gains and social functioning if his ADHD is treated.

RECOMMENDATIONS

Medical and Physical Interventions

1. Medication to alleviate Bill's ADHD is strongly recommended. Medication could make an enormous difference in his school performance. Obtain an evaluation and share this report with the prescribing professional. (Provide referrals.)
2. The most effective treatments for ADHD combine medication with therapeutic approaches.
3. Physical exercise is critical, both for health and for reducing Bill's impulsive ADHD symptoms. The best approach is to combine medication and vigorous physical exercise. An hour a day at least 5 times per week is recommended.
4. Bill's screen time (smartphone, TV, computer, tablets) should be kept to a minimum unless it is required for school.

Therapy/Counseling/Training

1. Obtain executive function coaching for Bill. (Provide referrals.) This training is important to help him develop organizational skills that will be increasingly important in middle school. It is also important in order for his mother not to be the sole homework support; to preserve the mother–son relationship, Bill needs to learn strategies to be able to monitor and regulate his own homework completion.
2. Social skills groups or social language groups would probably be valuable in developing Bill's social communication skills. (Provide referrals.) Work with him on understanding figures of speech, turns of phrase, and humor. Idiom dictionaries are widely available on the Internet.
3. Obtain working memory training. Computer-based solutions for attention problems related to working memory are available that combine cognitive neuroscience with close professional support to deliver substantial and lasting benefits. The treatment includes easy-to-use software. It has been empirically shown to improve attention, working memory, and academic performance in multiple studies. The training occurs over a 5-week period. It includes pre-evaluation, daily training sessions (in the privacy of one's own home via computer), weekly coaching sessions via video chat or phone, and a wrap-up. I will provide the appropriate literature and information.

Home Interventions

1. Teach Bill to guide himself through problems and to verbally list steps to follow.
2. Instruct him in strategies to organize himself at home and in transitions between environment. In particular, teach use of lists and personal planners.
3. Communicate frequently with teachers about school activities, equipment needed, and homework. Use the school's app daily to monitor grades and assignments.
4. Teach memory aids, such as verbal mediation or rehearsal and mnemonic strategies, for use with schoolwork.

School Accommodations

1. Allow Bill additional time to complete school tests and standardized tests (50%).
2. Minimize distractions during learning activities, and consider preferential seating arrangements close to the teacher and away from visual distractions and movement (e.g., doors opening and closing, windows).
3. Closely monitor Bill's math performance. There are some indications that math may present challenges in the future. After Bill's ADHD is treated, obtain academic achievement assessment in the mathematics domain to better understand his math skills.
4. Explore options to allow Bill activity so that he can focus better when seated. These may include a seating disk, desk bands, a standing-desk option, an under-desk cycle, and/or fidget toys.
5. Bill's auditory working memory and auditory immediate and delayed memory are superior to his visual working memory and visual immediate and delayed memory. Presentation of information in an auditory format to supplement visual material may facilitate his learning. Provide copies of images, tables, or charts used in lecture or class. Use visual materials sparingly, especially if they are abstract. Meaningful visuals (e.g., pictures that can be semantically encoded) will be more helpful than abstract ones.

6. Have Bill repeat back the directions and parts of an assignment or task, or have a peer with verbal strengths do so in front of him if they are in a group situation, to confirm understanding.
7. Teach reading of visual displays, drawings, diagrams, and charts, and coach Bill in how to talk them through. Encourage use of puzzles and other spatial games and toys.
8. For in-class assignments, pair all students in the class and enlist them to ensure that they have completed all necessary steps for in-class assignments.
9. Provide a written schedule on the board daily, crossing off each portion of the day as it occurs.
10. Make redirection discreet, clear, and concise. Use quiet auditory redirection or discreet visual direction (e.g., move near Bill) if possible.
11. Provide Bill with teacher notes or outlines in advance of or during lectures.

Thank you for the opportunity to work with Bill and your family. Please let me know if I can be of further assistance.

COMPLETE TEST DATA

Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) and WISC-V Integrated

Subtest Score Summary

Index	Subtest name	Scaled score
Verbal Comprehension	**Similarities**	**17**
	Vocabulary	**16**
	(Information)	13
	(Comprehension)	11
	Comprehension Multiple Choice[a]	11
Visual Spatial	**Block Design**	**6**
	Block Design Multiple Choice[a]	10
	Visual Puzzles	11
Fluid Reasoning	**Matrix Reasoning**	**14**
	Figure Weights	**10**
	Figure Weights Process Approach[a]	10
	(Picture Concepts)	13
	(Arithmetic)	10
	Arithmetic Process Approach Part A[a]	12
	Arithmetic Process Approach Part B[a]	12
	Written Arithmetic[a]	12
Working Memory	**Digit Span**	**9**
	Picture Span	4
	Spatial Span[a]	2
	(Letter–Number Sequencing)	8
	Sentence Recall[a]	8
Processing Speed	**Coding**	**11**
	Coding Copy[a]	11
	Symbol Search	16

Index	Subtest name	Scaled score
Complementary subtests		
Naming Speed	Naming Speed Literacy	116
	Naming Speed Quantity	107
Symbol Translation	Immediate Symbol Translation	113
	Delayed Symbol Translation	114
	Recognition Symbol Translation	114

Note. Subtests used to derive the Full Scale IQ are given in boldface. Secondary subtests are given in parentheses.

Results on most subtests are reported as scaled scores. Average scaled score is 10. About 68% of scores fall between 7 and 13. About 95% fall between 4 and 16. Complementary subtest scores are reported as standard scores. Average standard score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

[a]WISC-V Integrated subtest.

Composite Score Summary

Composite	Composite score	Percentile rank	95% confidence interval	Qualitative description
Verbal Comprehension Index (VCI)	**131**	**98.8**	**126–133**	**Extremely high**
Verbal (Expanded Crystallized) Index (VECI)	123	94	119–126	Very high
Visual Spatial Index (VSI)	**85**	**16**	**82–89**	**Low average**
Nonmotor Visual Spatial Index (NMVSI)[a]	102	57	99–105	Average
Fluid Reasoning Index (FRI)	**114**	**83**	**110–116**	**High average**
Expanded Fluid Index (EFI)	109	72	105–112	Average
Quantitative Reasoning Index (QRI)	100	50	97–103	Average
Working Memory Index (WMI)	**80**	**9**	**78–83**	**Low average**
Auditory Working Memory Index (AWMI)	100	50	97-103	Average
Visual Working Memory Index (VWMI)[b]	72	3	70-75	Very low
Symbol Translation Index (STI)	117	86	114–119	High average
Storage and Retrieval Index (SRI)	115	84	112–117	High average
Processing Speed Index (PSI)	**113**	**82**	**110–115**	**High average**
Naming Speed Index (NSI)	111	77	108–113	High average
Cognitive Proficiency Index (CPI)	96	38	94–99	Average
Full Scale IQ	*116*	*50*	*113–118*	*High average*
Nonmotor General Ability Index (NMGAI)[a]	*123*	*94*	*120–125*	*Very high*

Note. Boldface, primary index score. Italics, global composite score. Average composite score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

[a]WISC-V Integrated *Essentials Composites of WISC-V Integrated Assessment* (Raiford, 2017).

[b]WISC-V Integrated published composite.

Process Scores

Score	Scaled score
Block Design No Time Bonus	6
Digit Span Forward	10
Digit Span Backward	7
Digit Span Sequencing	8
Spatial Span Forward[a]	1
Spatial Span Backward[a]	2

Note. These process scores are reported as scaled scores. Average scaled score is 10. About 68% of scores fall between 7 and 13. About 95% fall between 4 and 16.

[a] WISC-V Integrated score.

Longest Span Scores

Score	Base rate
Longest Digit Span Forward	50
Longest Digit Span Backward	74
Longest Digit Span Sequence	77
Longest Letter–Number Sequence	74
Longest Spatial Span Forward[a]	100
Longest Spatial Span Backward[a]	100

Note. For the longest span scores, higher scores represent weaker performance, whereas lower scores represent stronger performance. Average base rate = 50; range = 0–100.

[a] WISC-V Integrated score.

WISC-V Integrated Coding Recall Scores

Score name	Percentile rank
Cued Symbol Recall	56
Free Symbol Recall	58
Cued Digit Recall	66
Pairing	62

Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Form A

Subtest	Standard score	95% confidence interval	Percentile rank	Descriptive category
Listening Comprehension	108	104–111	70	Average

Note. All KTEA-3 scores are reported as standard scores. Average standard score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

NEPSY: A Developmental Neuropsychological Assessment, Second Edition (NEPSY-II)

Domain	Score name	Scaled score	Percentile rank
Attention and Executive Functioning	Animal Sorting Combined	4	
	Animal Sorting Total Correct Sorts	6	
	Auditory Attention Total Correct	13	
	Auditory Attention Commission Errors	—	>75
	Auditory Attention Omission Errors	—	>75
	Auditory Attention Inhibitory Errors	—	>75
	Auditory Attention Combined	13	
	Response Set Total Correct	3	
	Response Set Commission Errors	—	>75
	Response Set Omission Errors	—	6–10
	Response Set Inhibitory Errors	—	>75
	Response Set Combined	10	
	Inhibition Total Errors	12	
	Inhibition Naming Completion Time	9	
	Inhibition Inhibition Completion Time	7	
	Inhibition Switching Completion Time	10	
	Inhibition Naming Combined	13	
	Inhibition Inhibition Combined	9	
	Inhibition Switching Combined	10	
Social Perception	Theory of Mind	—	26–50
	Affect Recognition	8	
	Affect Recognition Neutral Errors	—	<2
	Affect Recognition Angry Errors	—	11–25
	Affect Recognition Happy Errors	—	>75
	Affect Recognition Sad Errors	—	>75
	Affect Recognition Fear Errors	—	>75
	Affect Recognition Disgust Errors	—	2–5
Memory and Learning	Memory for Designs Immediate	6	
	Memory for Designs Delayed	6	
	Narrative Memory Free and Cued Recall Total	16	
	Narrative Memory Free Recall Total	15	
Language	Word Generation Semantic Total	13	
	Word Generation Initial Letter Total	10	

Note. Average scaled score is 10. About 68% of scores fall between 7 and 13. About 95% fall between 4 and 16. Percentile ranks less than or equal to 15% are considered unusual and potentially clinically relevant.

California Verbal Learning Test—Children's Version (CVLT-C)

Trials 1–5 Free Recall Total Correct: T score = 70; 95% confidence interval = 48–57.

Average T score is 50. About 68% of scores fall between 40 and 60. About 95% fall between 30 and 70.

Trial 1 Free Recall Total Correct: $z = 2.5$
Trial 5 Free Recall Total Correct: $z = 2.5$
Short-Delay Cued Recall Total Correct: $z = 2.0$
Long-Delay Free Recall Total Correct: $z = 1.5$
Long-Delay Cued Recall Total Correct: $z = 1.5$

Average z score is 0. About 68% of scores fall between –1 and 1. About 95% fall between –2 and 2.

Clinical Evaluation of Language Functioning—Fifth Edition (CELF–5), Pragmatics Activities Checklist

Pragmatics Profile scaled score: 6

Average scaled score is 10. About 68% of scores fall between 7 and 13. About 95% fall between 4 and 16.

Behavior Assessment System for Children, Third Edition (BASC-3)

Self-Report of Personality (Informant: Child), General Gender-Specific Norm Group

Composite Score Summary

Composite	*T* score	Percentile rank	90% confidence interval
School Problems	44	35	38–50
Internalizing Problems	50	50	47–53
Inattention/Hyperactivity	80	99	75–85
Emotional Symptoms Index	56	78	53–59
Personal Adjustment	50	50	45–55

Note. For this and all other BASC-3 tables, the average *T* score is 50. About 68% of scores fall between 40 and 59. About 95% fall between 30 and 69.

Scale Score Summary

Scale	*T* score	Percentile rank	90% confidence interval
Attitude to School	45	39	37–53
Attitude to Teachers	44	36	37–51
Atypicality	45	40	39–51
Locus of Control	57	79	49–65
Social Stress	44	33	38–50
Anxiety	58	60	55–61
Depression	51	70	45–57
Sense of Inadequacy	56	79	49–63
Attention Problems	85	99	78–92
Hyperactivity	71	97	65–77
Relations with Parents	54	55	48–60
Interpersonal Relations	48	28	41–55
Self-Esteem	58	86	51–65
Self-Reliance	41	18	32–50

Teacher Rating Scales (Informant: Teacher), General Gender-Specific Norm Group

Composite Score Summary

Composite	*T* score	Percentile rank	90% confidence interval
Externalizing Problems	61	86	57–65
Internalizing Problems	50	50	78–88
School Problems	58	70	71–79
Behavioral Symptoms Index	73	90	76–82
Adaptive Skills	40	38	32–38

Scale Score Summary

Scale	*T* score	Percentile rank	90% confidence interval
Hyperactivity	84	99	79–89
Aggression	45	33	39–51
Conduct Problems	51	69	45–57
Anxiety	56	71	53–58
Depression	40	38	38–43
Somatization	50	50	47–53
Attention Problems	77	99	73–81
Learning Problems	50	50	42–58
Atypicality	50	50	42–58
Withdrawal	35	20	32–39
Adaptability	48	37	42–54
Social Skills	41	22	37–45
Leadership	35	8	30–40
Study Skills	35	10	30–40
Functional Communication	50	50	47–53

Content Scales

Scale	*T* score	Percentile rank	90% confidence interval
Anger Control	52	77	46–58
Bullying	45	39	40–50
Developmental Social Disorders	50	50	47–53
Emotional Self-Control	50	50	47–53
Executive Functioning	73	98	69–77
Negative Emotionality	52	71	47–57
Resiliency	50	50	47–53

Parent Rating Scales (Informant: Mother), General Gender-Specific Norm Group

Composite Score Summary

Composite	*T* score	Percentile rank	90% confidence interval
Externalizing Problems	58	70	71–79
Internalizing Problems	73	90	76–82
Behavioral Symptoms Index	40	38	32–38
Adaptive Skills	50	50	45–55

Scale Score Summary

Scale	*T* score	Percentile rank	90% confidence interval
Hyperactivity	62	88	55–69
Aggression	48	55	41–55
Conduct Problems	49	51	43–55
Anxiety	56	71	53–58
Depression	40	38	38–43
Somatization	50	50	47–53
Atypicality	50	50	42–58
Withdrawal	35	20	32–39
Attention Problems	77	99	73–81
Adaptability	48	37	42–54
Social Skills	41	22	37–45
Leadership	55	70	52–58
Activities of Daily Living	35	10	30–40
Functional Communication	50	50	47–53

Content Scales

Scale	*T* score	Percentile rank	90% confidence interval
Anger Control	52	77	46–58
Bullying	45	39	40–50
Developmental Social Disorders	50	50	47–53
Emotional Self-Control	50	50	47–53
Executive Functioning	73	98	69–77
Negative Emotionality	52	71	47–57
Resiliency	50	50	47–53

REFERENCES

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Boxer, O., Jackson, K., & Kohlman, S. (2014). B-31 problem solving weaknesses in children with nonverbal learning differences. *Archives of Clinical Neuropsychology, 29*(6), 547.

Cardoso, C. O., Branco, L. D., Cotrena, C., & Fonseca, R. P. (2015). Correlational analysis of performance in executive function tasks after stroke. *Psychology and Neuroscience, 8*(1), 56–65.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Demakis, G. J., Sawyer, T. P., Fritz, D., & Sweet, J. J. (2001). Incidental recall on WAIS-R Digit Symbol discriminates Alzheimer's and Parkinson's diseases. *Journal of Clinical Psychology, 57*(3), 387–394.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment.* Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459–483). New York: Guilford Press.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Floyd, R. G., & Kranzler, J. H. (2012). Processing approaches to interpretation of information from cognitive ability tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 497–525). New York: Guilford Press.

Gagnon, L. G., & Belleville, S. (2011). Working memory in mild cognitive impairment and Alzheimer's disease: Contribution of forgetting and predictive value of complex span tasks. *Neuropsychology, 25*(2), 226–236.

Goldstein, F. C., & Green, R. C. (1995). Assessment of problem solving and executive functions. In R. L. Mapou & J. Spector (Eds.), *Clinical neuropsychological assessment: A cognitive approach* (pp. 49–81). New York: Plenum Press.

Groeger, J. A., Field, D., & Hammond, S. M. (1999). Measuring memory span. *International Journal of Psychology, 34,* 359–363.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.

Halleland, H. E., Sorensen, L., Posserud, M.-B., Haavik, J., & Lundervold, A. J. (2015). Occupational status is compromised in adults with ADHD and psychometrically defined executive function deficits. *Journal of Attention Disorders*. [Epub ahead of print]

Hoffmann, N., Donders, J., & Thompson, E. H. (2000). Novel learning abilities after traumatic head injury in children. *Archives of Clinical Neuropsychology, 15,* 47–58.

Joy, S., Fein, D., Kaplan, E., & Freedman, M. (1999). Information multiple choice among healthy older adults: Characteristics, correlates, and clinical implications. *The Clinical Neuropsychologist, 13,* 48–53.

Joy, S., Kaplan, E., & Fein, D. (2003). Digit Symbol–incidental learning in the WAIS-III: Construct validity and clinical significance. *The Clinical Neuropsychologist, 17,* 182–194.

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. J. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 129–167). Washington, DC: American Psychological Association.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). *Wechsler Intelligence Scale for Children—Third Edition as a process instrument.* San Antonio, TX: Psychological Corporation.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V.* Hoboken, NJ: Wiley.

Kramer, J. H., Knee, K., & Delis, D. C. (2000). Verbal memory impairments in dyslexia. *Archives of Clinical Neuropsychology, 15,* 83–93.

Kreiner, D. S., & Ryan, J. J. (2001). Memory and motor skill components of the WAIS-III Digit Symbol–Coding subtest. *The Clinical Neuropsychologist, 15,* 109–113.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Lichtenberger, E. O., & Kaufman, A. S. (2013). *Essentials of WAIS-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Mainela-Arnold, E., Misra, M., Miller, C., Poll, G. H., & Park, J. S. (2012). Investigating sentence processing and language segmentation in explaining children's performance on a sentence-span task. *International Journal of Language and Communication Disorders, 47*(2), 166–175.

Mayfield, A., Reyes, A., Mayfield, J., & Allen, D. (2014). C-44 improvement in executive function following traumatic brain injury in children. *Archives of Clinical Neuropsychology, 29*(6), 590.

McCloskey, G. (2009). The WISC-IV Integrated. In D. P. Flanagan & A. S. Kaufman, *Essentials of WISC-IV assessment* (2nd ed., pp. 310–467). Hoboken, NJ.: Wiley.

McCloskey, G., & Maerlender, A. (2005). The WISC–IV Integrated. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 101–149). Boston: Elsevier Academic Press.

McLean, R. L., Johnson, H. A., Zimak, E., Joseph, R. M., & Morrow, E. M. (2014). Executive function in probands with autism with average IQ and their unaffected first-degree relatives. *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(9), 1001–1009.

Milberg, W. P., Hebben, N., & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 65–80). New York: Oxford University Press.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D., & Jones, A. M. (2016). Interpreting the WISC-V from Dan Miller's integrated school neuropsychological/Cattell–Horn–Carroll model. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 459–492). Hoboken, NJ: Wiley.

Raiford, S. E. (2017). *Essentials of WISC-V Integrated assessment.* Hoboken, NJ: Wiley.

Raiford, S. E., Drozdick, L. W., Zhang, O., & Zhou, X. (2015). *Expanded index scores* (WISC-V Technical Report No. 1). Bloomington, MN: Pearson. Retrieved from *http://images.pearsonclinical.com/images/assets/WISC-V/WISC-VTechReport1_FNL_v3.pdf.*

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Author.

Sattler, J. M., Dumont, R., & Coalson, D. L. (2016). *Assessment of children: WISC-V and WPPSI-IV.* San Diego, CA: Jerome M. Sattler.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schroeder, P. J. (2014). The effects of age on processing and storage in working memory span tasks and reading comprehension. *Experimental Aging Research, 40*(3), 308–331.

Service, E., & Maury, S. (2015). Differential recall of derived and inflected word forms in working memory: Examining the role of morphological information in simple and complex working memory tasks. *Frontiers in Human Neuroscience, 8*(1064), 1–16.

Smyth, M. M., & Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 45A,* 479–501.

Wechsler, D. (1944). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D., & Kaplan, E. (2015). *Wechsler Intelligence Scale for Children—Fifth Edition Integrated.* Bloomington, MN: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., et al. (2004). *Wechsler Intelligence Scale for Children—Fifth Edition Integrated.* San Antonio, TX: Psychological Corporation.

Werner, H. (1937). Process and achievement: A basic problem of education and developmental psychology. *Harvard Educational Review, 7,* 353–368.

CHAPTER 12

The Kaufman Assessment Battery for Children—Second Edition and KABC-II Normative Update

Lisa Whipple Drozdick
Jennie Kaufman Singer
Elizabeth O. Lichtenberger
James C. Kaufman
Alan S. Kaufman
Nadeen L. Kaufman

This chapter provides an overview of the Kaufman Assessment Battery for Children—Second Edition (KABC-II; A. S. Kaufman & Kaufman, 2004a) and describes the KABC-II Normative Update (KABC-II NU; A. S. Kaufman & Kaufman with Drozdick & Morrison, 2018). The presentation of the KABC-II and KABC-II NU covers the following topics: theory and structure, description of subtests, administration and scoring, psychometric properties, interpretation, and clinical applications. A description of the normative update sample is also included in the chapter, and a special section on the innovations in measures of cognitive assessment is provided for the KABC-II. An illustrative case report is presented in an appendix to the chapter, to exemplify the KABC-II in practice.

THEORY AND STRUCTURE

Structure and Organization

The KABC-II (A. S. Kaufman & Kaufman, 2004a) measures the processing and cognitive abilities of children and adolescents between the ages of 3 years, 0 months and 18 years, 11 months (3:0 and 18:11). Like the original Kaufman Assessment Battery for Children (K-ABC; A. S. Kaufman & Kaufman, 1983), the KABC-II is an individually administered, theory-based, clinical instrument that is used worldwide (e.g., Malda, van de Vijver, Srinivasan, Transler, & Sukumar, 2010; Vannetzel, 2010). It has been translated into multiple languages and adapted for use in numerous countries, such as France, Belgium, Italy, Germany, Austria, Switzerland, South Korea, and Japan (e.g., A. S. Kaufman & Kaufman, 2008; A. S. Kaufman, Kaufman, Melchers, & Melchers, 2014; A. S. Kaufman, Kaufman, & Publication Committee of Japanese Version of KABC-II, 2013). The KABC-II also provides examiners with a Nonverbal scale, composed of subtests that may be administered in pantomime and responded to motorically, to permit valid assessment of children who have hearing impairments, limited English proficiency, and so forth.

Starting with the original K-ABC, the Kaufman approach to assessment has become popular internationally. For example, in Germany, the K-ABC "has become 'the' assessment tool for preschoolers, and . . . is widely used in all areas of assessment, including child psychiatry and early remediation facilities" (P. Melchers, personal communication, October 29, 2003). The K-ABC and KABC-II likewise have been popular in Japan for two decades (Fujita et al., 2011; Ishikuma, Matsuda, Fujita, & Ueno, 2016; Ishikuma, Shinohara, & Nakao, 2007). And, notably, foreign versions of the

K-ABC and KABC-II are more than translations from one language to another, but incorporate cultural and societal differences as well. For example,

> When the Japanese version of the KABC-II was developed, preparation of the developmental stage was more challenging. Since there were no individualized achievement tests in Japan, the Japanese KABC-II was expanded to include achievement scales: Vocabulary, Reading, Writing, and Arithmetic (A. S. Kaufman et al., 2013). . . . When translating and adopting the foreign tests such as Wechsler tests and KABC-II for Japanese children, regardless of the degree of adaptation, we "create" the tests to be useful to understand the child's intelligence and find the best way to help the learning process for Japanese children. (T. Ishikuma, personal communication, December 20, 2016)

The KABC-II is grounded in a dual theoretical foundation: Luria's (1966, 1970, 1973) neuropsychological model, featuring three blocks or functional units, and the Cattell–Horn–Carroll (CHC) approach to categorizing specific cognitive abilities (Carroll, 1993; Flanagan, McGrew, & Ortiz, 2000; Horn & Noll, 1997; see also Schneider & McGrew, Chapter 3, this volume). The dual theoretical model has been seen as a notable strength of the KABC-II (Bain & Gray, 2008; Lichtenberger, 2005), but, interestingly, it has been perceived as both an asset and a limitation by a reviewer of the French KABC-II: Vannetzel (2010) states that

> the use of multifold perspectives is an essential point from the view of epistemology of the psychological assessment and of precision in terms of practice and level of formation. This plural paradigm, very risky in regards to the theoretical pertinence that it demands from the practitioner, illustrates rather well the "new world" in which psychometrics entered at the turn of the twenty-first century. (p. 233)

The KABC-II includes both Core and Expanded Batteries, with only the Core Battery needed to yield a child's scale profile. The KABC-II Expanded Battery offers supplementary subtests to increase the breadth of the constructs measured by the Core Battery and to follow up hypotheses. Administration time for the Core Battery takes about 30–70 minutes, depending on the child's age and whether the examiner administers the CHC model of the KABC-II or the Luria model. One of the features of the KABC-II is the flexibility it affords the examiner in determining the theoretical model to administer to each child.

When interpreted from the Luria model, the KABC-II focuses on mental processing, excludes acquired knowledge to the extent possible, and yields a global standard score called the Mental Processing Index (MPI) with a mean of 100 and a standard deviation (*SD*) of 15. The Luria model measures four domains: *sequential processing, simultaneous processing, learning ability*, and *planning ability*.

From the vantage point of the CHC model, the KABC-II Core Battery includes all scales in the Luria system, but they are interpreted from an alternative perspective; for example, the scale that measures sequential processing from the Luria perspective is seen as measuring the CHC ability of *short-term memory* (Gsm), and the scale that measures planning ability (Luria interpretation) aligns with Gf or *fluid reasoning* (CHC interpretation). The CHC model includes one extra scale that is *not* in the Luria model—namely, a measure of *crystallized ability* (Gc), which is labeled Knowledge/Gc. The global standard score yielded by the CHC model is labeled the Fluid–Crystallized Index (FCI), also with a mean of 100 and *SD* of 15.

Table 12.1 summarizes the dual-model approach, showing the Luria process and the corresponding CHC ability measured by each scale. The use of two theoretical models allows examiners to choose the model that best meets the needs of the child or adolescent being evaluated. The dual labels for the scales reflect the complexity of what the cognitive tasks measure and how their scores are interpreted. Examiners must select either the Luria or CHC model *before* they administer the test, thereby determining which global score should be used—the MPI (Luria model) or FCI (CHC model).

- The CHC model is the model of choice—except in cases where including measures of acquired knowledge (crystallized ability) is believed by the examiner to compromise the validity of the FCI. In those cases, the Luria-based global score (MPI) is preferred.
- The CHC model is given priority over the Luria model because we believe that knowledge/Gc is, in principle, an important aspect of cognitive functioning. Therefore, the CHC model (FCI) is preferred for children with known or suspected disabilities in reading, written expression, or mathematics; for children assessed for giftedness or intellectual disabilities; for children assessed for emotional or behavioral dis-

orders; and for children assessed for attentional disorders such as attention-deficit/hyperactivity disorder (ADHD).

- Situations in which the Luria model (MPI) is preferred include, but are not limited to, the following:
    - A child from a bilingual background.
    - A child from any nonmainstream cultural background that may have affected knowledge acquisition and verbal development.
    - A child with known or suspected language disorders, whether expressive, receptive, or mixed receptive–expressive.
    - A child with known or suspected autism.
    - An examiner with a firm commitment to the Luria processing approach who believes that acquired knowledge should be excluded from any global cognitive score (regardless of reason for referral).

This set of recommendations does not imply that we consider one model to be theoretically superior to the other. Both theories are equally important as foundations of the KABC-II. The CHC psychometric theory emphasizes specific cognitive abilities, whereas the Luria neuropsychological theory emphasizes *processes*—namely, the ways children process information when solving problems. Both approaches are valid for understanding how children learn and solve new problems, which is why each scale has two names, one from Luria theory and the other from CHC theory. Regardless of the model of the KABC-II that is *administered* (Luria or CHC), the way in which psychologists *interpret* the scales will undoubtedly be influenced by their theoretical preference.

On the original K-ABC, the Sequential and Simultaneous Processing scales were joined by a separate Achievement scale. That concept is continued with the Luria model of the KABC-II, although conventional kinds of achievement (reading, arithmetic) are excluded from the KABC-II Knowledge/Gc scale.

At age 3, only a global score is offered, either the MPI or FCI. For ages 4–18, the global scale is joined by an array of scales (see Table 12.1). The Planning/Gf scale is included only at ages 7–18 because a factor corresponding to the high-level set of abilities did not emerge for younger children. All KABC-II scales have a mean of 100 and *SD* of 15.

Theory: Luria and CHC

Luria (1970) perceived the brain's basic functions to be represented by three main blocks or functional systems, which are responsible for arousal and attention (block 1); the use of one's senses to analyze, code, and store information (block 2); and the application of executive functions for formulating plans and programming behavior (block 3). Within block 2, Luria (1966) distinguished between "two basic forms of integrative activity of the cerebral cortex" (p. 74), which he labeled *successive* and *simultaneous*. Despite Luria's interpretation of three blocks, each with separate functions, his focus was on *integration* among the blocks to be capable of complex behavior. Block 3

TABLE 12.1. The Dual Theoretical Foundations of the KABC-II

	Interpretation of scale from Luria theory	Interpretation of scale from CHC theory	Name of KABC-II scale
	Learning ability	Long-term storage and retrieval (Glr)	Learning/Glr
	Sequential processing	Short-term memory (Gsm)	Memory/Gsm
	Simultaneous processing	Visual processing (Gv)	Simultaneous/Gv
	Planning ability	Fluid reasoning (Gf)	Planning/Gf
	—	Crystallized ability (Gc)	Knowledge/Gc
Name of global score	Mental Processing Index (MPI)	Fluid–Crystallized Index (FCI)	

Note. Knowledge/Gc is included in the CHC system for the computation of the FCI, but it is excluded from the Luria system for the computation of the MPI. The Planning/Gf scale is for ages 7–18 only. All other scales are for ages 4–18. Only the MPI and FCI are offered for 3-year-olds.

is very closely related to the functions of block 1, as both blocks are concerned with overall efficiency of brain functions; part of the role of block 2 is to establish connections with block 3 (Reitan, 1988). Indeed, "integration of these systems constitutes the real key to understanding how the brain mediates complex behavior" (Reitan, 1988, p. 333; see also Naglieri & Otero, Chapter 6, this volume).

In the development of the KABC-II, the test authors emphasized the integration of the three blocks, not the measurement of each block in isolation. The block 1 arousal functions are key aspects of successful test performance on any cognitive task, but attention and concentration per se do not fit Kaufman and Kaufman's definition of high-level, complex, intelligent behavior. The Learning/Glr scale requires much sustained attention and concentration (block 1), but depends more on the integration of the three blocks than on any one in isolation. The Sequential/Gsm and Simultaneous/Gv scales are deliberately targeted to measure the block 2 successive and simultaneous functions, respectively, but again the test authors have striven for complexity. Luria (1966) defined the block 2 functions of analysis and storage of incoming stimuli via successive and simultaneous processing as *coding* functions, not problem-solving functions. But because block 2 is responsible for establishing connections with block 3, the KABC-II measures of simultaneous processing require not just the analysis, coding, and storage of incoming stimuli, but also the block 3 executive functioning processes for success. In addition, block 2 requires the integration of the incoming stimuli; hence subtests like Word Order and Rover require synthesis of auditory and visual stimuli. Planning/Gf is intended to measure Luria's block 3; again, however, success on these complex tasks requires not just executive functioning, but also focused attention (block 1) and the coding and storage of incoming stimuli (block 2).

The CHC model is a psychometric theory that rests on a large body of research, especially factor-analytic investigations, accumulated over decades. The CHC theory represents a data-driven theory, in contrast to Luria's clinically driven theory. CHC theory has two separate psychometric lineages: (1) Raymond Cattell's (1941) original Gf-Gc theory, which was expanded and refined by Horn (1968, 1985, 1989) to include an array of abilities (not just Gf and Gc); and (2) John Carroll's (1943, 1993) half-century of rigorous efforts to summarize and integrate the voluminous literature on the factor analysis of cognitive abilities. Ultimately, Horn and Carroll agreed to merge their separate but overlapping models into the unified CHC theory. This merger was accomplished in a personal communication to Richard Woodcock in July 1999; the specifics of CHC theory and its applications have been articulated by Dawn Flanagan, Kevin McGrew, and their colleagues (Flanagan et al., 2000; Flanagan & Ortiz, 2001; see also Schneider & McGrew, Chapter 3, this volume).

Both the Cattell–Horn and Carroll models essentially started from the same point: Spearman's (1904) *g* factor theory. Though they took different paths, they ended up with remarkably consistent conclusions about the spectrum of broad cognitive abilities. Cattell built upon Spearman's *g* to posit two kinds of *g: fluid intelligence* (Gf), the ability to solve novel problems by using reasoning (believed by Cattell to be largely a function of biological and neurological factors); and *crystallized intelligence* (Gc), a knowledge-based ability that is highly dependent on education and acculturation.

Almost from the beginning of his collaboration with Cattell, Horn believed that the psychometric data, as well as neurocognitive and developmental data, were suggesting more than just these two general abilities. Horn (1968) quickly identified four additional abilities; by the mid-1990s, his model included 9–10 *broad abilities* (Horn, 1989; Horn & Hofer, 1992; Horn & Noll, 1997). The initial dichotomy had grown, but not in a hierarchy. Horn retained the name of Gf-Gc theory, but the diverse broad abilities were treated as equals, not as part of any hierarchy.

Carroll (1993) developed a hierarchical theory based on his in-depth survey of factor-analytic studies composed of three levels or strata of abilities: *stratum III (general)*, a Spearman-like *g*, which Carroll considered to be a valid construct based on overwhelming evidence from factor analysis; *stratum II (broad)*, composed of eight broad factors that correspond reasonably closely to Horn's broad abilities; and *stratum I (narrow)*, composed of numerous fairly specific abilities, organized by the broad factor with which each is most closely associated (many relate to level of mastery, response speed, or rate of learning).

To Horn, the *g* construct had no place in his Gf-Gc theory; consequently, Carroll's stratum III is not usually considered part of CHC theory. Nonetheless, the KABC-II incorporates stratum III in its theoretical model because it corresponds to the global measure of general cognitive ability, the FCI. However, the *g* level is intended more as a practical than a theoretical construct. The

KABC-II scales correspond to 5 of the 10 broad abilities that make up CHC stratum II—Glr, Gsm, Gv, Gf, and Gc. The test authors chose not to include separate measures of Gq (*quantitative knowledge*) or Grw (*reading and writing*) because they believe that reading, writing, and mathematics fit in better with tests of academic achievement, like the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; A. S. Kaufman & Kaufman, 2014); however, Gq is present in some KABC-II tasks. The Gq ability measured, though, is considered secondary to other abilities measured by these subtests.

The KABC-II assesses 15 of the approximately 70 CHC narrow abilities. Table 12.2 shows the relationship of the KABC-II scales and subtests to the three strata. For the KABC-II, the broad abilities are of primary importance for interpreting the child's cognitive profile. In developing the KABC-II, the test authors did not strive to develop "pure" tasks for measuring the five CHC broad abilities. In theory, for example, Gv tasks should exclude Gf or Gs. In practice, however, the goal of comprehensive tests of cognitive ability is to measure problem solving in different contexts and under different conditions, with complexity being necessary to assess high-level functioning. Consequently, the test authors constructed measures that featured a particular ability while incorporating aspects of other abilities. To illustrate, Rover is primarily a measure

TABLE 12.2. General (Stratum III), Broad (Stratum II), and Narrow (Stratum I) CHC Abilities Measured by the KABC-II

CHC ability	Measured on KABC-II by:
General ability (stratum III in Carroll's theory)	**Mental Processing Index (MPI)**—Luria model of KABC-II (excludes acquired knowledge), ages 3–18 **Fluid–Crystallized Index (FCI)**—CHC model of KABC-II (includes acquired knowledge), ages 3–18
Broad ability (stratum II in CHC theory)	
Long-term storage and retrieval (Glr)	**Learning/Glr Index** (ages 4–18)
Short-term memory (Gsm)	**Sequential/Gsm Index** (ages 4–18)
Visual processing (Gv)	**Simultaneous/Gv Index** (ages 4–18)
Fluid reasoning (Gf)	**Planning/Gf Index** (ages 7–18)
Crystallized ability (Gc)	**Knowledge/Gc Index** (ages 4–18)
Narrow ability (stratum I in CHC theory)	
Glr: Associative memory (MA)	*Atlantis, Rebus, Delayed Recall scale*
Glr: Learning abilities (L1)	*Delayed Recall scale*
Gsm: Memory span (MS)	*Word Order* (without color interference), *Number Recall, Hand Movements*
Gsm: Working memory (WM)	*Word Order* (with color interference)
Gv: Visual memory (MV)	*Face Recognition, Hand Movements*
Gv: Spatial relations (SR)	*Triangles*
Gv: Visualization (VZ)	*Triangles, Conceptual Thinking, Block Counting, Story Completion*
Gv: Spatial scanning (SS)	*Rover*
Gv: Closure speed (CS)	*Gestalt Closure*
Gf: Induction (I)	*Conceptual Thinking, Pattern Reasoning, Story Completion*
Gf: General sequential reasoning (RG)	*Rover, Riddles*
Gc: General information (K0)	*Verbal Knowledge, Story Completion*
Gc: Language development (LD)	*Riddles*
Gc: Lexical knowledge (VL)	*Riddles, Verbal Knowledge, Expressive Vocabulary*
Gq: Math achievement (A3)	*Rover, Block Counting*

Note. Gq, quantitative ability. KABC-II scales are in **bold**, and KABC-II subtests are in *italics*. All KABC-II subtests are included, both Core and supplementary. CHC stratum I categorizations are courtesy of D. P. Flanagan (personal communication, October 2, 2003).

of Gv because of its visualization component, but it also involves Gf; Story Completion emphasizes Gf, but Gc is also required to interpret the social situations that are depicted.

DESCRIPTION OF KABC-II SUBTESTS

Sequential/Gsm Scale

- *Word Order (Core for ages 3:0–18:11)*. The child touches a series of silhouettes of common objects in the same order as the examiner has said the names of the objects; more difficult items include an interference task (color naming) between the stimulus and response.
- *Number Recall (supplementary for ages 3:0–3:11; Core for ages 4:0–18:11)*. The child repeats a series of numbers in the same sequence as the examiner has said them, with series ranging in length from two to nine numbers; the numbers are single digits, except that 10 is used instead of 7 to ensure that all numbers are one syllable.
- *Hand Movements (supplementary for ages 4:0–18:11; Nonverbal scale for ages 4:0–18:11)*. The child copies the examiner's precise sequence of taps on the table with the fist, palm, or side of the hand.

Simultaneous/Gv Scale

- *Triangles (Core for ages 3:0–12:11; supplementary for ages 13:0–18:11; Nonverbal scale for ages 3:0–18:11)*. For most items, the child assembles several identical rubber triangles (blue on one side, yellow on the other) to match a picture of an abstract design; for easier items, the child assembles a different set of colorful rubber shapes to match a model constructed by the examiner.
- *Face Recognition (Core for ages 3:0–4:11; supplementary for ages 5:0–5:11; Nonverbal scale for ages 3:0–5:11)*. The child attends closely to photographs of one or two faces that are exposed briefly, and then selects the correct face or faces from a group photograph.
- *Conceptual Thinking (Core for ages 3:0–6:11; Nonverbal scale for ages 3:0–6:11)*. The child views a set of four or five pictures, and then identifies the one picture that does not belong with the others; some items present meaningful stimuli, and others use abstract stimuli.
- *Rover (Core for ages 6:0–18:11)*. The child moves a toy dog to a bone on a checkerboard-like grid that contains obstacles (rocks and weeds), and tries to find the "quickest" path—the one that takes the fewest moves.
- *Block Counting (supplementary for ages 5:0–12:11; Core for ages 13:0–18:11; Nonverbal scale for ages 7:0–18:11)*. The child counts the exact number of blocks in various pictures of stacks of blocks; the stacks are configured so that one or more blocks are hidden or partially hidden from view.
- *Gestalt Closure (supplementary for ages 3:0–18:11)*. The child mentally "fills in the gaps" in a partially completed "inkblot" drawing, and names (or describes) the object or action depicted in the drawing.

Planning/Gf Scale (Ages 7–18 Only)

- *Pattern Reasoning (Core for ages 7:0–18:11; Nonverbal scale for ages 5:0–18:11; Core for ages 5:0–6:11, but on the Simultaneous/Gv scale)*. The child is shown a series of stimuli that form a logical, linear pattern, but one stimulus is missing; the child completes the pattern by selecting the correct stimulus from an array of four to six options (most stimuli are abstract, geometric shapes, but some easy items use meaningful shapes).
- *Story Completion (Core for ages 7:0–18:11; Nonverbal scale for ages 6:0–18:11; supplementary for ages 6:0–6:11, but on the Simultaneous/Gv scale)*. The child is shown a row of pictures that tell a story, but some of the pictures are missing. The child is given a set of cards with pictures, selects the ones needed to complete the story, and places the cards with the missing pictures in their correct locations.

Learning/Glr Scale

- *Atlantis (Core for ages 3:0–18:11)*. The examiner teaches the child the nonsense names for fanciful pictures of fish, plants, and shells; the child demonstrates learning by pointing to each picture (out of an array of pictures) when it is named. Corrective feedback is provided for incorrect responses during the early learning trials.
- *Rebus (Core for ages 3:0–18:11)*. The examiner teaches the child the word or concept associated with each particular rebus (drawing), and the child then "reads" aloud phrases and sentences composed of these rebuses.

- *Delayed Recall (supplementary scale for ages 5:0–18:11)*. The child demonstrates delayed recall of paired associations learned about 20 minutes earlier during the Atlantis and Rebus subtests (this requires the examiner to administer the Atlantis—Delayed and Rebus—Delayed tasks).

Knowledge/Gc Scale (CHC Model Only)

- *Riddles (Core for ages 3:0–18:11)*. The examiner provides several characteristics of a concrete or abstract verbal concept, and the child has to point to a picture depicting the concept (early items) or name the concept (later items).
- *Expressive Vocabulary (Core for ages 3:0–6:11; supplementary for ages 7:0–18:11)*. The child provides the name of a pictured object.
- *Verbal Knowledge (supplementary for ages 3:0–6:11; Core for ages 7:0–18:11)*. The child selects from an array of six pictures the one that corresponds to a vocabulary word or answers a general information question read aloud by the examiner.

ADMINISTRATION AND SCORING

For the KABC-II, the Core Battery for the Luria model comprises five subtests for 3-year-olds, seven subtests for 4- and 5-year-olds, and eight subtests for 6- to 18-year-olds. The CHC Core Battery includes two additional subtests at each age, both measures of crystallized ability. Approximate average administration times for the Core Battery, by age, are given in Table 12.3. For the CHC test battery, the additional two Core subtests add about 10 minutes to the testing time for ages 3–6 and about 15 minutes for ages 7–18 years.

Examiners who choose to administer the entire Expanded Battery—all Core and supplementary subtests, the Delayed Recall scale, and all measures of crystallized ability—can expect to spend just under 60 minutes for 3- and 4-year-olds, about 90 minutes for ages 5 and 6, and about 100 minutes for ages 7–18. However, examiners who choose to administer supplementary subtests need not give all of the available subtests to a given child or adolescent—just the ones that are most pertinent to the reasons for referral.

TABLE 12.3. Average Administration Times (in Minutes) for the KABC-II Core Battery

Ages =	MPI (Luria model)	FCI (CHC model)
3–4	30	40
5	40	50
6	50	60
7–18	55	70

Sample and teaching items are included for all subtests, except those that measure acquired knowledge, to ensure that children understand what is expected of them to meet the demands of each subtest. Scoring of all subtests is objective. Even the Knowledge/Gc subtests require pointing or one-word responses rather than longer verbal responses, which often introduce subjectivity into the scoring process.

In their KABC-II review, Bain and Gray (2008) state that "generally, the KABC-II is easy to administer with practice and is inherently interesting for children, with several manipulative opportunities and brightly colored, well-designed stimuli. Some of the newly developed subtests are particularly attractive" (p. 100). They do note, however, that the numerous test materials "plus the manual make up a kit that is far from the lightest, most compact kit in the typical examiner's repertoire" (p. 100).

PSYCHOMETRIC PROPERTIES

The KABC-II is a psychometrically sound instrument. A brief review of the instrument's standardization sample, reliability, and validity is provided in this chapter. For a thorough description of the normative sample and reliability, stability, and validity data, see the KABC-II manual (A. S. Kaufman & Kaufman, 2004a) and *Essentials of KABC-II Assessment* (A. S. Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005).

Standardization Sample

The KABC-II standardization sample was composed of 3,025 children and adolescents. The sample matched the U.S. population on the stratification variables of gender, race/ethnicity, socioeconomic status (SES—parental education), region, and special education status. Each year of age between 3 and 18 was represented by 125–250 children, about equally divided between males and females, with most age groups consisting of exactly 200 children.

Reliability

KABC-II global scale (MPI and FCI) split-half reliability coefficients were in the mid-.90s for all age groups (only the value of .90 for the MPI at age 3 was below .94). The mean MPI coefficient was .95 for ages 3–6 and ages 7–18; the mean values for FCI were .96 (ages 3–6) and .97 (ages 7–18). Mean split-half reliability coefficients for the separate scales (e.g., Learning/Glr, Simultaneous/Gv) averaged .91–.92 for ages 3–6 and ranged from .88 to .93 (mean = .90) for ages 7–18. Similarly, the Nonverbal Index—the alternative global score for children and adolescents with hearing impairments, limited English proficiency, and the like—had an average coefficient of .90 for ages 3–6 and .92 for those ages 7–18. Mean split-half values for Core subtests across the age range were .82 (age 3), .84 (age 4), .86 (ages 5–6), and .85 (ages 7–18). Nearly all Core subtests had mean reliability coefficients of .80 or greater at ages 3–6 and 7–18. Stability data over an interval of about 1 month for three age groups (total *N* = 203) yielded coefficients of .86–.91 for the MPI and .90–.94 for the FCI. Stability coefficients for the separate scales averaged .81 for ages 3–6 (range = .74–.93), .80 for ages 7–12 (range = .76–.88), and .83 for ages 13–18 (range = .78–.95). Bain and Gray (2008) have concluded that "the reliability coefficients are consistently high" (p. 100).

Validity

Construct validity was given strong support by the results of confirmatory factor analysis (CFA). The CFA supported four factors for ages 4 and 5–6, and five factors for ages 7–12 and 13–18, with the factor structure supporting the scale structure for these broad age groups. The fit was excellent for all age groups; for ages 7–12 and 13–18, the five-factor solution provided a significantly better fit than the four-factor solution. In addition to the CFA evidence provided in the KABC-II manual, strong independent support for the theory-based structure of the KABC-II was provided in a reanalysis of standardization data for all age groups (Reynolds, Keith, Fine, Fisher, & Low, 2007), and for a new sample of 200 preschool children ages 4–5 years (Hunt, 2008; Morgan, Rothlisberg, McIntosh, & Hunt, 2009). In analyses of the KABC-II with preschool children, the authors found support for the four-factor structure of the KABC-II, but suggested that the data fit even better into a five-factor model, similar to the broad ability factors laid out for the older children (Potvin, Keith, Caemmerer, & Trundt, 2015). In addition, Reynolds and Keith (2007) demonstrated that the factor structure was invariant for high- and low-ability groups; Reynolds, Ridley, and Patel (2008) found invariant factor structure by gender for ages 6–18 years; and the construct validity of the KABC-II has also been demonstrated cross-culturally—for example, for 598 low-SES Kannada-speaking children ages 6–10 years in Bangalore, South India (Malda et al., 2010) and for children and adolescents tested in countries where the KABC-II has been translated and adapted (e.g., Fujita et al., 2011; A. S. Kaufman et al., 2013, 2014). Construct validity has also been demonstrated by ethnicity within the United States for 2001 white, black, and Hispanic children in grades 1–12 (Scheiber, 2016).

Correlation coefficients between the FCI and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991, 2003, 2014) Full Scale IQ (FSIQ), corrected for the variability of the normative sample, were .81 for the WISC-V (N = 89, ages 6–16), .89 for the WISC-IV (*N* = 56, ages 7–16), and .77 for the WISC-III (*N* = 119, ages 8–13). The FCI also correlated .77 with the Woodcock–Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014) General Intellectual Ability (GIA) composite (*N* = 50, ages 7–18), .72 with the K-ABC Mental Processing Composite (MPC) for preschool children (*N* = 67), and .84 with the K-ABC MPC for school-age children (*N* = 48). Correlations with the MPI were generally slightly lower (by an average of .05).

Fletcher-Janzen (2003) conducted a correlational study with the WISC-IV for 30 Native American children from Taos, New Mexico, who were tested on the KABC-II at an average age of 7:8 (range = 5–14) and on the WISC-IV at an average age of 9:3. As shown in Table 12.4, the two KABC-II global scores correlated about .85 with the WISC-IV FSIQ. This strong relationship indicates that the KABC-II global scores and the WISC-IV global score measure the same construct; nevertheless, the KABC-II yielded global scores that were about 0.5 *SD* higher than the FSIQ for this Native American sample.

Correlations were obtained between the KABC-II and achievement on the Woodcock–Johnson III (WJ-III; Woodcock, McGrew, & Mather, 2001), the Wechsler Individual Achievement Test—Second Edition (WIAT-II; Psychological Corporation,

TABLE 12.4. Means, *SD*s, and Correlations between KABC-II and WISC-IV Global Scores for 30 Native American Children and Adolescents from Taos, New Mexico

KABC-II and WISC-IV global scores	Mean	*SD*	*r* with WISC-IV FSIQ	Mean difference	
				MPI vs. FSIQ	FCI vs. FSIQ
KABC-II					
Mental Processing Index (MPI)	95.1	13.3	.86	+8.4	—
Fluid–Crystallized Index (FCI)	94.1	12.5	.84	—	+7.4
WISC-IV					
Full Scale IQ (FSIQ)	86.7	12.3	—	—	—

Note. Children were tested first on the KABC-II (age range = 5–14, mean = 7:8) and second on the WISC-IV (age range = 6–15, mean = 9:3). Data from Fletcher-Janzen (2003).

2001), and the Peabody Individual Achievement Test—Revised/Normative Update (PIAT-R/NU; Markwardt, 1998) for six samples with a total *N* of 401. Coefficients between the FCI and total achievement for the six samples, corrected for the variability of the norms sample, ranged from .67 on the PIAT-R/NU for grades 1–4 to .87 on the WIAT-II for grades 6–10 (mean *r* = .75). For these same samples, the MPI correlated .63–.83 (mean *r* = .71). In addition, the KABC-II FCI correlated .75 (and the MPI correlated .71) with the Academic Skills Battery on the KTEA-3 Comprehensive Form for 99 individuals ages 4½ through 18 years (A. S. Kaufman & Kaufman, 2014, Table 2.15). Correlations with the KTEA-3 Academic Skills Battery were as follows: FCI = .75 (and MPI = .71) for ages 4½ –18 (*N* = 99) (A. S. Kaufman & Kaufman, 2014, Table 2.12).

KABC-II Normative Update

As noted at the start of this chapter, a normative update of the KABC-II was released in 2018. The published KABC-II was used with no changes to allow application of existing clinical research and interpretation guidelines to be applied to the new normative data. A normative sample of 700 children and adolescents ages 3 years, 0 months to 18 years, 11 months was collected in 2017. Each age year between the ages of 3 and 14 includes 50 children and adolescents, and ages 15–16 and 17–18 each include 50 adolescents. The sample was matched to the U.S. population on the stratification variables of gender, race/ethnicity, parent education level, and geographic region. Individuals in special education were included in the sample in a similar manner as described in the KABC-II. Psychometric information on the KABC-II Normative Update can be found in the KABC-II Normative Update supplemental manual (A. S. Kaufman et al., 2018).

Reliability

KABC-II NU global scale (MPI and FCI) split-half reliability coefficients were in the .90s for all age groups. The mean MPI coefficient was .96 for ages 3–6 and .97 for ages 7–18; the mean values for FCI were .97 (ages 3–6) and .98 (ages 7–18). Mean split-half reliability coefficients for the separate scales (e.g., Learning/Glr, Simultaneous/Gv) averaged .91–.98 for ages 3–6 and ranged from .91 to .97 for ages 7–18. Similarly, the Nonverbal Index had an average coefficient of .94 for ages 3–6 and .95 for ages 7–18. Mean split-half values for Core subtests across the age range were .86 (age 3), .90 (age 4), .91 (age 5), .88 (age 6), and .90 (ages 7–18).

Validity

Construct validity is based on the data reported previously for the KABC-II. In addition to the evidence provided for the KABC-II, correlation coefficients between the KABC-II NU FCI and the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; 2014) Full Scale IQ (FSIQ), corrected for the variability of the normative sample, was .72 (*N* = 79, ages 6–16). In addition, the KABC-II NU FCI correlated .74 (and the MPI correlated .68) with the Academic Skills Battery on the KTEA-3 Comprehensive Form for 99 individuals ages 4½ through 18 years.

INTERPRETATION

What the Global Scores Provide

KABC-II interpretation is focused on the scale profile; however, the global scores provide a general cognitive picture of a child's performance and provide a context in which the child's strengths and weaknesses can be determined. The MPI provides a measure of the child's general mental processing ability from a Lurian perspective, while the FCI provides a measure of general cognitive ability as defined in CHC theory. The differences in these approaches have been described earlier in this chapter.

When an examiner is interpreting the global scores, it is important to review the specific scales and subtests that constitute the global score. It is important to understand the index and subtest scatter that contributes to a global score when interpreting a global score such as the FCI or MPI. However, scatter among index and subtest scores is fairly common, with over half the original KABC-II normative sample exhibiting significant scatter (McGill, 2016). But despite the scatter among index and subtest scaled scores, there is strong psychometric support for global scores: They consistently demonstrate the highest reliability and stability estimates, as well as the highest predictive validity estimates, among all scores yielded by intelligence tests (Canivez, 2013). Therefore, while it is important to consider the scatter of scores contributing to global scores, global scores are interpretable even when significant scatter in scores is present (McGill, 2016). Also, a large-scale study of the Kaufman tests (KABC-II and KTEA-II) and the Woodcock–Johnson tests (WJ III), using structural equation modeling and CFA, demonstrated that the *g* that underlies cognitive batteries is essentially the same *g* that underlies achievement batteries (S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012); furthermore, Gf and *g* are virtually identical.

What the Scales Measure

Sequential/Gsm (Ages 4–18)

- *CHC interpretation*. Short-term memory (Gsm) is the ability to apprehend and hold information in immediate awareness briefly, and then use that information within a few seconds, before it is forgotten.
- *Luria interpretation*. Sequential processing is used to solve problems, where the emphasis is on the serial or temporal order of stimuli. For each problem, the input must be arranged in a strictly defined order to form a chain-like progression; each idea is linearly and temporally related only to the preceding one.

Simultaneous/Gv (Ages 4–18)

- *CHC interpretation*. Visual processing (Gv) is the ability to perceive, manipulate, and think with visual patterns and stimuli, and to mentally rotate objects in space.
- *Luria interpretation*. Simultaneous processing demands a Gestalt-like, frequently spatial, integration of stimuli. The input has to be synthesized simultaneously, so that the separate stimuli are integrated into a group or conceptualized as a whole.

Learning/Glr (Ages 4–18)

- *CHC interpretation*. Long-term storage and retrieval (Glr) is the ability both to store information in long-term memory and to retrieve that information fluently and efficiently. The emphasis of Glr is on the *efficiency* of the storage and retrieval, not on the specific nature of the information stored.
- *Luria interpretation*. Learning ability requires an integration of the processes associated with all three of Luria's functional units. The attentional requirements for the learning tasks are considerable, as focused, sustained, and selective attention are requisites for success. However, for effective paired-associate learning, children need to apply both of the block 2 processes, sequential and simultaneous. Block 3 planning abilities help them generate strategies for storing and retrieving the new learning.

Planning/Gf (Ages 7–18)

- *CHC interpretation*. Fluid reasoning (Gf) refers to a variety of mental operations that a person can use to solve a novel problem with adaptability and flexibility—operations such as drawing inferences and applying inductive or deductive reasoning. Verbal mediation also plays a key role in applying fluid reasoning effectively.
- *Luria interpretation*. Planning ability requires hypothesis generation, revising one's plan of action, monitoring and evaluating the best hypothesis for a given problem (decision making), flexibility, and impulse control. This set of high-level skills is associated with block 3 executive functioning.

Knowledge/Gc (Ages 4–18)—CHC Model Only

- *CHC interpretation.* Crystallized ability (Gc) reflects a person's specific knowledge acquired within a culture, as well as the person's ability to apply this knowledge effectively. Gc emphasizes the *breadth* and *depth* of the specific information that has been stored.

- *Luria interpretation.* The Knowledge/Gc scale is not included in the MPI, but may be administered as a supplement if an examiner seeks a measure of a child's acquired knowledge. From a Lurian perspective, Knowledge/Gc measures a person's knowledge base, developed over time by applying block 1, block 2, and block 3 processes to the acquisition of factual information and verbal concepts. Like Learning/Glr, this scale requires an integration of the key processes, but unlike learning ability, acquired knowledge emphasizes the *content* more than the *process*.

Gender Differences

Analysis of KABC-II standardization data explored gender differences at four different age ranges: 3–4 years, 5–6 years, 7–12 years, and 13–18 years. At ages 3–4, females significantly outperformed males on the MPI, FCI, and Nonverbal Index by about 5 points (0.33 *SD*), but there were no other significant differences on the global scales at any age level (J. C. Kaufman, 2003). Consistent with the literature on gender differences, females tended to score higher than males at preschool levels. Females scored significantly higher than males at ages 3 and 4 years by about 3–4 points, with significant differences emerging on Learning/Glr (0.27 *SD*) and Simultaneous/Gv (0.34 *SD*). Also consistent with previous findings, males scored significantly higher than females on the Simultaneous/Gv scale at ages 7–12 (0.24 *SD*) and 13–18 (0.29 *SD*). Females earned significantly higher scores than males at ages 5–6 on the Sequential/Gsm scale (0.22 *SD*) and at ages 13–18 on the Planning/Gf scale (0.13 *SD*).

In a subsequent study of gender differences on the KABC-II at ages 6–18 years, also based on standardization data, Reynolds and colleagues (2008) applied multigroup higher-order analysis of mean and covariance structures and of multiple-indicator/multiple-cause models. They found that males consistently demonstrated a significant mean advantage on the latent visual–spatial ability (Gv) factor, even when *g* was controlled for; the same finding held true for the latent crystallized ability (Gc) factor at all ages except 17–18. Females scored higher on the latent higher-order *g* factor at all ages, although the difference was significant at only two ages. No significant age × gender interaction effect was identified (Reynolds et al., 2008). In a study of 137 Mexican American children ages 7–12 years, Gomez (2008) found a slight superiority of males over females on the MPI. In general, in all studies, gender differences on the KABC-II are small; even when statistically significant, they tend to be small in effect size (McLean, 1995).

Ethnicity Differences

Because the original K-ABC yielded considerably smaller ethnic differences than conventional IQ tests did, it was especially important to determine the magnitude of ethnic differences for the substantially revised and reconceptualized KABC-II. On the original K-ABC, the mean MPC for African American children ages 5:0–12:6 was 93.7 (A. S. Kaufman & Kaufman, 1983, Table 4.35). On the KABC-II, the mean MPI for African American children ages 7–18 in the standardization sample (N = 315) was 94.8, and the mean FCI was 94.0. On the two new KABC-II scales, African American children ages 7–18 averaged 94.3 (Planning/Gf) and 98.6 (Learning/Glr). At ages 3–6, African American children averaged 98.2 on Learning/Glr.

When standard scores were adjusted for SES and gender, European Americans scored 4.4 points higher than African Americans at 7–12 years on the MPI—smaller than an adjusted difference of 8.6 points on WISC-III FSIQ at ages 6–11 (J. C. Kaufman, 2003), and smaller than the 6-point difference at ages 6–11 on the WISC-IV. At the 13- to 18-year level, European Americans scored 7.7 points higher than African Americans on the adjusted MPI (J. C. Kaufman, 2003)—substantially smaller than the 14.1-point discrepancy on adjusted WISC-III FSIQ for ages 12–16, and also smaller than the 11.8-point difference at ages 12–16 on the WISC-IV. Similar results are observed in comparison to the WISC-V. The mean WISC-V FSIQ, adjusted for sex and parent education level, for European Americans is 8.7 points higher than the mean for African Americans at ages 6–16. (WISC-III data are from Prifitera, Weiss, & Saklofske, 1998; WISC-IV data are from Prifitera, Weiss, Saklofske, & Rolfhus, 2005; WISC-V data are from A. S. Kaufman, Raiford, & Coalson, 2016.) The adjusted discrepancies of

6.2 points (ages 7–12) and 8.6 points (ages 13–18) on the FCI, which includes measures of acquired knowledge, were also smaller than WISC-III and WISC-IV FSIQ differences. The KABC-II thus seems to continue in the K-ABC tradition of yielding higher standard scores for African Americans than are typically yielded on other instruments. In addition, as shown earlier in Table 12.4, the mean MPI and FCI earned by the 30 Native American children in Taos are about 7–8 points higher than this sample's mean WISC-IV FSIQ (Fletcher-Janzen, 2003).

Further ethnicity analyses of KABC-II standardization data were conducted for the entire age range of 3–18 years, which included 1,861 European Americans, 545 Hispanics, 465 African Americans, 75 Asian Americans, and 68 Native Americans. When adjusted for SES and gender, mean MPIs were 101.7 for European Americans, 97.1 for Hispanics, 96.0 for African Americans, 103.4 for Asian Americans, and 97.6 for Native Americans. Mean adjusted FCIs were about 1 point lower than the mean MPIs for all groups except European Americans (who had a slightly higher FCI) (J. C. Kaufman, 2003).

Dale, McIntosh, Rothlisberg, Ward, and Bradley (2011) examined ethnic differences for 49 African American and 49 European American preschool children from a U.S. Midwestern city (mean age = 5 years), who were matched on age, sex, and level of parental education. African American and European American preschool children had similar patterns of highs and lows in their profiles of scores and performed comparably on the different scales. Differences on the FCI (European American mean = 97.1, African American mean = 95.6) were a trivial and nonsignificant 1.5 points, "the smallest gap seen in the literature" (Dale, 2009, p. viii).

In a different approach to evaluating ethnic bias, Scheiber (2016, 2017) focused on differential construct validity and differential predictive validity of the KABC-II indexes in white, black, and Hispanic children in grades 1–12 (total *N* = 2,001). As discussed earlier in regard to construct validity, Scheiber (2016) demonstrated that the KABC-II produced the same five factors (i.e., constructs) for each of the three ethnic groups, thereby providing evidence of its construct validity for these three groups of students. Scheiber (2017) used structural equation modeling to further evaluate the differential validity of the KABC-II in predicting academic achievement on the KTEA-II (A. S. Kaufman & Kaufman, 2004b) for the same 2,001 students. Again, she found a lack of predictive bias in the slopes (i.e., similar magnitude of predictive validity coefficients) for the three separate samples of students in grades 1–12. When examining the intercepts in these analyses, Scheiber (2017) found consistent *over*prediction of the achievement scores for Hispanic and black students (notably, *under*prediction of achievement would have indicated test bias; this did not occur for any of the separate KABC-II indexes such as Planning/Gf or for the FCI). It is unclear why most KABC-II indexes overpredicted achievement for ethnic minority children. Scheiber (2017) suggested that the causes might be related to institutional or societal variables that have prevented black and Hispanic students from achieving in school at the level predicted by their KABC-II indexes.

In a variant of the differential prediction approach, Scheiber and Kaufman (2015) investigated the three KABC-II global scores—the FCI, MPI, and Nonverbal Index—to determine which one was the least biased in terms of its predictive validity for the three ethnic groups studied in grades 1–4, 5–8, and 9–12. Whereas the MPI and the Nonverbal Index were specifically developed to promote fair assessment for children with language difficulties and for children from ethnic minority groups, the surprising result of Scheiber and Kaufman's (2015) differential predictive validity study was that the FCI proved to be the least biased of the three global indexes. These findings do not minimize the importance of using the MPI or Nonverbal Index in clinical circumstances that dictate the need for fair assessment for children with linguistic or cultural differences. However, the results of the study provide a reinforcement of the notion that for students in general, across a wide grade span, the FCI (even with its inclusion of Knowledge/Gc tasks) provides an unbiased measure of general cognitive ability as a predictor of achievement for ethnically diverse children and adolescents.

CLINICAL APPLICATIONS

Like the original K-ABC, the KABC-II was designed to be a clinical and psychological instrument, not merely a psychometric tool. It has a variety of clinical benefits and uses:

1. The identification of process integrities and deficits for assessment of individuals with specific learning disabilities.

2. The evaluation of individuals with known or suspected neurological disorders, when the KABC-II (and KABC-II NU) is used along with other tests as part of a comprehensive neuropsychological battery.
3. The integration of the individual's profile of KABC-II (and KABC-II NU) scores with clinical behaviors observed during the administration of each subtest (Fletcher-Janzen, 2003)—identified as Qualitative Indicators (QIs) on the KABC-II record form (see A. S. Kaufman & Kaufman, 2004a; A. S. Kaufman et al., 2005).
4. The selection of the MPI to promote the fair assessment of children and adolescents from African American, Hispanic, Native American, and Asian American backgrounds (an application that has empirical support, as summarized briefly in the previous section on "Ethnicity Differences" and in Table 12.4).
5. Evaluation of individuals with known or suspected ADHD, intellectual disabilities/developmental delays, speech–language difficulties, emotional/behavioral disorders, autism, reading/math disabilities, intellectual giftedness, and hearing impairment (KABC-II data on all of these clinical samples are presented and discussed in the KABC-II manual and KABC-II NU manual supplement).

We believe that whenever possible, clinical tests such as the KABC-II should be interpreted by the same person who administered them—an approach that enhances the clinical benefits of the instrument and its clinical applications. The main goal of any evaluation should be to *effect change* in the person who was referred. Extremely competent and well-trained examiners are needed to best accomplish that goal; we feel more confident in a report writer's ability to effect change and to derive clinical benefits from an administration of the KABC-II when the professional who interprets the test data and writes the case report has also administered the test and directly observed the individual's test behaviors.

INNOVATIONS IN MEASURES OF COGNITIVE ASSESSMENT

Several of the features described here for the KABC-II (and, therefore, for the KABC-II NU) are innovative in comparison to the Wechsler and Stanford–Binet tradition of intellectual assessment, which has century-old roots. However, some of these innovations are not unique to the KABC-II; rather, several of them are shared by other contemporary instruments, such as the WJ IV (Schrank et al., 2014), the Cognitive Assessment System—Second Edition (CAS-2; Naglieri, Das, & Goldstein, 2014), and the most recent revisions of the Wechsler scales (Wechsler, 2008, 2012, 2014).

Integrates Two Theoretical Approaches

As discussed previously, the KABC-II utilizes a dual theoretical approach—Luria's neuropsychological theory and CHC theory. This dual model permits alternative interpretations of the scales, based on the examiner's personal orientation or based on the specific individual being evaluated. One of the criticisms of the original K-ABC was that the mental processing scales were interpreted solely from the sequential–simultaneous perspective, despite the fact that alternative interpretations are feasible (Keith, 1985). The KABC-II addressed that criticism and provided a strong theoretical foundation for the test by building the test on a dual theoretical model. In their review of the KABC-II, Bain and Gray (2008) have praised the theoretical model:

> In terms of construct validity, Kaufman and Kaufman's efforts to combine the two theoretical models, the Luria model and the CHC model, into one instrument is particularly attractive to those of us who favored the use of the original K-ABC for younger children but found ourselves scrambling for variations on subtests measuring broad and narrow abilities for cross-battery assessment. (p. 100)

(For a comprehensive review of cross-battery assessment, see Flanagan et al., Chapter 27, this volume.)

Provides the Examiner with Optimal Flexibility

The two theoretical models that underlie the KABC-II not only provide alternative interpretations of the scales, but also give the examiner the flexibility to select the model (and hence the global score) that is better suited to the individual's background and reason for referral. As mentioned earlier, the CHC model is ordinarily the model of choice, but examiners can choose to administer the Luria model when excluding measures of acquired knowledge from the global score promotes fairer assessment of a child's general cognitive ability. The MPI that results is an especially pertinent global score, for example, for individuals who have a receptive or expressive language disability or who are from a bilingual background. This flexibility of choice permits fairer assessment for anyone referred for an evaluation. Note, however, that the results of differential predictive validity studies with large samples of typically developing white, black, and Hispanic students found that all three KABC-II global indexes are unbiased, with the lowest differences observed with the FCI—at least regarding the prediction of academic achievement (Scheiber & Kaufman, 2015).

The examiner's flexibility is enhanced as well by the inclusion of supplementary subtests for most scales, including a supplementary Delayed Recall scale to permit the evaluation of a child's recall of paired associations that were learned about 20 minutes earlier. Hand Movements is a supplementary Sequential/Gsm subtest for ages 4–18, and Gestalt Closure is a supplementary task across the entire 3–18 range. Supplementary subtests are not included in the computation of standard scores on any KABC-II scales, but they do permit the examiner to follow up hypotheses suggested by the profile of scores on the Core Battery, to generate new hypotheses, and to increase the breadth of measurement on the KABC-II constructs.

Promotes Fairer Assessment of Minority Children

As mentioned earlier, children and adolescents from minority backgrounds—African American, Hispanic, Asian American, and Native American—earned mean MPIs that were close to the normative mean of 100, even prior to adjustment for SES and gender. In addition, there is some evidence that the discrepancies between European Americans and African Americans are smaller on the KABC-II than on the WISC-IV, and that Native Americans score higher on the KABC-II than the WISC-IV (see Table 12.4). These data, as well as additional data (Dale et al., 2011; Scheiber, 2016, 2017), suggest that the KABC-II (and KABC-II NU), like the original K-ABC, will be useful for promoting fairer assessment of children and adolescents from minority backgrounds.

Offers a Separate Nonverbal Scale

Like the K-ABC, the KABC-II (and KABC-II NU) offers a reliable, separate Nonverbal scale composed of subtests that can be administered in pantomime and responded to nonverbally. This special global scale, for the entire 3–18 age range, permits valid assessment of children and adolescents with hearing impairments, moderate to severe speech–language disorders, limited English proficiency, and so forth.

Permits Direct Evaluation of a Child's Learning Ability

The KABC-II Learning/Glr scale allows direct measurement of a child's ability to learn new information under standardized conditions. These tasks also permit examiners to observe the child's ability to learn under different conditions; for example, Atlantis gives the child feedback after each error, but Rebus does not offer feedback. In addition, Rebus involves meaningful verbal labels for symbolic visual stimuli, whereas Atlantis involves nonsensical verbal labels for meaningful visual stimuli. When examiners choose to administer the supplementary Delayed Recall scale to children ages 5–18, they are then able to assess the children's ability to retain information that was taught earlier in the evaluation.

USE WITH THE KTEA-3

The KTEA-3 can be used in tandem with the KABC-II to identify skill deficits and potential process deficits in children with learning difficulties. (See Morrison, Singer, & Raiford, Chapter 29, this volume, for detailed info on the KTEA-3.) When an achievement measure like the KTEA-3 is linked with a cognitive measure like the KABC-II, examiners can efficiently collect in-depth data that can help them test hypotheses, formulate a diagnosis, and make decisions about eligibility, as well as bridge the gap between assessment and intervention. The KTEA-3 is linked to the KABC-II and provides a cohesive theoretical basis for interpreting a comprehensive assessment battery. This integration extends to interpreting the KABC-II and KABC-II NU cognitive scales in conjunction with students' patterns of errors on the KTEA-3 (e.g., Liu et al., 2017).

Thus, to glean information about a wide spectrum of abilities within the CHC framework, examiners can integrate findings from the KTEA 3 and KABC-II. As described earlier in the chapter, the KABC-II addresses five of the CHC broad abilities: short-term memory (Gsm), visual processing

(Gv), long-term storage and retrieval (Glr), fluid reasoning (Gf), and crystallized ability (Gc). The KTEA-3 Comprehensive Form measures three additional broad abilities: auditory processing (Ga), reading and writing (Grw), and quantitative knowledge (Gq). It also measures Glr narrow abilities that increase the breadth of the Glr narrow abilities measured by the KABC-II when the two batteries are administered together. Moreover, the KABC-II indirectly measures one of the Gq narrow abilities (i.e., mathematics achievement, by virtue of the fact that Rover and Block Counting each require a child to count). Publications by Flanagan and her colleagues describe CHC theory in depth, and provide more detail in test interpretation from that theoretical perspective (see, e.g., Flanagan, Ortiz, & Alfonso, 2013). Lichtenberger and Breaux (2010) provide a CHC-based analysis of each KABC-II subtest that outlines the narrow CHC abilities measured by these tasks. This analysis is summarized below.

CHC ABILITIES MEASURED BY KABC-II SUBTESTS

Long-Term Storage and Retrieval (Glr) Narrow Ability

Associative Memory

- KABC-II Atlantis
- KABC-II Rebus
- KABC-II Atlantis—Delayed
- KABC-II Rebus—Delayed

Learning Abilities

- KABC-II Atlantis—Delayed
- KABC-II Rebus—Delayed

Short-Term Memory (Gsm) Narrow Ability

Memory Span

- KABC-II Word Order (without color interference)
- KABC-II Number Recall
- KABC-II Hand Movements

Working Memory

- KABC-II Word Order (with color interference)

Visual Processing (Gv) Narrow Ability

Visual Memory

- KABC-II Face Recognition
- KABC-II Hand Movements

Spatial Relations

- KABC-II Triangles

Visualization

- KABC-II Triangles
- KABC-II Conceptual Thinking
- KABC-II Block Counting
- KABC-II Pattern Reasoning
- KABC-II Story Completion

Spatial Scanning

- KABC-II Rover

Closure Speed

- KABC-II Gestalt Closure

Fluid Reasoning (Gf) Narrow Ability

Induction

- KABC-II Conceptual Thinking
- KABC-II Pattern Reasoning
- KABC-II Story Completion

General Sequential Reasoning

- KABC-II Story Completion
- KABC-II Rover
- KABC-II Riddles

Note. Success on KABC-II Rebus is dependent to some extent on Fluid Reasoning.

Crystallized Ability (Gc) Narrow Ability

General Information

- KABC-II Verbal Knowledge (items that measure general information)
- KABC-II Story Completion

Language Development

- KABC-II Riddles

Lexical Knowledge

- KABC-II Riddles
- KABC-II Verbal Knowledge (items that measure vocabulary)
- KABC-II Expressive Vocabulary

Grammatical Sensitivity

Note. Success on KABC-II Rebus is dependent to some extent on Grammatical Sensitivity.

Auditory Processing (Ga) Narrow Ability

Phonetic Coding—Synthesis

Note. Deficits in certain Ga narrow abilities, such as Speech Sound Discrimination (US), may affect performance negatively on such tests as KABC-II Riddles, Word Order, and Number Recall.

Quantitative Knowledge (Gq) Narrow Ability

Mathematical Achievement

- KABC-II Rover
- KABC-II Block Counting

Appendix 12.1 presents a case study that demonstrates how the KABC-II can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

APPENDIX 12.1

Illustrative Case Study

Name: Jacob W.
Age: 5 years, 10 months
Grade in school: Kindergarten (spring semester)

REASON FOR REFERRAL

Jacob attends kindergarten at a private school. He was referred for an evaluation by his parents, Mr. and Mrs. W., and his teacher, Mrs. A., due to concerns about Jacob's current cognitive and motor abilities. Jacob has been struggling in school and is not progressing at the expected pace in his kindergarten class. In addition, Mrs. A. reports that some skills Jacob acquired early in the school year have markedly declined. Mrs. A has also observed fine and gross motor difficulties, and his parents confirm that he has always been clumsier and less coordinated than his peers. Both Jacob's parents and his teacher and school are seeking recommendations for ways to assist Jacob to be successful in the classroom.

EVALUATION PROCEDURES

- Collateral interview with parent, Mrs. W.
- Collateral interview with teacher, Mrs. A.
- Review of school records
- Behavior Assessment Scale for Children, Third Edition (BASC-3; Reynolds & Kamphaus, 2015), parent and teacher versions
- Delis Rating of Executive Function (DREF; Delis, 2012), parent and teacher versions
- Kaufman Assessment Battery for Children—Second Edition (KABC-II NU)
- Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Form A
- NEPSY—Second Edition (NEPSY-II; Korkman, Kirk, & Kemp, 2007)
- Bruininks–Oseretsky Test of Motor Proficiency, Second Edition (BOT-2; Bruininks & Bruininks, 2005)

The KTEA-3 was administered on Q-interactive with a touch tablet (iPad).

BACKGROUND INFORMATION

Background information was obtained through separate interviews with Mrs. W. and with Mrs. A.

Family/Social History

Jacob lives with his mother, father, and two siblings. Mrs. W. reported that Jacob is a loving child who tends to "go with the flow" of the household. She also reported that he tends to be slow and takes his time to get ready in the morning, requiring frequent prompting to get going. Jacob's mother reports that he is "laid back" and relatively quiet. She indicated that he enjoys playing with his younger sister, but his sister tends to lead the play. When he doesn't want to do something, he will just sit still and not interact. However, he enjoys interacting with others and makes friends easily.

Mrs. W. noted that Jacob has problems with attention, and that previously it was a battle every evening to get him to complete homework assigned in kindergarten. He has difficulty with attention and working independently. He is easily distracted and loses track of what he is doing. When he is frustrated, he puts his hands on his head and doesn't continue. Usually after a short break, he will successfully continue with the task. He is working with a behavioral therapist weekly for 6 weeks, and this intervention has improved his behavior at home, with homework being completed more quickly and with less distress. Mrs. W. reported that the weekly intervention has also improved her relationship with Jacob because there is less stress and conflict in regard to schoolwork and behavior.

Jacob's mother also reported that he is very sensitive to others. He can get distressed over sad moments in shows he is watching, and he is very attuned to others' feelings. However, she reported that he recovers quickly from moments of distress and responds appropriately to support. There are no reports of educational, emotional, or social difficulties in Jacob's family members or relatives.

The interviewer noted in particular that when Mrs. W. was asked to indicate Jacob's strengths, she became emotional; she said that this is something that she never hears about Jacob. She indicated that reports from school are almost always negative, and that it was hard for her to think of his strengths. She was able to identify his mood and attitude as strengths, as well as his memory, comprehension, and social skills. The following day, a conference was held with the teacher about Jacob's mother's response, and it was agreed that daily communications could be changed to include more positive feedback to the family.

Medical/Developmental History

Jacob is small for his age, based on parental report. He was the result of a full-term, healthy pregnancy but did experience complications at birth. He had fluid in his lungs and spent 2 days in the NICU before he came home; there were no reports of anoxia, and his mother reported that there were no further complications. Jacob met most cognitive-developmental milestones as expected, although he was delayed in motor development. Although Jacob was an early talker, he speaks softly and mumbles at times. However, his mother reported that he has a large vocabulary and that he speaks and communicates well. She reported that his hearing and vision are within normal limits. Mrs. A. noted Jacob's quiet speech and confirmed his vocabulary and communication skills.

Jacob has a history of poor muscle tone. He has poor balance, and he is not as well developed motorically as his peers. For example, he cannot stand on one foot, and he runs with his arms in close to his body. Jacob's mother reports that he has been behind on motor development since birth. His teacher reported particular difficulty with eating and swallowing, and noted that Jacob eats finger foods by inserting his entire hand into his mouth. His mother reported that when he began eating, he frequently gagged and choked due to weak throat muscles. However, she stated that this has improved greatly, and that his swallowing and eating behavior is related to the foods he is eating. She cuts up his food into small pieces to accommodate his eating difficulties.

Educational History

Jacob started kindergarten in the fall and is currently in his sixth month of instruction. Mrs. A. reported that she has to frequently remind Jacob to focus and complete his work. She reported some regression of learned skills (e.g., spelling, months of the year) over the year. Jacob has not mastered his numbers and is not yet counting consistently above 20. She reported that reading fluency is also a problem, although she has seen some improvement in reading and spelling recently. In addition, he is slow in completing his work and takes longer to learn new material. Mrs. A. has placed Jacob at the front of the classroom to provide him with more support. She gives him instructions and has him repeat them back to her; often she repeats the instructions.

Jacob is working with a dyslexia therapist on his handwriting skills. His tracing and production of letters has improved, but reports indicate that he is not on grade level. Mrs. A. is recommending that he repeat kindergarten to allow him greater opportunity to learn and to catch up with his peers. His parents agree with this recommendation, but are concerned with his slower progress and with ways to help him.

Previous Test Results

Jacob completed a psychological evaluation through his local school district and had an evaluation with a developmental pediatrician. Parents reported that both evaluations showed that Jacob was in the average range. These prior reports were not available for this report.

BEHAVIORAL OBSERVATIONS

Classroom Observations

A classroom observation was conducted to observe Jacob in his classroom environment. At the beginning of the observation, he was sitting quietly at his desk. He had drawn over his paper instead of completing his work, and was now completing a “time out” while waiting his turn for individual reading instruction. He was observing everything in the room and fiddling with his papers, but he was not attempting to interact with other classmates or the teacher, or to get out of his seat and move around the classroom. He was dressed casually and wore a baseball cap. He had a friendly smile and made eye contact when introduced. During reading, he was very attentive to the teacher, but frequently had to be redirected to the reading task. His reading fluency was slow and did not increase when he was prompted; however, he was generally accurate in his reading, although he had more difficulty with blended word sounds. He utilized a finger-pointing technique to keep track of where he was on the page, and the teacher turned pages for him. Following reading, he went to a second center in the classroom, where he was required to use scissors. He had a loose gait and stumbled while moving between centers. He had mild difficulty with the scissors, and his cutting was imprecise. He placed all pieces of the paper he cut into a pile instead of sorting out the sections he needed to glue to his paper. He frequently looked to the teacher and his classmates for direction on what he should be doing.

Test Session Observations

Jacob was always happy and willing to participate in testing. Testing was completed in an empty, undecorated classroom, free from both visual and verbal distractions. He was talkative and repeatedly made conversation as he completed tasks. He frequently jumped from topic to topic and often reacted to stimuli on the tasks by relating them to his own experiences. For example, upon seeing a fish with headphones, he said, “He must be listening to music like I do.” He was hesitant to guess and often asked questions about the stimuli instead of answering the questions. He was concerned about his performance at times, asking if he got a question correct. However, he was more interested in completing the testing so he could go back to his friends.

Jacob was tested across four sessions, as he had difficulty maintaining performance beyond 45 minutes. He had difficulty maintaining attention to tasks and frequently asked to stop or do something else. However, he responded to redirection and prompting to continue on tasks. On lengthier tasks, he required frequent prompting to continue on the tasks. Several times when he stopped working, he stared blankly at the examiner or materials and had to be prompted to return to the task. When this occurred, he required the directions to be repeated before he completed an item. His pace throughout testing was fairly slow, and he frequently made mistakes. Jacob made a lot of self-corrections throughout testing, which increased the length of time it took to complete tasks, but did increase his accuracy.

TEST RESULTS AND INTERPRETATION

Behavior Reports

Jacob’s mother and teacher completed two behavioral rating scales assessing behavioral problems and executive functioning. The results are displayed in Table 12.A.1. Executive functioning is the general ability to plan, organize, and direct behavior. For most areas, the reports were very similar. Jacob exhibits difficulty with attention, executive functioning, working memory, and problem solving. The difficulties with attention and executive functioning are severely elevated both at home and school. In addition, Jacob exhibits normal levels of activity and impulsive behaviors in both environments. Similar behaviors are reported across environments as well, with the need for frequent prompting, difficulty following directions, and difficulty completing multiple tasks reported as the most stressful by both raters.

Jacob’s mother and teacher differed somewhat in their ratings of his behavior problems, with greater difficulty generally being reported in the classroom than at home. Jacob exhibits greater withdrawal symptoms (e.g., refuses to talk, prefers to play alone, isolates self from others) and greater social problems (e.g., poor adjustment to changes in routine, slow involvement in group activities) in the classroom. He demonstrates better adaptability and social skills at home, and his mother reported good resilience (e.g., recovers well from setbacks or changes).

TABLE 12.A.1. Jacob's Behavioral Rating Scores (Teacher and Parent Forms)

Scale/index	Teacher *T* score (percentile rank)	Parent *T* score (percentile rank)
BASC-3		
Hyperactivity	43 (30)	42 (24)
Aggression	43 (25)	41 (14)
Externalizing Problems	42 (25)	40 (14)
Anxiety	45 (37)	48 (50)
Depression	46 (41)	50 (58)
Somatization	*60 (88)*	37 (7)
Internalizing Problems	50 (58)	44 (28)
Attention Problems	*68 (96)*	*63 (89)*
Atypicality	59 (86)	45 (40)
Withdrawal	*64 (90)*	38 (8)
Behavioral Symptom Index	55 (73)	46 (38)
Adaptability	46 (37)	60 (84)
Social Skills	41 (23)	61 (87)
Functional Communication	*38 (14)*	54 (58)
Activities of Daily Living	—	40 (16)
Adaptive Skills	*40 (20)*	55 (64)
DREF		
Behavioral Functioning	50 (50)	48 (42)
Emotional Functioning	51 (54)	47 (38)
Executive Functioning	*78 (>99)*	*65 (93)*
Total Composite	*61 (86)*	54 (66)
Attention/Working Memory	*72 (99)*	*70 (98)*
Activity Level/Impulse Control	47 (38)	45 (31)
Compliance/Anger Management	52 (58)	54 (66)
Abstract Thinking/Problem Solving	*74 (99)*	*61 (86)*
Teacher top stressors		
Is off task when he is supposed to do classwork.		
Cannot do two or more tasks at the same time.		
Does not start classwork without extra prompting.		
Has trouble completing tasks like homework or class projects.		
Has trouble following directions.		
Parent top stressors		
Is off task when he is supposed to do homework or chores.		
Loses track of what he is supposed to do, due to noises or other things going on.		
Forgets what he is supposed to do.		
Cannot do two or more tasks at the same time.		
Has trouble following directions.		

Note. Italics indicate at-risk or clinically significant results.

In general, Jacob's behaviors at home and school fall into the normal or average range. However, attentional control was elevated on the BASC-3 both at home and at school. Moreover, his teacher and parent ratings on the DREF indicated severe elevations in attention/working memory and in executive functioning. The major stressors indicated by both his teacher and mother included being off task when he is supposed to be completing work/chores, having difficulty completing multiple tasks, and having trouble following directions.

Assessment of Cognitive Abilities

General cognitive ability refers to an individual's overall ability to reason, solve problems, and learn information in an efficient and timely manner. Jacob was administered the KABC-II NU, a comprehensive test of general cognitive abilities, to evaluate his overall level of functioning as well as provide a profile of cognitive abilities. The results are displayed in Table 12.A.2. He obtained a KABC-II NU Fluid–Crystallized Index (FCI) of 91 (95% confidence interval [CI] = 87–95), placing his general cognitive ability at the 27th percentile and classifying it in the average range.

Jacob's intellectual functioning is made up of a profile of specific cognitive abilities, which provides far greater information than a single overall score. Jacob's cognitive skills are in the average range for his age. His profile of scores on the KABC-II NU suggests some variability across specific cognitive abilities. Jacob's index scores ranged from 85 (16th percentile) to 104 (61st percentile). A closer review of his performance on the individual scales provides a better description of Jacob's abilities.

Jacob demonstrated a personal relative strength in crystallized ability, which reflects his breadth and depth of specific knowledge acquired within a culture, as well as his ability to apply this knowledge effectively. He obtained a Knowledge/Gc Index score of 104 (95% CI = 96–111; 61st percentile), which is also a measure of verbal ability and reflects Jacob's observed verbal skills. His strongest performance was observed on Riddles, a task requiring Jacob to listen to a series of verbal characteristics describing a concept, and then to point to or name the described concept. This task requires reasoning as well as knowledge of words and concepts.

Jacob's ability to learn new information, to store that information in long-term memory, and to retrieve it fluently and efficiently was average. He obtained a Learning/Glr scale score of 97 (95% CI = 94–100; 42nd percentile). His performance on two tasks requiring him to associate verbal labels with visual stimuli demonstrated good acquisition of and immediate recall of the information. However, his delayed recall of this learned information was in the lower extreme, with a Delayed Recall scale score of 66 (95% CI = 56–76; 1st percentile). On a task requiring him to recall verbal labels for visual stimuli, he was unable to produce any correct responses 20 minutes after learning the information. He did slightly better on a task requiring him to recognize and identify visual stimuli when provided with the verbal label.

Jacob demonstrated relative significant weaknesses in his ability to process visual and spatial information mentally and in his short-term memory abilities (when stimuli are presented in sequential fashion). His score of 85 (95% CI = 80–92; 16th percentile) on the Simultaneous/Gv scale reflects relatively lower ability to perceive, manipulate, and think with visual patterns and stimuli, and to mentally rotate objects in space. His weakest subtest performance was observed on a task requiring him to assemble plastic shapes or triangular blocks to match a picture of a design shown in a stimulus book. His relatively lower performance was influenced by his constructional abilities; however, most of his failures were due to misplaced shapes in his constructions. On tasks requiring reasoning utilizing visual stimuli, Jacob demonstrated slightly better performance.

Additionally, Jacob demonstrated a relative weakness in his short-term or working memory ability, which can be defined as the ability to take in information, hold it in immediate awareness briefly, and perform some mental manipulation or recall of that information within a few seconds. Jacob obtained a Sequential/Gsm Index score of 85 (95% CI = 79–93; 16th percentile). On a task requiring him to repeat a series of numbers, he had difficulty consistently recalling spans of greater than three digits. Interestingly, on the second task—which required him to listen to a series of words and then point to the correct pictures of the common objects named by the examiner—he also had difficulty recalling a span of longer than three words. Although short-term memory and visual processing were both relative weaknesses for Jacob, they were both in the average range for his age.

Assessment of Achievement

The KTEA-3 Comprehensive Form was administered to assess Jacob's current academic achieve-

TABLE 12.A.2. Jacob's Scores on Cognitive Testing

Index/scale	Index/standard score (mean = 100, *SD* = 15)	Subtest scaled score (mean = 10, *SD* = 3)	*T* score	Percentile rank
	KABC-II NU			
Sequential/Gsm	85			16
Simultaneous/Gv	85			16
Learning/Glr	97			42
Knowledge/Gc	104			61
Fluid–Crystallized Index (FCI)	91			27
Delayed Recall	66			*1*
	KTEA-3			
Reading	99			47
Math	*78*			*7*
Written Language	89			23
Academic Skills Battery	85			16
Oral Fluency	65			*1*
Comprehension	105			63
	NEPSY-II			
Auditory Attention Total Correct		6		*11–25%*
Auditory Attention Combined		*7*		*11–25%*
Omission Errors				*6–10%*
Commission Errors				51–75%
Inhibition Naming Completion Time		*2*		*≤2%*
Inhibition Naming Combined		*3*		*≤2%*
Naming Total Errors				*6–11%*
Naming Self-Corrected Errors				*<2%*
Inhibition Inhibition Completion Time		*4*		*3–10%*
Inhibition Inhibition Combined		*5*		*3–10%*
Inhibition Total Errors				*11–25%*
Inhibition Self-Corrected Errors				*2–5%*
Inhibition Total Errors		6		*11–25%*
	BOT-2 (male norms)			
Fine Manual Control			*38*	*12*
Manual Coordination			*34*	6
Body Coordination			*36*	8
Strength and Agility			*37*	*10*
Total Motor Composite			*34*	6

Note. Italics indicate at-risk or clinically significant results.

ment. The results are displayed in Table 12.A.2. Jacob scored in the average range on the Academic Skills Battery, reflecting his overall achievement. There was significant variability in his composite scores, ranging from a standard score of 99 (95% CI = 93–105; 47th percentile) on the Reading Composite to a standard score of 78 (95% CI = 73–83; 7th percentile) on the Math Composite. His Math Composite was a both a normative and relative weakness for him. His Written Language Composite of 89 (95% CI = 84–95; 23rd percentile) included notable score variability in its subtests. Jacob obtained a significantly higher score on spelling words spoken by the examiner (Spelling standard score = 106; 82nd percentile) than on expressing himself in writing sentences (Written Expression standard score = 74; 4th percentile).

Reading

Jacob obtained his highest performance on the Reading Composite. The reading skills measured include letter–sound correspondence (phonics), word recognition, decoding, fluency, and comprehension. Jacob's overall reading skills were in the average range. In addition to the reading measures included in the Reading Composite, Jacob completed measures of Phonological Processing and Oral Fluency. Both phonological processing and rapid automatic naming have been found to be critical to academic achievement. Jacob's performance on Phonological Processing was average for his age range (standard score = 94; 34th percentile). However, his performance on fluency measures was in the low average to low range (Oral Fluency Composite = 65; 1st percentile). His performance on these tasks was consistent with teacher and parent reports of Jacob's slowness on tasks in the classroom and at home.

Math

Jacob's math computation and arithmetic skills, and his math problem-solving skills, were below average. His performance was comparable on both subtests (standard scores of 79–80), indicating difficulty in his ability to apply mathematical principles to real-life problems and to perform mathematical calculations accurately. Many of the errors made in math appeared to be related to poor understanding of math structure. For example, on problems such as 1 + 1, he would respond by writing 11.

Writing

As noted earlier, Jacob demonstrated variability in his performance across writing tasks. When he was completing the tasks, he utilized an adaptive pencil grasp. As he produced letters, he used incorrect letter formation on multiple letters and frequently identified a letter out loud that did not match the letter he was writing (e.g., he stated *b* and madet he *b* sound while writing the letter *d*). Several times during spelling, he began a word and wrote the second letter to the left of the first letter instead of to the right. He usually caught this and self-corrected to spelling left to right. However, once he wrote the entire word in reverse before noticing his error. His handwriting was mostly legible, but generally poor. His spelling ability was average, but many of the letters were poorly formed, and it took him a long time to write each letter.

Comprehension

Jacob's comprehension ability both from reading passages and orally presented information was average (Comprehension standard score = 105; 63rd percentile). His Listening Comprehension (standard score = 110; 75th percentile) was higher than his Reading Comprehension (standard score = 98; 45th percentile), although both scores were in the average range.

Assessment of Neuropsychological Abilities

Additional testing was completed to assess Jacob's cognitive skills. He completed several subtests of the NEPSY-II; the results are displayed in Table 12.A.2. On measures of attention, he was in the borderline range overall, but produced a high rate of omission errors (percentile rank = 6–10%) in comparison to a relatively low rate of commission errors (percentile rank = 51–75%). Omission errors are failures to respond when a response is required (e.g., indicative of inattention), while commission errors are responses when a response is not required (e.g., indicative of impulsivity or poor inhibition). On measures of inhibition, Jacob also demonstrated difficulty, although his performance on these tasks was influenced by slow responding time and a high rate of errors. However, it is important to note that he corrected most of his errors, which added to his slow time. This finding is consistent with his performance on the fluency measures on the KTEA-3. Jacob's performance on measures of language, visuospatial processing, and memory and learning was consistent with his performance on the KABC-II NU and KTEA-3. He demonstrated relatively intact language and learning abilities, but difficulties in visual–spatial processing.

Assessment of Motor Abilities

Motor functioning includes fine motor skills such as handwriting and tracing, as well as gross motor skills such as balance, coordination, and gait. Jacob's fine and gross motor abilities are below what is expected for his age (BOT-2 Total Motor Composite = 34; below average). Again, the results are displayed in Table 12.A.2. The majority of his scores fell into the below-average range on the BOT-2. He had a relatively better performance on the fine motor integration task, which required him to copy shapes. As he did on the written tasks on the KTEA-3, he utilized an immature pencil grasp.

DIAGNOSTIC IMPRESSIONS

The following *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5) diagnoses are well supported:

F90.0 Attention-deficit/hyperactivity disorder, predominantly inattentive presentation
F82 Developmental coordination disorder (need to rule out medical or neurological cause)

The following DSM-5 diagnosis is suggested:

F81.2 Specific learning disorder, with impairment in mathematics

SUMMARY

Jacob, a 5-year old attending kindergarten in a private school, was referred for an evaluation by his parents and teacher due to concerns about his inattention, difficulties in school, and motor difficulties. Jacob was cooperative with testing, but required multiple testing sessions to accommodate his difficulty in maintaining effort and attention. He demonstrated difficulty maintaining attention and focus on tasks both during the testing sessions and during a classroom observation.

Jacob demonstrated several behavioral strengths during the assessment. He exhibited a positive attitude and strong social skills, which are also reflected in his good relationships with his family and peers. In addition, he was receptive to feedback and returned to work when prompted without struggle or complaint. However, parent and teacher reports as well as clinical observations indicated significant difficulties with sustaining attention, initiating work, and completing tasks in a timely manner.

Jacob demonstrated several cognitive and academic strengths. His general cognitive ability was in the average range, as were all of his KABC-II NU index scores. He exhibited a relative strength in his verbal crystallized ability, and relative cognitive weaknesses in visual processing and short-term memory; however, all his scores fell within the average range with the exception of his Delayed Recall index score, which was low. Jacob demonstrated significant variability in his composite scores on the KTEA-3, with a significant strength in his reading abilities and a normative and relative weakness in his math abilities. Additionally, he demonstrated a relative strength on tests of reading and listening comprehension. Further cognitive testing demonstrated difficulties with attention and executive functioning, and supported the strengths and weaknesses observed on the KABC-II NU. Finally, Jacob demonstrated significant motor delays in relation to his peers. Nonetheless, he can improve his achievement if he is shown how to use his cognitive strengths to facilitate school learning.

RECOMMENDATIONS

Further Assessment and Reevaluation

- Given Jacob's young age and developing skills, it is highly recommended that he be reevaluated in a year, to track his progress and develop appropriate supports if needed.
- It is recommended that a full medical and neurological evaluation be completed, to rule out any underlying neurological or medical causes for Jacob's motor and attentional difficulties.
- An oral–motor swallowing evaluation by a speech–language pathologist to determine any need for intervention with swallowing is also recommended.
- An evaluation by an occupational or physical therapist is strongly recommended, to create a

specific treatment plan for developing Jacob's motor abilities and to establish appropriate targets for intervention.

General Recommendations

- Utilize Jacob's strengths to help him on more difficult tasks or tasks that require his cognitive weaknesses.
    - Allow spoken responding if written responses are not required.
    - If he has difficulty initiating a response, provide a cue or choices for him to demonstrate learning.
    - For tasks in which speed is not essential, allow him additional time to complete tasks.
    - For tasks that require time, reward improvements in performance accuracy or completion.

Executive Functioning

- Use a consistent behavior system at both school and home.
- Develop predictable routines (e.g., regular school schedule, homework at same time).
- Develop his executive skills through appropriate modeling, chaining, and practice:
    - Modeling:
        - Pair Jacob with a peer for tasks requiring multiple steps or complex steps.
        - Use a talk-aloud procedure to help Jacob plan steps.
    - Chaining:
        - Use checklists for completing tasks (visual or written) that have steps in sequential order. Have him review a full list and then work on the first step. When possible, review the first step before moving to the next step, so he begins to put the steps together. When there are more steps, always review the steps he has completed. As he is learning a new task, it will help him to repeatedly go through all the steps he needs to complete a task.
    - Practice:
        - Have Jacob talk through his tasks as he learns them; this will take advantage of his verbal abilities and allow him to put steps together.
        - Utilize a daily behavioral self-report card. Include only a few key behaviors on this card, and have him evaluate his performance on these. This way, he can begin to direct his behavior instead of being evaluated by others only.

Attention/Working Memory

- Reduce distractions in classroom and other working environments when possible during learning and testing times.
- Seat Jacob in a location that maximizes his ability to pay attention (nearest to teacher, near kids with good attention skills, away from visual distractions).
- Utilize multiple modalities to teach constructs (e.g., use both verbal and visual teaching when introducing new concepts).
- Break tasks into short segments, and provide visual cues for steps in multistep tasks. Use memory aids (e.g., pictorial graphic organizers) as reminders of the instructions or to guide the completion of an activity.
- Provide incentives for sustained attention.
    - Begin with requiring attention for short lengths of time, and then extend these as Jacob achieves his goals.
    - Use a high-interest activity to reinforce attention on a low-interest activity.
- Provide cues to signal off-task behavior and improve Jacob's self-awareness. Consider nonverbal cues (such as tapping a finger on a reminder card on his desk). As he progresses, these cues can be minimized.

Motor Skills

- Have Jacob continue working with the intervention specialist on handwriting and fine motor skills.
- Encourage activities that involve fine motor abilities, such as coloring, puzzles, cutting, and drawing.
- Encourage other activities that involve gross motor activities, such as playing catch, throwing a ball, or physical exercises.
- Jacob will take longer than most children to complete physical tasks, including those required for schoolwork (e.g., cutting with scissors, writing). He should be given additional time to complete these activities.

Mathematics

- Create math games out of everyday activities (e.g., "How many socks do I have?," weighing things, counting together)

- Look into elementary learning games online that teach basic skills.

Language and Literacy

- Reading fluency:
    - Have Jacob practice reading new or challenging words in isolation prior to reading them in text.
    - Have him orally read "phrase cards," and track the time needed to read them accurately.
    - Use the *neurological impress* method (paired reading, where the student and an adult read the same text almost simultaneously).
- Writing:
    - Continue working with Jacob on handwriting and letter formation skills to improve his writing fluency. Encourage proper letter formation, and only work on speed once letter formation is well established and legible.

REFERENCES

Bain, S. K., & Gray, R. (2008). Test reviews: Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition*. Circle Pines, MN: American Guidance Service.

Bruininks, R. H., & Bruininks, B. D. (2005). *Bruininks–Oseretsky Test of Motor Proficiency, Second Edition*. Circle Pines, MN: American Guidance Service.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 84–112). New York: Oxford University Press.

Carroll, J. B. (1943). The factorial representation of mental ability and academic achievement. *Educational and Psychological Measurement, 3,* 307–332.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cattell, R. B. (1941). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Dale, B. A. (2009). *Profile analysis of the Kaufman Assessment Battery for Children, Second Edition with African American and Caucasian preschool children*. Unpublished doctoral dissertation, Ball State University, Muncie, IN.

Dale, B. A., McIntosh, D. E., Rothlisberg, B. A., Ward, K. E., & Bradley, M. H. (2011). Profile analysis of the Kaufman Assessment Battery for Children, Second Edition, with African American and Caucasian preschool children. *Psychology in the Schools, 48,* 476–487.

Delis, D. C. (2012). *Delis Rating of Executive Function*. Bloomington, MN: Pearson.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Fletcher-Janzen, E. (2003, August). Neuropsychologically-based interpretations of the KABC-II. In M. H. Daniel (Chair), *KABC-II: Theory, content, and interpretation*. Symposium presented at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

Fujita, K., Ishikuma, T., Aoyama, S., Hattori, K., Kumagai, K., & Ono, J. (2011). Theoretical foundation and structure of the Japanese version of the KABC–II. *Japanese Journal of K-ABC Assessment, 13,* 89–99. [In Japanese]

Gomez, M. T. (2008). *The performance evaluation of low achieving Mexican American students on the Kaufman Assessment Battery for Children II (KABC-II): The role of English language proficiency*. Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242–259.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 267–300). New York: Wiley.

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 61–116). New York: Freeman.

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 44–99). New York: Cambridge University Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hunt, M. S. (2008). A joint confirmatory factor analysis of the Kaufman Assessment Battery for Children, Second Edition, and the Woodcock–Johnson Tests of Cognitive Abilities, Third Edition, with preschool children. *Dissertation Abstracts International, 68*(11), 4605A.

Ishikuma, T., Matsuda, O., Fujita, K., & Ueno, K. (2016). Intellectual assessment of children and youth in Japan: Past, present, and future. *International Journal of School and Educational Psychology,4,* 241–246.

Ishikuma, T., Shinohara, Y., & Nakao, T. (2007). School psychology in Japan. In S. Jimerson, T. Oakland, & P. Farrell (Eds.), *Handbook of international school psychology* (pp. 217–227). Thousand Oaks, CA: SAGE.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children interpretive manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Manual for the Kaufman Assessment Battery for Children—Second Edition (KABC-II)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Manual for the Kaufman Test of Educational Achievement—Second Edition (KTEA-II), Comprehensive Form*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2008). *KABC-II Batterie pour L'Examen Psychologique de L'Enfant—Deuxième Edition*. Montreuil, France: Editions du Centre de Psychologie Appliqué.

Kaufman, A. S., & Kaufman, N. L. (2014). *Manual for the Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Comprehensive Form*. Bloomington, MN: Pearson.

Kaufman, A. S., Kaufman, N. L., Drozdick, L. W., & Morrison, J. (2018). *Kaufman Assessment Battery for Children—Second Edition Normative Update (KABC-II NU), manual supplement*. Bloomington, MN: NCS Pearson.

Kaufman, A. S., Kaufman, N. L., Melchers, P., & Melchers, M. (2014). *German-language adaptation of the Kaufman Assessment Battery for Children—Second Edition*. Frankfurt, Germany: Pearson.

Kaufman, A. S., Kaufman, N. L., & Publication Committee of Japanese Version of the KABC-II. (2013). *Japanese Version of Kaufman Assessment Battery for Children—Second Edition*. Tokyo: Maruzen. [In Japanese]

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: Wiley.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Hoboken, NJ: Wiley.

Kaufman, J. C. (2003, August). Gender and ethnic differences on the KABC-II. In M. H. Daniel (Chair), *The KABC-II: Theory, content, and administration*. Symposium presented at the annual meeting of the American Psychological Association, Toronto, Ontaio, Canada.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. (2012). Are cognitive *g* and academic *g* one and the same *g*?: An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123–138.

Keith, T. Z. (1985). Questioning the K-ABC: What does it measure? *School Psychology Review, 14,* 9–20.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY—Second Edition*. San Antonio, TX: Harcourt Assessment.

Lichtenberger, E. O. (2005). General measures of cognition for the preschool child. *Mental Retardation and Developmental Disabilities, 11,* 197–208.

Lichtenberger, E. O., & Breaux, K. C. (2010). *Essentials of WIAT-III and KTEA-II assessment*. Hoboken, NJ: Wiley.

Liu, X., Marchis, L., DeBiase, E., Breaux, K. C., Courville, T., Pan, X., et al. (2017). Do cognitive patterns of strengths and weaknesses differentially predict errors on reading, writing, and spelling? *Journal of Psychoeducational Assessment, 35,* 186–205.

Luria, A. R. (1966). *Human brain and psychological processes*. New York: Harper & Row.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222,* 66–78.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. Harmondsworth, UK: Penguin.

Malda, M., van de Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (2010). Traveling with cognitive tests: Testing the validity of a KABC-II adaptation in India. *Assessment, 17,* 107–115.

Markwardt, F. C., Jr. (1998). *Peabody Individual Achievement Test—Revised, Normative Update*. Circle Pines, MN: American Guidance Service.

McGill, R. J. (2016). Invalidating the Full Scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology, 6,* 49–79.

McLean, J. E. (1995). *Improving education through action research: A guide for administrators and teachers*. Thousand Oaks, CA: Corwin Press.

Morgan, K. E., Rothlisberg, B. A., McIntosh, D. E., & Hunt, M. S. (2009). Confirmatory factor analysis of the KABC-II in preschool children. *Psychology in the Schools, 46,* 515–525.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition (CAS2)*. Austin, TX: PRO-ED.

Potvin, D. C. H., Keith, T. Z., Caemmerer, J. M., & Trundt, K. M. (2015). Confirmatory factor structure of the Kaufman Assessment Battery for Children—Second Edition with preschool children: Too young for differentiation? *Journal of Psychoeducational Assessment, 33,* 522–533.

Prifitera, A., Weiss, L. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 1–38). San Diego, CA: Academic Press.

Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 3–32). Boston: Elsevier Academic Press.

Psychological Corporation. (2001). *Wechsler Individual Achievement Test—Second Edition*. San Antonio, TX: Author.

Reitan, R. M. (1988). Integration of neuropsychological theory, assessment, and application. *Clinical Neuropsychologist, 2,* 331–349.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior*

Assessment System for Children, Third Edition (BASC-3). Bloomington, MN: Pearson.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence, 35,* 267–281.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children—Second Edition: Consistency with Cattell–Horn–Carroll theory. *School Psychology Quarterly, 22,* 511–539.

Reynolds, M. R., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence, 26,* 236–260.

Scheiber, C. (2016). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment, 28,* 943–952.

Scheiber, C. (2017). Does the KABC-II display ethnic bias in the prediction of reading, math, and writing in elementary school through high school? *Assessment, 24*(6), 729–745.

Scheiber, C., & Kaufman, A. S. (2015). Which of the three KABC-II global scores is the least biased? *Journal of Pediatric Neuropsychology, 1*(4), 21–35.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Vannetzel, L. (2010). Le KABC-II. *Approche Neuropsychologique des Apprentissages Chez l'Enfant, 107–108,* 231–234.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. New York: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition*. San Antonio, TX: Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition*. Bloomington, MN: Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition*. Bloomington, MN: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III*. Itasca, IL: Riverside.

CHAPTER 13

The Differential Ability Scales—Second Edition

Colin D. Elliott
Joseph D. Salerno
Ron Dumont
John O. Willis

The late Colin D. Elliott (1937–2016) studied the ability profiles of children with learning disabilities (LDs) and the measurement of children's developmental stages in his work as an educational psychologist for 7 years, prior to his becoming a trainer of school psychologists for over 20 years at the University of Manchester in the United Kingdom. In 1983, he made a major contribution to the field of psychometric assessment in both his home country and the wider field with the publication of the British Ability Scales (BAS; Elliott, 1983a, 1983b). An American version followed, called the Differential Ability Scales (DAS; Elliott, 1990a, 1990b). The bright and attractive stimuli made the test engaging for children, while the selection of subtests and capacity to tailor testing to a child's age and ability level made it efficient for examiners to administer and score. New features from the DAS were incorporated into the second edition of the BAS (BAS II; Elliott, 1996), which in turn influenced the development of the current American version, the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007a, 2007b, 2007c, 2007d). A new version of the British version, the British Ability Scales 3 (BAS3; Elliott & Smith, 2012), has since been published.

Citing controversies and widespread popular misunderstandings in the definition of the terms *intelligence* and *IQ,* Elliott preferred to focus on ability and specific abilities in various cognitive domains. Elliott wrote:

> The DAS-II certainly produces a second-order composite score called the General Conceptual Ability score (GCA). However, *this is not a global composite score such as is found in other batteries.* It is not derived from a heterogeneous mix of all subtests, but is derived only from either six or four subtests that are the best measures of conceptual and reasoning abilities. (2007c, p. xiii; emphasis in original)

Therefore, the DAS-II was designed as a profile test that yields reliable and focused subtest and composite scores for various abilities and the higher-order General Conceptual Ability (GCA)

This chapter is a revision of Colin D. Elliott's chapter in the third edition of *Contemporary Intellectual Assessment* (Elliott, 2012). Shortly before his death, which was a tremendous loss to the fields of psychology and education as well as to his beloved family, his friends, and his many admirers, Dr. Elliott had, at the invitation of Dawn P. Flanagan, informally agreed to revise his previous chapter (in collaboration with Ron Dumont and John O. Willis) for the present edition. Much of this chapter was written by Dr. Elliott, and we have endeavored to follow both his style and his beliefs about assessment in the new material we have added.

composite, which is derived from only the subtests that are the best measures of higher-level conceptual and reasoning abilities; it is not an assortment of various cognitive abilities, as other tests' general composite scores are. Many of the features for which Elliott advocated have become more popular in other tests and have influenced the field of cognitive ability test construction. Although, sadly, Colin Elliott is no longer with us, his ideas and innovations live on—both in psychological tests influenced by his innovations in the five editions of the BAS and DAS, and in the daily practice of evaluators who have been educated and guided by the manuals for those tests.

STRUCTURE OF THE DAS-II

The DAS-II consists of 20 individually administered subtests divided into two overlapping batteries. The Early Years battery is normed for children ages 2 years, 6 months to 8 years, 11 months (2:6–8:11). The School-Age battery is normed for children ages 5:0–17:11. Thus the Early Years and School-Age batteries were co-normed on children ages 5:0–8:11. The overlap provides important advantages in terms of out-of-level testing because examiners assessing bright younger children or less able older ones are able to select a battery containing six tests that are appropriate for each child's ability and level of cognitive development.

This flexible design is rare, if not unique, among tests of cognitive abilities. The current Wechsler intelligence scales (Wechsler, 2008, 2012, 2014) offer slightly different instruments spanning the age range from 2:6 to 90, but examiners must use three separate test batteries with three different norming samples. The current edition of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV; Wechsler, 2012) covers the age range from 2:6 through 7:7 with seven subtests (five included in the Full Scale IQ) for ages 2:6 through 3:11, and 17 subtests (six for the Full Scale IQ) for ages 4:0 through 7:7. All of the subtests for younger children are included in the form for older children, and three of the Full Scale IQ subtests are the same for both age groups. However, like most tests with different subtests for different ages (and unlike the DAS-II), the WPPSI-IV does not provide norms that would allow a child older than 3:11 to take the younger version, or a child under age 4:0 to take the older version. The overlapping normative age ranges provided by the DAS-II norms constitute a tremendous asset for assessing low-scoring older children and high-scoring younger ones. Additional norms, permitting even greater ranges of our-of-level assessment can be found in Dumont, Willis, and Elliott (2009) and were being considered by Elliott for inclusion in the next edition of the DAS (C. D. Elliott, personal communication, May 15, 2014). The lack of norms beyond age 17:11 is a nuisance for examiners working with older high school students and adults. Elliott also hoped to extend the age range above age 17 in the next edition (C. D. Elliott, personal communication, May 15, 2015).

Six core subtests in each battery contribute a composite score: the GCA, focused on reasoning and conceptual abilities; a Special Nonverbal Composite (SNC); and three lower-level composite scores called *cluster scores*. All the core subtests are highly *g*-saturated.

In addition to the core subtests, up to 10 diagnostic subtests are available for children taking the Early Years battery, and up to eight are available for those taking the School-Age battery. These diagnostic subtests yield three additional cluster scores, and also measure other specific abilities that do not contribute to the composites.

The youngest children (ages 2:6–3:5) take a more limited range of subtests. This range is called the Early Years (Lower Level), and the 3:6–8:11 range is called the Early Years (Upper Level). The overall structure of the DAS-II is summarized in Table 13.1.

CLINICAL AND THEORETICAL UNDERPINNINGS

Users of cognitive ability tests typically compare an individual's performance and responses with those of a representative population sample of the same age as the person being tested. Paradoxically, however, while test scores are nearly always compared with the norm, the children, students, or adults whom users assess are often extremely atypical. The bread-and-butter work of a psychologist or psychological examiner involves individuals who are having difficulties or are failing to learn under normal conditions, or whose behavior gives cause for concern.

Because the individuals who are typically referred to psychologists manifestly have a huge range of individual special needs, the clinical priorities in the development of the DAS-II and its predecessors have been as follows:

TABLE 13.1. Number of DAS-II Subtests and Composites in Each Battery

Battery	Number of subtests	General composite	Cluster scores
Early Years (Lower Level), ages 2:6–3:5	4 core 3 diagnostic	1. GCA 2. SNC	*Core clusters* 1. Verbal 2. Nonverbal
Early Years (Upper Level), ages 3:6–8:11	6 core 10 diagnostic	1. GCA 2. SNC	*Core clusters* 1. Verbal 2. Nonverbal Reasoning 3. Spatial *Diagnostic clusters* 4. School Readiness 5. Working Memory 6. Processing Speed
School-Age, ages 5:0–17:11	6 core 8 diagnostic	1. GCA 2. SNC	*Core Clusters* 1. Verbal 2. Nonverbal Reasoning 3. Spatial Ability *Diagnostic clusters* 4. Working Memory 5. Processing Speed

Note. GCA, General Conceptual Ability; SNC, Special Nonverbal Composite.

1. Many children referred for psychological assessment have a history of problems with attending to adult instruction (e.g., distractibility or short attention span). For such children, test materials need to be varied and engaging, using different formats and types of tasks. A uniform approach to the administration of all subtests (e.g., using easels) has always been deliberately avoided in the DAS-II and its predecessors, in the interest of engaging children with variety.

2. Because many low-functioning children are likely to be assessed, there need to be plenty of test items that afford an opportunity for teaching and demonstrating what is required in each task. This is to ensure as far as possible that the children who are being assessed understand what they are supposed to be doing. Young children, and children with developmental disabilities and specific learning disabilities (SLDs), often have difficulty in warming up to a task and initially understanding the task requirements. For this reason, most of the DAS-II subtests have demonstration and teaching procedures built into the administration. Such teaching items enable an examiner to tell a child that he or she is correct (if this is the case), and, if the item has been failed, to give and explain the correct response or solution. The provision of feedback for designated teaching items even when an item is passed (an unusual provision in cognitive ability tests) ensures that a confused child who makes a lucky guess is not deprived of needed guidance.

3. Again, because many low-functioning children are likely to be assessed with the DAS-II, its subtests have been designed to have low floors. A *floor effect* in test norms is shown when a group of individuals in an age group all get raw scores of 0 or 1 point on the test. Such individuals find even the easiest items in the test to be too difficult; thus the test cannot discriminate among them, and an examiner cannot derive much useful information about such a child. This results in the normative scores' being *inflated* for those who obtain such low raw scores. For example, if, say, 10% of children in an age group obtain a raw score of 0 on a test, standard scores cannot reasonably go below 80. So a child who is actually at the 1st percentile in relation to his or her age group does not get a standard score that reflects this (i.e., a score of about 65), but instead gets a score of about 80. Later tables show that the DAS-II subtests and composites have low floors.

4. Because many children have had considerable experience of failing tasks set them by adults, the tests should minimize children's experience of failure to the greatest possible extent, and should maximize their enjoyment in success. The *item set* approach to test administration, which is used

throughout the DAS-II, meets this objective. The extent to which this approach is used in nearly every subtest in the battery is unique to the DAS-II and its predecessors.

5. At a practical level, the item set approach has two major additional advantages. First, it keeps the testing session moving briskly, with subtests ending and new ones starting more quickly than if traditional administration methods were employed. This has a positive impact on children's motivation and willingness to remain in the test situation. The second advantage is for examiners: The battery is relatively quick to administer. The Early Years battery has a median administration time for the core subtests of 31 minutes or less, while the core subtests of the School-Age battery have a median administration time of less than 40 minutes.

6. The normative overlap between the Early Years and School-Age batteries, described earlier, is 1 year wider than it was in the original DAS, and is designed to enable examiners to assess children whose developmental level is atypical for their age. Thus the Early Years battery may be given to children up to the age of 8:11, yielding composite scores with identical interpretation to those in the School-Age battery, if an examiner determines that the Early Years materials are developmentally more appropriate for a child. The choice of appropriate subtests for a child is left to the examiner's judgment, not decided by a rigid formula.

7. Out-of-level testing procedures and an extended GCA (taking the lowest GCA score down to 25) are provided for children with SLDs and developmental disabilities. Again, therefore, a school-age child may be assessed with the Early Years materials if these are considered by the examiner to be more developmentally appropriate. These procedures are described in detail elsewhere (Dumont et al., 2009).

8. Finally, because professionals assessing children with LDs and developmental disabilities need information at a finer level of detail than IQ scores, the DAS-II and its predecessors were designed to reflect modern knowledge on the nature and structure of cognitive abilities. A *pattern of strengths and weaknesses* (PSW) approach to assessing LDs has been broadly accepted by many professionals and by the U.S. government (Individuals with Disabilities Education Improvement Act of 2004 [IDEA 2004]; Daniel, Breaux, & Frey, 2010; Flanagan, Ortiz, & Alfonso, 2013; Hale & Fiorello, 2004). Because the DAS-II incorporates such an approach, examiners using it may obtain measures of seven broad abilities and a range of narrow abilities that reflect current theory on the structure of human cognitive abilities. This is discussed further below.

The major *technical* priority in the development of the DAS-II has been to produce a battery in which subtests and cluster scores are individually interpretable. For this, they need to have substantial reliability and need to be distinct measures of different cognitive functions. The high specificity of the DAS-II subtests and clusters (see later discussion in this chapter), which supports such interpretations of specific abilities, is a distinguishing feature of the battery.

What of the *theoretical* model? When the many theories of the structure of abilities were reviewed at the time of the DAS's initial development in the early 1980s, it was apparent that no single theory was entirely persuasive, and certainly no single theory had universal acceptance among theoreticians or practitioners. Because of this, the original DAS and BAS were not developed solely to reflect a single model of cognitive abilities, but reflected an eclectic number of theoretical perspectives. They were designed to address processes that often underlie children's difficulties in learning, as well as what we then knew about the neurological structures underlying these abilities.

During the years since the original DAS was published, a growing consensus has developed among factor theorists of human abilities. This centers on what became widely referred to as Gf-Gc theory, after the initial theory development by Cattell (1971) and Horn (Cattell & Horn, 1978; Horn & Blankson, 2012; Horn & Noll, 1997). The basic theory—that variance among multiple measures of cognitive ability can be accounted for by numerous first-order, narrow abilities and 8–10 broad, second-order factors—was also independently demonstrated by Carroll (1993, 2003, 2012), who showed that there are considerable similarities in the factor structures of cognitive test batteries. The contributions by Carroll and by Horn and Blankson make explicit the disagreement between Carroll and Horn about the reality of the general factor, *g*; Carroll strongly supported the construct, whereas Horn considered it to be a statistical artifact. Other than this disagreement, the similarities in the factors described by these two authors are considerable and impressive.

Because of the convergence between Carroll's and Horn's theoretical positions on the hierarchical structure of human abilities, Gf-Gc theory has more recently been referred to as Cattell–Horn-Carroll (CHC) theory (McGrew, 2005; see Flanagan et al., 2013; Schneider & McGrew, 2012 and Chapter 3, this volume). It also appears to be a unifying theory about which most workers in the area of the structure of human abilities broadly agree (Flanagan et al., 2013; Schneider & McGrew, 2012). As a result of the development of CHC theory, and before development of the DAS-II began, McGrew (1997, 2005), McGrew and Flanagan (1998), Flanagan, McGrew, and Ortiz (2000), Alfonso, Flanagan, and Radwan (2005), and Elliott (2005) were making links between CHC theory and findings on the factor structure of the DAS. And interestingly, at about the same time as the DAS-II was published, Sanders, McIntosh, Dunham, Rothlisberg, and Finch (2007) reported a joint factor analysis of the DAS and the Woodcock–Johnson III Tests of Cognitive Abilities (WJ III COG; Woodcock, McGrew, & Mather, 2001). Sanders and colleagues concluded that a three-stratum model provided the best fit to the data, with the DAS subtests measuring six of seven CHC broad ability factors.

Because of these theoretical developments, it was decided that the new DAS-II would be linked to CHC theory. Accordingly, all subtests are classified as measures of both narrow and broad CHC abilities. With one exception, all cluster scores have CHC broad ability classifications. The exception is the School Readiness cluster in the DAS-II Early Years battery, which is formed from a combination of three subtests measuring different CHC factors. It appears likely that in the early years at school, most teachers teach visual matching, early number concepts, and phonological awareness—skills that are defined by three of the DAS-II subtests. These three subtests "hang together" and form a factor that has a pragmatic rather than a theoretical basis, and that may prove useful for examiners who assess children in the early school years.

CHC theory continues to be a work in progress concerning (1) the number of factors representing independent abilities in the model; (2) the precise nature of each factor (see, e.g., the discussion by Schneider and McGrew (2012) on whether tests of rapid naming measure the broad ability of Gs or Glr); and (3) whether and to what extent subtests from different test batteries that purport to measure a given factor actually do so (see Keith & Reynolds, 2010).

ORGANIZATION AND FORMAT

Test Structure and Content

The subtests in the DAS-II Early Years (Upper Level) and School-Age cognitive batteries are listed in Tables 13.2 and 13.3, respectively. In each table, the subtests are grouped according to whether they are designated *core* subtests or *diagnostic* subtests. Each subtest has a brief description of the nature of its task, including its CHC broad ability classification. The core subtests are relatively strongly *g*-related and therefore measure complex processing and conceptual ability.

Subtests and Clusters in the DAS-II

All subtests that were in the original DAS are included in the DAS-II, as well as four new subtests. Recall of Sequential Order and Recall of Digits—Backward form a new diagnostic Working Memory cluster. A new Rapid Naming subtest combines with Speed of Information Processing to form a new diagnostic Processing Speed cluster. And Phonological Processing has been introduced to reflect the research done on this ability in relation to reading disability since the original DAS was published.

Moreover, the Matrices subtest has been extended downward to age 3:6 with a new set of colored pictorial items suitable for young children, thereby enabling the DAS-II Nonverbal Reasoning cluster (a measure of Gf) to extend down to age 3:6.

Block Building from the DAS has been merged into the Pattern Construction subtest, enabling that subtest to extend down to age 2:6. Technically, it was found during development that the goodness of fit to the Rasch model of the Block Building items was excellent when they and the Pattern Construction items were both included in the analysis, and they could therefore be considered to measure the same latent dimension.

The School-Age core subtests are identical to those in the original DAS. Five of the Early Years core subtests are identical, the exception being that Early Number Concepts is now designated a diagnostic subtest; its place as a core subtest has now been taken by the downward extension of Matrices. Some of the diagnostic subtests have a lower *g* saturation and measure such less cognitively complex functions as short-term memory and speed of information processing. However, the Working Memory subtests, the Phonological Processing subtest, and the Early Number Concepts subtests have substantial *g* loadings. Subtests have

TABLE 13.2. Subtests of the DAS-II Early Years (Upper Level) Battery (Ages 3:6–8:11), Showing Abilities Measured (and Relation of Measures to Broad CHC Ability Factors) and Their Contribution to Composites

Subtest	Description	CHC broad ability	Contribution to composite
Core subtests			
Verbal Comprehension	Using various materials, child gives a motor response to verbal commands	Gc	Verbal, GCA
Naming Vocabulary	Child sees pictures of objects and names them	Gc	Verbal, GCA
Picture Similarities	Child selects a picture or figure closest to a target picture or figure	Gf	Nonverbal Reasoning, GCA, SNC
Matrices	Child selects a picture or figure that completes a 2 × 2 or 3 × 3 matrix	Gf	Nonverbal Reasoning, GCA, SNC
Pattern Construction	Child replicates designs, using blocks or foam squares	Gv	Spatial, GCA, SNC
Copying	Child copies figure by drawing it on paper	Gv	Spatial, GCA, SNC
Diagnostic subtests			
Recall of Objects—Immediate	Examiner presents 20 objects on card; child recalls as many as possible; three trials	Glr	
Recall of Objects—Delayed	Child recalls as many objects as possible 10–30 minutes after third exposure	Glr	
Early Number Concepts	Child responds to questions requiring prenumerical and numerical concepts	Gc/Gf[a]	School Readiness
Matching Letter-Like Forms	Child matches a target figure to one of six alternatives	Gv	School Readiness
Recognition of Pictures	Child identifies previously seen pictures embedded in a larger display	Gv	
Phonological Processing	Child rhymes, blends, deletes, and identifies sounds in words	Ga	School Readiness
Recall of Digits—Forward	Child repeats spoken single-digit sequences	Gsm	
Recall of Digits—Backward	Child repeats spoken single-digit sequences in reverse order	Gsm	Working Memory
Recall of Sequential Order	Child reorganizes and repeats spoken lists of body parts and objects in correct order	Gsm	Working Memory
Speed of Information Processing	Child quickly selects the largest number of squares or the highest number in a row	Gs	Processing Speed
Rapid Naming	Child quickly names colors, objects, or colors and objects in a visual display	Gs	Processing Speed

Note. GCA, General Conceptual Ability; SNC, Special Nonverbal Composite; Gv, visual–spatial processing; Gc, crystallized intelligence or verbal ability; Gf, fluid reasoning; Gsm, auditory short-term memory; Glr, long-term storage and retrieval; Ga, auditory processing; Gs, processing speed.

[a] Note that in the DAS-II handbook (Elliott, 2007c, pp. 60–61), Early Number Concepts is characterized as measuring a mixture of Gc and Gf. However, Keith et al. (2010) have argued that Early Number Concepts is best conceived as a measure of fluid reasoning and not crystallized intelligence for ages 5 through 8. It appears that the influence of Gf on this subtest becomes stronger with increasing age.

normative scores in a *T*-score metric (mean = 50, standard deviation [*SD*] = 10).

Tables 13.2 and 13.3 also show the composites that can be derived from each of the core and the diagnostic subtests. Two types of composites are provided, all in a standard score metric (mean = 100, *SD* = 15). First are lower-order cluster scores. From age 3:6 onward, the core subtests yield three of these across the board in both Early Years and the School-Age batteries (Verbal, Nonverbal Reasoning, and Spatial). Three clusters are also derived from the diagnostic subtests (School Readiness, Working Memory, and Processing Speed). For the youngest children at the Lower Level of the Early Years battery (ages 2:6–3:5), there are just two cluster scores, Verbal and Nonverbal.

TABLE 13.3. Subtests of the DAS-II School-Age Battery (Ages 5:0–7:11), Showing Abilities Measured (and Relation of Measures to Broad CHC Ability Factors) and Their Contribution to Composites

Subtest	Description	CHC broad ability	Contribution to composite
Core subtests			
Word Definitions	Child tells the meaning of words given by the examiner	Gc	Verbal, GCA
Verbal Similarities	Child describes how three objects or concepts are similar	Gc	Verbal, GCA
Matrices	Child selects a picture or figure that completes a 2 × 2 or 3 × 3 matrix	Gf	Nonverbal Reasoning, GCA, SNC
Sequential and Quantitative Reasoning	Child completes a sequence of pictures, figures, or numbers	Gf	Nonverbal Reasoning, GCA, SNC
Pattern Construction	Child replicates designs, using blocks or foam squares	Gv	Spatial, GCA, SNC
Recall of Designs	Child draws figure after viewing it for 5 seconds	Gv	Spatial, GCA, SNC
Diagnostic subtests			
Recall of Objects—Immediate	Examiner presents 20 objects on card; child recalls as many as possible; three trials	Glr	
Recall of Objects—Delayed	Child recalls as many objects as possible 10–30 minutes after third exposure	Glr	
Recognition of Pictures	Child identifies previously seen pictures embedded in a larger display	Gv	
Phonological Processing	Child rhymes, blends, deletes, and identifies sounds in words	Ga	
Recall of Digits—Forward	Child repeats spoken single-digit sequences	Gsm	
Recall of Digits—Backward	Child repeats spoken single-digit sequences in reverse order	Gsm	Working Memory
Recall of Sequential Order	Child reorganizes and repeats spoken lists of body parts and objects in correct order	Gsm	Working Memory
Speed of Information Processing	Child quickly selects the largest number of squares or the highest number in a row	Gs	Processing Speed
Rapid Naming	Child quickly names colors, objects, or colors and objects in a visual display	Gs	Processing Speed

Note. GCA, General Conceptual Ability; SNC, Special Nonverbal Composite; Gv, visual–spatial processing; Gc, crystallized intelligence or verbal ability; Gf, fluid reasoning; Gsm, auditory short-term memory; Glr, long-term storage and retrieval; Ga, auditory processing; Gs, processing speed.

Both batteries provide two higher-order composites. For most children, the most general composite will be the GCA score. For children for whom it is judged that the verbal component of that score is inappropriate, the SNC score is provided. For the Early Years (Lower Level), this is identical to the lower-order Nonverbal cluster, formed from two subtests. For the Early Years (Upper Level) and School-Age batteries, this is formed from the four subtests in the Nonverbal Reasoning and Spatial clusters. Elliott (2007b) has left the choice between the GCA and the SNC as the better measure of an examinee's general ability up to the examiner, rather than imposing rigid rules or a formula.

One major change from the first DAS is that achievement tests are no longer part of the DAS-II battery. Instead, scores from the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009) have been linked to the DAS-II. In addition, correlational data have been provided between the DAS-II and the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014), and between the DAS-II and the WJ IV Tests of Cognitive Abilities, Achievement, and Oral Language (WJ IV COG, ACH, OL; McGrew, LaForte, & Schrank, 2014; Schrank, McGrew, & Mather, 2014). Discrepancies between ability (as measured by the DAS-II GCA or SNC) and WIAT-III or KTEA-3 achievement may be evaluated by taking either (1) the simple difference between the achievement score and the composite, a procedure we emphatically discourage; or (2) the difference between predicted and observed achievement, with predicted achievement being based on the GCA, Verbal, or Nonverbal score. The WIAT-III and KTEA-3 norms manuals (Breaux, 2010, pp. 462–473, 483–484; Kaufman & Kaufman with Breaux, 2014, pp. 535–539) provide information on the statistical significance of discrepancies (i.e., their reliability), and also on their frequency of occurrence (or unusualness) in the standardization sample.

Subtests as Specific Ability Measures

The chief aim in designing the content of the DAS-II was to produce subtests that are individually interpretable and can stand technically as separate, specific measures of various abilities. Once a specification was made of the desired tasks and dimensions to be measured, each subtest was designed to be unidimensional and homogeneous in content and distinct from other subtests, thus aiding the interpretation of children's performance. If a subtest score is to be interpreted as a measure of a specific, identifiable ability, the items within that subtest must be of similar content and must require the examinee to perform similar operations. For example, in each item of the Naming Vocabulary subtest, a child is asked to name an object in a picture. All items are therefore homogeneous. Naming Vocabulary is distinct from Verbal Comprehension, another verbal subtest because the former requires a verbal response and the latter does not.

Ideally, each subtest should be a clearly interpretable measure of a CHC narrow ability factor, and the clusters to which the subtests contribute should also be clearly interpretable measures of CHC broad ability factors. Subtests or clusters should not sit astride two factors, as it were, so that interpretation of performance becomes unclear (see "Factor Structure," below).

In addition to having homogeneous content that focuses on a distinct ability, each subtest should be reliable. Because the DAS-II emphasizes the identification of cognitive strengths and weaknesses, subtests must have a sufficient amount of reliable specificity to be separately interpretable (see "Accuracy and Reliability" and "Specificity," below).

Verbal Content

Although it was considered important to include measures of verbal ability in the DAS-II cognitive battery, too many verbal tasks would present problems for examiners wishing to assess children from multicultural or culturally disadvantaged backgrounds. Because of these considerations, measures of general knowledge, colloquialisms, or words with a specific meaning in the United States were eliminated as far as possible.

Because verbal abilities constitute a major component of cognition, it is certainly necessary to have some subtests that are purely verbally presented, particularly at the School-Age level. However, getting the balance right in a test battery is important, too. In development of the DAS-II content, only two core subtests were included with entirely verbal presentation and response (both at the School-Age level), plus the two verbally administered Recall of Digits subtests. Other than those subtests, the aim was to have subtests with varied tasks and materials. The "Test Materials" section below describes the range and variety of DAS-II stimulus materials.

Timed or Speeded Items

The DAS-II contains a diagnostic cluster called Processing Speed, contributed to by the Speed of Information Processing and Rapid Naming subtests, which are both timed. Apart from these two subtests, the DAS-II content minimizes the use of timed or speeded items. Of the other subtests, only one—Pattern Construction—gives extra points for correct completion of the designs within specified time limits. Of course, this feature of the subtest, which is appropriate for most individuals, is inappropriate for some children. Speed of response to the Pattern Construction items may not produce a valid measure for a child with a physical disability such as cerebral palsy, or one with an attentional problem, or one who takes an extremely deliberate approach to a task. For such children, an alternative procedure is provided, in which the score is based solely on accuracy within very liberal time limits. Confirmatory factor analyses (CFAs) reported in the DAS handbook (Elliott, 2007c) demonstrated the factorial equivalence of the standard and alternative versions of Pattern Construction, which are unchanged in the DAS-II. However, competing higher-order CFA models completed by Keith, Low, Reynolds, Patel, and Ridley (2010) found that the untimed version of Pattern Construction results in a cleaner measure of Gv. Elliott (2007b) has left the choice between using the standard or alternative score up to the examiner's professional judgment and knowledge of the examinee.

Test Materials

The DAS-II test kit includes three informational volumes: an administration and scoring manual (Elliott, 2007b), an introductory and technical handbook (Elliott, 2007c), and a manual of norms (Elliott, 2007d). Separate record forms are provided for the Early Years and School-Age batteries. The kit contains four stimulus books, as well as a variety of consumable booklets and manipulatives. Materials vary for each subtest. The materials were specifically designed to be colorful, varied, and engaging for children and students of all developmental levels, while also being easy to administer.

In the Early Years battery, only one subtest, Recall of Digits, is presented purely verbally, with no additional stimulus materials. In the School-Age battery, four subtests are presented purely verbally, with no additional stimulus materials. These are Word Definitions and Verbal Similarities, which constitute the Verbal cluster, and Recall of Digits—Forward and Recall of Digits—Backward, which measure aspects of verbal memory.

Translations

The DAS-II handbook (Elliott, 2007c, pp. 210–218) contains guidelines for the assessment of children with communication difficulties due to such causes as cultural differences, lack of proficiency in spoken English, and hearing impairments. To assist in assessing such children, the DAS-II has been published with two translations: Spanish and American Sign Language (ASL).

The manual contains an appendix in which are given the Spanish-language instructions needed to administer the subtests that do not require a verbal response from a child. The translated instructions are for the following Early Years core subtests: Copying, Matrices, Pattern Construction, and Picture Similarities. These subtests enable the Nonverbal Reasoning and Spatial cluster scores to be estimated, together with the SNC as a measure of *g*. For the School-Age battery, translated instructions are provided for Matrices, Pattern Construction, Recall of Designs, and Sequential and Quantitative Reasoning. Once again, these enable Nonverbal Reasoning, Spatial, and SNC scores to be obtained. Translations are also provided for the following diagnostic subtests: Matching Letter-Like Forms, Recognition of Pictures, and Speed of Information Processing. Of course, examiners should not attempt to use these Spanish instructions unless they are sufficiently fluent to understand and respond to a child's comments and questions. Examiners must also keep in mind that basing an assessment on tests of Gf, Gv, and Gs provides a seriously incomplete picture of a child's cognitive abilities.

The DAS-II kit includes a CD-ROM of signed administration directions in ASL for nine subtests. These include the same core subtests for the Early Years and the School-Age batteries as in the Spanish translation, thereby enabling Nonverbal Reasoning and Spatial cluster scores to be estimated, together with the SNC. In addition, three diagnostic subtests have been translated into ASL: Matching Letter-Like Forms, Recognition of Pictures, and Speed of Information Processing. The CD-ROM is intended to help an examiner or a certified interpreter to learn how to administer the subtests in ASL in a consistent, standardized format. The CD-ROM should never be used to administer the subtests.

TABLE 13.4. Subtest Floors for the DAS-II Early Years (Lower Level) Battery: *T* Scores, *z* Scores, and Percentiles Produced by a Raw Score of 1 on Each Subtest for Children Ages 2:6–2:9

Clusters and subtests	*T* score	*z* score	Percentile
Verbal cluster			
Verbal Comprehension	24	–2.6	0.5
Naming Vocabulary	24	–2.6	0.5
Nonverbal cluster			
Picture Similarities	30	–2.0	2
Pattern Construction	27	–2.3	1

PSYCHOMETRIC PROPERTIES

The DAS-II standardization sample, the forming procedures, and data on the reliability and validity of the battery are by now well known, and are described elsewhere (Elliott, 2007c). The DAS-II followed essentially the same procedures in sampling, and in obtaining a substantial bias oversample, that were employed in the standardization of the DAS (Elliott, 1990b, 1997).

Subtest Floors

As explained earlier, because many low-functioning children are likely to be assessed with the DAS-II, its subtests have been designed to have low floors.

Table 13.4 shows the *T* scores, *z* scores (showing the number of standard deviations below the mean), and percentiles for a raw score of 1 on each of the four core subtests at ages 2:6–2:9, which is the lowest age group for the Early Years (Lower Level), If a child of this age happened to obtain raw scores of 1 on every subtest, this would yield standard scores of 56 and 64 on the Verbal and Nonverbal clusters, respectively, and a GCA score of 58. These three scores are at or below the 1st percentile. Clearly, because of the development of abilities in childhood, the floors of the subtests are lower than those in the table for children between the ages of 2:10 and 3:5, with lower values in the table, and with lower cluster and GCA scores.

Similarly, Table 13.5 shows the floors for children at ages 3:6–3:9—the lowest age group for the Early Years (Upper Level). These are for the six core subtests, which yield three cluster scores. Note that the two subtests that start at age 3:6 have the highest *T* scores for a raw score of 1. If a very low-functioning child were unable to understand these subtests, it is still possible to give the subtests from the Early Years (Lower Level) bat-

TABLE 13.5. Subtest Floors for the DAS-II Early Years (Upper Level) Battery: *T* Scores, *z* Scores, and Percentiles Produced by a Raw Score of 1 on Each Subtest for Children Ages 3:6–3:9

Clusters and subtests	*T* score	*z* score	Percentile
Verbal cluster			
Verbal Comprehension	14	–3.6	<0.1
Naming Vocabulary	15	–3.5	<0.1
Nonverbal Reasoning cluster			
Picture Similarities	21	–2.9	0.2
Matrices	31	–1.9	3
Spatial cluster			
Copying	32	–1.8	4
Pattern Construction	18	–3.2	<0.1

tery to obtain estimates of Verbal and Nonverbal ability, together with the GCA. If a child age 3:6 obtained raw scores of 1 on all six subtests, this would yield cluster standard scores of 39 (Verbal), 61 (Nonverbal Reasoning), and 60 (Spatial), with a GCA score of 47. All these scores would indicate that the child's ability in each area was below the 1st percentile.

Finally, Table 13.6 shows the floors of the DAS-II Early Years and School-Age batteries for children ages 7:0–7:5. At this age, the School-Age battery would normally be given. On this battery, three of the six core subtests and three diagnostic subtests have *T* scores above the lowest possible level of 10. The *z* scores and percentile columns show that a child who obtained a raw score of 1 on any of these six subtests would be estimated to be at or below the 1st percentile for his or her age. The subtest *T* scores in Table 13.6 yield cluster standard scores as follows: Verbal, 44; Nonverbal Reasoning, 46; Spatial, 38; School Readiness, 37; Working Memory, 59; Processing Speed, 42; and GCA, 40. However, as discussed earlier, the normative overlap between the Early Years and School-Age

TABLE 13.6. Subtest Floors for the DAS-II Early Years and School-Age Batteries: *T* Scores, *z* Scores, and Percentiles Produced by a Raw Score of 1 on Each Subtest for Children Ages 7:0–7:5

Clusters and subtests	*T* score	*z* score	Percentile
School-Age battery			
Verbal cluster			
Word Definitions	23	–2.7	0.4
Verbal Similarities	10	–4.0	<0.1
Nonverbal Reasoning cluster			
Matrices	10	–4.0	<0.1
Seq. and Quant. Reasoning	25	–2.5	1
Spatial cluster			
Recall of Designs	16	–3.4	<0.1
Pattern Construction	10	–4.0	<0.1
Early Years (Upper Level) battery			
Verbal cluster			
Verbal Comprehension	10	–4.0	<0.1
Naming Vocabulary	10	–4.0	<0.1
Nonverbal Reasoning cluster			
Picture Similarities	10	–4.0	<0.1
Matrices	10	–4.0	<0.1
Spatial cluster			
Copying	10	–4.0	<0.1
Pattern Construction	10	–4.0	<0.1
Diagnostic clusters			
School Readiness			
Early Number Concepts	14	–3.6	<0.1
Matching Letter-Like Forms	10	–4.0	<0.1
Phonological Processing	10	–4.0	<0.1
Working Memory			
Recall of Sequential Order	24	–2.6	0.5
Recall of Digits—Backward	24	–2.6	0.5
Processing Speed			
Speed of Information Processing	27	–2.3	1
Rapid Naming	10	–4.0	<0.1

batteries means that relatively low-functioning School-Age children ages 7:0–8:0 may be given the core subtests of the Early Years battery, which clearly has the lowest possible floors for children of that age—*T* scores of 10, four standard deviations below the mean for all subtests, which would yield a GCA score of 30.

The three age groups referred to in Tables 13.4, 13.5, and 13.6 are the lowest ages of children who would normally be given the Early Years (Lower Level), Early Years (Upper Level), and School-Age batteries. Low-functioning children in these young age groups would be most likely to have the greatest difficulty with the easiest items, and would show the greatest floor effects. Children in older age groups would show lesser effects the older they became. The tables show that floor effects are minimal, and that one of the goals of test development—to provide subtests with low floors—has been met.

Accuracy and Reliability

The DAS-II uses what is termed an *item set* approach to test administration, as noted earlier. This is a form of tailored testing that makes the assessment time-efficient while maintaining a high level of accuracy. This approach, and the procedures used in the DAS-II to achieve accuracy and reliability, are described in the DAS-II handbook (Elliott, 2007c).

Two diagnostic subtests (Recall of Objects—Immediate and Recognition of Pictures) have adequate mean internal-reliability coefficients of .79 at the Early Years level. All other Early Years subtests have internal reliabilities of .80 and above, nine of them being between .80 and .89, and five of them being .90 and over. At the School-Age level, Recognition of Pictures once again has an adequate but lower mean internal-reliability coefficient (.74) than other subtests. All other subtests have mean coefficients of .80 and above, nine of them being between .80 and .89, and six over .90.

The mean internal reliabilities of DAS-II cluster scores range from .89 to .95 for both the Early Years and School-Age levels. The mean internal reliabilities of the GCA and SNC are .95 (Early Years) and .96 (School-Age).

Without exception, the reliability coefficients improved for the DAS-II subtests retained from the original DAS; identical methods of estimation were used. Extensive further information on the reliability of the DAS-II is provided in the DAS-II handbook (Elliott, 2007c, pp. 121–140).

Specificity

The variance of test scores can be partitioned into a number of components. The *proportion of error variance* may be estimated and is defined as the value of 1 minus the reliability of a test. The *proportion of reliable variance* (i.e., the reliability of the test) may itself be partitioned into two components: *reliable common variance*, which is shared or overlapping with other tests in the battery, and *reliable specific variance*, which is not shared and does not overlap with other tests.

The proportion of common variance (often termed *communality*) may be estimated by the squared multiple correlation between a subtest and all others in the battery (Kaufman, 1979; Silverstein, 1976). The proportion of specific reliable variance is usually termed the *specificity* of a test and is estimated by subtracting the communality from the reliability coefficient of the test.

McGrew and Murphy (1995) consider test specificity to be high when it is (1) .25 or more (indicating that it accounts for 25% or more of the total variance of the test), and (2) greater than the proportion of error variance. Analyses of the specificity of the DAS-II (reported in detail in Elliott, 2007c, pp. 141–142) have shown every subtest to be of high specificity. For both the Early Years and the School-Age batteries, about 42% of subtest score variance is reliable specific variance. The range of subtest specificity is .31–.65 in the Early Years battery, and .31–.66 in the School-Age battery.

With one exception, the cluster scores in both batteries also show very high specificity, ranging from .37 to .57 (mean .47) for the Early Years battery and from .34 to .66 (mean .46) for the School-Age battery. The exception is the School Readiness cluster, which has a moderate specificity of .23 (falling just under the .25 criterion for high specificity). All subtest and cluster specificities substantially exceed the proportion of error variance.

Such values of specificity are very consistent with those previously found in the DAS and the BAS-II. These findings support the view that the original development goal of a battery with reliable, specific, individually interpretable subtests has been achieved. The results support the use of the DAS-II for the analysis of cognitive processing strengths and weaknesses, not just for measuring *g*, despite the recommendation of Canivez and McGill (2016) that "the DAS-II provides strong measurement of general intelligence but clinical interpretation should be primarily at that level"

(p. 1475). We believe that there is a sound foundation for interpreting patterns of strengths and weaknesses, following the guidelines and data for statistical significance and uncommonness in Elliott (2007c, 2007d).

Validity

The DAS-II handbook (Elliott, 2007c) contains extensive information on the instrument's validity. This can be broadly categorized into correlational studies, including CFAs of the structure of the battery, and studies on defined clinical samples of children with varying special needs. This section of the present chapter gives a brief description of studies on the factor structure of the DAS-II, and some data on the variety of significant strengths and weaknesses in cognitive abilities shown by children who are poor readers.

Factor Structure

The downward extension of Matrices in the DAS-II was expected to produce three core cluster scores in the Early Years battery that would be equivalent to those in the School-Age battery—namely, Verbal (Gc), Nonverbal Reasoning (Gf), and Spatial (Gv). The two new measures of working memory included in the DAS-II were also expected to form a cluster, and in addition it was thought that the new Rapid Naming subtest might cluster with Speed of Information Processing rather than with Phonological Processing when the final data were analyzed, as did similar subtests in the WJ III (McGrew & Woodcock, 2001). Thus it was anticipated that there would be two additional diagnostic clusters in the DAS-II—Working Memory and Processing Speed, measuring the CHC factors of Gsm and Gs, respectively.

Finally, the inclusion of the Phonological Processing subtest was expected to provide a measure of auditory processing—the CHC broad ability of Ga. This subtest would thereby fill a gap in the coverage of CHC broad ability factors that McGrew (1997, p. 160) noted in the original DAS. More importantly, it would also meet a clinical need for such a test that is relevant to reading acquisition and the assessment of reading disability.

Confirmatory Factor Analyses

Although the background of the DAS-II has remained eclectic, it was clear at the time of development that CHC theory had become the most dominant and widely accepted theory of the structure of human abilities. Accordingly, the DAS-II handbook (Elliott, 2007c) discusses the relation of the subtests and clusters to CHC theory in some detail. The major emphasis in conducting factor analyses was on using CFAs to test the correspondence of DAS-II subtests and clusters to the CHC model.

The DAS-II handbook contains details of the CFAs that were conducted to test the factor structure of the battery (Elliott, 2007c, pp. 153–162). The three clusters formed by the core subtests were confirmed as robust factors throughout the age range from 3:6 through 17:11, Elliott therefore continued to call these clusters Verbal, Nonverbal Reasoning, and Spatial—names given them in the first DAS. In CHC terms, they measure the broad abilities of Gc (crystallized intelligence/knowledge), Gf (fluid reasoning), and Gv (visual–spatial ability), respectively. Unfortunately, the name Nonverbal Reasoning for the Gf factor occasionally causes confusion with other nonverbal measures, such as visual–spatial ability or the SNC. Examiners should make clear in their reports that Nonverbal Reasoning is a measure of fluid reasoning (Gf).

Among the diagnostic subtests, a Working Memory factor was confirmed, formed by the subtests Recall of Digits—Forward, Recall of Digits—Backward, and Recall of Sequential Order. In CHC terms, this factor measures the broad ability of Gsm or Gwm (short-term working memory; the CHC terminology is continually evolving). This is clearly a verbal short-term memory factor because visual short-term memory tasks (CHC narrow ability MV) are always found under the Gv factor. Moreover, because working memory tasks requiring mental manipulation (Gwm WM) are cognitively more complex than simple digit recall or memory span (Gwm MS), Elliott found that the working memory subtests requiring mental manipulation (Recall of Digits—Backward, and Recall of Sequential Order) consistently had higher loadings on the factor than did Recall of Digits—Forward (Gwm MS). In order to avoid any ambiguity in interpretation (because Recall of Digits—Forward is not a measure of working memory capacity with mental manipulation or WM), only the two working memory subtests form the Working Memory cluster in the DAS-II.

The CFAs also confirmed the Processing Speed factor, formed by the Speed of Information Processing and Rapid Naming subtests.

As an example of the analyses that were conducted, Figure 13.1 shows the factor structure of the DAS-II School-Age battery for children ages 6:0–12:11. This is the operating age range of the Phonological Processing subtest, and is the age range in which all seven CHC broad factors are found and confirmed. Figure 13.1 represents the final model (the full CHC model), which fits the standardization data significantly better than any alternative model of one, two, three, or five factors (Elliott, 2007c, p. 156).

The robustness of the structure across age levels was confirmed in an independent study by Keith and colleagues (2010), using both DAS-II batteries (Early Years and School-Age) across the 4- to 17-year age range. Two of those authors commented that this detailed study demonstrated remarkable consistency of the DAS-II with CHC theory (Keith & Reynolds, 2010, p. 638). The chief difference between this study and those reported in the DAS-II handbook is that Keith and colleagues dropped the Phonological Processing subtest from the analysis because it is the only representative of the CHC G*a* factor in the battery.

As noted above, Canivez and McGill (2016) reported that results of their "exploratory factor analyses, multiple factor extraction criteria, and hierarchical exploratory factor analyses (Schmid & Leiman, 1957)" of the DAS-II standardization data

> indicated that most DAS–II subtests were properly associated with the theoretically proposed first-order factors. Hierarchical exploratory analyses with the Schmid and Leiman (1957) procedure, however, found that the hierarchical g factor accounted for large portions of total and common variance, while the two or three first-order factors accounted for small portions of total and common variance. It was concluded that the DAS–II provides strong measurement of general intelligence but clinical interpretation should be primarily at that level. (p. 1475)

We believe, however, that there is a sound foundation for interpreting patterns of strengths and weaknesses, following the guidelines and data for statistical significance and uncommonness provided by Elliott (2007c, 2007d).

Just as the Keith and colleagues (2010) study looked at the consistency and invariance of factor structure across a wide age range, so CFA was used to investigate construct bias in the instrument. A CFA study on the original DAS (Keith, Quirk, Schartzer, & Elliott, 1999) used the standardization sample and the bias oversample to test for construct bias. A hierarchical, multisample CFA was used to examine the constructs measured by the DAS in black, Hispanic, and white children. Results showed that the DAS measured the same constructs across all three ethnic groups across all age levels of the battery. The authors concluded that the DAS showed no construct bias, and that users of the battery could have confidence that the battery measures the same abilities for black, Hispanic, or white children and youth.

INTERPRETATION

Recommendations for Interpreting General, Broad, and Specific Cognitive Abilities

Chapters 4 and 5 in the DAS-II handbook (Elliott, 2007c) give detailed suggestions about cognitive processes underlying scores on the various DAS-II subtests and composites. The interpretive guidelines are largely (but not solely) based on the interpretation of subtests and clusters in terms of CHC broad and narrow abilities.

The DAS-II handbook also gives a systematic procedure for test interpretation. The procedure is partly based on the identification of scores that are significantly high or low at the .05 probability level. This is greatly facilitated by the user-friendly design of the summary page of the record form, which shows the size of differences that are significant at $p < .05$ between achievement tests, composites, and subtest scores.

Other comparisons are made possible by tables in the DAS-II handbook. In particular, the handbook provides tables enabling evaluation of discrepancies between observed and predicted achievement, as well as tables showing the frequency or unusualness of discrepancies. Because the development of individually interpretable subtests was a primary goal of the DAS-II, the handbook contains extensive interpretive guidelines for subtest scores.

Studies Conducted with Samples of Students with LDs

Research on DAS-II Score Profiles

DAS-II score profiles of 12 special populations have been reported in the handbook (Elliott, 2007c). Children were selected for each special-group sample according to specified inclusion and exclusion criteria. These studies provided the means and standard deviations of standard scores of each group on each cluster and subtest.

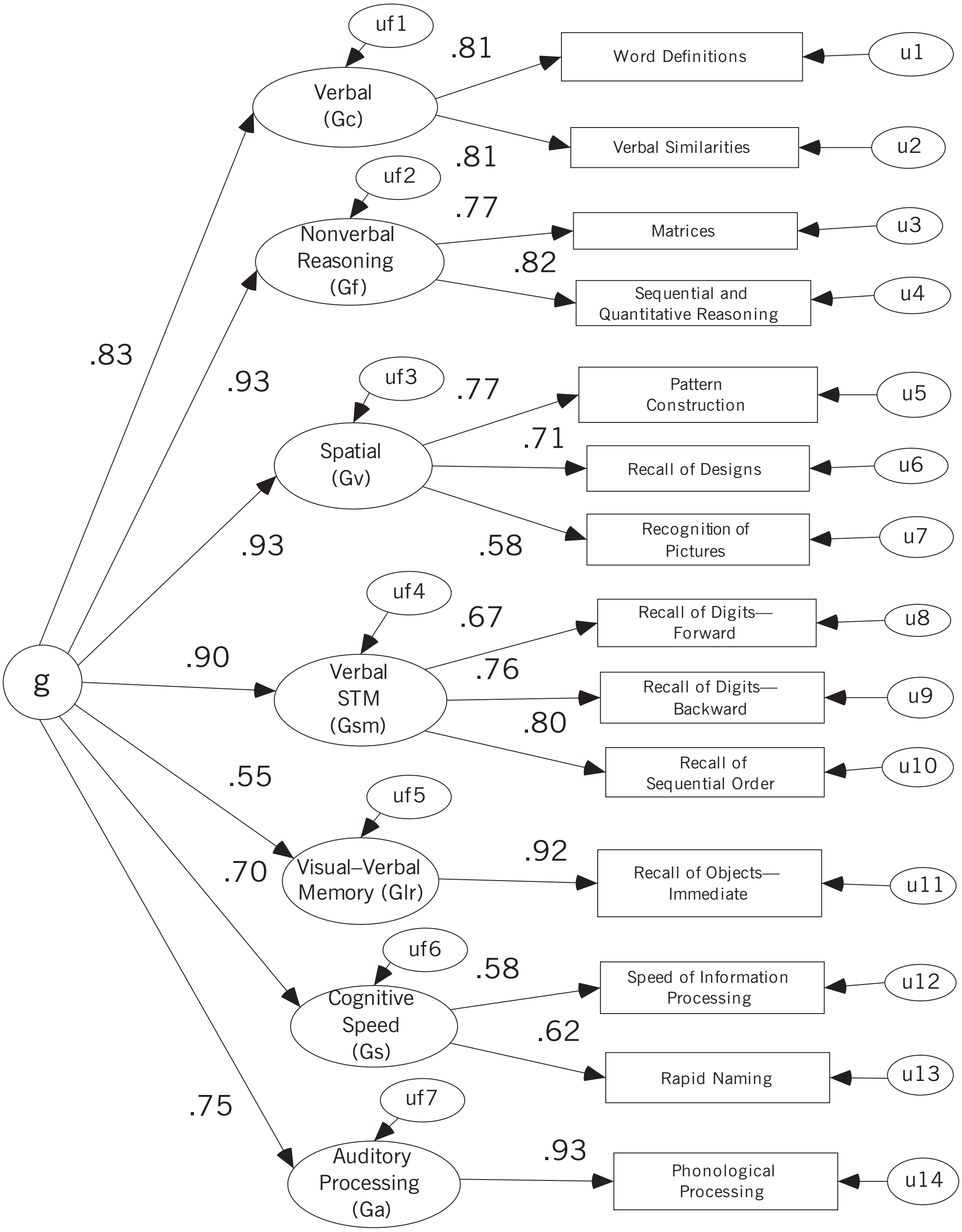


FIGURE 13.1. Factor structure of DAS-II School-Age battery for children ages 6:0–12:11, showing seven CHC broad ability factors.

Although such studies are of some interest—particularly when relatively homogeneous groups are being studied, such as children who are gifted and talented, or those with mild to moderate intellectual disabilities—such studies can be misleading when applied to heterogeneous groups such as those with LDs, where there are likely to be a wide range of causal influences for the disorders, resulting in a wide range of score profiles.

This is illustrated in Table 13.7, which shows that students who were poor readers and those with an LD in math had a wide range of cognitive profiles on the DAS-II. First, a sample of 293 poor readers was drawn from the DAS-II standardization sample, together with extra individuals who were tested at the time of standardization. Some of these children were surplus to the sampling requirements of the project; others belonged to the special groups referred to above and also had an LD in literacy. All 293 poor readers showed a significant discrepancy between their obtained Word Reading scores on the WIAT-II (Psychological Corporation, 2001) and their predicted Word Reading scores based on their GCA scores. The second sample of 43 children consisted of the special group of children who had been identified as having an LD in math.

Score profiles on the five DAS-II cluster scores were defined as follows:

- *Low Spatial, High Verbal:* The Verbal cluster score was significantly higher ($p < .05$) than the Spatial cluster. Also, the Nonverbal Reasoning score was intermediate, being lower than Verbal, and at or above the level of the Spatial score. This pattern might possibly suggest a nonverbal LD.
- *Low Verbal, High Spatial:* The Verbal cluster score was significantly lower than the Spatial cluster. Again, the Nonverbal Reasoning score was intermediate, being lower than Spatial, and at or above the level of the Verbal score. This has been a typically reported pattern for poor readers (e.g., British Psychological Society, 1999; Snow, Burns, & Griffin, 1998).
- *High Nonverbal Reasoning:* The Nonverbal Reasoning cluster score was higher than both the Verbal and Spatial scores, and significantly higher than at least one of them. This pattern might signify good ability to process complex auditory–visual information.
- *Low Nonverbal Reasoning:* The Nonverbal Reasoning cluster score was lower than both the Verbal and Spatial scores, and significantly lower than at least one of them. This core cluster profile might suggest difficulty in processing complex auditory–visual information. Elliott (2005) analyzed the scores obtained from various samples of children who had been identified

TABLE 13.7. Percentage of Students with Significantly High or Low Scores on DAS-II Clusters: Comparison of Poor Readers (*n* = 293), Those with LDs in Math (*n* = 43), and the Standardization Sample for Ages 6:0–17:11 (*N* = 2,600)

Type of profile	Poor readers (with significant discrepancy)	LDs in math	DAS-II standardization sample
No significant differences between clusters	14.3	14.0	11.5
Low Spatial, high Verbal	10.2	16.3	13.3
Low Verbal, high Spatial	17.1	7.0	13.3
High Nonverbal Reasoning	17.1	16.3	18.5
Low Nonverbal Reasoning	19.8	27.9	18.8
High Processing Speed	28.0	20.9	25.3
Low Processing Speed	22.2	34.9	25.0
High Working Memory	16.7	16.3	22.1
Low Working Memory	25.9	27.9	24.2

Note. The term *poor readers* refers to children with standard scores below 85 on WIAT-II Word Reading. The term *discrepancy* refers to the presence or absence of a statistically significant difference ($p < .05$) between obtained and predicted WIAT-II Word Reading scores (the prediction was based on the GCA score).

as having an LD in reading, and reported that approximately one-third of them had this low Nonverbal Reasoning profile.

- *High Processing Speed:* The Processing Speed cluster score was significantly higher than the GCA score.
- *Low Processing Speed:* The Processing Speed cluster was significantly lower than the GCA score.
- *High Working Memory:* The Working Memory cluster score was significantly higher than the GCA score.
- *Low Working Memory:* The Working Memory cluster score was significantly lower than the GCA score.

Table 13.7 shows the percentage of children in each sample who showed each profile. The table also shows the frequency of each profile in the total DAS-II standardization sample for ages 6:0 through 17:11. What we can conclude from an initial inspection of these profile frequencies is that there was no common profile for poor readers or for children with an LD in math. Some children in each sample clearly had profiles that were the exact opposites of those shown by other children in the same sample.

Table 13.7 indicates that over 85% of children in the population would be expected to have one or more significantly high or significantly low cluster scores. There are also interesting differences in profile frequencies between the two samples of poor achievers.

It should be noted that the poor readers were in general not children who had been formally identified as having an LD in reading. About 20% had the low Nonverbal Reasoning profile previously reported to have been found in about one-third of previous samples who had an LD in reading. However, over a quarter of the sample with an LD in math had such a profile. Also, about one-third of the children in this sample also had significantly low Processing Speed scores, while very few (7%) had high Spatial ability. Both samples had a higher percentage of children with significantly low Working Memory than those with high Working Memory.

It appears that there are a number of contrasting subgroups within each group of children with reading or math difficulties. If one subgroup has significantly low mean scores on a DAS-II cluster, and the other subgroup has significantly high ones on the same cluster, the resulting mean for the total group would be attenuated, tending toward some midvalue. Correlations between this cluster and other variables would also be attenuated. Despite such likely problems, correlational studies have been conducted on samples of children with achievement problems in reading and math.

Correlational Studies

The interpretation of intelligence tests has been and remains one of the most controversial and divisive issues in cognitive assessment. Some authors (e.g., Canivez & McGill, 2016; Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998; Glutting, Watkins, Konold, & McDermott, 2006; Kahana, Youngstrom, & Glutting, 2002; Kotz, Watkins, & McDermott, 2008; Watkins & Smith, 2013) dispute the validity, stability, and utility of patterns of performance, and affirm that there is little value in interpreting cognitive scores beyond a global ability estimate such as IQ or GCA.

Forty years of work on the development of the DAS-II and its predecessors run counter to this suggestion. Statistically significant intraindividual differences between subtest and cluster scores are, by definition, reliable; they indicate the presence of strengths and weaknesses in processing information that are not artifacts of measurement error. The DAS and the DAS-II standardization data indicate that a large proportion of children show such differences (see Table 13.7). The *raison d'être* of the DAS-Il is that significant intraindividual differences between cluster and subtest scores should lead us to consider whether and how they illuminate processing strengths and weaknesses that may be related to the problem for which the child has been referred for assessment.

The evidence in favor of this approach is strong. CFAs conducted on the DAS-II standardization data, reviewed above, show that a model with multiple factors fits the data highly significantly better than a single-factor *g* model. The analyses demonstrate that the general factor g is not sufficient to explain the relationships between the subtests and clusters. Similarly, Vanderwood, McGrew, Flanagan, and Keith (2001) showed that specific cognitive abilities provide a better-fitting model for predicting reading achievement than does general cognitive ability.

Additional evidence on these issues has been provided by two studies, summarized below, which used the DAS-II (Elliott, Hale, Fiorello, Dorvil, & Moldovan, 2010; Hale et al., 2008). Interested readers can also see references in these articles to methodological criticisms of the Glutting, Watkins, and McDermott group's use of multicollinear datasets.

Prediction of WIAT-II Math Scores

Hale and colleagues (2008) used regression commonality analysis to examine the unique and shared variance components among DAS-II CHC factors in the prediction of WIAT-II Numerical Operations and Math Reasoning skills for the DAS-II normative sample and for a sample of children with a math LD. Because of the likely attenuation of correlations, it is possible (even probable) that this would reduce the number of variables found to have significant interrelationships. However, it was considered to be important to demonstrate, even from correlational data derived from heterogeneous samples, that the broad and narrow abilities represented in the DAS-II clusters and subtests would explain significantly more variance in math achievement than the GCA alone.

Results showed that the DAS-II predictors accounted for more achievement variance in typical children than in children with a math LD. The reason for this is very likely to be a restriction of range in the Math scores of the latter group. For typical children, DAS-II predictors accounted for 46% of variance in Numerical Operations and 58% of variance in Math Reasoning. On the other hand, the DAS-II predictors accounted for 33% of variance in Numerical Operations and 50% of variance in Math Reasoning for the children with a math LD. There was substantial loss of predictive validity when the GCA was used instead of cluster or subtest scores: 13% loss for the typical group on both math tests, and 56% loss on Numerical Operations and 20% loss on Math Reasoning for the group with a math LD.

Prediction of WIAT-II Reading Scores

Elliott and colleagues (2010) used both commonality analysis and structural equation modeling (SEM) to investigate the effect of broad CHC abilities measured by the DAS-II, together with the effect of the general factor (*g*) on reading achievement measured by the WIAT-II.

The SEM analyses indicated that for typical children drawn from the standardization sample, four CHC-related measures were significant direct predictors of Reading Decoding (a combination of WIAT-II Word Reading and Pseudoword Decoding). These predictors were the Verbal (Gc), Nonverbal Reasoning (Gf), and Working Memory (Gsm) clusters, together with the Phonological Processing (Ga) subtest.

A similar analysis was conducted for a sample of 230 poor readers drawn from the standardization sample; children who were tested for standardization but who were surplus to requirements; and children with reading disorder, reading and written expression disorder, mathematics disorder, attention-deficit/hyperactivity disorder (ADHD), or ADHD with LD—samples gathered for studies of special populations at the time of standardization, and referred to above. For this sample, Phonological Processing was again found to be a significant predictor, together with the Spatial Ability (Gv) and Processing Speed (Gs) clusters, and the Recall of Objects—Immediate subtest (Glr).

Although Phonological Processing had a significant large effect on Reading Decoding for both samples, no other effects were significant for both samples. Each sample had three significant but different CHC factors that produced significant effects on Reading Decoding. When both analyses were considered together, the results showed that every CHC broad ability factor was a significant predictor in one or the other analysis. In both analyses, the effect of the general factor (*g*) was indirect. In other words, its effect was mediated through the three first-order factors measuring it (Verbal, Nonverbal Reasoning, and Spatial). The results demonstrated that children with reading problems have different cognitive predictor–reading achievement relationships than adequate readers.

The commonality analyses examined DAS-II predictors of WIAT-II Word Reading and Reading Comprehension scores. Once again, as might be expected, different commonalities were found for the typical sample and a sample of children with an LD in reading. Across all analyses, Verbal Ability (Gc), Nonverbal Reasoning Ability (Gf), Spatial Ability (Gv), Working Memory (Gsm), and Phonological Processing (Ga) showed important and significant effects, and explained significant amounts of variance over and above that explained by estimates of *g*. The data from the studies on both math and reading suggested that practitioners should not emphasize global GCA, but should instead interpret cluster and subtest scores and their interrelationships in developing hypotheses about an individual's processing strengths and weaknesses.

Appendix 13.1 presents a case study that demonstrates how the WIAT-III and DAS-II can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

CONCLUSIONS

This chapter has outlined various ways in which the DAS-II has been designed to be appealing and accessible to children of all abilities across a very wide age range, from 2:6 to 17:11. Its good floors, and its procedures to help the least able examinees understand what is required in each task, make the battery highly appropriate for use with clinically referred populations. It has also been designed for speed and efficiency in administration. Its cluster and subtest scores have essential qualities of reliability and interpretability. Its consistency and clarity in measuring the constructs of CHC theory make it an ideal instrument for cross-battery assessment (Flanagan et al., 2013). When the DAS was first published in 1990, it was at first virtually unknown and seemed so different in its procedures that professionals probably feared it was difficult to learn. Now in its second edition, it is widely used, and is accepted as an instrument that enjoyably engages children and helps to identify their processing strengths and weaknesses with efficiency and precision.

APPENDIX 13.1

Brief Case Study

Michelle (age 8:11), a white female in grade 3, was referred for assessment when her child study team reported their concern that she had not responded to the reading interventions provided as part of her regular instructional program. She struggled to name high-frequency words, and her team felt she needed additional targeted instructional recommendations.

Michelle's achievement scores on the Wechsler Individual Achievement Test—Third Edition (WIAT-III) and her cognitive ability scores on the Differential Ability Scales—Second Edition (DAS-II) are shown in Table 13.A.1. Examination of her WIAT-III scores shows that she has a specific difficulty with Word Reading. Her higher score on Pseudoword Decoding suggests that she finds decoding phonetically regular made-up words (such as *flarp*) easier than decoding irregular real

TABLE 13.A.1. WIAT-III and DAS-II Subtest and Cluster Scores for Michelle (Age 8:11)

Subtest or cluster	Score[a]
WIAT-III	
Total Reading	69
Word Reading	58
Pseudoword Decoding	86
Reading Comprehension	72
Oral Reading Fluency	66
Written Expression	77
Sentence Composition	84
Essay Composition	79
Spelling	79
Mathematics	99
Math Problem Solving	97
Numerical Operations	101
DAS-II core clusters (School-Age battery)[b]	
Verbal	102
Nonverbal Reasoning[c]	89
Spatial	102
{GCA	97}
DAS-II diagnostic clusters and subtests	
Working Memory	93
Recall of Sequential Order	*T* = 40
Recall of Digits—Backward	*T* = 52
Processing Speed	89
Speed of Information Processing	*T* = 51
Rapid Naming	*T* = 38
Additional DAS-II diagnostic subtests	
Recall of Objects—Immediate	*T* = 40
Recall of Objects—Delayed	*T* = 41
Recall of Digits—Forward	*T* = 53
Recognition of Pictures	*T* = 54
Phonological Processing	*T* = 51

[a] Standard scores except where indicated.

[b] There are no significant differences between subtests within each cluster

[c] The Nonverbal Reasoning cluster score is significantly lower than *both* the Verbal and Spatial cluster scores.

words. Such irregular words make demands on visual memory; they are often called *sight words,* because they have to be remembered as a whole rather than being solvable phonetically. Michelle's oral reading fluency for passages was only slightly stronger than her reading of real words, and much weaker than her decoding of made-up words. The initial hypothesis is that Michelle may have visual information-processing difficulties.

On the DAS-II, the core clusters show a significant difference between Nonverbal Reasoning and both of the other clusters. Any hypothesis concerning poor processing of purely visual information is disconfirmed by Michelle's Spatial cluster score, which is average and at the same level as her Verbal cluster score. The subtests within each cluster have consistent scores, so the cluster scores may be interpreted without further qualification. Because of Michelle's significantly low Nonverbal Reasoning score, it makes little sense to report her General Conceptual Ability (GCA) score, which has little or nothing to offer in terms of describing Michelle's cognitive processing.

The DAS-II Nonverbal Reasoning Gf tasks are called *nonverbal* because they are presented visually. However, to solve the problems effectively, the individual needs to use internal language to encode the components of the visual stimulus, and to generate hypotheses, to test them, and to identify the correct solution. These tasks are therefore characterized in the DAS-II handbook (Elliott, 2007c) as requiring integrated analysis and complex transformation of both visual and verbal information. Problems in this type of processing may be at the root of Michelle's problems in learning to read.

The diagnostic clusters of Working Memory and Processing Speed offer further information. Michelle's Working Memory score is not significantly lower than her GCA score. However, this conclusion is qualified by a significant difference between the two component subtest scores. Recall of Digits—Backward is a purely verbal subtest, and Michelle's score on this is average for her age. On the other hand, her score on Recall of Sequential Order is 12 *T*-score points below her Recall of Digits—Backward score. This difference is highly significant ($p < .01$). Recall of Sequential Order is verbally presented, but requires the examinee to visualize the position of various parts of the body.

Based on the results from the Nonverbal Reasoning cluster, and supported by the results from the Working Memory subtests, it appears that Michelle's processing difficulties do not appear to be shown when she is working with purely auditory–verbal or purely visual–spatial information. Our revised hypothesis is that she seems to have particular difficulties in processing auditory–visual information.

Such a view is confirmed by Michelle's scores on the Processing Speed cluster. Once again, her overall cluster score of 89 is not significantly lower than her GCA score. However, her score on the Rapid Naming subtest is 13 *T*-score points lower than her score on Speed of Information Processing, a statistically significant difference ($p < .05$). Rapid Naming presents colors and pictures that have to be named quickly, and this is another example of a subtest that requires auditory–visual processing.

The diagnostic subtest Recall of Objects is another visual–verbal task, which that yields separate scores for Immediate and Delayed recall. The subtest presents a visual array of pictures, which are named by the examiner and then removed from view. The student is asked to recall them verbally and then recall them two more times after seeing them again, but not hearing them named. Michelle's scores on Recall of Objects are below average, and significantly below the average level of her scores on the core subtests ($p < .05$). The other diagnostic subtests, on which Michelle has achieved average scores for her age, require either purely verbal or purely visual processing.

As a result of these analyses, there is now strong support for the hypothesis that Michelle's cognitive processing difficulties center on problems with auditory–visual materials. Her reading scores support this hypothesis. Reading requires a high level of visual–verbal integration in order to convert visual printed codes into sounds and words. For fluent reading, and for recognition of common words or letter strings, an individual needs information in the auditory–verbal and visual processing systems to be effectively integrated. Similarly, to perform well on the DAS-II Nonverbal Reasoning tasks (or, indeed, any good measures of fluid reasoning), and on the Recall of Sequential Order, Rapid Naming, and Recall of Objects subtests, one needs good integration of the visual and verbal processing systems. These tasks, like the task of reading, present visual information—but to solve the problems effectively, the use of internal language to label and to mediate the solution of the problems is generally essential. In the case of an individual who has excellent verbal and spatial abilities, if the two brain processing systems specialized for those abilities do not "talk" to each other effectively, this may have an adverse effect on performance both in reasoning and in reading acquisition.

The question now arises about appropriate intervention methods for students who have a consistent pattern of difficulties with tasks requiring auditory–visual integration. For many years, teachers of children with dyslexia have actively advocated multisensory teaching methods, despite research evidence that appeared to discredit auditory–visual integration as a cause of poor reading acquisition (e.g., Bryant, 1968). Teachers appear to have long held the view that children with dyslexia have difficulty integrating visual and verbal information. Thus it has been recommended that multisensory teaching methods should be used as much as possible in teaching Michelle basic literacy skills. Useful references to multisensory teaching approaches are given by Thomson and Watkins (1998), Augur and Briggs (1992), Walker and Brooks (1993), Birsh (1999), and Walker (2000).

REFERENCES

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185–202). New York: Guilford Press.

Augur, J., & Briggs, S. (Eds.). (1992). *The Hickey multisensory language course* (2nd ed.). London: Whurr.

Birsh, J. R. (1999). *Multisensory teaching of basic language skills*. Baltimore: Brookes.

Breaux, K. C. (2010). *Wechsler Individual Achievement Test—Third Edition: Technical manual with adult norms*. Bloomington, MN: Pearson.

British Psychological Society. (1999). *Dyslexia, literacy and psychological assessment* (Report by a working party of the Division of Educational and Child Psychology). Leicester, UK: Author.

Bryant, P. B. (1968). Comments on the design of developmental studies of cross-modal matching and cross-modal transfer. *Cortex, 4,* 127–137.

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales—Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment, 28* (11), 1475–1488.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). Amsterdam: Pergamon Press.

Carroll, J. B. (2012). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 883–890). New York: Guilford Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139–164.

Daniel, M. H., Breaux, K. C., & Frey, F. E. (2010, March). *Patterns of strengths and weaknesses (PSW) models for identifying SLD.* Paper presented at the annual meeting of the National Association of School Psychologists, Chicago, IL.

Dumont, R., Willis, J. O., & Elliott, C. D. (2009). *Essentials of DAS-II assessment*. Hoboken, NJ: Wiley.

Elliott, C. D. (1983a). *The British ability scales: Manual 1. Introductory handbook*. Windsor, UK: NFER-Nelson.

Elliott, C. D. (1983b). *The British ability scales: Manual 2. Technical handbook*. Windsor, UK: NFER-Nelson.

Elliott, C. D. (1990a). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (1990b). *Differential Ability Scales: Introductory and technical handbook*. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (1996). *British Ability Scales—Second Edition*. Windsor, UK: NFER-Nelson.

Elliott, C. D. (1997). The Differential Ability Scales. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 183–208). New York: Guilford Press.

Elliott, C. D. (2005). The Differential Ability Scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 402–424). New York: Guilford Press.

Elliott, C. D. (2007a). *Differential Ability Scales—Second Edition*. San Antonio, TX: Harcourt Assessment.

Elliott, C. D. (2007b). *Differential Ability Scales—Second Edition: Administration and scoring manual.* San Antonio, TX: Harcourt Assessment.

Elliott, C. D. (2007c). *Differential Ability Scales—Second Edition: Introductory and technical handbook*. San Antonio, TX: Harcourt Assessment.

Elliott, C. D. (2007d). *Differential Ability Scales—Second Edition: Normative data tables manual.* San Antonio, TX: Harcourt Assessment.

Elliott, C. D. (2012). The Differential Ability Scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 336–356). New York: Guilford Press.

Elliott, C. D., Hale, J. B., Fiorello, C. A., Dorvil, C., & Moldovan, J. (2010). Differential Ability Scales–II prediction of reading performance: Global scores are not enough. *Psychology in the Schools, 47,* 698–720.

Elliott, C. D., & Smith, P. (2012). *British Ability Scales—Third Edition*. London: GL Assessment.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 599–612.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40,* 103–114.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. D. (2008). Differential Ability Scales—Second Edition (neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools, 45,* 838–858.

Horn, J. L., & Blankson, N. (2012). Foundation for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73–98). New York: Guilford Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. 108-446, 118 Stat. 2647 (2004).

Kahana, S. Y., Youngstrom, E. A., & Glutting, J. J. (2002). Factor and subtest discrepancies on the Differential Ability Scales: Examining prevalence and validity in predicting academic achievement. *Assessment, 9,* 82–93.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement—Third Edition*. Bloomington, MN: Pearson.

Kaufman, A. S., & Kaufman, N. L., with Breaux, K. C. (2014). *Technical and interpretive manual: Kaufman Test of Educational Achievement—Third Edition*. Bloomington, MN: Pearson.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales–II: Consistency across ages 4 to 17. *Psychology in the Schools, 47,* 676–697.

Keith, T. Z., Quirk, K. J., Schartzer, C., & Elliott, C. D. (1999). Construct bias in the Differential Ability Scales?: Confirmatory and hierarchical factor structure across three ethnic groups. *Journal of Psychoeducational Assessment, 17,* 249–268.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 30 years of research. *Psychology in the Schools, 47,* 635–650.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review, 37,* 261–278.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York: Guilford Press.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (IDTR): Gf-Gc cross-battery assessment*. Boston: Allyn & Bacon.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Murphy, S. (1995). Uniqueness and general factor characteristics of the Woodcock–Johnson Tests of Cognitive Ability—Revised. *Journal of School Psychology, 33,* 235–245.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III: Technical manual*. Itasca, IL: Riverside.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition*. San Antonio, TX: Author.

Psychological Corporation. (2001). *Wechsler Individual Achievement Test—Second Edition*. San Antonio, TX: Author.

Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the Differential Ability Scales and the Woodcock–Johnson Tests of Cognitive Abilities—Third Edition. *Psychology in the Schools, 44,* 119–138.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.

Schneider, J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 99–144). New York; Guilford Press.

Schrank, F. A., McGrew, K. S. & Mather, N. (2014). *Woodcock–Johnson IV*. Rolling Meadows, IL: Riverside.

Silverstein, A. B. (1976). Variance components in the subtests of the WISC-R. *Psychological Reports, 39,* 1109–1110.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Thomson, M. E., & Watkins, F. J. (1998). *Dyslexia: A teaching handbook* (2nd ed.). London: Whurr.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2001). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences, 13,* 159–188.

Walker, J. (2000). Teaching basic reading and spelling. In J. Townend & M. Turner (Eds.), *Dyslexia in practice: A guide for teachers* (pp. 93–129). New York: Kluwer Academic/Plenum.

Walker, J., & Brooks, L. (1993). *Dyslexia Institute literacy programme.* London: James & James.

Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment, 25,* 477–483.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition.* San Antonio, TX: Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition.* Bloomington, MN: Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

CHAPTER 14

The Woodcock–Johnson IV

Tests of Cognitive Abilities, Tests of Oral Language, Tests of Achievement

Fredrick A. Schrank
Barbara J. Wendling

The Woodcock–Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014a) is a comprehensive, co-normed psychoeducational assessment system that includes three test batteries: the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b), the Woodcock–Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b), and the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a). Although some of the WJ IV tests can be used with children as young as 24 months (particularly the WJ IV OL), the WJ IV COG and ACH are best suited for use with individuals from 5 to 95 years of age. Scores from all three batteries can be compared and contrasted with various intra-ability variations and ability–achievement comparison procedures, which can help examiners determine the presence and severity of any intraindividual strengths and weaknesses or performance discrepancies. Normative information was obtained on a large, nationally representative sample of 7,416 individuals ranging in age from 2 to 90+ years of age. A co-normed but separate battery of early development tests called the Woodcock–Johnson IV Tests of Early Cognitive and Academic Development (ECAD; Schrank, McGrew, & Mather, 2015b) is designed primarily for use with children from 30 months through 7 years of age, and is also appropriate for use with children ages 8 through 9 who have cognitive-developmental delays. We have discussed the ECAD in Chapter 10 of this volume.

The WJ IV COG, WJ IV OL, and WJ IV ACH can be used independently or in any combination. Additional interpretive options are available when the batteries are used together. Following a broad overview of general administration procedures and an update to the theoretical model that applies to all three batteries, this chapter is organized into three major sections that address each battery individually. Each section includes information on the organization of the tests and clusters that constitute the battery, any unique administration and scoring rules for some tests, a summary of psychometric properties, and interpretive information that includes an emphasis on what we refer to as *consequential validity*—how the WJ IV can be used to relate any measured limitations in test performance to evidence-based interventions, strategies, accommodations, or instructional planning.

GENERAL ADMINISTRATION AND SCORING PRINCIPLES

Administration and scoring of the WJ IV tests requires knowledge of the exact administration and scoring procedures, as well as an understanding of the importance of adhering to standardized procedures. Each examiner's manual provides guidelines for this purpose. Additionally, the test books contain an abbreviated set of instructions for administering and scoring items on each test; these are found on the introductory page of each test (the first printed page after the tabbed page). Additional instructions may appear on the test pages, such as in boxes with special instructions.

Typically, it is not necessary to administer all tests in a battery. The WJ IV COG, WJ IV OL, and WJ IV ACH each include a small set of core tests that provide some of the most useful information and form the foundation of the intra-ability analytic procedures for each battery. The core tests are placed in a head-initial position in each battery, so that examiners can begin testing with the first test in the test book and discontinue administering tests at preferred junctures. After the core tests are administered, additional tests can be administered as desired, and can also be included in an analysis of relative strengths and weaknesses. Alternatively, examiners can use the Selective Testing Table for each battery to determine which tests to administer; the tests may be administered in any order deemed appropriate, and testing can be discontinued after completion of any test.

Administration of some tests requires the examiner to use a response booklet and an audio recording. Directions for using the response booklet are provided in the test book. The audio recording is used to ensure standardized presentation of certain auditory processing, short-term working memory, oral language, and some language-related tasks. The audio equipment used must have a good speaker, be in good working order, and produce a faithful, clear reproduction of the test items. Using headphones is recommended, as they were used in the standardization. Examiners can wear a monaural earphone or wear only one headphone over one ear to monitor the audio recording, while the examinee is also listening through his or her headphones.

To help examiners document general testing observations, each test record includes a brief, seven-category behavior rating scale on the first page. The categories include levels of conversational proficiency, cooperation, level of activity, attention and concentration, self-confidence, care in responding, and response to difficult tasks. Each category has a range of possible responses, in order to help an examiner identify whether the behavior is typical or atypical for the age or grade of the individual being assessed. This checklist should be completed at the end of the testing session.

Examiners will need to learn how to establish a basal and a ceiling level for several tests. Basal and ceiling criteria are included in the test book for each test requiring them. If an examinee fails to meet the basal criterion for any test, the examiner is directed to test backward until the examinee has met the basal criterion or until item 1 has been administered. When stimuli are visible to the examinee, the examiner is required to test by complete pages. The complete-page rule may affect the process of establishing a basal or a ceiling. For example, if the basal is not established on the first page administered, the examiner tests backward full page by full page until the basal is established or item 1 is administered. Conversely, if the ceiling appears to be reached on a page, the examiner must complete that page. If in the process of completing the page, the examinee responds correctly to an item, testing must continue until the ceiling rule is met and the page is completed. Note that the basal rule is based on the lowest-numbered correct responses specified, and the ceiling is based on the highest-numbered incorrect responses specified. For some tests, examinees begin with item 1 and test until they reach their ceiling level; these tests do not require a basal. Other tests are arranged with item sets or groups; for these tests, specific criteria are outlined in the corresponding manual, test book, and test record.

Individual test items are scored during test administration. Many tests use a 1 (correct) or 0 (incorrect) scoring rule for determining raw scores. Examiners will need to learn how to score items and calculate the raw score for each test. Generally, raw scores are determined by adding the number of correctly completed items to the number of test items below the basal. Scores for sample or practice items should not be included when calculating raw scores. The correct and incorrect keys in the test books are intended to be guides to demonstrate how certain responses are scored, and not all possible or acceptable responses are included in the keys. Completion of the scoring procedure requires using the Woodcock–Johnson Online Scoring and Reporting System (Schrank & Dailey, 2014). Additionally, examiners may purchase and use the WJ IV Interpretation and Instructional Interventions Program (WIIIP; Schrank & Wendling, 2015a), an online option for generating narrative interpretive text and related interventions. The WIIIP simplifies the process of linking assessment to instruction by using WJ IV test results to generate evidence-based interventions.

Although both age and grade norms are provided for the WJ IV, age norms are often preferred for cognitive assessments. The associated scoring and interpretive programs provide a number of different interpretive report options, profiles, and a full array of interpretive scores, including age- and grade-equivalent scores, relative proficiency indexes (RPIs), and peer comparison scores (e.g., standard scores and percentile ranks). Different types of scores provide different perspectives on test performance and are not interchangeable. Schrank, Decker, and Garruto (2016) provide some perspective on recommended scores for inclusion in a report:

- Age-equivalent (AE) scores express the individual's performance in terms of the age equivalent in the standardization sample that is associated with the raw score. As such, these scores provide information about an individual's developmental age.
- Proficiency-level terms (e.g., *very advanced, advanced, average, limited, very limited, extremely limited*) describe the individual's ability with the measured tasks in relation to individuals of the same age. These terms have objective meaning because they are tied directly to task proficiency on the WJ IV and are not related to standard score classifications. The associated RPI scores express proficiency as a ratio that compares the individual's task proficiency to the relative proficiency of a typical, or median, individual of the same age (e.g., 82/90).
- Standard scores and percentile ranks provide information about the individual's relative position among peers. Standard scores are often required and can be useful for many comparative purposes. However, because standard scores lack objective meaning, the associated percentile ranks are recommended: They provide information on how common or rare the individual's ability level is within the general population, in a metric that is more commonly understood by both professional and nonprofessional consumers of report information.

EVOLUTION OF THE WJ IV INTERPRETIVE MODEL

The WJ IV is associated with the contemporary version of the Cattell–Horn–Carroll (CHC) theory of cognitive abilities, sometimes referred to as *CHC version 2* (McGrew, LaForte, & Schrank, 2014; Schneider & McGrew, 2012 and Chapter 3, this volume). Analysis of the Woodcock–Johnson—Revised (WJ-R; Woodcock & Johnson, 1989), Woodcock–Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001), and WJ IV standardization samples (none of which were analyzed by Carroll, 1993) provided three large, multiple-ability datasets to either confirm or revise prior CHC construct specifications. Additional impetus for changes to some of the interpretive constructs or terminology was gleaned from other sources of neuroscience research, such as that reported in McGrew and colleagues (2014) and Schrank and colleagues (2016). The concept of *cognitive complexity* (Lohman & Lakin, 2011) was employed in test construction to increase ecological validity and diagnostic relevance. Perhaps the most significant changes to the previous CHC constructs are derived from external neurocognitive and psychometric research in the domains of working memory and attentional control, memory consolidation and retrieval, and auditory intelligence.

Working Memory and Attentional Control

Within the past 35-plus years, the term *working memory* became increasingly cited in neuroscience research (Unsworth, 2016), often with operational definitions that either replaced or included the construct of *short-term memory*. In its broadest sense, *working memory* refers to a dynamic, temporary storage system that allows information—either sensory inputs or prior knowledge—to be held in *primary memory* (James, 1890) (sometimes referred to as *immediate awareness*) and manipulated (Atkinson & Shiffrin, 1968; 1971; Baddeley & Hitch, 1974; Goldman-Rakic, 1992; Miller, Galanter, & Pribram, 1960). In contrast, the term *short-term memory* has come to refer more narrowly to "tasks that involve significant storage but only minimal processing" (Gathercole & Alloway, 2008, p. 21).

In contemporary professional nomenclature, working memory refers to a broader, more complex construct than short-term memory span. In contrast to the WJ IV narrow Memory Span (MS) cluster, which consists solely of tasks that measure passive holding onto verbal information, the WJ IV Short-Term Working Memory (Gwm) clusters comprise tasks that involve both temporary storage *and* active review or manipulation of words and numbers, and, as a consequence of task complexity, rely more heavily on increased control of attention for successful completion of any retrieval, recoding, or reorganization subprocesses. The term *working* suggests that working memory is the active "hub" of human cognition (Camos, 2015) and can invoke other cognitive functions or abilities to effect goal attainment, such as visualization, phonology, reasoning, and retrieval or secondary memory. In addition, in both neuroscience and contemporary CHC theory, the construct of working memory includes the attentional control processes that bring representations into heightened activation and maintain the representations free from distraction (Unsworth, 2016).

Klingberg (2009) suggested that working memory is important for at least two reasons. First, it is the mechanism by which we focus our attention. Second, its *capacity* places one of the most important limitations on the brain's ability to process information and is "fundamental and vital to numerous mental tasks, from attention control to solving logical problems" (p. 33). All stimulus-driven attention requires working memory; consequently, limitations in working memory capacity have important educational and functional implications. As noted by Klingberg, "Working memory is used to control attention, to remember instructions, to keep in mind a plan of things to do, and to solve complex problems" (p. 45).

Working memory capacity appears to be domain-general, and its influence cuts across a wide variety of complex learning tasks (Kyllonen & Stephens, 1990; Unsworth, 2016). Limited working memory capacity can act as a bottleneck for learning and may be related to learning disabilities (Gathercole, 2004; Gathercole, Lamont, & Alloway, 2006). Many children with identified reading, writing, and mathematics disabilities show related impairments in working memory capacity (Bull & Scerif, 2001; de Jong, 1998; Mayringer & Wimmer, 2000; Siegel & Ryan, 1989; Swanson, 1994; Swanson, Ashbaker, & Lee, 1996). Limited working memory capacity may also be related to learning disabilities that are not domain-specific. Gathercole and Alloway (2008) stated, "Learning difficulties that extend across both reading and mathematics, or language, therefore appear to be characteristic of children with poor working memory function" (p. 25).

Research beyond the initial specifications of CHC theory suggested that working memory, working memory capacity, and attentional control appear to be valid interpretive constructs. In contemporary CHC theory, the term *working* was added to the name of the former short-term memory factor to become *short-term working memory* (Gwm). The WJ IV Gwm clusters are designed to contain different measures of working memory capacity (WM), a quantifiable aspect of the more broadly encompassing working memory construct used in contemporary neuroscience. In addition, attentional control (AC) has attained greater specification in CHC theory as a facet, or perhaps a correlate, of working memory capacity (Unsworth, 2016).

Updates to the constructs of working memory and attentional control in contemporary CHC theory are reflected in the WJ IV. The WJ IV COG includes three different types of complex working memory capacity tasks and one verbal memory span test. COG Test 3: Verbal Attention measures an individual's temporary storage of verbal information and assesses the ability to retrieve and reactivate information that might not be actively maintained in primary memory (Unsworth, 2016) via a cue-dependent search process (Unsworth & Engle, 2007a). COG Test 10: Numbers Reversed taps the recoding function of working memory. COG Test 16: Object–Number Sequencing evaluates the reorganization and assembly functions of working memory. In contrast, COG Test 18: Memory for Words measures verbal short-term memory span for unrelated words. Finally, because working memory capacity is related to vocabulary learning (Daneman & Green, 1986), reading, and language comprehension (Daneman & Carpenter, 1980), the WJ IV OL includes two ecologically relevant tests for evaluating working memory for language processing. OL Test 5: Sentence Repetition is an expressive language measure of verbal, auditory memory span. OL Test 6: Understanding Directions assesses the ability to follow directions from orally imparted instructions, and is thus a receptive language measure requiring working memory.

All of the WJ IV short-term working memory tests require varying parameters of attentional control. Complex working memory tasks (those involving a storage function in the context of processing) require higher allocation of attentional resources than pure memory span tasks (those involving passive storage only) do, and individuals with higher working memory capacities are better able to sustain attention on complex tasks than are individuals with limited working memory capacities (Unsworth, 2016). For example, individuals with lower levels of attentional control may be subject to the influences of external distraction or may let their minds wander when performing complex tasks. The different WJ IV working memory tests provide opportunities for highly knowledgeable professionals to draw inferences or make clinical hypotheses about the relationships between examinees' levels of working memory capacities and attentional control under different conditions.

Memory Consolidation and Retrieval

In contemporary CHC theory and in the WJ IV, an important distinction is made between tests and clusters that measure storage *and* retrieval functions on the one hand, and tests and clusters that solely measure retrieval functions on the other. This distinction was initially posited by Carroll (1993), whose three-stratum theory articulated separate and distinct stratum II factors for (1) general memory and learning and (2) broad retrieval ability. However, in the initial conceptualization of CHC theory (McGrew, 1997), the Glr factor included both types of functions.

In retrospect, the initial CHC misspecification of both storage and retrieval and retrieval-only cognitive functions into one common broad factor might have been a classic "wrong turn at Albuquerque."[1] In addition to the misspecification error, many professionals routinely dropped the word *storage* from the name of the Glr construct and simply referred to the factor or cluster as *long-term retrieval.* This tendency may have caused some confusion with the neurocognitive construct of long-term memory. Consequently, because of an initial "wrong turn" and the simultaneous verbal (*long-term retrieval*) shortcut, changes to contemporary CHC theory nomenclature as represented in the WJ IV provide cleaner distinctions between retrieval measures that do, and those that do not, involve the storage function.

The WJ IV COG Long-Term Storage and Retrieval (Glr) cluster measures the integrity of consolidation (encoding) of semantic (meaning-based) representations into long-term memory. The tests that constitute the Glr cluster measure two different processes by which semantic memories (Tulving, 1972, 1985) are created, using standardized formats for evaluating the integrity of semantic memory consolidation functions in working memory. COG Test 6: Story Recall measures the development, consolidation, and reconstruction of mental representations from orally imparted discourse. In CHC theory, this is referred to as the narrow ability of *meaningful memory* (MM), an aspect of semantic storage and retrieval. Measures of MM may have practical, incremental validity for predicting success in learning situations, beyond what an individual's general intellectual ability might suggest (Cucina, Su, Busciglio, & Peyton, 2015). COG Test 13: Visual–Auditory Learning measures the ability to associate words with rebus representations. In CHC theory, this narrow ability is called *associative memory* (MA).

In the WJ IV ACH, Test 12: Reading Recall is also a measure of Glr and MM. This task requires reading a short story and then retelling the principal components. The Reading Recall test measures the development and consolidation of mental representations from textual material. Although successful performance on this task is dependent on background knowledge and foundational basic reading skills, the central cognitive operation utilized is called *mapping,* wherein key elements of the story are added to a mental representation being built in working memory during the process of reading (Ashcraft, 2002; Zhou & Black, 2000). The recall or retelling phase of each story merely provides a test of whether encoding (e.g., storage of story elements) has occurred during the reading phase.

The WJ IV COG Glr cluster and associated tests represent critically important links between the constructs of working memory and long-term memory. The storage aspect of the Glr tests and clusters refers to the consolidation of semantic memories, a cognitive function that is critically important to learning. Thus Glr, in contemporary CHC theory and in the WJ IV, may be described as the gateway between *information to be learned* and *learned information.*

In contrast, the WJ IV OL Speed of Lexical Access (LA) cluster measures the *retrieval-only* function that Carroll (1993) would probably have associated with his broad retrieval ability factor. Schneider and McGrew (Chapter 3, this volume) have suggested the initialism Gr to represent this function as a broad factor of retrieval ability; this helps distinguish it from the WJ IV COG Glr cluster, which includes the memory consolidation function. WJ IV OL Test 4: Rapid Picture Naming and OL Test 8: Retrieval Fluency, as well as Part B Word Fluency of WJ IV COG Test 5: Phonological Processing, all measure retrieval of names and words from previously stored knowledge. A type of verbal efficiency (Perfetti, 1985) or automaticity for lexical access (LaBerge & Samuels, 1974; Neely, 1977) may be a common overt characteristic of these tasks, contrasting its nature with the memory consolidation function that is the primary distinguishing characteristic of the WJ IV Glr tests.

Cognitive Complexity and Auditory Intelligence

In the design of several new tests and in revision of some other tests, cognitive complexity was infused by increasing or expanding the breadth of information-processing demands. Cognitively complex information processing typically requires greater allocation of key cognitive resources (such as working memory and attention control) and involves more executive functions (Arend et al., 2003; Jensen, 2011; Lohman & Lakin, 2011; Marshalek, Lohman, & Snow, 1983: McGrew, 2012). As a consequence of the conjoint integration or hierarchical application of other critical cognitive processes and capacities, storage and retrieval functions, or basic fluency mechanisms, cognitively complex tasks often correlate with other tests that tap similar abilities, and will thus load on more than one cognitive factor. Although cognitive complexity is not the same as factorial complexity (e.g., a mixed measure of Gf and Gv), task requirements of cognitively complex tests often require the interplay of more than one CHC ability, just as typical classroom or occupational task demands do.

Cognitively complex tests tend to correlate highly with psychometric *g*, making them well suited for evaluation of general intellectual ability and prediction of academic achievement. Consequently, the concept of cognitive complexity was an important criterion in selection of the new tests that compose the WJ IV COG General Intellectual Ability (GIA) score and the Scholastic Aptitude clusters, including in particular COG Test 2: Number Series, COG Test 3: Verbal Attention, and COG Test 5: Phonological Processing. These tests correlate highly with measures of *g*—not necessarily because the tasks require high-level abstract reasoning or problem solving, but because each task requires the efficient interaction of simple cognitive operations (e.g., searching a person's mental lexicon network based on sounds) or conjointly taxes the parameters of cognitive efficiency (e.g., working memory capacity and attentional control).

The tests that compose the WJ IV COG Auditory Processing (Ga) cluster are cognitively complex, in terms of both information-processing demands and factor composition. For example, COG Test 5: Phonological Processing measures three aspects of speech sound processing that lead to the construction of sound-based lexical representations; the three aspects invoke parameters of language development, word fluency, and phonological sensitivity. In terms of contemporary CHC theory, COG Test 5: Phonological Processing involves comprehension–knowledge as a broad cognitive ability and invokes applications of phonetic coding (PC), language development (LD), word fluency (FW), and speed of lexical access (LA). COG Test 12: Nonword Repetition is a complex task that is best explained by a combination of two factors—auditory short-term working memory and "a further ability that is specific to the repetition of novel multisyllabic phonological forms" (Archibald & Gathercole, 2007, p. 923). Although primarily phonological in nature, nonword repetition ability is constrained by phonological storage capacity (Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006). COG Test 12: Nonword Repetition measures both span in short-term working memory and phonological sensitivity in auditory processing (McGrew et al., 2014).

As a factor of intelligence, auditory processing (Ga) has historically received far less attention (and was less likely to be measured in intelligence tests) than its parallel cognitive construct, visual processing (Gv). Although broad auditory perception was firmly established as a factor of general intelligence by Carroll (1993), seminal works by Stankov and Horn (1980) and Horn and Stankov (1982) were foundational to a recognition of the importance of auditory abilities to academic performance, particularly orthographic skill development. A relative latecomer to the table of intellectual abilities, the construct of auditory processing is still being refined in terms of scope and defining nomenclature (McGrew et al., 2014).

Another body of research, not previously integrated into CHC theory, posited the existence of a phonological storage ability or short-term memory function for speech sounds (Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992, 1994). Although Carroll (1993) identified a narrow auditory processing ability apparently involving memory as a critical component, his datasets were small and consisted primarily of tests utilizing patterns of nonspeech sounds. However, it is of interest to note that Carroll was able to distinguish an aspect of auditory processing involving memory, and he was intrigued with the possibility of distinguishing an auditory factor analogous to the narrow visual memory (MV) factor (Carroll, 1993, p. 391). Consequently, a growing body of research that relates auditory abilities to the development of language and cognition, including the studies of Gathercole and colleagues, has suggested a need to expand the initial description of the CHC narrow ability called *memory for sound patterns* (UM) to include memory for speech sounds; it also supports the inclusion of the auditory short-term working memory tasks as part of the construct of auditory processing.

Related research led to the suggestion that auditory abilities can be involved in reasoning as well as memory functions. Conzelmann and Süß (2015) credited the research of Seidel (2007) in a proposed definition of auditory intelligence as "the ability to discriminate, remember, reason, and work creatively on auditory stimuli, which may consist of tones, environmental sounds, and speech units" (p. 28). The expansion of the construct of auditory processing to include aspects of reasoning with auditory stimuli is supported by the phonological mediation hypothesis (Van Orden, 1987; Van Orden & Goldinger, 1996), which suggests that phonological codes (or cues) can convey or transmit information that facilitates word identification. Consequently, in contemporary CHC theory (see Schneider & McGrew, Chapter 3, this volume), auditory processing includes the activation and restructuring of information in primary (working) memory and retrieval of information from secondary (long-term) memory based on phonological codes.

Phonological codes play an important role in the cognitive processes involved in identifying word meaning and in word production (Vitevitch & Goldstein, 2014). The ability to reason with speech sounds and retrieve words based on phonological cues is particularly applicable to reading, including identifying words and understanding passages (Leinenger, 2014). For example, when a person is reading silently, orthographic information is often recoded into phonological information. The process of phonological coding produces the sensation of an inner voice that can activate—by making a connection to—candidate words in the individual's lexicon. In fact, some researchers have argued that phonological codes are the primary means by which readers achieve lexical access and understand what they read (Lukatela & Turvey, 1994a, 1994b; Van Orden, 1987). As Van Orden and Goldinger (1996) suggested, "We read in order to 'hear,' in order to understand" (p. 206).

Although auditory processing abilities are measured in several different ways in several WJ IV tests, the WJ IV COG Auditory Processing (Ga) cluster was designed to best represent the broader construct of auditory intelligence that includes the phonological memory and reasoning functions alluded to by Conzelmann and Süß (2015). As a component of general intelligence and predictor of achievement, COG Test 5: Phonological Processing measures three aspects of speech sound processing that lead to the construction of lexical representations from phonological codes. Also referred to as tasks of *phonological awareness* or *phonological sensitivity,* these types of tasks are closely associated with learning new words (de Jong, Seveke, & van Veen, 2000), learning to read (Liberman, Shankweiler, & Liberman, 1989; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993; Wagner, Torgesen, & Rashotte, 1994), and the ability remains critical even for mature readers when approaching difficult, infrequent, or unfamiliar words in text (Jared, Levy, & Rayner, 1999; McCusker, Hillinger, & Bias, 1981).

Although changes to CHC constructs in the domains of working memory and attentional control, memory consolidation and retrieval, and auditory intelligence are not the only modifications to CHC theory since its debut in the WJ III (Woodcock et al., 2001), these examples demonstrate how the evolving interpretation of the WJ IV has been based on an accumulation of knowledge about the structure of, and interplay among, cognitive abilities. In each subsequent section of this chapter, information is provided about what each WJ IV test or cluster measures according to contemporary CHC theory, as well as relationships of the WJ IV tests to other measures that assess the same or similar interpretive constructs. Very important, we think, is the concept of *consequential validity*—the idea that valid measures of cognitive, linguistic, and academic abilities should be relevant for purposes of an evaluation, and that scores from the WJ IV should have meaning for planning and decision making. For the WJ IV tests and clusters, the links to practical suggestions and educational interventions or accommodations provide evidence of consequential validity.

WJ IV TESTS OF COGNITIVE ABILITIES

The WJ IV COG includes 18 tests for measuring intellectual ability, broad and narrow cognitive abilities, and related aspects of cognitive functioning. The WJ IV COG is best suited for use with school-age children, adolescents, and adults.

Organization of the WJ IV COG

Figure 14.1 presents the Selective Testing Table for the WJ IV COG. Tests 1–10 comprise the Standard Battery; Tests 11–18 comprise the Extended Battery. Combinations of tests comprise clusters for interpretive purposes. Three cognitive composites are available: General Intellectual Ability (GIA), Brief Intellectual Ability (BIA), and the Gf-Gc Composite. The WJ IV COG has the following clusters that measure seven broad CHC factors: Comprehension–Knowledge (Gc), Fluid Reasoning (Gf), Short-Term Working Memory (Gwm), Cognitive Processing Speed (Gs), Auditory Processing (Ga), Long-Term Storage and Retrieval (Glr), and Visual Processing (Gv). Measures of several narrow CHC abilities or otherwise clinically useful clusters are also available, including Quantitative Reasoning (RQ), Number Facility (N), Perceptual Speed (P), and Cognitive Efficiency. When selected tests from the WJ IV OL are administered with designated tests from the WJ IV COG, clusters that measure additional narrow abilities are available: Auditory Memory Span (MS) and Vocabulary (VL/LD). This section of the chapter contains definitions of what is measured by each of the WJ IV COG tests and clusters, and, based on the cognitive processing demands

			Cognitive Composites			CHC Factors							Narrow Ability and Other Clinical Clusters					
			General Intellectual Ability (GIA)	Brief Intellectual Ability	Gf-Gc Composite	Comprehension-Knowledge (Gc)	Fluid Reasoning (Gf)	Short-Term Working Memory (Gwm)	Cognitive Processing Speed (Gs)	Auditory Processing (Ga)	Long-Term Retrieval (Glr)	Visual Processing (Gv)	Quantitative Reasoning (RQ)	Auditory Memory Span (MS)	Number Facility (N)	Perceptual Speed (P)	Vocabulary (VL/LD)	Cognitive Efficiency
Standard Battery	COG 1	Oral Vocabulary	■	■	■	■											■	
	COG 2	Number Series	■	■	■		■						■					
	COG 3	Verbal Attention	■	■				■										□
	COG 4	Letter-Pattern Matching	■						■							■		■
	COG 5	Phonological Processing	■							■								
	COG 6	Story Recall	■								■							
	COG 7	Visualization	■									■						
	COG 8	General Information			■	■												
	COG 9	Concept Formation			■		■											
	COG 10	Numbers Reversed						■							■			■
Extended Battery	COG 11	Number-Pattern Matching													■	■		□
	COG 12	Nonword Repetition								■								
	COG 13	Visual-Auditory Learning									■							
	COG 14	Picture Recognition										■						
	COG 15	Analysis-Synthesis					□						■					
	COG 16	Object-Number Sequencing						□										
	COG 17	Pair Cancellation							■									
	COG 18	Memory for Words												■				
Oral Language Battery	OL 1	Picture Vocabulary				□											■	
	OL 5	Sentence Repetition												■				

■ Tests required to create the cluster listed.

□ Additional tests required to create an extended version of the cluster listed.

FIGURE 14.1. Selective Testing Table (tests and clusters) for the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG). The Glr cluster is more accurately defined as Long-Term Storage and Retrieval. From *Woodcock–Johnson IV™* (*WJ IV™*). Copyright © The Riverside Publishing Company. All rights reserved. Used by permission of the publisher.

required for performance on the tests, provides links to interventions or accommodations that may be helpful in many situations for developing a targeted educational plan that is based in part on an individual's cognitive profile. Figure 14.1 can be used to help examiners determine which tests to administer for the desired level of interpretive information in an evaluation.

How to Administer the WJ IV COG Tests

The WJ IV COG tests are placed in an order that facilitates both administration and interpretation. For example, examiners can begin testing with Test 1, administer additional tests in sequential order, and discontinue testing at a point that provides the level of information desired or required. Administration of the core tests (Tests 1–7) allows examiners to obtain an evaluation of any intra-individual differences among component scores. Administering the core tests often yields the most essential information in the least amount of testing time because each of the first seven tests is the best single indicator of a broad ability. Any information or clinical observations obtained during or from administration of the core tests can prompt an examiner to administer additional tests (e.g., to obtain a cluster score). An important point to remember is that administering the core tests meets the minimal requirement for calculation of intra-individual variations, and that additional tests and clusters can also be included in the same analysis.

Many examiners routinely administer Tests 1–10 because of the depth of interpretive information obtained, including clusters scores for Gc, Gf, Gwm, Cognitive Efficiency, the Gf-Gc Composite, and an automated comparison of the Gwm and Cognitive Efficiency scores to the Gf-Gc Composite. This is called a *standard administration protocol* for the WJ IV COG (Schrank et al., 2016). Any additional WJ IV clusters that are not primarily Gc or Gf can also be compared to the Gf-Gc Composite to determine whether other intraindividual strengths and weaknesses exist relative to the Gf-Gc Composite. The Gf-Gc Composite may have particular relevance as a measure of intellectual development or intellectual level for some individuals who have a limitation in one or more areas of cognitive processing or efficiency. This can be true for individuals whose Gf-Gc Composite scores are markedly higher than their scores from other cognitive, oral language, or achievement clusters, which may help define the nature of a disability (Schrank, McGrew, & Mather, 2015a; Schrank et al., 2016). In fact, the WJ IV authors recommend use of the Gf-Gc Composite as the most appropriate index of intellectual level for comparison to measures of other cognitive processes, cognitive–linguistic abilities, and areas of academic performance for identification of a specific learning disability (Schrank et al., 2015a).

Selected tests from the WJ IV COG can be used in conjunction with the WJ IV ACH to determine if an individual is achieving as well as would be predicted from scores on a small set of closely related cognitive tests. In the WJ IV, these are called *scholastic aptitude–achievement comparisons*. However, the Scholastic Aptitude clusters are not meant to determine the presence of a specific learning disability, and using them in this way is counter to the meaning of determining an ability–achievement discrepancy as described in federal guidance and state eligibility guidelines.

Administration Time

As a general rule, experienced examiners require about 35–40 minutes to administer the seven core tests, or 50–60 minutes to administer all of the tests in the Standard Battery. As a general guideline, an examiner should allow approximately 5 minutes per additional cognitive or oral language test administered.

Testing Materials

Testing materials required for administration of the WJ IV COG include the test book, a test record, the audio CD, appropriate audio equipment, pencils, and a stopwatch.

Summary of Key Administration and Scoring Points

Most tests use suggested starting points and basal and ceiling rules, and responses are scored 1 or 0. Table 14.1 summarizes the key administration and scoring points for each test.

Special Administration and Scoring Considerations

Although the WJ IV COG examiner's manual (Mather & Wendling, 2014b) and test book provide detailed rules for test administration, this section presents important reminders about tests that have special administration or scoring rules.

TABLE 14.1. Summary of WJ IV COG Test Administration and Scoring Rules

Test name	Item scoring rule	Sample items?	Basal/ceiling rules?[a]	Administration notes	Extra materials required for administration
			Standard Battery tests		
Test 1: Oral Vocabulary				Both subtests must be administered.	
1A: Synonyms	1, 0	Yes	Yes; 6/6	Suggested starting points are available after sample items.	
1B: Antonyms	1, 0	Yes	Yes; 6/6	Suggested starting points are available after sample items.	
Test 2: Number Series	1, 0	Yes	Yes; 5/5	Suggested starting points are available after sample item.	Response booklet; stopwatch to monitor response time
Test 3: Verbal Attention	1, 0	Yes	Yes; 6/6	Suggested starting points available.	Audio
Test 4: Letter–Pattern Matching	1 for each correct pair in 3 minutes	Yes	No; 3-minute time limit	All begin after sample items and practice exercise; use scoring guide.	Stopwatch (timed test); response booklet
Test 5: Phonological Processing				All three subtests must be administered.	
5A: Word Access	1, 0	Yes	Yes; 6/6	Begin with sample item based on estimated ability.	Audio
5B: Word Fluency	1 for each correct word in 1 minute	No	No; 1-minute time limit per item	All examinees are administered both items.	Stopwatch (timed test)
5C: Substitution	1, 0	Yes	Yes; 6/6	Locate suggested starting point based on estimated ability; then begin with sample items.	Audio
Test 6: Story Recall	1 for each correct element	No	No; ceiling determined by continuation instructions	Suggested starting points are available; scoring is based on administration of specific groups of stories.	Audio
Test 7: Visualization				Both subtests must be administered.	
7A: Spatial Relations	1, 0	Yes	Item 1 is basal; ceiling is 5	All begin with the introduction.	
7B: Block Rotation	1, 0	Yes	Item 1 is basal; ceiling is 5	Suggested starting points are available; all begin with an introduction.	Stopwatch to monitor response time

Test 8: General Information				Both subtests must be administered.	
8A: Where	1, 0	No	Yes; 4/4	Suggested starting points are available.	
8B: What	1, 0	No	Yes; 4/4	Suggested starting points are available.	
Test 9: Concept Formation	1, 0	Yes	No; ceiling determined by cutoffs	Suggested starting points are available; proceed to introduction and sample items.	Stopwatch to monitor response time
Test 10: Numbers Reversed	1, 0	Yes	Yes; 5/5	Suggested starting points are available; then begin with sample item.	Audio
			Extended Battery tests		
Test 11: Number–Pattern Matching	1 for each correct pair in 3 minutes	Yes	No; 3-minute time limit	All begin after sample items and practice exercise; use scoring guide.	Stopwatch (timed test); response booklet
Test 12: Nonword Repetition	1, 0	Yes	Yes, 6/6	All begin with sample items; then use suggested starting points.	Audio
Test 13: Visual–Auditory Learning	Count of errors	No	All begin with Introduction 1; ceiling determined by cutoffs	Record number of errors for each story on test record; scoring is based on stories administered.	Stopwatch to monitor response time
Test 14: Picture Recognition	1, 0	Yes	Yes; 6/6	Suggested starting points are available; then begin with sample item.	Stopwatch to monitor time to view stimulus
Test 15: Analysis–Synthesis	1, 0	Yes	All begin with item 1; ceiling determined by cutoffs	All begin with color pretest, then Introduction 1 and sample item; scoring is based on items administered.	Stopwatch to monitor response time
Test 16: Object–Number Sequencing	1, 0	Yes	Yes; 5/5	All begin with sample item; then use suggested starting points.	Audio
Test 17: Pair Cancellation	1 for each correct pair in 3 minutes	Yes	No; 3-minute time limit	All begin with sample items and practice exercise; use scoring guide.	Stopwatch (timed test); response booklet
Test 18: Memory for Words	1, 0	Yes	Yes; 4/4	All begin with sample items; then use suggested starting points.	Audio

---

[a] May be modified by complete-page rule.

Timed Tests

The WJ IV COG contains four timed tests. Test 4: Letter–Pattern Matching, Test 11: Number–Pattern Matching, and Test 17: Pair Cancellation each have a 3-minute time limit. Part B Word Fluency of Test 5: Phonological Processing is also timed; this subtest has two items that each have a 1-minute time limit. The time limits are noted on the appropriate test book pages as well as on the test record. Administering these tests requires using a stopwatch or a watch with a second hand.

For Part B Word Fluency of Test 5: Phonological Processing, duplicate responses are not accepted, but variations of a word do receive credit (e.g., *do, doing, does*). Words presented as examples do not count toward credit (e.g., *milk* and *dog*). An examiner should record a tally mark for each correct response, grouping by fives. For each item, the examiner counts the correct responses and records the total in the Number Correct box. The maximum number that can be entered for each item is 99.

Audio Tests

The standardized audio recording should be used to present Test 3: Verbal Attention, Part A Word Access of Test 5: Phonological Processing, Test 6: Story Recall, Test 10: Numbers Reversed, Test 12: Nonword Repetition, Test 16: Object–Number Sequencing, and Test 18: Memory for Words. An examiner should not repeat or replay any item unless the directions permit it (e.g., on sample items).

Test 6: Story Recall

The score on Test 6: Story Recall is based on the number of correct responses the examinee has on the set of stories administered. Each story element is based on a keyword set in boldface type in the test record, which must be recalled exactly to receive a score of 1. Synonyms that preserve the meaning of the boldface word are also scored 1 (e.g., *speak* for *talk*). Derivations of or possessive case added to proper names are also scored 1 (e.g., *Rick* for *Rick's*). Responses that vary in verb tense or number (e.g., singular–plural) are scored 1 (e.g., *star* for *stars*). Nonboldface words do not need to be recalled exactly, but can be recalled as a general concept. In the score entry section of the test record, an examiner should enter the number of points for each set of stories administered and enter an X if a set was not administered. To obtain estimated age and grade equivalents, the examiner locates the number of points in the appropriate column corresponding to the group of stories administered.

Test 13: Visual–Auditory Learning

In Test 13: Visual–Auditory Learning, the score is based on the number of errors the examinee makes. In the Test Record, the examiner circles each word the examinee misses or is told after a 5-second pause. Omitted words, inserted words, synonyms are scored as errors. For words with two symbols (e.g., *ride . . . ing*), each symbol is considered separately in scoring. If one part of a two-symbol word is incorrect, only the incorrect part is scored as an error. The total number of errors at each cutoff point is entered in the box provided in the test record. The examiner enters the number of errors and the letter corresponding to the set of test stories administered in the score entry box in the test record, and in the corresponding field in the Online Scoring and Reporting System.

Reliability and Validity

Examiners can also selectively administer any tests and clusters to meet an assessment purpose. Each test and cluster should be interpreted from the perspective of available validity evidence, including coefficients of stability. Median reliability coefficients for the WJ IV COG tests are reported in Table 14.2. With the exception of the timed tests, coefficients of stability are reported as median internal-consistency (r_{11}) and median standard error of measurement *(SEM)* values. The internal-consistency reliabilities were calculated with the split-half procedure (odd and even items) and corrected for length with the Spearman–Brown formula. *SEM* values provide information about the precision of each test's scores. For the three timed tests, test–retest values (r_{12}) were used as the most appropriate coefficient of stability. Mosier's (1943) formula was used to calculate reliabilities for tests with subtests, such as COG Test 5: Phonological Processing. Mosier's formula uses individual subtest reliabilities that were obtained from either, or a combination of, the split-half procedure or the test–retest procedures.

Table 14.3 contains the median cluster reliability coefficients (r_{cc}) and median standard errors of measurement for obtained cluster standard scores, or *SEM* (SS), as calculated via Mosier's formula. All of the WJ IV COG CHC-based broad and nar-

TABLE 14.2. WJ IV COG Median Test Reliability Statistics

Test	Median r_{11}	Median *SEM* (SS)	Median r_{12}
	Standard Battery		
Test 1: Oral Vocabulary	.89	4.97	—
Test 2: Number Series	.91	4.64	—
Test 3: Verbal Attention	.86	5.70	—
Test 4: Letter–Pattern Matching	—	—	.91
Test 5: Phonological Processing	.84	6.00	—
Test 6: Story Recall	.93	3.90	—
Test 7: Visualization	.85	5.81	—
Test 8: General Information	.88	5.20	—
Test 9: Concept Formation	.93	4.04	—
Test 10: Numbers Reversed	.88	5.15	—
	Extended Battery		
Test 11: Number–Pattern Matching	—	—	.85
Test 12: Nonword Repetition	.91	4.55	—
Test 13: Visual–Auditory Learning	.97	2.65	—
Test 14: Picture Recognition	.74	7.70	—
Test 15: Analysis–Synthesis	.93	4.02	—
Test 16: Object–Number Sequencing	.89	4.95	—
Test 17: Pair Cancellation	—	—	.89
Test 18: Memory for Words	.82	6.39	—

Note. r_{11}, internal-consistency reliability; r_{12}, test–retest reliability for speeded tests.

row ability clusters, with the exception of Visual Processing (Gv), exceed the commonly referenced professional standard of .90 for decision-making purposes. The reliability coefficients for the GIA and Gf-Gc Composite are very high. Most of the WJ IV COG tests exceed the .80 value, with some test reliabilities exceeding .90. The reported test reliabilities support use and interpretation of the test-level scores as single indicators of a broad cognitive ability for inferential and qualitative analysis, but the higher cluster reliabilities support the recommendation that the WJ IV COG clusters are the preferred indexes for most decision-making purposes, such as determining a deficit or impairment in cognitive processing. Note also that reliabilities for most of the WJ IV COG Scholastic Aptitude clusters do not exceed .90 (although they approach that value), and that these clusters should not be used to supplant use of the WJ IV COG GIA or Gf-Gc Composite for purposes of determining an ability–achievement discrepancy or determining whether a pattern of strengths and weaknesses exists. The Scholastic Aptitude clusters are intended for descriptive analysis only.

Table 14.4 contains the correlations for the WJ IV COG GIA, BIA, and Gf-Gc Composite with other composite measures of intellectual ability, including the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003), the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV; Wechsler, 2008), the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004a), and the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003). As noted by Reynolds and Niileksela (2015), “Correlations between corresponding scores from the WJ IV COG and the Wechsler scales were especially strong” (p. 389). The consistently strong correlations of the WJ IV GIA, BIA, and Gf-Gc Composite scores with Full Scale IQ scores from the WISC-IV, WAIS-IV, and SB5 supports the validity of the WJ IV indexes as valid measures of psychometric intelligence. Of particular interest is the notably lower correlation of the WJ IV COG Gf-Gc Composite with the KABC-II Mental Processing Index; this supports the interpretation of the WJ IV COG Gf-Gc Composite as a measure of intellectual level for use in an evaluation of learning disabilities because it is less influenced by mental processing abilities or disabilities than the WJ IV COG GIA or BIA scores (Schrank et al., 2015a).

TABLE 14.3. WJ IV COG Median Cluster Reliability Statistics

Cluster	r_{cc}	Median *SEM* (SS)
General Intellectual Ability	.97	2.60
Brief Intellectual Ability	.94	3.35
Gf-Gc Composite	.96	2.12
Comprehension–Knowledge (Gc)	.93	3.00
Comprehension–Knowledge—Ext.[a]	.94	2.60
Fluid Reasoning (Gf)	.94	3.00
Fluid Reasoning—Ext.	.96	3.00
Short-Term Working Memory (Gsm)	.91	5.20
Short-Term Working Memory—Ext.	.93	4.50
Cognitive Processing Speed (Gs)	.94	3.67
Auditory Processing (Ga)	.92	4.74
Long-Term Storage and Retrieval (Glr)	.97	3.00
Visual Processing (Gv)	.86	4.97
Quantitative Reasoning (RQ)	.94	3.00
Auditory Memory Span (MS)[b]	.90	5.41
Number Facility (N)	.90	4.50
Perceptual Speed (P)	.93	4.24
Cognitive Efficiency	.93	3.67
Cognitive Efficiency—Ext.	.95	3.97
Reading Aptitude A	.89	5.20
Reading Aptitude B	.90	5.20
Math Aptitude A	.89	5.20
Math Aptitude B	.89	5.20
Writing Aptitude A	.89	5.20
Writing Aptitude B	.90	3.35

[a] Requires WJ IV OL Test 1: Picture Vocabulary.

[b] Requires WJ IV OL Test 5: Sentence Repetition.

Interpretation

The WJ IV COG can be interpreted at the cluster and test levels, depending on the purposes of an evaluation. The broad and narrow abilities described in contemporary CHC theory provide the basic interpretive architecture for the validity of the clusters and tests. Interpretation is based on extensive psychometric data and analyses provided by McGrew and colleagues (2014), as well as a trove of related research outside the WJ IV literature (see Schrank et al., 2016) that provides support for our interpretation of the underlying cognitive processes required for performance on each test. Table 14.5 identifies the broad and narrow cognitive abilities measured by each test in the WJ IV COG; it also includes brief test descriptions. The WJ IV COG tests are organized into clusters for interpretive purposes; these clusters are outlined in Table 14.6. The WJ IV COG clusters and tests are described in this section. The WJ IV COG options for measuring intellectual level or ability are also described. Finally, we describe several variation and discrepancy procedures that are available to assist examiners in comparing scores, determining profiles of abilities, and making diagnostic decisions.

TABLE 14.4. Correlations for Select WJ IV COG Measures and Other Measures of Cognitive Abilities

Other measure	General Intellectual Ability (GIA)	Brief Intellectual Ability (*g*)	Gf-Gc Composite
Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV)[a]	.86	.83	.83
Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)[a]	.84	.74	.78
Kaufman Assessment Battery for Children—Second Edition (KABC-II)[b]	.72	.67	.57
Stanford–Binet Intelligence Scales, Fifth Edition (SB5)[a]	.80	.79	.82

[a] The measure reported is the Full Scale IQ (*g*) score.

[b] The measure reported is the Mental Processing Index score.

TABLE 14.5. WJ IV COG Tests, CHC Broad and Narrow Abilities Measured, and Brief Test Descriptions

Test name	Primary broad CHC ability *Narrow ability*	Brief test description
Test 1: Oral Vocabulary	Comprehension–knowledge (Gc) *Lexical knowledge* (VL) *Language development* (LD)	Measures knowledge of words and word meanings.
Test 2: Number Series	Fluid reasoning (Gf) *Quantitative reasoning* (RQ) *Inductive reasoning* (I)	Assesses quantitative reasoning.
Test 3: Verbal Attention	Short-term working memory (Gwm) *Working memory capacity* (WM) *Attentional control* (AC)	Measures temporary storage of verbal information and the cue-dependent search function of primary memory.
Test 4: Letter–Pattern Matching	Cognitive processing speed (Gs) *Perceptual speed* (P)	Evaluates orthographic visual-perceptual discrimination ability under timed conditions.
Test 5: Phonological Processing	Auditory processing (Ga) *Phonetic coding* (PC) *Speed of lexical access* (LA) *Word fluency* (FW)	Measures word activation, fluency of word access, and word restructuring via phonological codes.
Test 6: Story Recall	Long-term storage and retrieval (Glr) *Meaningful memory* (MM) *Listening ability* (LS)	Assesses listening ability with attention to orally imparted details; formation of mental representations in the stimulus phase; story reconstruction in the response phase.
Test 7: Visualization	Visual processing (Gv) *Visualization* (Vz)	Evaluates size and shape perception, part-to-whole analysis, and mentally transforming two- and three-dimensional images.
Test 8: General Information	Comprehension–knowledge (Gc) *General (verbal) information* (K0)	Measures general object knowledge.
Test 9: Concept Formation	Fluid reasoning (Gf) *Inductive reasoning* (I)	Evaluates verbal (language-based) inductive reasoning.
Test 10: Numbers Reversed	Short-term working memory (Gwm) *Working memory capacity* (WM)	Measures temporary storage and recoding of numerical information in primary memory
Test 11: Number–Pattern Matching	Cognitive processing speed (Gs) *Perceptual speed* (P)	Assesses numerical visual-perceptual discrimination ability under timed conditions.
Test 12: Nonword Repetition	Auditory processing (Ga) *Phonetic coding* (PC) *Memory for sound patterns* (UM) *Auditory memory span* (MS)	Assesses phonological short-term working memory, sensitivity, and capacity.
Test 13: Visual–Auditory Learning	Long-term storage and retrieval (Glr) *Associative memory* (MA)	Measures visual–auditory paired-associate encoding in the learning phase; identification and word retrieval in the response phase.
Test 14: Picture Recognition	Visual processing (Gv) *Visual memory* (MV)	Evaluates formation (passive storage) of images in the visual cache and recognition of previously presented visual stimuli.
Test 15: Analysis–Synthesis	Fluid reasoning (Gf) *General sequential (deductive) reasoning* (RG)	Measures algorithmic, deductive reasoning.

(continued)

TABLE 14.5. *(continued)*

Test name	Primary broad CHC ability *Narrow ability*	Brief test description
Test 16: Object–Number Sequencing	Short-term working memory (Gwm) *Working memory capacity* (WM)	Assesses assembly of new cognitive structures out of information maintained in working memory.
Test 17: Pair Cancellation	Cognitive processing speed (Gs) *Perceptual speed* (P) *Spatial scanning* (SS) *Attentional control* (AC)	Evaluates symbolic visual-perceptual discrimination ability requiring cognitive control under timed conditions.
Test 18: Memory for Words	Short-term working memory (Gwm) *Auditory memory span* (MS)	Measures storage capacity for unrelated words in primary memory.

TABLE 14.6. WJ IV COG Clusters and Brief Cluster Descriptions

Cluster	Brief cluster description
General Intellectual Ability (GIA)	Measures psychometric *g* as distilled from a broad spectrum of important cognitive abilities and functions: knowledge of words and their meanings; quantitative reasoning; temporary storage of verbal information, including the cue-dependent search function of primary memory; orthographic visual-perceptual discrimination ability under timed conditions; word activation, fluency of word access, and word restructuring via phonological codes; listening and attention to orally imparted details that facilitate formation and reconstruction of mental representations; and size and shape perception, part-to-whole analysis, and mentally transforming two- and three-dimensional images.
Gf-Gc Composite	Assesses intellectual level or development based solely on fluid reasoning (Gf) and comprehension–knowledge (Gc) tests.
Brief Intellectual Ability (BIA)	Provides a short but highly reliable measure of intellectual ability based on a sampling of verbal comprehension–knowledge (Gc) and quantitative fluid reasoning (Gf) abilities, and the executive information-lookup and storage capacity functions in short-term working memory (Gwm).
Comprehension-Knowledge	Assesses comprehension of words and general object knowledge.
Comprehension-Knowledge—Ext.	Assesses comprehension of words and general object knowledge, including knowledge of object names.
Fluid Reasoning	Measures quantitative and verbal reasoning.
Fluid Reasoning—Ext.	Measures quantitative, algorithmic, and verbal reasoning.
Short-Term Working Memory	Assesses cue-dependent search and recoding functions from temporary stores of verbal and numerical information in primary memory.
Short-Term Working Memory—Ext.	Assesses cue-dependent search, recoding, and assembly functions from temporary stores of verbal and numerical information in primary memory.
Cognitive Processing Speed	Evaluates orthographic and symbolic visual-perceptual discrimination ability and attentional control under timed conditions.
Auditory Processing	Measures word activation, access, restructuring via phonological codes, and phonological sensitivity capacity in working memory.

(continued)

TABLE 14.6. *(continued)*

Cluster	Brief cluster description
Long-Term Storage and Retrieval	Assesses consolidation (encoding) of semantic (meaning-based) representations into secondary memory.
Visual Processing	Measures visual–spatial analysis, formation of internal visual images, mental transformation of images in working memory, passive storage, and recognition of images.
Cognitive Efficiency	Evaluates efficiencies of orthographic visual-perceptual discrimination ability under timed conditions. and temporary storage and recoding of numeric information in primary memory.
Cognitive Efficiency—Ext.	Measures a combination of orthographic visual-perceptual discrimination ability under timed conditions, and cue-dependent search and recoding functions from temporary stores of verbal and numeric information.
Perceptual Speed	Evaluates orthographic visual-perceptual discrimination ability under timed conditions.
Quantitative Reasoning	Measures quantitative and non-numerical algorithmic reasoning.
Number Facility	Evaluates efficiencies of visual-perceptual discrimination, temporary storage, and processing of numerical information in working memory.
Vocabulary[a]	Measures knowledge of object names and words and their meanings.
Auditory Memory Span[b]	Assesses storage capacity for unrelated words and connected discourse in primary memory.

[a] Requires WJ IV OL Test 1: Picture Vocabulary.

[b] Requires WJ IV OL Test 5: Sentence Repetition.

Comprehension–Knowledge (Gc)

Comprehension–knowledge (Gc) is a CHC broad ability that is also sometimes referred to as either *verbal ability* or *crystallized intelligence* (Keith & Reynolds, 2010). Although Cattell (1941, 1943, 1963) suggested that the constructs of verbal ability and crystallized intelligence are not synonymous, Carroll (1993) argued that it is purely a matter of individual preference whether Gc is interpreted as verbal ability or crystallized intelligence. To help bridge this chasm, the WJ IV COG Gc cluster is broadly based, measuring both comprehension of words (an aspect of verbal ability) and general knowledge (an aspect of crystallized intelligence). The two tests that compose the Gc cluster are Test 1: Oral Vocabulary and Test 8: General Information.

Test 1: Oral Vocabulary

Oral Vocabulary measures knowledge of words and word meanings (Schrank et al., 2016). The CHC narrow abilities are described as lexical knowledge (i.e., vocabulary knowledge) and language development (i.e., general development of spoken language skills that do not require reading ability). Oral Vocabulary includes two subtests, Synonyms and Antonyms. An analysis of the task demands from an information-processing perspective suggests that a stimulus word is connected to a concept via semantic access, which then activates or primes its meaning in the lexicon and consequently activates closely associated words (Caplan, 1992; Gazzaniga, Irvy, & Mangun, 1998). Synonym and antonym associations from known stimulus words are matched more or less directly and automatically in the neural semantic memory networks of individuals with large vocabularies (Martin, 1998). When a stimulus word is unknown or unfamiliar, a response can be attempted by parsing or segmenting the stimulus word into any recognizable phonological, orthographic, or morphological units (Van Orden & Goldinger, 1994, 1996) for clues that can be used to support an information search within working memory.

Knowledge of words is critically important to school success (Anderson & Nagy, 1992; Baker, Simmons, & Kame'enui, 1998), and no area of cognition is more responsive to intervention than vocabulary knowledge. Many interventions for vocabulary development exist (Schrank & Wendling, 2015b; Wendling & Mather, 2009). To be maximally effective, vocabulary development instruction must be multifaceted and appropriate to the individual's age and ability level (National Reading Panel, 2000). Table 14.7 includes a summary of suggested interventions related to the WJ IV COG tests.

Test 8: General Information

The General Information test measures general object knowledge (Schrank et al., 2016). In this task, an individual is asked to identify where an object is typically found or what people typically do with an object. As suggested by Cattell's (1987) thesis, performance on general information tests is a reflection of background knowledge---the level of general knowledge that an individual brings to the learning situation (Schrank et al., 2016). Background knowledge is extremely important for new learning and complex problem solving because the integration of prior knowledge with new information is the basis for constructing the higher-order cognitive representations required for greater levels of knowledge (Hannon & Daneman, 2014; Oberauer, Süß, Wilhelm, & Wittman, 2008).

Limitations observed in the General Information test may suggest the need for building background knowledge. This involves developing a connection between the topic of instruction and what the learner already knows (sometimes referred to as activating prior knowledge). Activating what a student already knows from personal experience, prior schooling, or family history will aid in understanding of current instructional tasks. When students are encouraged to relate their background knowledge to the material being studied, they become more engaged in the learning process. Learners of any age will benefit from connecting new learning to prior knowledge and/or building background knowledge if necessary (Moje et al., 2004). See Table 14.7.

Fluid Reasoning (Gf)

Fluid reasoning (Gf) is a complex aspect of intelligence that can rely on many cognitive processes, depending on the nature and requirements of the task (Gray, Chabris, & Braver, 2003; Kosslyn & Smith, 2000). Reasoning ability is considered to be a scaffold for the development of many other cognitive abilities, particularly those in the acquired knowledge domain (Blair, 2006; Cattell, 1987; McArdle, 2001). Consequently, measured deficits in Gf may provide insights into observed learning difficulties and may suggest that reasoning skills might need to be specifically modeled, taught, and practiced in an academic context.

Test 2: Number Series, a measure of deductive and inductive quantitative reasoning, and Test 9: Concept Formation, a measure of verbal, language-based inductive reasoning, are two tests that compose the Gf cluster. An additional Gf test, Test 15: Analysis–Synthesis, can be described as a measure of algorithmic reasoning. All three tests combine to create the Fluid Reasoning—Extended cluster.

Test 2: Number Series

Number Series measures the ability to identify and apply an analogue or rule to complete a numerical sequence. The task is cognitively complex (Holzman, Pellegrino, & Glaser, 1983) and requires the application of several narrow abilities, cognitive processes, and executive functions. To be able to deduce a pattern in an item sequence, an examinee must be able to count and (depending on the difficulty level of the item) carry out foundational arithmetic operations, including addition, subtraction, multiplication, and division—so the task also requires application of mathematics knowledge (KM). To perform successfully, an individual must have a solid foundation in math facts (Geary, 1990; Geary & Brown, 1991). The test is a measure of reasoning with numbers, or quantitative reasoning (RQ); general sequential reasoning (deduction) (RG) is required to determine the analogue or rule that solves the task; and induction (I) is required to determine the value that completes the numeric analogy. The level and integrities of other broad and narrow CHC abilities can play supporting or inhibiting roles in performance on Number Series, including retrieval of counting sequences and/or math facts from semantic memory (Temple, 1991), cognitive processing speed (Gs), working memory capacity (WM), and attentional control (AC). Finally, the executive function of place keeping (Hambrick & Altmann, 2015) supports the systematic exploration of hypotheses to generate a problem solution.

TABLE 14.7. Examples of Interventions, Strategies, and Accommodations Related to Limitations in Performance on WJ IV COG Tests

Test	Brief intervention, strategy, or accommodation
Test 1: Oral Vocabulary	Creating a vocabulary-rich environment; reading aloud to a young child and discussing new words; text talks; directed vocabulary thinking activities; explicit teaching of specific words; direct instruction in morphology; semantic feature analysis; semantic maps; association of key words to prior knowledge; reexposure to words; reading for a variety of purposes; independent word-learning strategies; use of computer technology to develop word knowledge.
Test 2: Number Series	Teaching number patterns and core math concepts; counting by increments; use of manipulatives; developing a sense of numerical quantity; math talk.
Test 3: Verbal Attention	Active learning environments; complex computer games; rehearsal; reducing distractions; reducing amount of material to be remembered at any one time; simplifying linguistic structures of verbal material; restructuring complex tasks into separate and independent steps; repeating important information.
Test 4: Letter–Pattern Matching	Focus on learning and recognition of specific English-language orthographic patterns; emphasizing speediness; building perceptual speed via repetition; extended time; reducing quantity of work; eliminating or limiting copying activities.
Test 5: Phonological Processing	Activities that focus on the sounds in words; games to increase production of words with the same sound; rhyming and alliteration games; pictures to stimulate fluent production of object names; oral practice with onset–rime patterns of spoken words; explicit modeling of word pronunciations; intensive practice on word pronunciation in the learning phase; explicit, systematic, synthetic phonics instruction; targeted small-group instruction in syllable dividing and rehearsal; use of semantics in word instruction.
Test 6: Story Recall	Opportunities to hear and practice language; active learning; elaborative rehearsal; direct instruction in semantics; use of visual representations; use of mnemonics.
Test 7: Visualization	Activities designed to develop ability to discriminate visual features, mentally manipulate visual images, and match visual information; increased exposure to graphs, charts, maps; video games; three-dimensional sketching exercises.
Test 8: General Information	Text talks; reading for different purposes; use of relevant, real-word examples for learning; cooperative learning environments.
Test 9: Concept Formation	Categorizing real objects; developing skills in drawing conclusions; performing hands-on problem-solving tasks; making meaningful associations; providing concrete examples of grouping objects; commercially available reasoning games; think-aloud procedures; cooperative learning groups; reciprocal teaching; slowing down in response to difficult reasoning tasks.
Test 10: Numbers Reversed	Chunking strategies; rehearsal; providing visual cues; complex computer games; reducing distractions; reducing working memory loads (reducing amount of material to be remembered at any one time, simplifying linguistic structures of verbal material, restructuring complex tasks into separate and independent steps); repeating important information.

(continued)

TABLE 14.7. *(continued)*

Test	Brief intervention, strategy, or accommodation
Test 11: Number–Pattern Matching	Focus on learning and expression of place value in mathematics; speed drills; building speed via repetition; computer matching games; extended time; reducing quantity of work; eliminating or limiting copying activities.
Test 12: Nonword Repetition	Increased exposure to new words; repeating (uttering) new words when they are learned; activities that require repetition and construction of nonwords; vocalization strategy for new words; increased time spent reading.
Test 13: Visual–Auditory Learning	Active, successful learning experiences; rehearsal; overlearning; mnemonics; illustrating or visualizing content.
Test 14: Picture Recognition	Activities designed to discriminate/match visual features and recall visual information.
Test 15: Analysis–Synthesis	Deductive reasoning activities using concrete objects; hands-on problem-solving tasks; metacognitive strategies; think-aloud procedures.
Test 16: Object–Number Sequencing	Focusing attention on one thing at a time; avoiding multitasking.
Test 17: Pair Cancellation	Emphasizing speediness; slowing down if errors are caused by working too quickly; increasing perceptual speed with computer games.
Test 18: Memory for Words	Listening and repeating games; chunking strategies; mnemonics.

Difficulties with the lowest-numbered Number Series items may suggest a need for the development of foundational knowledge in mathematics, perhaps with manipulatives or board games that require counting and help develop a sense of numerical quantity, patterns, core mathematical concepts, and the relationships among numbers (Gersten et al., 2008; National Council of Teachers of Mathematics, 2000; Ramani & Siegler, 2005). See Table 14.7. For young children, a key ability is the concept of *counting on* (Gersten et al., 2009; National Council of Teachers of Mathematics, 2000). For older children and adolescents, interventions include the development of seriation, pattern recognition, and geometric sequencing skills (National Council of Teachers of Mathematics, 2000)—particularly when used in conjunction with *math talk,* which requires the teacher to listen and provide corrective feedback as the student explains his or her reasoning, step by step, to obtain solutions to problems (High/Scope Educational Research Foundation, 2003; Kroesbergen & Van Luit, 2003).

Test 9: Concept Formation

In contrast to the Number Series test, Concept Formation measures verbal, language-based fluid reasoning, or the ability to use language concepts to categorize and compare—a basis for abstracting universal or rational concepts (Andrewes, 2001; Wang, 1987). The first five items measure categorical perception (Goldstone & Hendrickson, 2010), which requires understanding of the word *different* and the more abstract concept of *most different.* These items measure the ability to recognize and identify the essential, rather than any secondary, characteristic of target objects. This is the first stage in the development of an object concept. Item 6 and beyond require familiarity with and understanding of a number of terms, including *red, yellow, big, little, round, square, one,* and *two.* Concept Formation also requires rule application and frequent switching from one rule to another.

Interventions that are designed to develop skills in categorization and drawing conclusions, that involve connecting new concepts to prior knowledge, that use teacher demonstrations and guided practice, and that provide feedback on performance may positively influence the development of reasoning abilities (Klauer, Willmes, & Phye, 2002). In addition, there are many different board and computer games designed to help develop patterning skills (Willis, 2008), such as finding similarities and differences, sorting, matching, and categorizing (Mackey, Hill, Stone, & Bunge, 2010). See Table 14.7.

Test 15: Analysis–Synthesis

Analysis–Synthesis measures algorithmic, deductive reasoning through the process of analyzing puzzles (using symbolic formulations) to determine missing components. This test requires drawing correct conclusions from stated conditions or premises, often from a series of sequential steps. Because of its use of specific solution keys that, if followed correctly, furnish the correct answer to each test item, Analysis–Synthesis can be described as a measure of algorithmic reasoning. Algorithmic reasoning can be specifically taught (see Table 14.7). When Analysis–Synthesis is administered in conjunction with Number Series, the narrow Quantitative Reasoning (RQ) cluster is obtained.

Short-Term Working Memory (Gwm)

The WJ IV COG Gwm cluster is operationally defined as measuring cue-dependent search and recoding functions from temporary stores of verbal and numeric information in primary memory (Schrank et al, 2016). This cluster consists of Test 3: Verbal Attention and Test 10: Numbers Reversed. Because a combination of qualitatively different working memory tasks creates a cluster that is more broadly predictive of outcomes on a wide variety of cognitive tasks (Oberauer, Süß, Schulze, Wilhelm, & Wittman, 2000; Unsworth & Engle, 2006), a three-test Short-Term Working Memory—Extended cluster that includes Test 16: Object–Number Sequencing is also available; this cluster is operationally defined as measuring a combination of cue-dependent search, recoding, and assembly functions from temporary stores of verbal and numeric information in primary memory (Schrank et al., 2016). Because working memory is a very broad construct, the WJ IV COG includes four working memory tests, each measuring a different function of working memory or aspect of working memory capacity.

Test 3: Verbal Attention

The Verbal Attention test measures the size of a person's attentional focus in primary memory (Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999; Unsworth & Engle, 2007b). This test includes the executive search and updating processes that are characterized by the adjective *working* as a modifier for *memory*. In this task, the individual is required to answer a series of questions, each based on a verbally presented string of unrelated object names and numbers. Verbal Attention represents the Unsworth and Engle (2007b) dual-component model of working memory capacity, in which information is maintained in primary memory through the controlled allocation of attention where a focus of attention is retrieved through a cue-dependent search process.

Performance on this test can be described as maintaining a focus of attention and/or the ability to retrieve information that has been momentarily displaced from attention (Shipstead, Lindsey, Marshall, & Engle, 2014). The cued-recall questions tap the real-time updating function of an individual's working memory (Bunting, Cowan, & Saults, 2006; Dahlin, Stigsdotter Neely, Larsson, Bäckman, & Nyberg, 2008; Miyake et al., 2000). Kosslyn, Alpert, and Thompson (1995) suggested that this "information-lookup" process plays a critical role in working memory. Limitations in the ability to actively maintain and/or retrieve information from working memory may require additional evaluation and compensatory strategies. See Table 14.7.

Test 10: Numbers Reversed

In the Numbers Reversed task, the individual is asked to repeat a series of digits backward. This requires the ability to temporarily store and recode orally presented numerical information in primary memory—a complex span task (Daneman & Carpenter, 1980). Typically, performing complex span tasks reflects the executive, controlled-attention aspect of working memory (Engle, 2002; Kane, Brown, et al., 2007; Kane, Conway, Hambrick, & Engle, 2007); requires intense allocation of attention (Kane, Brown, et al., 2007; Kane, Conway, et al., 2007; Unsworth & Engle, 2006); and is strongly predictive of an individual's attentional control in the service of complex cognitive processing (Hutchison, 2007; Unsworth & Spillers, 2010). Strategies (Schrank & Wendling, 2015b; see Table 14.7) can be employed to extend capacity limits by recoding information into fewer chunks (Hardiman, 2003).

Test 16: Object–Number Sequencing

The Object–Number Sequencing test measures the ability to assemble new cognitive structures out of information maintained in working memory (Schrank et al., 2016). This cognitively complex task is best described by the model of working memory developed by Cowan (1995), Oberauer (2002, 2009), and Oberauer and Hein (2012). Oberauer and Hein distinguish between a *region of direct access* and a *focus of attention*. The region of direct access provides a broad focus (sometimes called a *blackboard*) and has a limited capacity, about four chunks of information (Cowan, 2001; Oberauer & Kliegel, 2006). The primary function of the broad focus is to bind or chunk multiple representations or units together for information processing, such as assembling new structures out of the selected representations. In Object–Number Sequencing, the objects and numbers presented for each item must be retained as chunks in the broad (blackboard) focus of working memory. Then a list of objects must be assembled in sequential order, requiring a single focus of attention. After the sequentially ordered object list is created, the individual must return to the broad focus blackboard and create a new focus of attention by creating a sequential list of numbers from the broad focus blackboard. Individuals who have extreme difficulties with this type of task may benefit by learning to avoid multitasking and focus attention on one thing at a time (Alloway, 2011; Gathercole & Alloway, 2007). See Table 14.7.

Test 18: Memory for Words

Memory for Words measures an individual's storage capacity for unrelated words (Schrank et al., 2016). It is a simple running memory span task that is valuable as a measure of absolute working memory capacity when complex processing is not required (Broadway & Engle, 2010; Bunting et al., 2006). Many individuals who perform well on this type of task employ a strategy to encode and/or recall the words (Mueller, 2002). For example, some individuals use rehearsal as a strategy to maintain or refresh the string of words in immediate awareness, and other individuals employ a chunking strategy to encode the unrelated words into distinct blocks of words or create compound nonwords (Hardiman, 2003; Logie, Della Sala, Laiacona, & Chalmers, 1996). Chunking strategies enable a person to group related items into units, making the information more manageable for understanding, storage, and recall. See Table 14.7.

Perceptual Speed (P) and Cognitive Processing Speed (Gs)

In contemporary CHC theory, perceptual speed (P) is a narrow ability included under the broad domain of cognitive processing speed (Gs). The WJ IV COG Perceptual Speed cluster measures orthographic visual-perceptual discrimination ability under timed conditions (Schrank et al., 2016); this cluster is designed to measure rapid processing with two educationally relevant types of stimuli, representing critical visual inspection efficiencies that are closely related to reading, writing, and mathematics fluency. In contrast, the WJ IV COG Cognitive Processing Speed (Gs) cluster includes more breadth; it measures orthographic and symbolic visual discrimination ability and attentional control under timed conditions (Schrank et al., 2016) and may be more representative of overall cognitive speediness.

For many years, cognitive speediness, or mental quickness, has been considered an important aspect of intelligence (Nettelbeck, 1994; Vernon, 1983). "In the face of limited processing resources, the speed of processing is critical because it determines in part how rapidly limited resources can be reallocated to other cognitive tasks" (Kail, 1991, p. 152). In the WJ IV COG, Test 4: Letter–Pattern Matching and Test 11: Number–Pattern Matching compose the Perceptual Speed (P) cluster. Test 4: Letter–Pattern Matching and Test 17: Pair Cancellation compose the Cognitive Processing Speed (Gs) cluster. Limitations in cognitive speediness, particularly perceptual speed, may have implications for the provision of educational accommodations (Geary & Brown, 1991; Hayes, Hynd, & Wisenbaker, 1986; Kail, 1990, 1991, 2003; Kail, Hall, & Caskey, 1999; Ofiesh, 2000; Shaywitz, 2003; Wolff, Michel, Ovrut, & Drake, 1990).

Test 4: Letter–Pattern Matching

The WJ IV COG Letter–Pattern Matching test measures orthographic visual perceptual discrimination ability under timed conditions (Schrank et al., 2016). Performance on this test is facilitated by well-developed sublexical orthographic recognition and chunking efficiencies. Sublexical orthographic recognition efficiency allows the individual to recognize a pattern as a single chunk for purposes of comparison to other letter strings, rather than attempting to hold a string of letters in primary memory while searching for a match. Learning to quickly recognize and subvocally process orthographic chunks of information is thought to play a critical role in the development of automatic word recognition skill, which supports the development of reading fluency (Apel, 2009) and reading speed (O'Brien, Wolf, Miller, Lovett, & Morris, 2011). Limitations in letter pattern recognition may also be related to spelling problems (Gerber, 1984) such as an overreliance on phonology rather than orthographic knowledge when spelling (Cornelissen, Bradley, Fowler, & Stein, 1994). If so, targeted interventions that focus on learning and recognition of specific English-language spelling patterns may be beneficial (Blevins, 2001; Moats, 2004, 2009). See Table 14.7. An accommodation, such as extended time, may be required if a well-substantiated orthographic visual-perceptual discrimination limitation impairs the individual's ability to demonstrate knowledge under time constraints. For example, some individuals with dyslexia have been observed to show deficits on tasks that require rapid detection of letter position (Cornelissen & Hansen, 1998; Cornelissen, Hansen, Hutton, Evangelinou, & Stein, 1998; Katz, 1977; Pammer, Lavis, Hansen, & Cornelissen, 2004) and may require more time to overcome any adverse impact inherent in timed tasks that require rapid orthographic visual-perceptual discrimination.

Test 11: Number–Pattern Matching

Number–Pattern Matching measures numerical visual perceptual discrimination ability under timed conditions (Schrank et al., 2016); this test is the numerical counterpart to Test 4: Letter–Pattern Matching. Just as Letter–Pattern Matching ability is facilitated by well-developed recognition and chunking strategies for English-language regular letter patterns, Test 11: Number–Pattern Matching ability is influenced by well-developed number-chunking mechanisms. As stated by Schrank and colleagues (2016), "Individuals who are able to mentally represent a string of numbers as a 1-item chunk will possess a processing advantage over individuals who perceive each string of numbers as an unassociated series of numerals" (p. 172). Chunking is an aspect of cognitive processing efficiency (Oberauer & Hein, 2012), and number chunking allows an individual to perceive and process multidigit numbers rapidly and efficiently. Selection of interventions is dependent on whether any limitations are due to poorly developed number-chunking abilities or a more generalized limitation in cognitive speediness. Development of number chunking efficiency is one possible intervention for low performance on Number–Pattern Matching, and perceptual speediness can sometimes be improved with repetitive practice, speed drills, and use of computer games that require quick perceptual decisions (Klingberg, 2009; Mahncke, Bronstone, & Merzenich, 2006; Tallal et al., 1996). See Table 14.7.

Test 17: Pair Cancellation

Pair Cancellation measures speeded visual-perceptual attention (Schrank et al., 2016), an aspect of cognitive control that is responsible for preferential concentration on stimuli of relative importance (Andrewes, 2001). The ability to sustain one's attention is sometimes called *vigilance* (Bunge, Mackey, & Whitaker, 2009; Posner & DiGirolamo, 2000). Good cognitive control—or vigilance—is required for tasks where prior knowledge alone is insufficient to meet task demands, such as learning something new. This is because well-developed cognitive control supports working memory, selective attention, long-term retrieval, response inhibition, and response selection (Cohen, Dunbar, & McClelland, 1990; Desimone & Duncan, 1995; Miller & Cohen, 2001; Race, Kuhl, Badre, & Wagner, 2009). Selection of interventions requires careful analysis of related abilities and test session observations. For example, if analysis of the response worksheet reveals several errors of commission (e.g., circling an incorrect set or pair of pictures), the individual may benefit from interventions designed to increase cognitive control, such as slowing down to increase response accuracy. See Table 14.7.

Auditory Processing (Ga)

The WJ IV COG Auditory Processing (Ga) cluster measures word activation, word access, word restructuring via phonological codes, and phonological sensitivity capacity in working memory (Schrank et al., 2016). In the WJ IV COG, the operational definition of auditory processing has been redefined and distinguished from measures of phonological awareness by an increased emphasis on tasks that involve memory and reasoning processes with auditory stimuli (Conzelmann & Süß, 2015). More so than in prior editions of the COG, the WJ IV Ga cluster accurately reflects the concept of auditory intelligence (Conzelmann & Süß, 2015; Seidel, 2007). The WJ IV COG Ga tests are cognitively complex, requiring a mix of cognitive functions and parameters of cognitive efficiency. Performance on each test requires a combination of narrow abilities that span one or more broad ability. The two component Ga tests are Test 5: Phonological Processing and Test 12: Nonword Repetition.

Test 5: Phonological Processing

Phonological Processing measures word activation, fluency of word access, and word reconstruction via phonological codes (Schrank et al., 2016). This test is based on a growing body of evidence that phonological codes are routes to word access (Leinenger, 2014) and are the initial and primary ways that a word accesses a semantic representation in memory (Lukatela & Turvey, 1994b). Test 5: Phonological Processing is highly correlated with psychometric *g* (McGrew et al., 2014), perhaps because phonological codes can activate, integrate, restructure, and/or sustain information in working memory (Baddeley, 1979; Baddeley, Eldridge, & Lewis, 1981; Klatt, 1979; Levy, 1978; McCusker et al., 1981; Slowiaczek & Clifton, 1980).

Test 5: Phonological Processing involves the reasoning and memory functions required to tap long-term phonological knowledge—an important link between primary (working) memory and long-term memory (Jones, Gobet, & Pine, 2007). Although Phonological Processing is primarily a measure of auditory processing and the narrow ability of phonetic coding (PC), its factor structure is complex: It includes aspects of comprehension–knowledge such as semantic memory and other narrow ability variance, specifically language development (LD), speed of lexical access (LA), and word fluency (FW). In addition, this test measures multiple cognitive operations and parameters of cognitive efficiency. Part A Word Access measures the depth of word access from phonemic cues; Part B Word Fluency measures the breadth and fluency of word activation from phonemic cues; Part C Substitution measures lexical substitution from phonemic cues in working memory.

Interventions for limitations in phonological processing depend on the impact of any delay or disability on learning. See Table 14.7. For example, well-developed phonological awareness is foundational in learning to read because phonology is mapped onto orthography when words are sounded out (Liberman et al., 1989; Wagner et al., 1993, 1994).

Test 12: Nonword Repetition

Nonword Repetition is a cognitively complex test that measures phonemic sensitivity and phonological short-term working memory capacity (Schrank et al., 2016). Although the test's primary psychometric factor loading is on short-term working memory, the constituent ability is interpreted as memory for sound patterns (UM), a narrow ability of auditory processing in contemporary CHC theory (McGrew et al., 2014).

Nonword repetition tests have gained wide acceptance because the task demands closely match the phonological processes involved in learning new words (Coady & Evans, 2008; Gathercole, 2006; Gathercole, Hitch, Service, & Martin, 1997), and because many individuals who perform poorly on nonword repetition tasks often have difficulties learning the phonological form of language (Archibald & Gathercole, 2007) and learning new words (Edwards, Beckman, & Munson, 2004; Gathercole, 2006; Michas & Henry, 1994). In addition, nonword repetition tests may be time-efficient and reliable tools to identify individuals with or at risk for language impairments (Bishop, North, & Donlan, 1996; Coady & Evans, 2008; Conti-Ramsden, 2003; Conti-Ramsden, Botting, & Farragher, 2001; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Gray, 2003; Horohov & Oetting, 2004; Taylor, Lean, & Schwartz, 1989). Interventions for limited proficiency in nonword repetition focus on phonologically based language development activities. See Table 14.7.

Long-Term Storage and Retrieval (Glr)

The CHC ability of long-term storage and retrieval (Glr) involves the cognitive processes of storing and retrieving information. *Storage* refers to the process by which semantic memories (Tulving, 1972, 1985) are created. Whether it is through the development and consolidation of mental representations from orally imparted discourse in Test 6: Story Recall, or through the association of words with rebus representations in Test 13: Visual–Auditory Learning, each test in the WJ IV COG Glr cluster is a standardized experiment for assessment of memory consolidation. Retrieval is a function of whether consolidation (encoding) of meaning-based representations has occurred. Consequently, the WJ IV COG Glr cluster measures consolidation (encoding) of semantic (meaning-based) representations into secondary memory (Schrank et al., 2016).

Test 6: Story Recall

Test 6: Story Recall is a cognitively complex task that requires listening ability, as well as background knowledge for the words, objects, or situations that are described in the stimulus phase; it is also influenced by working memory capacity.

Attention to orally imparted details supports the *formation of mental representations* (storage) during the stimulus phase; the response phase requires *reconstruction of the story details* (retrieval) through meaningful memory. Because the Story Recall task is so complex, interventions may be related to the development of listening ability (LS), background knowledge (K0), and/or meaningful memory (MM). Also, examiners should consider whether the ability to retain the story elements might be influenced by the individual's working memory capacity (WM), which places limits on the volume of information that can be reconstructed into a coherent and connected representation of the objects, events, or situations in the story-retelling phase (van den Broek, 1989). See Table 14.7.

Test 13: Visual–Auditory Learning

Test 13 measures visual–auditory paired-associate encoding in the learning phase; identification and word retrieval in the response phase (Schrank et al., 2016). The initial task requires associating a visual rebus symbol with a verbal label. The controlled-learning format of this test employs *directed-spotlight attention* (Gazzaniga et al., 1998)—the mental, attention-focusing process that prepares the examinee to encode the stimulus (Brefczynski & DeYoe, 1999; Klingberg, 2009; Sengpiel & Hubener, 1999). The retrieval phase requires the examinee to match a rebus presentation with its stored representation; this process is called *identification*. The directed-spotlight attention mechanism provides a cue to an intervention known as *active learning* (Marzano, Pickering, & Pollock, 2001). Active learning is required for the creation of meaning-based codes that are subsequently used to relate new information or task requirements to previously acquired knowledge. See Table 14.7.

Visual Processing (Gv)

The WJ IV COG Visual Processing (Gv) cluster measures visual–spatial analysis, formation of internal visual images, mental transformation of images in working memory, and passive storage for subsequent recognition of images (Schrank et al., 2016). Visual–spatial abilities are important in academic areas where reasoning with figures, patterns, and shapes is essential (Lubinski, 2010). Test 7: Visualization and Test 14: Picture Recognition are the two tests that create the Gv cluster.

Test 7: Visualization

Visualization measures size and shape perception, part-to-whole analysis, and mentally transforming two- and three-dimensional images (Schrank et al., 2016). Visualization ability may be related to the ability to construct mental representations, which is of fundamental importance across many cognitive domains and for the development of academic skills, including number sense (Gunderson, Ramirez, Beilock, & Levine, 2012) and reading comprehension (De Koning & van der Schoot, 2013). Although there are many specific interventions for the development of visualization skills, integrating spatially challenging activities into the curriculum is an easy-to-implement suggestion for improving visualization skills (Federation of American Scientists, 2006; Foreman et al., 2004; Gee, 2003; McAuliffe, 2003; Newcombe, Uttal, & Sauter, 2013). See Table 14.7.

Test 14: Picture Recognition

Picture Recognition is a visual memory task that measures the recognition of previously presented visual stimuli from images or icons held in passive storage (Schrank et al., 2016). The passive storage of visual images is called the *visual cache* (Baddeley & Hitch, 1994), part of the working memory system. The visual cache is of limited capacity, typically three to four items (Luck & Vogel, 1997). Images can be retained in the visual cache for several seconds (Vogel, Woodman, & Luck, 2001). Individuals with limited performance on Picture Recognition may benefit from interventions designed to develop skills in attending to and discriminating visual features, matching, and recalling visual information. See Table 14.7.

Other Narrow Ability or Clinical Clusters

In addition to the Perceptual Speed (P) cluster previously discussed, some WJ IV COG tests can be combined (sometimes with tests from the WJ IV OL) to provide measures of CHC narrow cognitive abilities. These clusters include Quantitative Reasoning (RQ), Number Facility (N), Auditory Memory Span (MS), and Vocabulary (VL/LD). An additional cluster, Cognitive Efficiency, is a cognitively complex score that has interpretive value for some evaluation purposes.

Quantitative Reasoning

Composed of Test 2: Number Series and Test 15: Analysis–Synthesis, the Quantitative Reasoning cluster measures both quantitative and non-numerical algorithmic reasoning (Schrank et al., 2016).

Number Facility

Composed of Test 10: Numbers Reversed and Test 11: Number–Pattern Matching, the Number Facility cluster measures the efficiencies of visual-perceptual discrimination, temporary storage, and processing of numerical information in working memory (Schrank et al., 2016).

Auditory Memory Span

Composed of Test 18: Memory for Words and WJ IV OL Test 5: Sentence Repetition, the Auditory Memory Span cluster measures storage capacity for both unrelated words and connected discourse in primary memory (Schrank et al., 2016).

Vocabulary

Composed of Test 1: Oral Vocabulary and WJ IV OL Test 1: Picture Vocabulary, the Vocabulary cluster measures knowledge of object names, as well as knowledge of words and their meanings (Schrank et al., 2016).

Cognitive Efficiency

Composed of Test 4: Letter–Pattern Matching and Test 10: Numbers Reversed, the Cognitive Efficiency cluster measures efficiencies of orthographic visual-perceptual discrimination ability under timed conditions, as well as temporary storage and recoding of numeric information in primary memory (Schrank et al., 2016). The Cognitive Efficiency—Extended cluster includes the same two tests, and also Test 3: Verbal Attention and Test 11: Number–Pattern Matching; this cluster measures a combination of orthographic visual-perceptual discrimination ability under timed conditions, and cue-dependent search and recoding functions from temporary stores of verbal and numerical information (Schrank et al., 2016). The Cognitive Efficiency clusters are often used for descriptive comparison to clusters measuring other abilities or combinations of other abilities, such as the Gf-Gc Composite.

Intellectual Level and Scholastic Aptitude Clusters

As noted earlier, the WJ IV COG includes three score options to determine intellectual level: the BIA score, the GIA score, and the Gf-Gc Composite. In addition, differential Scholastic Aptitude clusters can be created from combinations of selected tests.

Brief Intellectual Ability

The BIA score is a short but highly reliable measure of intellectual ability. The BIA is an equally weighted score derived from Test 1: Oral Vocabulary, Test 2: Number Series, and Test 3: Verbal Attention. The BIA includes markers for three important CHC abilities for estimating general intellectual ability, or *g*—comprehension–knowledge (G*c*), fluid reasoning (G*f*), and short-term working memory (G*wm*).

General Intellectual Ability

The GIA score is a measure of psychometric *g* that is broadly derived from the first seven tests in the WJ IV COG Standard Battery. Each of the seven tests was selected for inclusion in the GIA score because it is a highly representative, single measure of one of the seven primary broad CHC intellectual abilities; has high loadings on general intelligence (*g*); is relatively high in cognitive complexity; and is a strong predictor of achievement in the WJ IV ACH (McGrew et al., 2014). The GIA score provides assessment professionals with the best predictor score in the WJ IV COG—*across individuals*—of overall school achievement and other life outcomes that have some relationship to general intelligence. However, the GIA score may not be the best predictor score for any given individual, particularly if it is characterized by remarkable profile scatter and is attenuated by discrepantly low performance on one or more of the component tests (Schrank et al., 2015a).

Gf-Gc Composite

The Gf-Gc Composite is a measure of intellectual level that is derived from two tests of fluid reasoning and two tests of comprehension–knowledge—the two highest-order (*g*-loaded or *g*-saturated) factors included in the CHC theory of cognitive abilities (McGrew, 2005, 2009; McGrew et al., 2014; Schneider & McGrew, 2012 and Chapter 3, this volume). The composite can be operationally defined as a combined measure of verbal, inductive reasoning; quantitative, deductive reasoning; knowledge of words; and general object knowledge. The G*f*-Gc Composite is a measure of intellectual level or development that provides an alternative to the WJ IV GIA or any full-scale intelligence score when an individual has a significant limitation in one or more of the basic cognitive processes, memory storage and retrieval functions, or mechanisms of cognitive efficiency that are included in calculation of the GIA or a full-scale intelligence score. The Gf-Gc Composite represents the level of intelligence the individual has developed, regardless of any limitations in Gwm, Ga, Gv, Glr, Gs, or any of their associated narrow abilities, such as perceptual speed, working memory capacity, or phonological processing (Schrank et al., 2015a).

Scholastic Aptitude Clusters

The WJ IV COG includes a number of Scholastic Aptitude clusters. Each Scholastic Aptitude cluster is a small set of WJ IV cognitive tests that is statistically associated with performance on different WJ IV achievement clusters. Figure 14.2 contains the Selective Testing Table for the Scholastic Aptitude clusters and the associated target tasks for scholastic aptitude–achievement comparisons. These clusters are primarily intended to be used for short, academic-domain-focused assessments (McGrew, 2012; McGrew & Wendling, 2010), to determine if an individual is performing academically as well as would be expected based on a small set of closely associated cognitive abilities; these clusters are not intended for use in determining the presence of a specific learning disability. Students with a specific learning disability may not exhibit a scholastic aptitude–achievement discrepancy because a low achievement score may be reflected in low cognitive test scores that are closely related to the achievement area.

			Target Tasks for Scholastic Aptitude/ Achievement Comparisons													
			Reading						**Mathematics**				**Writing**			
			Reading	Broad Reading	Basic Reading Skills	Reading Comprehension	Reading Fluency	Reading Rate	Mathematics	Broad Mathematics	Math Calculation Skills	Math Problem Solving	Written Language	Broad Written Language	Basic Writing Skills	Written Expression
Standard Battery	COG 1	Oral Vocabulary	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	COG 2	Number Series							■	■	■					
	COG 3	Verbal Attention			■										■	
	COG 4	Letter-Pattern Matching														
	COG 5	Phonological Processing	■	■	■	■	■	■					■	■	■	■
	COG 6	Story Recall											■	■		■
	COG 7	Visualization							■	■	■	■				
	COG 8	General Information														
	COG 9	Concept Formation	■	■		■	■	■								
	COG 10	Numbers Reversed										■				
Extended Battery	COG 11	Number-Pattern Matching	■	■	■	■	■	■					■	■	■	■
	COG 12	Nonword Repetition														
	COG 13	Visual-Auditory Learning														
	COG 14	Picture Recognition														
	COG 15	Analysis-Synthesis										■				
	COG 16	Object-Number Sequencing														
	COG 17	Pair Cancellation							■	■	■					
	COG 18	Memory for Words														

FIGURE 14.2. Selective Testing Table for the WJ IV COG Scholastic Aptitude clusters and the associated target tasks for scholastic aptitude–achievement comparisons. From *Woodcock–Johnson IV™* (*WJ IV™*). Copyright © The Riverside Publishing Company. All rights reserved. Used by permission of the publisher.

Score Comparison Procedures

The WJ IV Online Scoring and Reporting System (Schrank & Dailey, 2014) includes a number of procedures that are useful for making comparisons between and among scores within the WJ IV COG, as well as comparisons that include scores from the WJ IV OL and WJ IV ACH.

Intracognitive Variation Procedure

The first procedure compares performance among tests and clusters to determine the presence and severity of any within-individual profile variations. Based on a selected standard error of the estimate (*SEE*) cutoff score, a profile-discrepant test or cluster can be defined as a relative (within individual) strength or weakness. Tests 1–7 must be administered, at a minimum, to evoke this procedure.

Gf-Gc Composite–Other Ability Comparison Procedure

The second procedure compares the Gf-Gc Composite score to other WJ IV COG, OL, or ACH cluster scores that primarily measure other (non-Gf and non-Gc) abilities. Although calculated via the traditional ability–achievement discrepancy model (with attendant cutoff score criteria), a target cluster score that is significantly lower than the Gf-Gc Composite is also interpreted as an intraindividual weakness, and a target cluster score that is significantly higher than the Gf-Gc Composite is interpreted as an intraindividual strength. The Gf-Gc Composite–Other Ability comparison procedure is appropriate for analyzing a pattern of strengths and weaknesses, as well as an ability–achievement discrepancy. This procedure is particularly useful when an individual's Gf and Gc abilities are relatively intact, compared to other cognitive abilities that may be a source of a learning disability (Schrank et al., 2016). Documentation of a relative weakness in an area of cognitive processing—relative to the Gf-Gc Composite of intellectual level—can be an important component of a specific learning disability determination (Schrank et al., 2015a).

GIA–Achievement Discrepancy Procedure

The third procedure compares the GIA score to selected WJ IV OL and WJ IV ACH cluster scores, to determine the presence and significance of any discrepancies between current levels of general intellectual ability and achievement or oral language ability. This procedure is based on the traditional intellectual ability–achievement discrepancy model, but uses actual discrepancy norms (McGrew et al., 2014), which reduce the error inherent in comparisons made between tests derived from different normative samples.

Scholastic Aptitude–Achievement Discrepancy Procedure

The fourth procedure compares a predicted score that is based on a small set of cognitive tests to selected WJ IV ACH cluster scores, to determine the presence and significance of any discrepancies between current levels of achievement and a set of cognitive tests that is highly predictive of the academic area of interest. This procedure is sometimes useful for short, selective, focused assessment purposes, where the comparison to achievement is used descriptively (not diagnostically). Alternatively, as part of a comprehensive evaluation, a scholastic aptitude–achievement comparison can provide a highly knowledgeable professional with additional descriptive information that may be useful in evaluating a clinical hypothesis of any concordance or discordance between clusters of highly predictive cognitive tests and expected academic performance in a particular domain. However, it is important to note that the term *aptitude* may suggest that these clusters measure something they do not. Although these clusters can be useful for comparing an individual's current academic performance levels to his or her levels of domain-associated cognitive abilities for inferential analysis, the Scholastic Aptitude clusters are not intended to predict an individual's potential for scholastic success (Schrank et al., 2016).

WJ IV TESTS OF ORAL LANGUAGE

The WJ IV OL (Schrank, Mather, & McGrew, 2014b) is a set of tests that may be administered independently or in conjunction with the WJ IV COG or the WJ IV ACH.

Organization of the WJ IV OL

The WJ IV OL has 12 tests measuring varied aspects of receptive and expressive oral language. There are nine English tests (Tests 1–9) and three Spanish tests (Tests 10–12). The three Spanish tests are adaptations of Test 1: Picture Vocabulary, Test 2: Oral Comprehension, and Test 6: Understanding Directions. All 12 tests are contained within one easel test book. Figure 14.3 presents the Selective Testing Table of the WJ IV OL. Table 14.8 provides an overview of the primary broad and narrow CHC abilities measured, as well as task demands for each of the oral language tests.

How to Administer the WJ IV OL Tests

The WJ IV OL tests may be administered in any order. However, it is generally recommended that the core tests, Tests 1–4, be administered first. Not all tests need to be administered, as test choice is guided by referral concerns and selective testing principles (see Figure 14.3).

Administration Time

As a general rule, experienced examiners require about 40 minutes to administer the first eight tests. This varies depending on the age and ability level of the examinee. Test 9: Sound Awareness is a screening measure for younger students or those who have poor phonological awareness so it is not administered in every case. About 15–20 minutes should be allowed for administration of the Spanish tests.

Testing Materials

The testing materials required for administration of the WJ IV OL include the test book, a test record, the audio CD, appropriate audio equipment, pencils, and a stopwatch.

			Oral Language Clusters									OL + COG	
			Oral Language	Broad Oral Language	Oral Expression	Listening Comprehension	Phonetic Coding	Speed of Lexical Access	Lenguaje oral	Amplio lenguaje oral	Comprensión auditiva	Vocabulary (VL/LD)	Auditory Memory Span (MS)
Oral Language Battery	OL 1	Picture Vocabulary	■	■	■							■	
	OL 2	Oral Comprehension	■	■		■							
	OL 3	Segmentation					■						
	OL 4	Rapid Picture Naming						■					
	OL 5	Sentence Repetition			■								■
	OL 6	Understanding Directions		■		■							
	OL 7	Sound Blending					■						
	OL 8	Retrieval Fluency						■					
	OL 9	Sound Awareness[1]											
	OL 10	Vocabulario sobre dibujos							■	■			
	OL 11	Comprensión oral							■	■	■		
	OL 12	Comprensión de indicaciones								■	■		
Cognitive Abilities Battery	COG 1	Oral Vocabulary										■	
	COG 18	Memory for Words											■

■ Tests required to create the cluster listed.

[1] This is a screening test and does not contribute to a cluster.

FIGURE 14.3. Selective Testing Table (tests and clusters) for the Woodcock–Johnson IV Tests of Oral Language (WJ IV OL). From *Woodcock–Johnson IV*™ (*WJ IV*™). Copyright © The Riverside Publishing Company. All rights reserved. Used by permission of the publisher.

TABLE 14.8. Broad and Narrow CHC Abilities Measured and Task Demands of WJ IV OL Tests

Test name	Primary broad CHC ability *Narrow ability*	Task demands
Test 1: Picture Vocabulary; Test 10: Vocabulario sobre dibujos	Comprehension–knowledge (Gc) *Lexical knowledge* (VL) *Language development* (LD)	Requires naming familiar to less familiar pictured objects in English (or Spanish).
Test 2: Oral Comprehension; Test 11: Comprensión oral	Comprehension–knowledge (Gc) *Listening ability* (LS)	Requires listening to a short passage in English (or Spanish) and supplying a key missing word.
Test 3: Segmentation	Auditory processing (Ga) *Phonetic coding* (PC)	Requires listening to a word and breaking it into its parts (compound words, syllables, phonemes).
Test 4: Rapid Picture Naming	Long-term storage and retrieval (Glr) *Naming facility* (NA) *Speed of lexical access* (LA)	Requires naming simple pictures quickly.
Test 5: Sentence Repetition	Short-term working memory (Gwm) *Memory span* (MS) Comprehension–knowledge (Gc) *Listening ability* (LS)	Requires listening to a word, phrase, or sentence and repeating it verbatim.
Test 6: Understanding Directions; Test 12: Comprensión de indicaciones	Short-term working memory (Gwm) *Working memory capacity* (WM) Comprehension–knowledge (Gc) *Listening ability* (LS)	Requires listening to a sequence of directions in English (or Spanish), and then following those directions by pointing to objects in a picture.
Test 7: Sound Blending	Auditory processing (Ga) *Phonetic coding* (PC)	Requires listening to a series of syllables or phonemes and then blending the sounds into a word.
Test 8: Retrieval Fluency	Long-term storage and retrieval (Glr) *Speed of lexical access* (LA) *Ideational fluency* (FI)	Requires naming as many items as possible in a given category within a 1-minute time limit.
Test 9: Sound Awareness	Auditory processing (Ga) *Phonetic coding* (PC)	Requires providing a rhyming word in 9A and deleting a word part or sound to form a new word in 9B.

Summary of Key Administration and Scoring Points

Most tests use suggested starting points and basal and ceiling rules, and responses are scored 1 or 0. Table 14.9 summarizes key administration and scoring points for each test.

Special Administration and Scoring Considerations

Although the WJ IV OL examiner's manual (Mather & Wendling, 2014c) and test book provide detailed rules for test-by-test administration, this section presents important reminders about tests that have special administration or scoring rules.

Spanish Tests

If the Spanish-language tests are administered, the examiner must be bilingual in Spanish and English, or a primary–ancillary examiner procedure must be employed. This procedure is described in the WJ IV OL examiner's manual. When evaluating an English-language learner or when using the Spanish tests in the WJ IV OL, an examiner may wish to complete the Language Exposure and Use Questionnaire on the last page of the test record, which includes a number of questions about the history of the examinee's language use.

Timed Tests

The WJ IV OL contains two timed tests: Rapid Picture Naming and Retrieval Fluency. The time

TABLE 14.9. Summary of WJ IV OL Test Administration and Scoring Rules

Test name	Item scoring rule	Sample items?	Basal/ceiling rules?[a]	Administration notes	Extra materials required for administration
Test 1: Picture Vocabulary	1, 0	Yes (PreK)	Yes; 6/6	Suggested starting points are available.	
Test 2: Oral Comprehension	1, 0	Yes	Yes; 6/6	Suggested starting points are available after sample items.	Audio
Test 3: Segmentation	1, 0	Yes	Yes; 5/5	All begin with either Introduction 1 or 2.	
Test 4: Rapid Picture Naming	1, 0	Yes	No; 2-minute time limit	All begin with sample items, then proceed to item 1.	Stopwatch (timed test)
Test 5: Sentence Repetition	1, 0	Yes	Yes; 4/4	Suggested starting points are available; then begin with sample item for pre-grade 3 starting point.	Audio
Test 6: Understanding Directions	1, 0	No	No; ceiling determined by continuation instructions	Suggested starting points are available; scoring is based on administration of specific group of pictures.	Audio
Test 7: Sound Blending	1, 0	Yes	Yes; 6/6	Suggested starting points are available after sample items.	Audio
Test 8: Retrieval Fluency	1 for each correct response	No	No; 1-minute time limit per item	All examinees are administered all three items.	Stopwatch (timed test)
Test 9: Sound Awareness				Both subtests must be administered.	
9A: Rhyming	1, 0	Yes	Yes; 6/6	Suggested starting points are available; all begin with either Introduction 1 or 2.	
9B: Deletion	1, 0	Yes	Item 1 is basal; ceiling is 6	All begin with sample items, then proceed to item 1.	Audio
Test 10: Vocabulario sobre dibujos	1, 0	Yes (Pre)	Yes; 6/6	Suggested starting points are available; examiner must be fluent in Spanish; parallel to Test 1: Picture Vocabulary.	
Test 11: Comprensión oral	1, 0	Yes	Yes; 6/6	Suggested starting points are available after sample items; examiner must be fluent in Spanish; parallel to Test 2: Oral Comprehension.	Audio
Test 12: Comprensión de indicaciones	1, 0	No	No; ceiling determined by continuation instructions	Suggested starting points are available; scoring is based on administration of specific group of pictures; examiner must be fluent in Spanish; parallel to Test 6: Understanding Directions.	Audio

[a] May be modified by complete-page rule.

limits are noted on the test page, as well as on the test record. Administration of these tests requires a stopwatch or a watch with a second hand. When administering Rapid Picture Naming, the examiner must be prepared to turn each page as soon as the examinee names the last picture on the page.

For Retrieval Fluency, duplicate responses should not be accepted, although variations are acceptable (e.g., *Bob, Bobby*). The raw score for Retrieval Fluency is based on the total number of tally marks recorded for each of the three items. The examiner records a tally mark for each correct response, grouping them by fives. For each item, the correct responses are counted, and the total is recorded in the Number Correct box. The maximum number that can be entered for each item is 99. To obtain estimated age and grade equivalents, the examiner adds the three item totals together and locates that sum in the scoring table.

Audio Tests

The standardized audio recording should be used to present Oral Comprehension, Sentence Repetition, Understanding Directions, Sound Blending, Sound Awareness (Deletion), Comprensión oral, and Comprensión de indicaciones. The examiner should not repeat or replay any items unless directions permit it (e.g., on sample items).

Test 3: Segmentation

Examiners must say the sound a letter makes when it is presented between slash marks (e.g., /s/). On items 11–20, a response is correct if the correct number of segments is provided even if pronunciation is not perfect. On items 21–37, all sounds must be given in order to score the response as correct. This test can be difficult to score correctly for some examiners who have trouble hearing sounds in words. In these cases, the examiner may wish to have a speech–language professional or reading specialist administer the test.

Test 6: Understanding Directions and Test 12: Comprensión de indicaciones

Examiners should familiarize themselves with each picture in order to follow examinees' pointing responses. Examinees must be given 10 seconds to study each picture before any item is administered. The score on Understanding Directions is based on the number of correct responses the examinee has on the set of pictures administered. Each correct response is scored 1, and each incorrect response is scored 0. On the test record, an examiner writes the number of points earned for each picture in the space provided. When indicated on the test record, the examiner records the cumulative total for the two pictures specified. In the score entry section on the test record, the examiner enters the number of points for each picture or set of pictures administered, or an X if the set was not administered. To obtain estimated age and grade equivalents, the examiner locates the number of points in the appropriate column corresponding to the group of pictures administered. If more than one group of pictures has been administered, the last group administered following the continuation instructions is used to estimate age and grade equivalents.

Reliability and Validity

Following is a summary of reliability and validity information for the WJ IV OL. More complete information is available in the WJ IV technical manual (McGrew et al., 2014) and the discussion of CHC theory in the introductory section of this chapter.

Median reliability coefficients for the WJ IV OL tests are reported in Table 14.10. With the exception of the timed tests, coefficients of stability are reported as median internal-consistency (r_{11}) and median *SEM* values. For the two timed tests, test–retest values (r_{12}) are used as the most appropriate coefficient of stability. Mosier's (1943) formula was used to calculate reliabilities for tests with subtests, such as Test 9: Sound Awareness. Table 14.11 contains the median cluster reliability coefficients (r_{cc}) and median *SEM* (SS), as calculated via Mosier's formula. All of the WJ IV OL clusters, with the exception of Oral Expression (.89) and Speed of Lexical Access (.89), exceed the commonly referenced professional standard of .90 for decision-making purposes. Although the test reliabilities support use and interpretation of the test-level scores as single indicators of an ability, the WJ IV OL clusters are preferred for most decision-making purposes, due to their higher reliabilities.

Table 14.12 contains a summary of correlations for the primary WJ IV OL clusters with other measures of oral language abilities, including the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003), the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007), the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999), and the

TABLE 14.10. WJ IV OL Median Test Reliability Statistics

Test	Median r_{11}	Median *SEM* (SS)	Median r_{12}
Test 1: Picture Vocabulary	.88	5.47	—
Test 2: Oral Comprehension	.82	7.06	—
Test 3: Segmentation	.94	4.14	—
Test 4: Rapid Picture Naming	—	—	.90
Test 5: Sentence Repetition	.83	6.48	—
Test 6: Understanding Directions	.87	6.42	—
Test 7: Sound Blending	.89	5.99	—
Test 8: Retrieval Fluency	.80	6.00	—
Test 9: Sound Awareness	.82	9.49	—

Note. r_{11}, internal-consistency reliability; r_{12}, test–retest reliability for speeded tests.

Oral and Written Language Scales: Listening/Comprehension/Oral Expression (OWLS; Carrow-Woolfolk, 1995). The significant correlations of the WJ IV Oral Language, Oral Expression, and Listening Comprehension clusters with the CELF-4 composites, the total PPVT-4 score, the CASL Core Composite, and the OWLS composites support the validity of the WJ IV OL clusters as valid measures of general oral language abilities. The lower correlations for the WJ IV OL Speed of Lexical Access cluster indicate that it is measuring unique abilities not measured in the CELF-4, PPVT-4, CASL, or OWLS.

Interpretation

Critically important for success in life and academics, oral language is the foundation for and the primary means of communicating and learning. Even thinking and metacognition rely on language. Key aspects of oral language are measured in the WJ IV OL, which aids in early identification of weaknesses in these building blocks of learning and helps explore the role of oral language in an individual's achievement performance.

TABLE 14.11. WJ IV OL Median Cluster Reliability Statistics

Cluster	r_{cc}	Median *SEM* (SS)
Oral Language	.90	4.74
Broad Oral Language	.92	4.50
Oral Expression	.89	5.20
Listening Comprehension	.90	5.61
Phonetic Coding	.95	3.97
Speed of Lexical Access	.89	4.97
Vocabulary[a]	.93	4.97

[a] Requires WJ IV COG Test 1: Oral Vocabulary.

An individual's performance on various linguistic and cognitive abilities can be evaluated by using the WJ IV OL tests and clusters, including two of the Individuals with Disabilities Education Improvement Act (IDEA, 2004) eligibility areas for specific learning disabilities: oral expression and listening comprehension. Several different linguistic abilities, including phonological awareness, speed of lexical access, memory, vocabulary, and listening comprehension, are evaluated. In terms of CHC theory, the WJ IV OL tests measure aspects of comprehension–knowledge, short-term working memory, long-term storage and retrieval, cognitive processing speed, and auditory processing. Additionally, if the parallel English and Spanish clusters are administered, examiners can explore language dominance and proficiency (using the Comparative Language Index or CLI).

Interpretation requires knowledge of all derived scores and profiles, as well as the intra-ability variation and ability–achievement comparison procedures. While cluster-level information is preferred for decision making, the individual tests provide important insights into functioning and aid in program planning. Examples of evidence-based interventions are included for each test, and selected interventions are summarized in Table 14.13.

There are nine English oral language clusters available (Oral Language, Broad Oral Language, Phonetic Coding, Speed of Lexical Access, Listening Comprehension, and Oral Expression) and

TABLE 14.12. Correlations for Select WJ IV OL Measures and Other Measures of Oral Language Abilities

Other measures	Oral Language	Oral Expression	Listening Comprehension	Speed of Lexical Access
Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4)[a]				
Core Language	.63	.74	.64	.31
Expressive Language	.64	.72	.64	.32
Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)[a]	.74	.70	.69	.43
Comprehensive Assessment of Spoken Language (CASL) (Core Composite)[b]	.85	.72	.76	.57
Oral and Written Language Scales: Listening/Comprehension/Oral Expression (OWLS)[b]				
Oral Composite	.68	.62	.64	.41
Oral Expression	.67	.60	.59	.45
Listening Comprehension	.53	.50	.56	.25

Note. Samples used in studies varied by age as noted.

[a] Ages 5–8.

[b] Ages 7–17.

three Spanish clusters (Lenguaje oral, Amplio lenguaje oral, and Comprensión auditiva). Two additional clusters are available when tests from the WJ IV COG are also used (Vocabulary and Auditory Memory Span).

Oral Language

Two clusters, Oral Language and Broad Oral Language, provide estimates of the individual's verbal ability. The first cluster is composed of two tests, Test 1: Picture Vocabulary and Test 2: Oral Comprehension. The Broad cluster is composed of three tests—the two in Oral Language plus Test 6: Understanding Directions. The two parallel Spanish clusters are Lenguaje oral and Amplio lenguaje oral. Broad Oral Language or Amplio lenguaje oral can be used as the predictor in the oral language–achievement comparison procedure.

Test 1: Picture Vocabulary (Test 10: Vocabulario sobre dibujos)

Picture Vocabulary measures vocabulary, verbal ability, and cultural knowledge, all aspects of comprehension–knowledge. It is primarily a single-word expressive language task.

An individual may have difficulty with this test due to limited knowledge of word meanings, word retrieval difficulties, English as a second language, limited experiences and opportunities, or cultural differences. Observing performance and analyzing errors will help determine whether poor performance is a result of limited vocabulary or retrieval problems. When an error is related to the correct response (e.g., describes an attribute or function), it may indicate a retrieval or word-finding problem. An error that is not directly associated with the correct response may indicate a weakness in vocabulary knowledge. Although an error that is associated with the word meaning is scored as incorrect, it does suggest some knowledge or understanding is present. Sometimes an individual may know the function of an object, but may not be able to come up with the exact name.

One intervention for building vocabulary in younger children is reading aloud to them (Adams, 1990). Effectiveness is increased when an interactive read-aloud method such as *dialogic reading,* which requires the reader to actively engage the child with the text, is used. For older children and adolescents, possible interventions for building vocabulary include reading for different purposes (National Reading Panel, 2000) and intentional explicit word instruction (teaching synonyms, antonyms, multiple-meaning words) (Graves, Juel, & Graves, 2004; McKeown & Beck, 2004; National Reading Panel, 2000). See Table 14.13 for additional intervention suggestions.

Test 2: Oral Comprehension (Test 11: Comprensión oral)

Oral Comprehension measures listening ability and language development, both aspects of comprehension–knowledge. The individual must use previously acquired knowledge, syntax, and context clues to identify the missing word. Oral Comprehension is primarily a receptive language task.

Low performance may result from limited semantic or syntactic knowledge, limited exposure to English, or poor attention. Similar in format to the WJ IV ACH Passage Comprehension test, this oral test requires the examinee to listen to a passage instead of read a passage. The results of these two tests may be compared to help determine whether limited oral language or limited decoding skill are impairing reading comprehension performance. Several of the same abilities, such as vocabulary, reasoning, and background knowledge, are involved with both reading and listening comprehension. Individuals who have oral language impairments or intellectual disabilities may obtain similar scores on both measures of listening and reading comprehension. Individuals with reading disabilities will frequently score higher on measures of listening comprehension than on measures of reading comprehension. However, when an individual scores higher on Passage Comprehension than on Oral Comprehension, it suggests that comprehension improves when reading. The permanence of the text facilitates comprehension because it reduces the effects of memory and gives the individual greater control over the rate of input. Consideration of

TABLE 14.13. Examples of Interventions, Strategies, and Accommodations Related to Limitations in Performance on WJ IV OL Tests

Test	Brief intervention, strategy, or accommodation
Test 1: Picture Vocabulary; Test 10: Vocabulario sobre dibujos	Creating a vocabulary-rich environment; text talks; intentional, explicit word instruction; preteaching vocabulary; use of semantic maps.
Test 2: Oral Comprehension; Test 11: Comprensión oral	Rehearsal and oral elaboration; use of echo activities; reducing amount of material to be remembered at any one time; simplifying linguistic structures of verbal material.
Test 3: Segmentation	Explicit, systematic instruction in phonics; teaching the six syllable types; practice in segmenting words for spelling; use of letter tiles to teach segmentation skills.
Test 4: Rapid Picture Naming	Increasing fluency through speed drills and monitoring progress.
Test 5: Sentence Repetition	Playing listening and repeating games; teaching chunking strategies; use of mnemonics; asking student to paraphrase directions to ensure understanding; providing visual cues such as outlines of key points.
Test 6: Understanding Directions; Test 12: Comprensión de indicaciones	Playing barrier games (taking turns giving and following directions); presenting tasks in separate and independent steps; repeating important information.
Test 7: Sound Blending	Providing early exposure to language sounds; prompting phonological awareness; direct instruction in sound blending; use of manipulatives to teach blending skills.
Test 8: Retrieval Fluency	Oral elaboration; use of cues; allowing more time for thinking and responding.
Test 9: Sound Awareness	Use of manipulatives to teach adding, deleting, substituting, and rearranging sounds; practicing rhyming skills (e.g., discriminating words that rhyme or don't rhyme; producing a rhyming word).

the discrepancies between listening and reading comprehension can help determine the presence of a specific reading disability or a more generalized language problem.

Use of directed vocabulary thinking activities (Graves, 2000) helps develop listening comprehension skills. School-age children who have difficulties following a teacher's oral discourse in the classroom may benefit from an outline of key points (Wallach & Butler, 1994) on the board or overhead projector prior to the beginning of each instructional unit. See Table 14.13.

Test 6: Understanding Directions (Test 12: Comprensión de indicaciones)

Understanding Directions measures listening ability and language development, both aspects of comprehension–knowledge. Understanding Directions also measures memory span and working memory, two aspects of short-term working memory. Some of the items are simple memory span tasks (e.g., "Point to the cat"); others engage working memory because they involve rearranging or reordering the sequence (e.g., "Before you point to the cat, point to the tree and then the flower"). An individual may have low scores on this test for several reasons: poor attention, limited receptive vocabulary knowledge, or weaknesses in listening comprehension, attention, or memory.

Interventions related to limited proficiency on Understanding Directions include opportunities to practice listening and following directions (Galda & Cullinan, 1991; Leung & Pikulski, 1990), and echo activities (Clay, 1991). Accommodations include modifying the listening environment (Hardiman, 2003).

Phonetic Coding

The Phonetic Coding cluster has two tests, Test 3: Segmentation and Test 7: Sound Blending. Both measure aspects of *phonetic coding,* a narrow ability of auditory processing. Phonetic coding is the ability to hear phonemes, to blend sounds into words, and to segment words into parts or phonemes. Test 3: Segmentation involves the ability to break apart the sounds in words which is a fundamental skill underlying spelling. Test 7: Sound Blending involves pushing together sounds to form words which underlies the application of phonics.

Phonological awareness provides the foundation for learning to apply phonics knowledge to reading and spelling. The two most important phonological awareness skills are blending and segmenting, both of which are measured in this cluster.

Test 3: Segmentation

Segmentation is a critical phonological ability that underlies aspects of spelling—that is, the ability to break apart the sounds within a word. Poor performance on the Segmentation test suggests that the individual may have difficulty recognizing the individual phonemes and then putting the sounds in the correct order when spelling words. An individual could have low scores on this test for several reasons: poor phonological awareness, English as a second language, articulation difficulties, weak memory, or inadequate instruction. Providing explicit code instruction (e.g., mapping phonemes to graphemes) facilitates development of the segmenting skills required for spelling (Berninger & Amtmann, 2003).

Test 7: Sound Blending

Sound blending is a key ability of phonemic processing, a narrow auditory processing ability. An individual could have low scores on the Sound Blending test for several reasons: poor phonological awareness, English as a second language, articulation difficulties, weak memory, or inadequate instruction. When an individual has a low score on Sound Blending, it is also likely that he or she may have a low score on the WJ IV ACH Word Attack test. In order to read phonically regular nonsense words, the individual has to push together the sounds, or blend the sounds, to form the word. Conversely, an individual with a high score on Sound Blending is likely to have good phonics skills, unless this type of instruction has been limited. With appropriate instruction, blending skills can be developed.

For school-age children and some adolescents with limited phonemic awareness, interventions include explicit, systematic instruction in phonics (National Reading Panel, 2000) and use of decodable texts for daily practice (Meyer & Felton, 1999). See Table 14.13.

Speed of Lexical Access

Speed of lexical access is an individual's ability to retrieve words rapidly from his or her lexicon by using orthographic, phonological, or semantic characteristics of words. Initially thought of as a narrow aspect of long-term storage and retrieval, Schneider and McGrew (Chapter 3, this volume) now suggest that the Speed of Lexical Access cluster is distinct from measures that involve the storage of information because it measures the retrieval function only.

This cluster consists of Test 4: Rapid Picture Naming and Test 8: Retrieval Fluency, both timed measures of word retrieval. A difference in performance on these two tests may occur because of the task demands and the underlying narrow abilities. Rapid Picture Naming is a confrontational naming task that requires the person to label specific pictures, a sequential naming facility task. Retrieval Fluency is a measure of ideational fluency and requires the person to produce as many words as possible belonging to a specific category.

A considerable body of research suggests that speed of lexical access, or how quickly individuals can name highly familiar visual stimuli, is a strong predictor of reading performance and a cognitive marker of developmental dyslexia (Georgiou & Parrila, 2013). This type of task appears most related to reading fluency. In remediating reading fluency, it is important to make sure that the reader is accurate on the text before focusing on building speed. Deficits in rapid automatized naming have also been found in individuals with speech and language disorders (Georgiou & Parrila, 2013; Windsor & Kohnert, 2008). If the problem is severe enough, an individual may be classified as having an expressive language impairment.

Test 4: Rapid Picture Naming

Rapid Picture Naming measures the narrow abilities of naming facility and speed of lexical access (i.e., the speed of producing names for objects). An individual could have low scores on this test for several reasons: poor attention, slow articulation speed, word-finding difficulties, or slow word retrieval speed. Observing the pace and accuracy of an individual's performance (e.g., slow and accurate, fast and accurate, slow and inaccurate, or fast and inaccurate), can help identify appropriate instruction and accommodations. See Table 14.13.

Test 8: Retrieval Fluency

The Retrieval Fluency test does not include the encoding and storage processes, but rather measures the rate or automaticity of word retrieval. In CHC theory, the cognitive abilities measured by Retrieval Fluency are ideational fluency and speed of lexical access. Martin (2009) described this type of retrieval task as *associative or encyclopedic knowledge,* which has three principal characteristics: Retrieval is explicit (e.g., specific names); there is no intrinsic limit on the amount of information that can be retrieved; and this type of knowledge is idiosyncratic (some people will produce many responses; others will not). An individual may have a low score on this test for several reasons: poor attention, limited vocabulary, weaknesses in word retrieval, or limited use of categorical strategies. Oral elaboration (Wolf, Bowers, & Biddle, 2000; Wolfe, 2001) may be an effective intervention to improve fluency of retrieval.

Oral Expression

The Oral Expression cluster is composed of two tests, Test 1: Picture Vocabulary and Test 5: Sentence Repetition. It measures expressive vocabulary and primary memory span in connected oral discourse. A significant difference between the two tests of this cluster may indicate a specific strength or weakness in vocabulary or memory span. (See the information given above in the section on the Oral Language cluster for Test 1: Picture Vocabulary.)

Test 5: Sentence Repetition

Sentence Repetition measures auditory memory span, a narrow ability of short-term working memory. It is the ability to encode and maintain verbal information in memory and then accurately reproduce it in sequence. Performance on this task can be aided by meaning and use of the sentence context. However, in some cases, the context and syntax of language may interfere with performance. As a result, performance may be better on a task that has less language context, such as repeating a list of unrelated words (see Auditory Memory Span). An individual could obtain low scores on this test for several reasons, such as limited attention, poor memory, or limited oral language.

An appropriate intervention linked to limited proficiency on Sentence Repetition is rehearsal of the information to be remembered (Medina, 2008; Squire & Schacter, 2003). Accommodations may be needed to compensate for limitations in short-term memory or working memory capacity, such as keeping oral directions short and simple, asking the student to paraphrase directions to ensure understanding, and providing visual cues for directions or steps to be followed (Gathercole & Alloway, 2008).

Listening Comprehension

The Listening Comprehension cluster is composed of two tests, Test 2: Oral Comprehension and Test 6: Understanding Directions. The cluster measures listening ability and verbal comprehension. Both of these tasks involve receptive language, as well as memory span and working memory. The parallel Spanish cluster is Comprensión auditiva. (See information for both Test 2 and Test 6 in the discussion of the Oral Language cluster, above.)

In the field of reading disabilities, listening comprehension is an important construct in an evaluationof dyslexia, based on the assumption thatan individual's reading comprehension should be comparable to his or her listening comprehension. When performance is better on listening comprehension than on reading comprehension, it suggests the presence of underdeveloped reading competence or a specific reading disability. When performance is low on listening comprehension, it suggests a more generalized language problem and accommodations may be needed such as providing shortened directions, writing directions on the board, or asking the individual to repeat or paraphrase directions.

Vocabulary

The Vocabulary cluster includes two expressive vocabulary measures, Test 1: Picture Vocabulary and (from the WJ IV COG) Test 1: Oral Vocabulary. In some cases, performance may vary between Picture Vocabulary and Oral Vocabulary, due to differences in task demands. Picture Vocabulary measures knowledge of object names using pictured objects, and Oral Vocabulary measures knowledge of synonyms and antonyms for vocabulary words. Vocabulary knowledge is highly related to reading comprehension and to school success in general (Anderson & Nagy, 1992). The importance of word knowledge increases as reading skill develops, progressing from learning to read to reading to learn.

Auditory Memory Span

Auditory memory span can be considered a narrow ability within the broader construct of working memory because the tasks require controlled, focal attention to retain a string of orally presented information, often the first requirement of a working memory task. Additionally, memory span capacity, which reflects the amount of information that can be actively maintained in immediate awareness, is required for the two tests in this cluster, Test 5: Sentence Repetition and (from the WJ IV COG) Test 18: Memory for Words. A difference may exist in a person's performance on these two tests. Sentence Repetition has more linguistic context than Memory for Words, which requires repeating a string of unrelated words. Comparing the results of these two tests can help identify whether the context of language helps or interferes with performance.

Screening Test

Test 9: Sound Awareness does not contribute to any cluster. Its primary use is as a screening measure for children in kindergarten through third grade. However, it can be useful in evaluating both beginning readers and older individuals who are experiencing reading difficulty because it does not require reading and measures two specific aspects of phonological awareness: rhyming and deletion. Difficulties with these types of tasks serve as early warning signs of risk for reading problems. Additional information regarding phonological awareness can be obtained by administering Test 7: Sound Blending and (from the WJ IV COG) Test 5: Phonological Processing, which includes the Word Access, Word Fluency, and Substitution subtests. Comparing the results of these tests to those of the WJ IV ACH Word Attack test may help determine whether problems are due to limitations in phonological awareness, phoneme/grapheme knowledge, or difficulties with both.

Score Comparison Procedures

There are several score comparison procedures available for the WJ IV OL. The intra-oral language variation procedure explores significant differences among a person's oral language abilities, and thus is helpful in determining a pattern of strengths and weaknesses in language functions. The oral language-achievement comparison procedure documents any significant discrepancies between an oral language predictor and achievement performance. The Comparative Language Index (CLI) is generated when the parallel Spanish and English tests have been administered so that a comparison can be made between English and Spanish language proficiency. In addition, Cognitive Academic Language Proficiency (CALP) levels for specific clusters can be selected for inclusion in the score report. All of these options are available when the Online Scoring and Reporting System is used.

Intra-Oral-Language Variation Procedure

The intra-oral-language variation procedure is a norm-based method for evaluating the presence of significant strengths or weaknesses among an individual's linguistic abilities. To calculate this variation procedure, a minimum of four tests (Tests 1–4) must be administered. The individual's performance on one test is then compared to his or her predicted performance, which is based on the average performance on the other three tests. For example, when a person's performance on Test 1: Picture Vocabulary is being evaluated, the average of the other three tests (Test 2: Oral Comprehension, Test 3: Segmentation, and Test 4: Rapid Picture Naming) is used to obtain the predicted score. The actual score is then compared to the predicted score. If the actual score is higher than predicted, a positive difference is obtained. If the actual score is lower than predicted, a negative difference is obtained. Depending on the magnitude of the positive (or negative) difference, the person may have a relative strength (or weakness) in a specific area. If, for example, Picture Vocabulary is significantly higher than predicted, the person exhibits a relative strength in vocabulary knowledge. If the person's actual Picture Vocabulary score is significantly lower than predicted, he or she exhibits a relative weakness in vocabulary knowledge. This type of information is valuable in documenting a pattern of strengths and weaknesses or determining a specific area of weakness rather than generalized low performance. Additionally, Tests 5–8 can be included in the variation procedure, along with any cluster that results from the tests administered. Three tests from the WJ IV COG (Test 1: Oral Vocabulary, Test 5: Phonological Processing, and Test 12: Nonword Repetition) and the Vocabulary and Auditory Processing clusters are also entered into the intra-oral-language variation procedure when administered. Whenever additional tests or clusters are included in the variation procedure, it is called the Extended variation.

Oral Language Ability–Achievement Comparison Procedure

The WJ IV OL provides an option to use oral language as a predictor score in the ability/achievement comparison calculation. The Broad Oral Language cluster may be used to predict levels of achievement from the individual's level of oral language development.

If testing is completed in Spanish, the Amplio lenguaje oral cluster may be used as the predictor. This allows the individual's oral language ability in Spanish to predict his or her ability to perform academically in English. For Spanish-dominant bilingual students, this may be an important comparison to make because oral language ability in Spanish can suggest the level of academic performance a student may be able to attain, in English, if provided with intensive English language instruction and appropriate Spanish-language supports (Gersten et al., 2007; Lindholm & Aclan, 1991; Short & Fitzsimmons, 2007).

Within this comparison procedure, the standard score from the Broad Oral Language cluster, composed of three tests, may be used to predict achievement on any of the achievement cluster scores. Individuals with a significant negative discrepancy (e.g., ≥ -1.50) between oral language ability and achievement exhibit relative strengths in oral language with weaknesses in one or more achievement areas (for this example, a discrepancy of this size will occur about 6 out of 100 times).

As noted by Stanovich (1991a, 1991b), use of an oral language measure to predict reading and writing is often preferable to use of a general intelligence score because it is more in line with the concepts of so-called "potential" and "unexpected" failure. He further explains that using oral language ability as the aptitude measure moves us closer to a more principled definition of reading disability because it provides a more accurate estimate of what the person could achieve if the reading problem were entirely resolved. Essentially, what distinguishes the individual with a reading disability from other poor readers is that listening comprehension ability is higher than ability to decode words (Rack, Snowling, & Olson, 1992), and thus the difficulty is "unpredicted." However, it is important to note that an individual with a learning disability may or may not exhibit an oral language ability–achievement discrepancy. For example, an older student with reading difficulties may have depressed performance in oral language because of his or her limited experiences with text. This lack of exposure to print contributes to reduced knowledge and vocabulary.

Comparative Language Index

The CLI is a unique comparison procedure available when any of the three parallel English and

Spanish clusters have been administered. The CLI is available when the following clusters are compared: Oral Language to Lenguaje oral, Broad Oral Language to Amplio lenguaje oral, and Listening Comprehension to Comprensión auditiva.

The CLI is presented as a ratio of the individual's RPI numerators on the two clusters being compared, one in Spanish and one in English. The CLI is helpful in determining language dominance and proficiency. For example, when an individual is administered Tests 1, 2, and 6 in English and Tests 10–12 in Spanish, the Broad Oral Language and Amplio lenguaje oral clusters are obtained. Here is an example: Antonio, a fifth-grade student, had an RPI for Broad Oral Language of 45/90 and an RPI for Amplio lenguaje oral of 90/90. The CLI is expressed with the Spanish numerator first and the English numerator second, resulting in a CLI of 90/45. This CLI indicates that Antonio has 90% proficiency on the Spanish cluster and 45% proficiency on the parallel English cluster.

Cognitive Academic Language Proficiency

Another interpretive feature of the WJ IV involves the CALP levels that may be reported. If selected in the Online Scoring and Reporting System, all administered clusters that yield a CALP score are reported in a separate section of the score report. For the WJ IV OL, CALP levels can be reported for the following clusters: Oral Language, Broad Oral Language, Oral Expression, Listening Comprehension, and the three Spanish-language clusters (Lenguaje oral, Amplio lenguaje oral, and Comprensión auditiva). Labels attached to the CALP levels describe the individual's proficiency. For example, a person's proficiency on a task may be described as *advanced*, *fluent*, or *limited*. In all, six levels of CALP are available, plus two regions that fall between the levels.

WJ IV TESTS OF ACHIEVEMENT

The WJ IV ACH (Schrank, Mather, & McGrew, 2014a) provides examiners with a comprehensive instrument for evaluating academic achievement. The WJ IV ACH is a companion battery to the WJ IV COG and the WJ IV OL.

Organization of the WJ IV ACH

The WJ IV ACH has 20 tests that are organized into five main areas: reading, mathematics, written language, academic knowledge, and cross-domain clusters. The cross-domain clusters include tests from three different curricular areas (reading, mathematics, and writing). For example, the Academic Fluency cluster includes Sentence Reading Fluency, Math Facts Fluency, and Sentence Writing Fluency.

All of the tests are contained in two easel test books, the Standard Battery (Tests 1–11) and the Extended Battery (Tests 12–20). The Standard Battery has three forms (Forms A, B, and C), and there is one form of the Extended Battery. The Standard Battery includes the most commonly administered tests, so having three forms provides alternate and equivalent tests to avoid overexposure to the same set of test items. The Extended Battery, which can be used with any of the three forms of the Standard Battery, includes tests that provide greater breadth and depth of coverage. Figure 14.4 shows the organization of the WJ IV ACH, which applies to all forms of the Standard Battery and the Extended Battery.

The areas of reading, mathematics, and written language each include measures of basic skills, fluency or automaticity, and application or higher-level skills. The Academic Knowledge cluster includes individual tests of Science, Social Studies, and Humanities, which respectively sample an individual's knowledge of the biological and physical sciences; history, geography, government, psychology, and economics; and art, music, and literature. Table 14.14 provides an overview of the content and task demands of each of the 20 achievement tests.

How to Administer the WJ IV ACH

In many cases, tests will be administered in the order they are presented in the easel test books, particularly the core set of tests (Tests 1–6). The tests have been ordered so that tasks alternate between different formats and achievement areas (e.g., writing vs. math). However, the tests may be administered in any order.

Administration Time

As a general rule, experienced examiners require about 40 minutes to administer the core set of tests (Tests 1–6). The Writing Samples test requires approximately 15 minutes to administer, whereas the other tests, on average, require about 5 minutes each. The tests in the Extended Battery require an additional 5–10 minutes each. The amount of time varies, depending on an examinee's particular characteristics, age, and speed of response.

			Reading						Mathematics				Writing				Cross-Domain Clusters					
			Reading	Broad Reading	Basic Reading Skills	Reading Comprehension	Reading Fluency	Reading Rate	Mathematics	Broad Mathematics	Math Calculation Skills	Math Problem Solving	Written Language	Broad Written Language	Basic Writing Skills	Written Expression	Academic Skills	Academic Fluency	Academic Applications	Academic Knowledge		Brief (or Broad) Achievement
Standard Battery	ACH 1	Letter-Word Identification	■	■	■												■					■
	ACH 2	Applied Problems							■	■		■							■			■
	ACH 3	Spelling											■	■	■		■					■
	ACH 4	Passage Comprehension	■	■		■													■			●
	ACH 5	Calculation							■	■	■						■					●
	ACH 6	Writing Samples											■	■		■			■			●
	ACH 7	Word Attack			■																■	
	ACH 8	Oral Reading					■															
	ACH 9	Sentence Reading Fluency		■			■	■										■				●
	ACH 10	Math Facts Fluency								■	■							■				●
	ACH 11	Sentence Writing Fluency												■		■		■				●
Extended Battery	ACH 12	Reading Recall				■																
	ACH 13	Number Matrices										■										
	ACH 14	Editing													■							
	ACH 15	Word Reading Fluency						■														
	ACH 16	Spelling of Sounds																			■	
	ACH 17	Reading Vocabulary				□																
	ACH 18	Science																		■		
	ACH 19	Social Studies																		■		
	ACH 20	Humanities																		■		

■ Tests required to create the cluster listed.
□ Additional test required to create an extended version of the cluster listed.
● Additional tests required to create the Broad Achievement cluster.

FIGURE 14.4. Selective Testing Table (tests and clusters) for the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH). From *Woodcock–Johnson IV™* (*WJ IV™*). Copyright © The Riverside Publishing Company. All rights reserved. Used by permission of the publisher.

Testing Materials

The testing materials required for administration of the WJ IV ACH include the test books (for the Standard Battery and Extended Battery), a test record, a response booklet, the audio CD, appropriate audio equipment, pencils, and a stopwatch.

Summary of Key Administration and Scoring Points

Most tests use suggested starting points and basal and ceiling rules; responses are scored 1 or 0. Table 14.15 summarizes key administration and scoring points for each test.

Special Administration and Scoring Considerations

Although the WJ IV ACH examiner's manual (Mather & Wendling, 2014a) and test book provide detailed rules for test by test administration, this section presents important reminders about tests that have special administration or scoring rules.

Tests Requiring a Response Booklet

The following tests require a response booklet: Test 3: Spelling, Test 5: Calculation, Test 6: Writing Samples, Test 9: Sentence Reading Fluency, Test 10: Math Facts Fluency, Test 11: Sentence

TABLE 14.14. Content and Task Demands of the 20 WJ IV ACH Tests

Test name	Description	Task demands
	Reading	
Test 1: Letter–Word Identification	Measures an aspect of reading decoding.	Requires identifying and pronouncing isolated letters and words.
Test 4: Passage Comprehension	Measures reading comprehension of contextual information.	Requires reading a short passage and supplying a key missing word.
Test 7: Word Attack	Measures aspects of phonological and orthographic coding.	Requires applying phonic and structural analysis skills to pronounce phonically regular nonsense words.
Test 8: Oral Reading	Measures word-reading accuracy and prosody.	Requires reading sentences aloud that gradually increase in difficulty.
Test 9: Sentence Reading Fluency	Measures reading rate.	Requires reading and comprehending simple sentences, and then deciding if each statement is true or false by marking "Yes" or "No" (3-minute time limit).
Test 12: Reading Recall	Measures reading comprehension and meaningful memory.	Requires reading a passage silently one time and then retelling the story orally.
Test 15: Word Reading Fluency	Measures vocabulary knowledge and semantic fluency.	Requires marking two words that go together in a row of four words (3-minute time limit).
Test 17: Reading Vocabulary	Measures reading vocabulary and comprehension.	Requires reading and providing synonyms or antonyms.
	Mathematics	
Test 2: Applied Problems	Measures the ability to analyze and solve practical math problems, mathematical reasoning.	Requires comprehending the nature of the problem, identifying relevant information, performing calculations, and recording solutions.
Test 5: Calculation	Measures the ability to perform mathematical computations.	Requires calculation of simple to complex mathematical facts and equations.
Test 10: Math Facts Fluency	Measures aspects of number facility and math achievement.	Requires rapid calculation of single-digit addition, subtraction, and multiplication facts (3-minute time limit).
Test 13: Number Matrices	Measures quantitative reasoning.	Requires providing the missing number from a matrix.
	Written Language	
Test 3: Spelling	Measures the ability to spell dictated words.	Requires writing the correct spelling of words presented orally.
Test 6: Writing Samples	Measures quality of meaningful written expression and ability to convey ideas.	Requires writing sentences in response to a series of demands that increase in difficulty.
Test 11: Sentence Writing Fluency	Measures aspects of automaticity with syntactic components of written expression.	Requires formulating and writing simple sentences rapidly (5-minute time limit).

(continued)

TABLE 14.14. *(continued)*

Test name	Description	Task demands
Test 14: Editing	Measures the ability to identify and correct errors in spelling, usage, punctuation, and capitalization.	Requires identifying errors in short written passages and correcting them orally.
Test 16: Spelling of Sounds	Measures aspects of phonological/orthographic coding.	Requires spelling nonsense words that conform to conventional English spelling rules.
	Academic Knowledge	
Test 18: Science	Provides a survey of knowledge in science, including biology, chemistry, geology, and physics.	Requires providing an oral response to orally presented questions. Many items provide visual stimuli, and early items require a pointing response only.
Test 19: Social Studies	Provides a survey of knowledge in social studies, including history, psychology, geography, government, and economics.	Requires providing an oral response to orally presented questions. Many items provide visual stimuli, and early items require a pointing response only.
Test 20: Humanities	Provides a survey of knowledge in humanities, including art, music, and literature.	Requires providing an oral response to orally presented questions. Many items provide visual stimuli, and early items require a pointing response only.

Writing Fluency, Test 15: Word Reading Fluency, and Test 16: Spelling of Sounds. In addition, the response booklet contains a worksheet to use with Test 2: Applied Problems and Test 13: Number Matrices. The examiner must maintain control of the response booklet, presenting it to the examinee and removing as directed by instructions in the test book.

Timed Tests

The following tests are timed: Test 9: Sentence Reading Fluency, Test 10: Math Facts Fluency, Test 11: Sentence Writing Fluency, and Test 15: Word Reading Fluency. The time limit is 3 minutes each for Sentence Reading Fluency, Math Facts Fluency and Word Reading Fluency, and 5 minutes for Sentence Writing Fluency.

A stopwatch or the stopwatch feature on a smartphone is required to administer these tests. If a stopwatch is unavailable, a watch or clock with a second hand should be used. In this case, the examiner should write down the exact starting and stopping times in minutes and seconds in the spaces provided in the test record, rather than depending on memory. The examiner enters the times and numbers correct in the Online Scoring and Reporting System to generate the scores for these tests. The exact finishing time for each test must be entered because earlier finishers who do well will receive a higher score than individuals who continue to work for the full time limit.

Audio-Recorded Tests

Test 16: Spelling of Sounds is the only WJ IV ACH test that is presented with an audio recording. Use of the test's audio recording is expected, and headphones are recommended unless the person being tested resists wearing headphones or has difficulty attending to a recorded presentation.

Qualitative Observation Checklists

Each of the 11 tests in the Standard Batteries (Forms A, B, C) has a Qualitative Observation Checklist in the test record. These checklists are designed to document an examinee's performance on the test through qualitative observations, or in the case of Test 8: Oral Reading, a quantitative observation. Although optional, use of these checklists can provide important insights about how the individual completed the task.

TABLE 14.15. Summary of WJ IV ACH Test Administration and Scoring Rules

Test name	Item scoring rule	Sample items?	Basal/ceiling rules?[a]	Administration notes	Extra materials required for administration
			Standard Battery tests (Forms A, B, C)		
Test 1: Letter–Word Identification	1, 0	No	Yes; 6/6	Suggested starting points are available; word must be pronounced smoothly to receive credit.	
Test 2: Applied Problems	1, 0	No	Yes; 5/5	Suggested starting points are available.	Response booklet
Test 3: Spelling	1, 0	No	Yes; 6/6	Suggested starting points are available.	Response booklet
Test 4: Passage Comprehension	1, 0	Yes (PreK)	Yes; 6/6	Suggested starting points are available.	
Test 5: Calculation	1, 0	Yes (Pre-G1)	Yes; 6/6	Suggested starting points are available.	Response booklet
Test 6: Writing Samples	2, 1.5, 1, 0.5, 0	No	No; ceiling determined by score on block of items	Suggested starting points are available; scoring is based on specific block of items administered	Response booklet
Test 7: Word Attack	1, 0	Yes (G3–adult)	Yes; 6/6	Suggested starting points are available; word must be pronounced smoothly to receive credit.	
Test 8: Oral Reading	2, 1, 0	No	No; ceiling determined by continuation instructions	Suggested starting points are available; be familiar with the types of reading mistakes that count as errors.	
Test 9: Sentence Reading Fluency	1, 0	Yes	No; 3-minute time limit	All begin with sample items, then proceed to item 1; use scoring guide overlay.	Stopwatch (timed test); response booklet
Test 10: Math Facts Fluency	1, 0	No	No; 3-minute time limit	All begin with item 1; use scoring guide overlay.	Stopwatch (timed test); response booklet
Test 11: Sentence Writing Fluency	1, 0	Yes	No; 5-minute time limit	All begin with sample items, then proceed to item 1.	Stopwatch (timed test); response booklet

			Extended Battery tests		
Test 12: Reading Recall	1 for each correctly recalled element	No	No; ceiling determined by continuation instructions	Suggested starting points are available; direct examinee to read the story once silently.	
Test 13: Number Matrices	1, 0	Yes	Yes; 6/6	Choose appropriate sample item; then use suggested starting points.	Response booklet; stopwatch to monitor response time
Test 14: Editing	1, 0	Yes	Yes; 6/6	All begin with sample items; then use suggested starting points.	
Test 15: Word Reading Fluency	1, 0	Yes	No; 3-minute time limit	All begin with sample items; then proceed to item 1.	Stopwatch (timed test); response booklet
Test 16: Spelling of Sounds	1, 0	Yes	Yes; 6/6	Suggested starting points are available; scoring is based on written response only.	Audio; response booklet
Test 17: Reading Vocabulary				Both subtests must be administered.	
17A: Synonyms	1, 0	Yes	Yes; 5/5	All begin with sample items; then use suggested starting points.	
17B: Antonyms	1, 0	Yes	Yes; 5/5	All begin with sample items; then use suggested starting points.	
Test 18: Science	1, 0	No	Yes; 6/6	Suggested starting points are available.	
Test 19: Social Studies	1, 0	No	Yes; 6/6	Suggested starting points are available.	
Test 20: Humanities	1, 0	No	Yes; 6/6	Suggested starting points are available.	

[a] May be modified by complete-page rule.

Test 6: Writing Samples

The examiner administers the appropriate block of items as indicated on the table in the test book on the page after the Writing Samples tab. If it is apparent that the examinee is experiencing undue ease or difficulty with the assigned block of items, it may be necessary to administer additional items to obtain a better estimate of writing ability. This can be done immediately, or because the test is scored after testing is completed, it may be necessary to administer the additional items at a convenient time within the next few days.

The Writing Samples scoring table in the test record allows the examiner to determine whether the most appropriate block of items has been administered. If the individual's raw score falls within one of the seven shaded areas on the scoring table, the additional items noted in the Adjusted Item Block chart on page 7 in the test record should be administered. This chart also indicates the block of items to use for calculating the raw score.

If an examinee's response to an item is illegible or difficult to read, the person can be asked to write as neatly as possible—*but not, however, to read aloud what was written for scoring purposes*. Illegible responses are scored as 0. If requested by the examinee, the examiner may read any words during this test or repeat the instructions. When an examinee asks whether spelling is important or how to spell a word, the examiner should encourage the examinee just to do the best he or she can.

Writing Samples is scored after the testing is completed. Items 1–6 are scored 1 or 0. Items 7 and higher may be scored 2, 1.5, 1, 0.5, or 0 points, according to a modified holistic procedure that requires examiner judgment. Because scoring of this test is more involved and subjective than the scoring of other WJ IV ACH tests, special rating and scoring procedures are provided in Appendix B of the examiner's manual that accompanies the test.

Test 8: Oral Reading

Oral Reading is composed of a set of sentences that gradually increase in difficulty. The examinee reads each sentence orally from the test book. The examiner uses the sentences printed in the test record and marks each error with a slash mark at the point in a sentence the error occurs. Examiners must familiarize themselves with the types of errors that are marked: mispronunciations, omissions, insertions, substitutions, hesitations of 3 seconds or more, repetitions, transpositions, and ignoring punctuation. Oral Reading uses continuation instructions to determine when to continue or discontinue testing. A multipoint scoring system is used to score the examinee's oral reading of each sentence: 2 (no errors), 1 (one error), or 0 (2 or more errors).

Test 9: Sentence Reading Fluency

The score for Sentence Reading Fluency is based on both the number of correct responses and the number of incorrect responses. Both totals must be entered into the Online Scoring and Reporting System.

Test 12: Reading Recall

Reading Recall requires the examinee to read a story silently one time and then retell the story. The scoring is based on the number of story elements the examinee recalls correctly during the retelling. The test is administered in two-story sets, and continuation instructions are used to determine when to administer additional stories and when to discontinue testing. The administration and scoring of this test are similar to those for WJ IV COG Test 6: Story Recall.

Test 16: Spelling of Sounds

Spelling of Sounds requires the examinee to write the spelling of a given sound or nonsense word. The examinee is asked to say the sound or word aloud before writing, but the scoring is based solely on the written response. All correct responses are shown in the test book and represent the only spellings that can be scored a 1.

Reliability and Validity

Following is a summary of reliability and validity information for the WJ IV ACH. For more complete information, consult the WJ IV technical manual (McGrew et al., 2014) and the discussion of CHC theory in the introductory section of this chapter.

Median reliability coefficients for the WJ IV ACH tests are reported in Table 14.16. With the exception of the timed tests, coefficients of stability are reported as median internal-consistency (r_{11}) and median *SEM* values. For the four timed tests, test–retest values (r_{12}) were used as the most appropriate coefficient of stability. Mosier's (1943) formula was used to calculate reliabilities for tests

TABLE 14.16. WJ IV ACH Median Test Reliability Statistics

Test	Median r_{11}	Median *SEM* (SS)	Median r_{12}
	Standard Battery		
Test 1: Letter–Word Identification	.94	5.21	—
Test 2: Applied Problems	.92	4.75	—
Test 3: Spelling	.92	4.86	—
Test 4: Passage Comprehension	.89	6.08	—
Test 5: Calculation	.93	4.99	—
Test 6: Writing Samples	.90	4.74	—
Test 7: Word Attack	.90	5.03	—
Test 8: Oral Reading	.96	3.35	—
Test 9: Sentence Reading Fluency	—	—	.93
Test 10: Math Facts Fluency	—	—	.95
Test 11: Sentence Writing Fluency	—	—	.83
	Extended Battery		
Test 12: Reading Recall	.92	2.62	—
Test 13: Number Matrices	.92	3.75	—
Test 14: Editing	.91	4.85	—
Test 15: Word Reading Fluency	—	—	.92
Test 16: Spelling of Sounds	.88	5.89	—
Test 17: Reading Vocabulary	.88	6.36	—
Test 18: Science	.84	7.39	—
Test 19: Social Studies	.87	6.76	—
Test 20: Humanities	.87	7.96	—

Note. r_{11}, internal-consistency reliability; r_{12}, test–retest reliability for speeded tests.

with subtests, such as ACH Test 17: Reading Vocabulary. Because the WJ IV ACH Standard Battery has three parallel forms (A, B, and C), alternate-form equivalence was studied. The empirical evidence supports the equivalence of the alternate forms.

Table 14.17 contains the median cluster reliability coefficients (r_{cc}) and median standard errors of measurement for obtained cluster standard scores, or *SEM* (SS), as calculated via Mosier's formula. All of the WJ IV ACH clusters exceed the commonly referenced professional standard of .90 for decision-making purposes. Although the test reliabilities support use and interpretation of the test-level scores as single indicators of an ability, the WJ IV ACH clusters are preferred for most decision-making purposes, due to their higher reliabilities.

Table 14.18 contains a summary of correlations for the WJ IV ACH reading, math, and written language clusters with other measures of achievement, including the Kaufman Test of Educational Achievement—Second Edition (KTEA-2; Kaufman & Kaufman, 2004b), the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), and the Oral and Written Language Scales: Written Expression (OWLS-WE; Carrow-Woolfolk, 1996). The significant correlations support the validity of the WJ IV ACH clusters as valid measures of reading, mathematics, and written language.

Interpretation

The WJ IV ACH provides measures for evaluating six of the eligibility areas for specific learning disabilities listed in IDEA (2004): basic reading skills, reading comprehension, reading fluency, math calculation skills, math problem solving, and written expression. The other two areas, oral expression and listening comprehension, are evaluated with the WJ IV OL.

Within the framework of CHC theory, the WJ IV ACH tests are primarily measures of reading–writing ability (Grw) or quantitative knowledge (Gq). The reading and written language tests provide measures of Grw, defined by Newton and McGrew (2010) as "the breadth and depth of a person's acquired store of declarative and procedure reading and writing skills and knowledge" (p. 628). The WJ IV ACH mathematics tests provide measures of a broad quantitative knowledge (Gq) construct, or "the breadth and depth of a person's acquired store of declarative and procedural quantitative or numerical knowledge" (Newton & McGrew, 2010, p. 628). The three tests in the Academic Knowledge cluster, Science, Social Studies, and Humanities, are primarily measures of domain-specific knowledge (Gkn), defined as depth, breadth, and mastery of specialized knowledge (Schneider & McGrew, 2012).

However, the WJ IV ACH tests are cognitively complex tasks requiring the application of one or more narrow abilities and dynamic cognitive processes (McGrew, Schrank, & Woodcock, 2007). Table 14.19 outlines the narrow abilities defined by CHC theory for each of the WJ IV ACH tests. Table 14.20 provides a summary of selected interventions and accommodations for each WJ IV ACH test. As in the WJ IV COG and WJ IV OL, the identified narrow abilities and inferred cognitive processes are used to provide a link between WJ IV ACH test scores and evidence-based interventions.

TABLE 14.17. WJ IV ACH Median Cluster Reliability Statistics

Cluster	r_{cc}	Median *SEM* (SS)
Reading	.95	4.24
Broad Reading	.97	3.35
Basic Reading Skills	.95	3.97
Reading Comprehension	.93	4.24
Reading Comprehension—Ext.	.96	3.00
Reading Fluency	.96	3.00
Reading Rate	.96	3.00
Mathematics	.96	3.67
Broad Mathematics	.97	3.00
Math Calculation Skills	.97	3.35
Math Problem Solving	.95	3.35
Written Language	.94	3.97
Broad Written Language	.95	3.67
Basic Writing Skills	.95	3.67
Written Expression	.92	4.50
Academic Skills	.97	3.35
Academic Applications	.96	3.67
Academic Fluency	.97	2.60
Academic Knowledge	.95	4.24
Phoneme–Grapheme Knowledge	.94	4.24
Brief Achievement	.97	3.35
Broad Achievement	.99	2.12

Basic Reading Skills

Test 1: Letter–Word Identification and Test 7: Word Attack constitute the Basic Reading Skills cluster and provide a broad view of an individual's basic word-reading skills, including sight word recognition and phonic skills. The dual-route theory (Coltheart, 1980, 2007) suggests that two pathways are involved in learning words: a *lexical* procedure and a *nonlexical* procedure. The lexical procedure is engaged when the person is reading real words, whereas the nonlexical procedure involves converting letters to phonemes and aids in using phonics and reading nonwords. Comparing the results of the two tests in this cluster helps determine whether word recognition skills, phonic skills, or both are limited and require remediation. Analysis of errors made on both tests can help target specific instructional elements. Table 14.20 provides examples of interventions.

If an individual has trouble decoding words, few resources are left for comprehension. Slow, labored reading with many errors will have a negative impact on comprehension. Analyzing an individual's performance on the Basic Reading Skills tests is important to the interpretation of the Reading Comprehension tests. The examiner must be sure to consider the effect of decoding problems before identifying a problem in reading comprehension.

Test 1: Letter–Word Identification

Letter–Word Identification requires the individual to read aloud isolated letters and words presented in a list. It is a measure of reading decoding (word recognition), including reading readiness skills. Knowledge of word meanings is not required to perform this task. Individuals with good word

TABLE 14.18. Correlations for Select WJ IV ACH Measures and Other Measures of Achievement

Other measure	Broad Reading	Reading Fluency	Broad Math	Broad Written Language
Kaufman Test of Educational Achievement—Second Edition (KTEA-2)[a]				
Reading	.92	.86	.78	.85
Decoding Fluency	.83	.80	.62	.79
Math	.73	.70	.91	.68
Written Language	.79	.75	.71	.84
Wechsler Individual Achievement Test—Third Edition (WIAT-III)[b]				
Total Reading	.89	.87	.67	.83
Reading Comprehension and Fluency	.82	.80	.58	.68
Mathematics	.73	.71	.90	.72
Written Expression	.75	.76	.63	.77
Oral and Written Language Scales: Written Expression (OWLS-WE)[c]				.70

Note. Samples used in studies varied by age or grade as noted.

[a] Ages 8–12.

[b] Grades 1–8.

[c] Ages 7–17.

recognition skills recognize the letters and words rapidly and with little effort. In other words, they demonstrate automaticity at the word-reading level, which facilitates reading performance.

Low performance on this test may be due to inefficient or limited strategies for word identification. Individuals who lack automaticity of word identification skills may identify several words accurately, but require increased time and greater attention to phonic analysis to determine the correct response. These individuals tend to read slowly and make numerous errors. They may be easily frustrated or unwilling to try for fear of making an error. The results of this test can be compared with those of Test 7: Word Attack, the other test in the Basic Reading Skills cluster. Other helpful comparisons include Test 4: Passage Comprehension, Test 8: Oral Reading, Test 9: Sentence Reading Fluency, Test 12: Reading Recall, Test 15: Word Reading Fluency, and Test 17: Reading Vocabulary. These comparisons can provide insights into the individual's level of reading skills with and without the context of meaning. Although all of the tests require word identification, these other tests also require knowledge of word meanings, sentence structure, and comprehension.

Use of an explicit, systematic, synthetic phonics program is one of the most effective interventions for developing basic reading skills (Brady, 2011; National Reading Panel, 2000). High-frequency words should be taught in conjunction with a structured phonics program to help develop word recognition skills.

Test 7: Word Attack

Word Attack requires the individual to read phonically regular nonsense words orally. It measures aspects of both phonological and orthographic coding. Phoneme–grapheme knowledge is necessary to perform well on this test because the words are not real words and cannot be recognized or recalled from memory. The reader must apply phoneme–grapheme knowledge in order to translate a written word that is not stored in memory into speech (Berninger & Richards, 2010).

Low performance on Word Attack may result from poor phonological processing, limited phoneme–grapheme knowledge, poor decoding skills and strategies, or a lack of fluency. Impaired decoding is frequently thought to be the basis of reading problems. Analyzing the types of errors

TABLE 14.19. WJ IV ACH Tests: CHC Broad and Narrow Abilities Measured

Test name	Primary broad CHC ability	*Narrow ability*
Test 1: Letter–Word Identification	Reading and writing ability (Grw)	*Reading decoding* (RD)
Test 2: Applied Problems	Quantitative knowledge (Gq)	*Mathematical achievement* (A3)
	Fluid reasoning (Gf)	*Quantitative reasoning* (RQ)
Test 3: Spelling	Reading and writing ability (Grw)	*Spelling ability* (SG)
Test 4: Passage Comprehension	Reading and writing ability (Grw)	*Reading comprehension* (RC)
Test 5: Calculation	Quantitative knowledge (Gq)	*Mathematical achievement* (A3)
Test 6: Writing Samples	Reading and writing ability (Grw)	*Writing ability* (WA)
Test 7: Word Attack	Reading and writing ability (Grw)	*Reading decoding* (RD)
	Auditory processing (Ga)	*Phonetic coding* (PC)
Test 8: Oral Reading	Reading and writing ability (Grw)	*Reading comprehension* (RC) *Verbal (print) language comprehension* (V)
Test 9: Sentence Reading Fluency	Reading and writing ability (Grw)	*Reading comprehension* (RC) *Reading speed* (RS)
	Processing speed (Gs)	
Tet 10: Math Facts Fluency	Quantitative knowledge (Gq)	*Mathematical achievement* (A3)
	Processing speed (Gs)	*Number facililty* (N)
Test 11: Sentence Writing Fluency	Reading and writing ability (Grw)	*Writing ability* (WA) *Writing speed* (WS)
	Processing speed (Gs)	
Test 12: Reading Recall	Reading and writing ability (Grw)	*Reading comprehension* (RC)
	Long-term storage and retrieval (Glr)	*Meaningful memory* (MM)
Test 13: Number Matrices	Fluid reasoning (Gf)	*Quantitative reasoning* (RQ)
Test 14: Editing	Reading and writing ability (Grw)	*English usage* (EU)
Test 15: Word Reading Fluency	Reading and writing ability (Grw)	*Reading comprehension* (RC) *Reading speed* (RS)
	Processing speed (Gs)	
Test 16: Spelling of Sounds	Reading and writing ability (Grw)	*Spelling ability* (SG)
	Auditory Processing (Ga)	*Phonetic coding* (PC)
Test 17: Reading Vocabulary	Reading and writing ability (Grw)	*Reading comprehension* (RC)
	Comprehension–knowledge (Gc)	*Lexical knowledge* (VL)
Test 18: Science	Domain-specific knowledge (Gkn)	*General science information* (K1)
	Comprehension–knowledge (Gc)	*General (verbal) information* (K0)
Test 19: Social Studies	Domain-specific knowledge (Gkn)	*Knowledge of culture* (K2) *Geography achievement* (A5)
	Comprehension–knowledge (Gc)	*General (verbal) information* (K0)
Test 20: Humanities	Domain-specific knowledge (Gkn)	*Knowledge of culture* (K2)
	Comprehension–knowledge (Gc)	*General (verbal) information* (K0)

TABLE 14.20. Examples of Interventions, Strategies, and Accommodations Related to Limitations in Performance on WJ IV ACH Tests

Test	Brief intervention, strategy, or accommodation
Test 1: Letter–Word Identification	Explicit, systematic synthetic phonics instruction; teaching word recognition strategies; teaching high-frequency words.
Test 2: Applied Problems	Combining direct instruction and strategy instruction; teaching schema-based problem solving; use of modeling and verbal rehearsal.
Test 3: Spelling	Use of the write–say method; teaching common irregular words; teaching morphology; having student keep a dictionary of words commonly used in his or her writing; reducing number of words on spelling tests.
Test 4: Passage Comprehension	Activating prior knowledge; use of graphic organizers; teaching self-monitoring strategies; use of books on audio to ensure access to content.
Test 5: Calculation	Developing number sense; sequential direct instruction; employing concrete–representational–abstract teaching sequence; use of computer-assisted instruction.
Test 6: Writing Samples	Creating a literate, motivating, risk-free environment; teaching the writing process; providing daily practice; teaching text structures; increasing time to complete writing tasks; decreasing number of writing assignments; use of technology to support writing (e.g., word processor, spell-checker, voice recognition).
Test 7: Word Attack	Teaching phoneme–grapheme mapping, using letter tile or Elkonin boxes; explicit, systematic, synthetic phonics instruction; use of decodable texts.
Test 8: Oral Reading	Modeling fluent oral reading; for practice, use of texts on which the student is accurate; cueing of phrase boundaries (e.g., use of slash marks or scoops to identify phrases).
Test 9: Sentence Reading Fluency	Use of repeated readings with error correction procedure; providing frequent practice; previewing passages and practicing words in isolation.
Test 10: Math Facts Fluency	Practicing with math facts charts; use of explicit timings and graph progress; use of computer programs for drill and practice.
Test 11: Sentence Writing Fluency	Use of word, phrase, and sentence fluency-building activities; teaching the mechanics of writing; permitting use of word processor.
Test 12: Reading Recall	Teaching use of mental imagery strategies while reading; providing instruction and practice with a self-monitoring strategy; reciprocal teaching; having student summarize passages after reading.
Test 13: Number Matrices	Teaching number patterns and core math concepts; counting by increments; use of manipulatives; developing a sense of numerical quantity; use of math talk.
Test 14: Editing	Teaching proofreading skills; employing peer editing; use of technology.
Test 15: Word Reading Fluency	Conducting 1-minute speed drills of word lists; teaching word recognition skillls; use of word webs.
Test 16: Spelling of Sounds	Instruction in orthography and morphology; use of multi-sensory techniques; teaching use of computerized spell-checker.
Test 17: Reading Vocabulary	Use of semantic feature analysis; encouraging reading for different purposes; teaching synonyms, antonyms, and multiple-meaning words.
Test 18: Science	Frequent exposure to and practice with words and concepts related to science; increasing experiences (e.g., visits to museums, conducting experiments); use of semantic maps.
Test 19: Social Studies	Frequent exposure to and practice with words and concepts related to social studies; increasing experiences (e.g., visits to museums, community services); reading in the content area.
Test 20: Humanities	Frequent exposure to and practice with words and concepts related to humanities; increasing experiences (e.g., visits to art museums/concerts, reading different genres aloud).

the individual makes on this test not only helps to plan the most appropriate instructional program, but also helps determine whether further testing is indicated. In addition to explicit and systematic instruction in phonics, teaching structural analysis skills (e.g., six syllable types or onsets–rimes) improves the individual's ability to decode words (Ehri, 2000).

Reading Comprehension

The Reading Comprehension cluster is composed of Test 4: Passage Comprehension and Test 12: Reading Recall. This cluster provides a broad view of the individual's reading comprehension skill. Reading comprehension is a complex cognitive process that requires intentional interaction between the reader and the text to construct meaning (Durkin, 1993). Both tests in this cluster measure comprehension in the context of connected discourse. Reading Recall also has a meaningful memory component (the ability to encode, consolidate, or store meaningful representations of the story details), requiring subsequent reconstruction of the passage that was read. Because of the retelling aspect of Reading Recall, expressive language demands are increased and must be considered when interpreting results.

The Reading Comprehension—Extended cluster includes Passage Comprehension and Reading Recall, but adds a third test, Test 17: Reading Vocabulary. Reading Vocabulary measures comprehension in a decontextualized format—that is, understanding of word meanings in isolation. Comparing the results of the three tests in this cluster helps determine whether a meaningful context helps or interferes with an individual's comprehension.

Other helpful comparisons include comparing the results of the Reading Comprehension cluster to those of the Basic Reading Skills cluster, the Academic Knowledge cluster, the WJ IV OL Listening Comprehension cluster, and the WJ IV COG General Information test. Low performance on Reading Comprehension tasks may result from low basic reading skills or limited oral language or background knowledge. Considering the impact of these various factors helps determine the most appropriate instructional program for the individual. See Table 14.20 for examples of interventions.

Test 4: Passage Comprehension

Passage Comprehension requires the individual to read a short passage silently, comprehend the information, and provide a missing word. It is a measure of reading comprehension and lexical knowledge. This modified cloze task requires the ability to use both syntactic and semantic clues in comprehending text.

Low performance on Passage Comprehension may result from limited basic reading skills, comprehension difficulties, or both. Analysis of the types of errors made helps determine the most appropriate instructional plan. One effective intervention for improving reading comprehension is to teach a variety of comprehension strategies (National Reading Panel, 2000). The seven most effective strategies identified by the National Reading Panel are these: comprehension monitoring, cooperative learning, use of graphic and semantic organizers, question answering, question generating, story structures, and summarization.

The results from Passage Comprehension can be compared to those of WJ IV OL Test 2: Oral Comprehension, a similar task that does not require reading, to help determine whether the problem is with reading alone or involves limited language comprehension. If the individual does well on the oral test, then language comprehension is not likely to be the reason for poor performance on the reading test. If the individual does poorly on the oral test, then any limitations in language comprehension must also be considered as a contributing factor to poor reading performance.

Test 12: Reading Recall

Reading Recall requires the individual to read a story and then reconstruct the elements of that story. Both reading and expressive language skills are required to perform this story-retelling task. Reading Recall measures reading comprehension, meaningful memory, and language development. Poor attention, poor memory consolidation, poor decoding, limited vocabulary, low expressive language, or limited comprehension may negatively affect performance on this test. Performance on the Reading Recall test can be compared to that on WJ IV COG Test 6: Story Recall, a similar task that does not require reading. If performance is higher on Reading Recall than Story Recall, this suggests that comprehension is stronger when the examinee is reading silently than when attending to orally imparted discourse. If performance is higher on Story Recall than on Reading Recall, this suggests that the problem is not with memory consolidation or attention, but rather with reading skills.

One intervention to improve comprehension is to teach the reader to visualize while reading.

Visualization requires the reader to be actively engaged with the text and to think deeply about the content. Individuals who have difficulty visualizing the information provided by the text struggle with comprehension (Gersten, Fuchs, Williams, & Baker, 2001). Explicit instruction in how to visualize while reading may be required. Teachers can ask specific questions about a text to evoke imagery, and can also model how to use visualization through a think-aloud process.

Test 17: Reading Vocabulary

Reading Vocabulary has two parts: 17A Synonyms and 17B Antonyms. The individual is required to read words and to supply (orally) synonyms and antonyms, respectively. Low performance on this test may result from poor basic reading skills, limited vocabulary, or both. An individual who reads the stimulus words correctly, but provides incorrect responses, may be better at decoding than comprehending. An individual who misreads the stimulus words, but provides the correct responses, may be better at comprehending than decoding.

Reading Vocabulary performance can be directly compared to the individual's performance on WJ IV COG Test 1: Oral Vocabulary, a similar task that does not require reading. If the individual's performance is higher on the oral task than on the reading task, then the focus of instruction should be on developing basic reading skills. If the individual's performance is low on both the oral task and the reading task, then the focus of instruction should be on developing oral vocabulary, as well as basic reading skills. Use of semantic feature analysis (Anders & Bos, 1986; Pittelman, Heimlich, Berglund, & French, 1991) helps increase vocabulary by exploring similarities, differences, and connections between words and concepts.

Reading Fluency

Reading fluency is often described as the bridge between word identification and comprehension, and is considered the best predictor of reading comprehension (Chall, 1996; Chard, Vaughn, & Tyler, 2002). The National Reading Panel (2000) identified reading fluency as one of the five critical areas of reading assessment and instruction, which led to the addition of reading fluency as one of the eight areas of eligibility for specific learning disabilities (IDEA, 2004). The Reading Fluency cluster contains two tests: Test 8: Oral Reading, an untimed measure of oral reading, and Test 9: Sentence Reading Fluency, a timed measure of silent reading. Although the Oral Reading test is not timed, a person's ease of reading, accuracy, and expression (prosody) can be observed as the sentences are read aloud. The number and type of errors the person makes while reading can also be recorded in the Qualitative Observation Tally located in the test record. This type of information helps inform instructional planning. Sentence Reading Fluency is a timed silent reading task that requires comprehending simple sentences quickly. An examiner comparing the individual's performance on the two tests in this cluster should consider how the differences in task and response demands may have influenced performance. See Table 14.20 for suggested interventions.

Test 8: Oral Reading

The Oral Reading test requires oral reading of a set of sentences that gradually increase in complexity. The test provides information about the individual's decoding skill, automaticity with reading, and prosody (reading with appropriate expression and attention to punctuation). The Qualitative Observation Tally in the Test Record can be used to code the type of errors an individual makes while reading the sentences and can lead to specific instructional implications. Both explicit modeling of fluent reading and repeated reading practice with corrective feedback have been identified as highly effective interventions for improving reading fluency (Chard et al., 2002).

Test 9: Sentence Reading Fluency

Sentence Reading Fluency requires the individual to read simple sentences quickly and indicate if each statement is true or false by circling "Yes" or "No." It is a measure of reading speed and automaticity. Low performance on Sentence Reading Fluency may be a result of difficulty sustaining attention, limited basic reading skills, slow processing speed, or comprehension difficulties. An individual's orthographic processing, perceptual speed, or processing speed (as assessed with the WJ IV COG Cognitive Processing Speed cluster or Perceptual Speed cluster) may facilitate or inhibit performance on this test.

The speed and fluency with which an individual performs basic skills can influence performance on higher-level skills. Sentence Reading Fluency can be compared with Test 15: Word Reading Fluency, to see whether performance is similar on both these timed tests. If the Sentence Reading Fluency score is higher than the Word Reading

Fluency score, it suggests that increased context improves word recognition and comprehension. Speed drills and monitoring progress are two interventions for building reading speed. When instruction is focused on rate, it is important to use materials on which the individual has a 95–98% accuracy level.

Other Reading Clusters

There are three additional reading clusters available: Reading Rate, Reading, and Broad Reading.

Reading Rate

The Reading Rate cluster consists of Test 9: Sentence Reading Fluency and Test 15: Word Reading Fluency. Both of these tests are timed and read silently. Sentence Reading Fluency requires comprehension of simple sentences, whereas Word Reading Fluency involves the understanding of vocabulary and the semantic relationships among pairs of words. Results on this cluster can be compared to those on Basic Reading Skills, as well as Reading Comprehension. In addition, the examiner may wish to investigate how the person performs on other timed measures (e.g., Math Facts Fluency and Sentence Writing Fluency), as well as on measures of processing and perceptual speed on the WJ IV COG. In some cases, the results of this cluster, when substantiated with additional information, can be used to help document a need for extended time on tests.

Test 15: Word Reading Fluency. Test 15 is a timed test that requires the examinee to silently and quickly read groups of four words and mark the two in each group that share a semantic relationship. Low performance on this test may result from poor basic reading skills, limited vocabulary, or slow processing speed. An individual's processing speed may facilitate or inhibit performance on this test.

Reading

The Reading cluster consists of two tests: Test 1: Letter–Word Identification, a measure of word recognition skill; and Test 4: Passage Comprehension, a measure of reading comprehension. The cluster does not include a timed reading test, so it provides an overall measure of reading accuracy and comprehension that does not include the effects of reading speed.

Broad Reading

The Broad Reading cluster is composed of three tests—Test 1: Letter–Word Identification, Test 4: Passage Comprehension, and Test 9: Sentence Reading Fluency. This cluster provides a broad overview of the individual's overall reading level. Because it is a mix of three different aspects of reading (basic skills, comprehension, and fluency), interpretation of this cluster is most meaningful when performance is similar on all three tests.

Math Calculation Skills

The Math Calculation Skills cluster is composed of two tests, Test 5: Calculation and Test 10: Math Facts Fluency. It provides a measure of basic math skills, including computational skills and automaticity with basic math facts. Relative ease with computations is an important factor in predicting math performance. The results of the two tests in this cluster can be compared, to help determine whether fluency with basic math facts or calculation skill is affecting performance. Table 14.20 provides a selection of interventions.

Test 5: Calculation

The Calculation test requires the individual to perform a variety of calculations ranging from simple addition and subtraction to complex calculus. Tasks progress in this order: (1) basic addition, subtraction, multiplication, and division; (2) advanced calculations of each operation with regrouping; (3) advanced calculations of each operation with negative numbers (except division); (4) fractions; (5) percentages; (6) algebra; (7) trigonometry; (8) logarithms; and (9) calculus. Calculation measures the ability to perform mathematical computations that are fundamental to more complex math reasoning and problem solving. Fluency with basic math skills is fundamental to more complex math (Geary, 2007). Low performance may result from limited basic math skills, weaknesses in short-term working memory or memory span, limited fluency or automaticity with math facts, poor or limited instruction, or difficulty with attention.

The examiner should observe the examinee's behaviors; categorize errors that are made; note which concepts are known and unknown; and interview the examinee, if needed. Effective instruction should be direct and explicit, with cumulative review (Fuchs et al., 2008). Combining direct instruction with strategy instruction yields more positive results than using either method alone (Karp & Voltz, 2000; Swanson, 2001).

Test 10: Math Facts Fluency

Math Facts Fluency requires the individual to solve simple addition, subtraction, and multiplication tasks quickly. Poor math fact retrieval is often described as a primary characteristic of individuals with a specific math disability (Andersson, 2008; Geary, 2007). Low performance on this test may result from poor attention, limited basic math skills, lack of automaticity, or slow perceptual speed. Some individuals work slowly and accurately, whereas other individuals work very quickly and make many errors. If numerous errors are made, an examiner should attempt to determine the reasons for the mistakes. As examples, a student may be confused or not pay careful attention to the signs, may not know multiplication facts, or may be confused about the properties of zero (e.g., 6 + 0 = 0). Related interventions for low performance on math facts fluency include developing number sense (Berch, 2005; Griffin, 2004) and use of explicit timings to increase speed of performance (Miller & Hudson, 2007; Rathvon, 1999).

Math Problem Solving

The Math Problem Solving cluster is composed of two tests, Test 2: Applied Problems and Test 13: Number Matrices. It provides a measure of mathematical knowledge and quantitative reasoning. Results of this cluster can be compared to those of the Math Calculation Skills cluster, to help determine whether or not limited basic math skills are affecting performance in math problem solving. In addition, the results of this cluster can be compared to those of the Oral Language cluster if the WJ IV OL is used, and to those of the Comprehension–Knowledge (Gc) and Fluid Reasoning (Gf) clusters if the WJ IV COG is used. Individuals with low oral language skills may have difficulty with quantitative terminology or math vocabulary. Individuals with low comprehension–knowledge may lack prerequisite knowledge for acquiring and identifying mathematical concepts. Individuals with low fluid reasoning may have difficulty identifying and thinking through the various steps of a mathematical problem. See Table 14.20 for suggested interventions.

Test 2: Applied Problems

Applied Problems requires the individual to analyze and solve practical math problems. It is a measure of quantitative reasoning, math achievement, and math knowledge. Applied Problems requires the construction of mental models (Johnson-Laird, Byrne, & Schaeken, 1992) to solve problems through the application of insight or quantitative reasoning. Solutions to these problems require access to complex cognitive processes and the calculation abilities that depend on them (Ashcraft, 1995). Because no reading is required, low performance will most likely be related to limitations in mathematical knowledge, attention, fluid reasoning, basic math skills, or oral language comprehension.

Difficulties with vocabulary, working memory, and perceptual speed can also impair performance on Applied Problems. Results on this test can be compared to those on reading and writing clusters, to determine whether the individual does better when no reading is required. Analyzing the individual's errors may help provide ideas for instructional planning. Interventions related to low performance on Applied Problems include schema-based strategy instruction (Fuchs & Fuchs, 2007; Xin, Jitendra, & Deatline-Buchman, 2005) and use of the concrete–representational–abstract (CRA) instructional sequence (e.g., Morin & Miller, 1998).

Test 13: Number Matrices

The task in Number Matrices requires the individual to look at a matrix of numbers, figure out the pattern, and then provide the missing number. This test requires a foundation in mathematics knowledge (i.e., access to the category-specific verbal and visual codes, such as knowledge of the number line). The solution to each item is obtained by mapping the relationship deduced from the completed part of the item onto the incompleted part of the item, thereby completing a serial relationship. Because the solution must work both ways (horizontally and vertically), a correct response must also produce a number matrix. It is a measure of deductive and inductive quantitative reasoning (aspects of fluid reasoning) and also requires attentional control in working memory and perceptual speed.

Low performance on this test may result from limited fluid reasoning, especially the narrow ability of quantitative reasoning. In addition, limited attention, working memory capacity, and perceptual speed may have an impact on performance. Related interventions involve explicit instruction in seriation and number reasoning skills (High/Scope Educational Research Foundation, 2003; Kroesbergen & Van Luit, 2003).

Other Math Clusters

There are two additional math clusters: Mathematics and Broad Mathematics.

Mathematics

The Mathematics cluster consists of two tests, Test 2: Applied Problems and Test 5: Calculation. The cluster does not include a timed test, so it provides a level of math development without the influence of speed.

Broad Mathematics

The Broad Mathematics cluster is composed of three tests—Test 2: Applied Problems, Test 5: Calculation, and Test 10: Math Facts Fluency. It provides a broad, comprehensive view of the individual's math achievement level. The Broad Mathematics cluster measures computational skill, automaticity with math facts, and problem solving and reasoning. Because this cluster measures three different aspects of math ability, interpretation of the cluster is most accurate when performance is similar on all three of the tests.

Written Expression

The Written Expression cluster includes two tests, Test 6: Writing Samples and Test 11: Sentence Writing Fluency. It provides a measure of meaningful written expression and automaticity with writing. Expressing oneself in writing is a highly complex task that requires executive functions, such as attention, working memory, planning, and self-regulating behaviors. Ineffective or inconsistent use of executive function capacities can affect any aspect of the writing process and may be at the core of many written language problems (Dehn, 2008; Hale & Fiorello, 2004; McCloskey, Perkins, & Van Divner, 2009). Multiple cognitive abilities are related to written expression, including auditory processing, long-term storage and retrieval, cognitive processing speed, crystallized intelligence, short-term working memory, and fluid reasoning (Flanagan, Ortiz, & Alfonso, 2013).

Comparing the results of this cluster to those of the Reading Comprehension and Basic Writing Skills clusters, or the WJ IV OL Oral Expression cluster, can be helpful in pinpointing the underlying cause of limits in written expression as well as determining instructional objectives. Low performance may result from low oral language skills or limited basic writing skills. See Table 14.20 for suggested interventions for each writing test.

Test 6: Writing Samples

The Writing Samples test requires the individual to produce meaningful written sentences in response to a variety of tasks. In several items, the individual must make bridging inferences in working memory to integrate the initial and final sentences into a well-formed passage. These items require planning, or tailoring the target sentence to the lexical and semantic information that is conveyed in the other portions of the sample (Ferreira, 1996). Low performance may result from limits in oral language, vocabulary, organizational ability, working memory, or spelling skill. Although errors in spelling are not penalized on this test unless the response is illegible, an individual's spelling difficulties may limit the quality of his or her written output. For example, the person's writing may be restricted to words he or she can spell. The individual's attitude toward writing may also influence performance. Weaknesses in key abilities required for writing may result in a poor attitude toward and avoidance of writing tasks, whereas strengths in those abilities may create a positive attitude and a willingness to complete writing tasks. Important qualitative information can be gained through a careful analysis of the individual's responses. Writing Samples results can be compared to results on measures of oral language, to see whether writing ability is commensurate with levels of receptive and expressive language ability.

Interventions for limited proficiency on Writing Samples include teaching of text structures (Englert, 2009) and use of strategy instruction, such as the self-regulated strategy development approach developed by Harris, Graham, and colleagues over the past two decades.

Test 11: Sentence Writing Fluency

Sentence Writing Fluency requires the individual to produce, in writing, legible, simple sentences with acceptable English syntax. This test measures the narrow abilities of writing ability and writing speed. It requires quick formulation of constituent structures, or fluency of combining words into phrases (Gazzaniga et al., 1998).

Low performance on this test may result from poor attention, poor motor control/handwriting, limited spelling or reading skills, slow cognitive processing speed, or a response style that interferes with performance (e.g., slow and accurate, fast and inaccurate, slow and inaccurate). The results on this test can be compared to those on other timed measures, to see if the person generally works slowly or has a particular difficulty with writing speed. Possible interventions related to limited performance on Sentence Writing Fluency include explicit instruction in the mechanics of writing (Graham, Berninger, Abbott, Abbott, & Whitaker, 1997) and use of technology (MacArthur, 1996).

Other Written Language Clusters

There are three additional clusters available in written language: Basic Writing Skills, Written Language, and Broad Written Language.

Basic Writing Skills

The Basic Writing Skills cluster consists of two tests, Test 3: Spelling and Test 14: Editing. It provides a measure of basic writing skills in isolated and context-based formats. Task demands include writing letters, spelling single words, and identifying and correcting errors in spelling, usage, capitalization, and punctuation. Because Editing requires the examinee to read the items and then identify and correct the error, the impact of the examinee's reading ability on performance must be considered. The results of this cluster can be compared with those of the Written Expression cluster to determine if basic writing skills are affecting performance. Mastery of basic skills is a fundamental part of complex, meaningful written expression. Error analysis of the two tests in this cluster may also help identify any problems related to spelling.

Test 3: Spelling. The Spelling test requires the individual to produce, in writing, single letters or words in response to oral prompts. Several factors that may influence performance include fine motor skill, handwriting, phonological coding, and orthographic coding. This test measures prewriting skills and spelling. A careful analysis of errors can often result in specific instructional recommendations.

Test 14: Editing. The Editing test requires the ability to identify and correct errors in punctuation, capitalization, spelling, and usage in short written passages read by the examinee. Low performance may result from limited instruction, lack of knowledge of writing conventions, failure to self-monitor or self-correct errors, or poor reading skill.

Written Language

The Written Language cluster contains two tests, Test 3: Spelling and Test 6: Writing Samples. Because it includes a measure of basic writing skills and a measure of written expression, this cluster may be used to estimate a person's general writing ability. This cluster does not include a timed test, so it provides a measure of written language that does not include the effects of speed.

Broad Written Language

The Broad Written Language cluster includes three tests—Test 3: Spelling, Test 6: Writing Samples, and Test 11: Sentence Writing Fluency. It provides a broad, comprehensive view of the individual's written language achievement. Task demands include spelling single-word responses, expressing ideas to various tasks, and writing simple sentences quickly. The results from this cluster can be compared to the results on the Broad Reading cluster and on the Oral Language cluster from the WJ IV OL, to help determine the impact of oral language and/or reading skills on written language performance.

Cross-Domain Clusters

The cross-domain clusters are made up of tests from different achievement domains. The seven clusters are Academic Skills, Academic Fluency, Academic Applications, Academic Knowledge, Phoneme–Grapheme Knowledge, Brief Achievement, and Broad Achievement.

Academic Skills

The Academic Skills cluster is composed of the three basic skill tests—Test 1: Letter–Word Identification, Test 3: Spelling, and Test 5: Calculation. It provides a general, basic skills achievement composite and can help determine whether the individual's level of basic skills is similar or variable across the three academic areas. The examiner should consider whether basic skills facilitate or inhibit the individual's performance. Low performance may suggest particular curricular adaptations, such as use of books on CD during reading time, or a calculator during math activities. Examinees with low performance in one or more areas often require direct, explicit instruction to improve the accuracy of skills.

Academic Fluency

The Academic Fluency cluster is composed of the three timed tests—Test 9: Sentence Reading Fluency, Test 10: Math Facts Fluency, and Test 11: Sentence Writing Fluency. It provides a general academic fluency composite and can help determine if the individual's level of automaticity with basic skills is facilitating or inhibiting academic performance. The examiner should note whether the individual's speed of performance is similar or variable across the three academic areas. For example, a person may work slowly on reading and writing tasks, but at an average rate on measures of mathematics. In many cases, low performance on this cluster when contrasted with measures of Gc, such as the Academic Knowledge cluster, may suggest a need for extended time or shortened assignments. Additional information can be obtained by comparing the Academic Fluency cluster to the Perceptual Speed (P) and the Cognitive Processing Speed (Gs) clusters in the WJ IV COG. This may help determine if any limitation in academic fluency is related to a cognitive function or if performance in one or more academic areas is slowed because of low basic skills.

Academic Applications

The Academic Applications cluster is composed of the three application tests—Test 2: Applied Problems, Test 4: Passage Comprehension, and Test 6: Writing Samples. It provides a general measure of an individual's ability to reason and apply academic knowledge. As with the other cross-domain clusters, the examiner should note whether the individual's performance on the tests in this cluster is similar or variable, and should also consider the impact of basic skills, fluency, and oral language proficiency when interpreting this cluster. Low performance may suggest a need for adjusting the difficulty level of the instructional materials. Individuals with low Academic Applications scores often need adjustment in the difficulty levels of instructional tasks.

Academic Knowledge

The Academic Knowledge cluster contains three tests—Test 18: Science, Test 19: Social Studies, and Test 20: Humanities. In the WJ IV ACH, these three tests were expanded to be full-length tests, rather than subtests as they were in the WJ III ACH. Because of this change, individual test scores are available in addition to the cluster score. Each of these tests measures acquired curricular knowledge, an aspect of comprehension–knowledge. The tasks require the individual to respond orally to questions, although early items require only a pointing response.

Low performance on any of these tests may result from limits in vocabulary, reading, or exposure to curricular areas, or from lack of experiences related to science, social studies, or humanities. This cluster can provide valuable insights into the individual's interests, as well as the level of crystallized intelligence. Because no reading is required, the results can help determine whether the individual's acquired knowledge base is affecting performance in the other academic areas. If the WJ IV COG has been administered, results from this cluster can be compared to those from the Comprehension–Knowledge (Gc) cluster. If the WJ IV OL has been administered, results can be compared to those from the Oral Language and Listening Comprehension clusters. In addition, the Academic Knowledge cluster score can be used as the ability score in an ability–achievement comparison procedure available within the WJ IV ACH. This comparison procedure is designed to help determine whether an individual's performance in reading, writing, and/or mathematics is consistent with his or her level of acquired knowledge.

Test 18: Science. The Science test measures an individual's knowledge of the sciences including anatomy, biology, chemistry, geology, medicine, and physics.

Test 19: Social Studies. The Social Studies test contains items testing an individual's knowledge of history, economics, geography, government, and psychology.

Test 20: Humanities. The Humanities test measures an individual's knowledge of art, music, and literature.

Phoneme/Grapheme Knowledge

The Phoneme/Grapheme Knowledge cluster is a combination of two tests, Test 7: Word Attack and Test 16: Spelling of Sounds. It may be used to evaluate the individual's proficiency in using phonics for reading and spelling. The tasks require the individual to decode (read) and encode (spell) pseudowords or nonsense words. Although they have no meaning, these words all use possible English spelling patterns (orthography). Low performance may result from poor phonological skills, limited orthographic skills, or both. A low score may also be attributable to poor or limited instruction. To help understand the reasons for an individual's low performance, the examiner should compare this cluster to the Auditory Processing (Ga) cluster in the WJ IV COG or the Phonetic Coding cluster in the WJ IV OL. The Phoneme/Grapheme cluster can be especially useful in documenting specific reading disabilities (dyslexia).

Test 16: Spelling of Sounds. The Spelling of Sounds test requires the individual to spell nonsense words that conform to conventional phonics and spelling rules. Both phonological coding and orthographic coding are measured by this test. Low performance may result from poor attention, poor phonological processing, poor orthographic awareness, or low phoneme–grapheme knowledge. An examiner can compare results from this test to results from the Word Attack test or various WJ IV OL tests (Segmentation, Sound Blending, and Sound Awareness), to help determine whether difficulties result from phonological problems or limited knowledge of phoneme–grapheme relationships. An analysis of errors can be particularly valuable on this test.

Brief Achievement

The Brief Achievement cluster is a combination of three tests—Test 1: Letter–Word Identification, Test 2: Applied Problems, and Test 3: Spelling. This cluster provides a quick screening of the person's levels of performance in reading, writing, and math that does not include the effects of performance under timed conditions

Broad Achievement

The Broad Achievement cluster is composed of nine tests in the Standard Battery. These nine tests are used to create the Broad Reading, Broad Mathematics, and Broad Written Language clusters. The purpose of the Broad Achievement cluster is to provide a general index of an examinee's academic proficiency and fluency with academic tasks. The cluster can be used to identify students with very limited or advanced performance levels across curricular areas. It is also helpful when a global view of an individual's overall performance across the various achievement domains is needed. The tests required for obtaining the Broad Achievement cluster also yield the four additional cross-domain clusters (Academic Skills, Academic Fluency, Academic Applications, and Brief Achievement) described previously.

Score Comparison Procedures

The WJ IV ACH provides two variation procedures and one comparison procedure. The primary purpose of the variation procedures is determination of relative strengths and weaknesses, whereas the primary function of the ability–achievement comparison procedure is prediction of achievement.

Intra-Achievement Variation

The intra-achievement variation allows an examiner to analyze an individual's academic test and cluster scores and to explore relative strengths and weaknesses. For example, a person may have a relative strength in math, but a relative weakness in reading. These variations within achievement can help determine the person's present educational needs. The examiner can calculate intra-achievement variations for three curricular areas (reading, mathematics, and writing) if Tests 1–6 have been administered. Additional tests and clusters can be added to this variation procedure. Individuals who have relative achievement strengths or weaknesses, such as significantly higher math skills relative to the average of all other achievement areas, exhibit an intra-achievement variation.

Academic Skills–Fluency–Applications Variation

Nine WJ IV ACH tests are required for this variation procedure: three in reading (Letter–Word Identification, Passage Comprehension, Sentence Reading Fluency), three in written language (Spelling, Writing Samples, Sentence Writing Fluency), and three in mathematics (Applied Problems, Calculation, Math Facts Fluency). The person's performance in skills, fluency, and applications is compared across the academic areas of reading, written language, and mathematics.

Three more clusters can be added to this variation procedure: Reading Rate from the WJ IV ACH, and Cognitive Processing Speed or Perceptual Speed from the WJ IV COG.

Academic Knowledge–Achievement Comparison Procedure

Because Academic Knowledge is a strong measure of acquired knowledge and does not require reading, writing, or math, it serves as a good predictor of academic ability. This procedure allows the examiner to determine whether current achievement levels are commensurate with the individual's store of acquired curricular knowledge. As described above, the Academic Knowledge cluster consists of Test 18: Science, Test 19: Social Studies, and Test 20: Humanities. The standard score for this cluster is used to calculate predicted scores for expected achievement. The individual's predicted achievement is then compared to his or her actual achievement. Examinees with predicted scores significantly higher than their actual achievement scores exhibit an ability–achievement discrepancy in one or more areas of achievement. Additionally, two clusters from the WJ IV OL (Phonetic Coding and Speed of Lexical Access) can be included in this comparison procedure.

SUMMARY

The WJ IV consists of three co-normed batteries, the WJ IV COG, the WJ IV OL, and WJ IV ACH, providing a comprehensive system for identifying an individual's cognitive, linguistic, and academic levels and determining relative strengths and weaknesses. There are 18 tests in the WJ IV COG, 12 tests in the WJ IV OL, and 20 tests in the WJ IV ACH. Interpretation of the WJ IV is based on contemporary CHC theory and an accumulation of knowledge about the structure of, and interplay among, cognitive abilities and neurocognitive functions—particularly in the domains of working memory and attentional control, memory consolidation and retrieval, and auditory intelligence. To enhance ecological validity, cognitive complexity is increased in several new tests and clusters.

The narrow abilities identified by CHC theory and external neurocognitive research combine to form a theoretical and conceptual basis for suggested links between the WJ IV tests and evidence-based educational interventions. The WIIIP operationalizes the link between WJ IV test results and instructional interventions, providing a seamless method for making evaluations instructionally relevant. Together, the WJ IV and the WIIIP represent a comprehensive evaluation system that is useful for diagnosis, eligibility determinations, and instructional planning.

NOTE

1. A "wrong turn at Albuquerque" is a catchphrase popularized in the Bugs Bunny™ cartoons; it humorously refers to an incorrectly perceived juncture (sometimes caused by trying to follow an overly complicated set of directions or a poor map) that lands a traveler in an unintended place. Bugs Bunny is a registered trademark of Warner Brothers Entertainment, Inc., Burbank, California.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Alloway, T. P. (2011). *Improving working memory: Supporting students' learning.* Thousand Oaks, CA: SAGE.

Anders, P., & Bos, C. (1986). Semantic feature analysis: An interactive strategy for vocabulary development and text comprehension. *Journal of Reading, 9,* 610–616.

Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator, 16*(4), 14–18, 44–47.

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology, 100,* 48–66.

Andrewes, D. (2001). *Neuropsychology: From theory to practice*. New York: Psychology Press.

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly, 31*(1), 42–52.

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin and Review, 14*(5), 919–924.

Arend, I., Colom, R., Botella, J., Contreras, M. J., Rubio, V., & Santacreu, J. (2003). Quantifying cognitive complexity: Evidence from a reasoning task. *Personality and Individual Differences, 35,* 659–669.

Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition, 1,* 3–34.

Ashcraft, M. H. (2002). *Cognition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225,* 82–90.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 355–370). New York: Plenum Press.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33,* 439–454.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology,* 8, 485–493.

Baker, S., Simmons, D., & Kame'enui, E. (1998). *Vocabulary acquisition: Synthesis of the research.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38,* 333–339.

Berninger, V., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). New York: Guilford Press.

Berninger, V., & Richards, T. (2010). Inter-relationships among behavioral markers, genes, brain and treatment in dyslexia and dysgraphia. *Future Neurology,* 5, 597–617.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37,* 391–403.

Blair, C. (2006). How similar are fluid cognition and general intelligence?: A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*(2), 109–125, discussion 125–160.

Blevins, W. (2001). *Building fluency: Lessons and strategies for reading success.* New York: Scholastic.

Brady, S. A. (2011). Efficacy of phonics teaching for reading outcomes: Indications from post-NRP research. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Individual differences in reading: Theory and evidence* (pp. 69–96). New York: Psychology Press.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nature Neuroscience, 2,* 370–374.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42,* 563–570.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, task switching, and working memory. *Developmental Neuropsychology, 19,* 273–293.

Bunge, S. A., Mackey, A. P., & Whitaker, K. J. (2009). Brain changes underlying the development of cognitive control and reasoning. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 73–85). Cambridge, MA: MIT Press.

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59,* 1691–1700.

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science, 24*(6), 440–445.

Caplan, D. (1992). *Language: Structure, processing, and disorders.* Cambridge, MA: MIT Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies.* New York: Cambridge University Press.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales.* Circle Pines, MN: American Guidance Service.

Carrow-Woolfolk, E. (1996). *Oral and Written Language Scales: Written Expression.* Circle Pines, MN: American Guidance Service.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language.* Circle Pines, MN: American Guidance Service.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Cattell, R. B. (1987). *Advances in psychology: Vol. 35. Intelligence: Its structure, growth, and action.* Amsterdam: Elsevier.

Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Fort Worth, TX: Harcourt Brace.

Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35,* 386–406.

Clay, M. M. (1991). *Becoming literate: The construction of inner control.* Portsmouth, NH: Heinemann.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders, 43*(1), 1–40.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Coltheart, M. (1980). Reading phonological recording and deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 197–226). London: Routledge & Kegan Paul.

Coltheart, M. (2007). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Malden, MA: Blackwell.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46,* 1029–1037.

Conti-Ramsden, G., Botting, N., & Farragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42,* 741–748.

Conti-Ramsden, G., & Hesketh, L. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language and Communication Disorders, 38,* 251–263.

Conzelmann, K., & Süß, M. H. (2015). Auditory intelligence: Theoretical considerations and empirical findings. *Learning and Individual Differences, 40,* 27–40.

Cornelissen, P., Bradley, L., Fowler, S., & Stein, J. (1994). What children see affects how they spell. *Developmental Medicine and Child Neurology, 36*(8), 716–726.

Cornelissen, P., & Hansen, P. (1998). Motion detection, letter position encoding, and single word reading. *Annals of Dyslexia, 48*(1), 155–188.

Cornelissen, P., Hansen, P., Hutton, J., Evangelinou, V., & Stein, J. (1998). Magnocellular visual function and children's single word reading. *Vision Research, 38*(3), 471–482.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of the mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70,* 1082–1097.

Cucina, J. M., Su, C., Busciglio, H. H., & Peyton, S. T. (2015). Something more than *g*: Meaningful Memory uniquely predicts training performance. *Intelligence, 49,* 192–206.

Dahlin, E., Stigsdotter Neely, A., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320,* 1510–1512.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Green, I. (1986). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

de Jong, P. F. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology, 170,* 75–76.

de Jong, P. F., Seveke, M. J., & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. *Journal of Experimental Child Psychology, 76,* 275–301.

De Koning, B. B., & van der Schoot, M. (2013). Becoming part of the story!: Refueling the interest in visualization strategies for reading comprehension. *Educational Psychology Review, 25,* 261–287.

Dehn, M. J. (2008). *Working memory and academic achievement: Assessment and learning.* Hoboken, NJ: Wiley.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Dollaghan, C. A., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition.* Minneapolis, MN: Pearson Assessments.

Durkin, D. (1993). *Teaching them to read* (6th ed.). Boston: Allyn & Bacon.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421–436.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20,* 19–36.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43,* 865–878.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Englert, C. S. (2009). Connecting the dots in a research program to develop, implement, and evaluate strategic literacy interventions for struggling readers and writers. *Learning Disabilities Research and Practice, 24,* 104–120.

Federation of American Scientists. (2006). Summit on educational games: Harnessing the power of video games for learning. Retrieved from *www.fas.org/*

gamesummit/Resources/Summit%20on%20Educational%20Games.pdf.

Ferreira, V. S. (1996). Is it better to give than to donate?: Syntactic flexibility in language production. *Journal of Memory and Language, 35,* 724–755.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Foreman, J., Gee, J. P., Herz, J. C., Hinrichs, R., Prensky, M., & Sawyer, B. (2004). Game-based learning: How to delight and instruct in the 21st century. *EDUCAUSE Review, 39*(5), 50–66. Retrieved from *www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume39/GameBasedLearningHowtoDelighta/157927.*

Fuchs, L. S., & Fuchs, D. (2007). Mathematical problem solving: Instructional intervention. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children?: The nature and origins of mathematical learning difficulties and disabilities* (pp. 397–414). Baltimore: Brookes.

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31,* 79–92.

Galda, L., & Cullinan, B. E. (1991). Literature for literacy: What research says about the benefits of using trade books in the classroom. In J. Flood, J. M. Jensen, D. Lapp, & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts* (pp. 529–535). New York: Macmillan.

Gathercole, S. E. (2004). Working memory and learning during the school years. *Proceedings of the British Academy, 125,* 365–380.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513–543.

Gathercole, S. E., & Alloway, T. P. (2007). *Understanding working memory: A classroom guide.* York, UK: University of York. Retrieved July 8, 2014, from *www.york.ac.uk/res/wml/Classroom%20guide.pdf.*

Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers.* London: SAGE.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children. *Journal of Memory and Language, 28,* 200–213.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966–979.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In S. J. Pickering (Ed.), *Working memory and education* (pp. 220–241). Burlington, MA: Elsevier.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887–898.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103–127.

Gazzaniga, M. S., Irvy, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind.* New York: Norton.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49,* 363–383.

Geary, D. C. (2007). An evolutionary perspective on learning disabilities in mathematics. *Developmental Neuropsychology, 32,* 471–519.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27,* 398–406.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy.* New York: Palgrave Macmillan.

Georgiou, G. K., & Parrila, R. (2013). Rapid automatized naming and reading. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 169–185). New York: Guilford Press.

Gerber, M. (1984). Techniques to teach generalizable spelling skills. *Academic Therapy, 20*(1), 49–58.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007-4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education. Retrieved from *http://ies.ed.gov/ncee/wwc/practiceguides.*

Gersten, R., Beckmann, S., Clarke, B., Foeggen, A., March, L., Star, J. R., et al. (2009). *Assisting students struggling with mathematics: Response to interventions (RtI) for elementary and middle schools* (MCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from *http://ies.ed.gove/ncee/wwc/publications/practiceguides.*

Gersten, R., Ferrini-Mundy, J., Benbow, D., Clements, D. H., Loveless, T., Williams, V., et al. (2008). Report of the task group on instructional practices. In *National Mathematics Advisory Panel: Reports of the task groups and subcommittees* (pp. 45–53). Washington, DC: U.S. Department of Education.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of the research. *Review of Educational Research, 71,* 279–320.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American, 267*(3), 110–117.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 69–78.

Graham, S., Berninger, V., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89,* 170–182.

Graves, M. F. (2000). A vocabulary program to complement and bolster a middle-grade comprehension program. In B. M. Taylor, M. F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp. 116–135). New York: Teachers College Press.

Graves, M. F., Juel, C., & Graves, B. B. (2004). *Teaching reading in the 21st century* (3rd ed.). Boston: Allyn & Bacon.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6,* 316–322.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36,* 129–151.

Griffin, S. (2004). Teaching number sense. *Educational Leadership, 61*(6), 39–42.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48*(5), 1229–1241.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin and Review, 22,* 1104–1110.

Hannon, B., & Daneman, M. (2014). Revisiting the construct of "relational integration" and its role in accounting for general intelligence: The importance of knowledge integration. *Intelligence, 47,* 175–187.

Hardiman, M. M. (2003). *Connecting brain research with effective teaching.* Lanham, MD: Rowman & Littlefield.

Hayes, F. B., Hynd, G. W., & Wisenbaker, J. (1986). Learning disabled and normal college students' performance on reaction time and speeded classification tasks. *Journal of Educational Psychology, 78,* 39–43.

High/Scope Educational Research Foundation. (2003). *Classification, seriation, and number.* Ypsilanti, MI: High/Scope Press.

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75,* 603–618.

Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6,* 165–185.

Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25,* 43–65.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 645–662.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108-446, 20 U.S.C. 1400 (2004).

James, W. (1890). *Principles of psychology* (2 vols.). New York: Holt.

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128,* 219–264

Jensen, A. R. (2011). The theory of intelligence and its measurement. *Intelligence, 39,* 171–177.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Prepositional reasoning by model. *Psychological Review, 99,* 418–439.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*(6), 853–873.

Kail, R. (1990). More evidence for a common, central constraint on speed of processing. In J. Enns (Ed.), *The development of attention: Research and theory* (pp. 159–173). Amsterdam: Elsevier.

Kail, R. (1991). Development of processing speed in childhood and adolescence. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 23, pp. 151–185). San Diego, CA: Academic Press.

Kail, R. (2003). Information processing and memory. In M. Bornstein, L. Davidson, C. L. M. Keyes, & K. A. Moore (Eds.), *Well-being: Positive development across the life course* (pp. 269–279). Mahwah, NJ: Erlbaum.

Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics, 20,* 303–314.

Kane, M. J., Brown, L. E., Little, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18,* 614–621.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Karp, K. S., & Voltz, D. L. (2000). Weaving mathematical instructional strategies into inclusive settings. *Intervention in School and Clinic, 35,* 206–215.

Katz, L. (1977). Reading ability and single-letter orthographic redundancy. *Journal of Educational Psychology, 69*(6), 653–659.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children—Second Edition*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement—Second Edition*. Circle Pines, MN: American Guidance Service.

Keith, T. Z., & Reynolds, M. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650.

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology, 27,* 1–25.

Klingberg, T. (2009). *The overflowing brain: Information overload and the limits of working memory* (N. Betteridge, Trans.). New York: Oxford University Press.

Kosslyn, S. M., Alpert, N. M., & Thompson, W. L. (1995). Identifying objects at different levels of hierarchy: A positron emission tomography study. *Human Brain Mapping, 3,* 107–132.

Kosslyn, S. M., & Smith, E. E. (2000). Introduction to higher cognitive functions. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 961–962). Cambridge, MA: MIT Press.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematical interventions for children with special educational needs. *Remedial and Special Education, 24,* 97–114.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2,* 129–160.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin, 140*(6), 1534–1555.

Leung, C. B., & Pikulski, J. J. (1990). Incidental learning of word meanings by kindergarten and first-grade children through repeated read aloud events. *National Reading Conference Yearbook, 39,* 231–239.

Levy, B. A. (1978). Speech processing during reading. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 123–151). New York: Plenum Press.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor: University of Michigan Press.

Lindholm, K., & Aclan, Z. (1991). Bilingual proficiency as a bridge to academic achievement: Results from bilingual/immersion programs. *Journal of Education, 173*(2), 99–113.

Logie, R. H., Della Sala, S., Laiacona, M., & Chalmers, P. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition, 24,* 305–321.

Lohman, D. F., & Lakin, J. (2011). Reasoning and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (2nd ed., pp. 419–441). New York: Cambridge University Press.

Lubinski, D. (2010). Neglected aspects and truncated appraisals in vocational counseling: Interpreting the interest–efficacy association from a broader perspective. *Journal of Counseling Psychology, 57,* 226–238.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 107–128.

Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331–353.

MacArthur, C. A. (1996). Using technology to enhance the writing processes of students with learning disabilities. *Journal of Learning Disabilities, 29,* 344–354.

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2010). Differential effects of reasoning and speed training in children. *Developmental Science, 14*(3), 582–590.

Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Brain plasticity and functional losses in the aged: Scientific bases for a novel intervention. In A. R. Moller (Ed.), *Progress in brain research: Vol. 157. Reprogramming the brain* (pp. 81–109). Amsterdam: Elsevier.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radix and hierarchical models of intelligence. *Intelligence, 7*(2), 107–127.

Martin, A. (1998). The organization of semantic knowledge and the origin of words in the brain. In N. Jablonski & L. Aiello (Eds.), *The origins and diversifications of language* (pp. 69–98). San Francisco: California Academy of Sciences.

Martin, A. (2009). Circuits in mind: The neural foundations for object concepts. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 1031–1045). Cambridge, MA: MIT Press.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mather, N., & Wendling, B. J. (2014a). *Examiner's manual: Woodcock–Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Mather, N., & Wendling, B. J. (2014b). *Examiner's manual: Woodcock–Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Mather, N., & Wendling, B. J. (2014c). *Examiner's manual: Woodcock–Johnson IV Tests of Oral Language.* Rolling Meadows, IL: Riverside.

Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology, 75,* 116–133.

McArdle, J. J. (2001). Growth curve analysis. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of social and behavioral sciences* (pp. 6439–6445). New York: Pergamon Press.

McAuliffe, C. (2003). *Visualizing topography: Effects of presentation strategy, gender, and spatial ability* (Doctoral dissertation, Arizona State University). *Dissertation Abstracts International, 64*(10), 3653A. (UMI No. 3109579)

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties.* New York: Routledge.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89,* 217–245.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York: Guilford Press.

McGrew, K. S. (2009). CHC theory and the Human Cognitive Abilities Project: Standing on the shoulders of the giants of psychometric intelligence research [Editorial]. *Intelligence, 37,* 1–10.

McGrew, K. S. (2012, September). *Implications of 20 years of CHC cognitive–achievement research: Back-to-the-future and beyond CHC.* Paper presented at the Richard Woodcock Institute, Tufts University, Medford, MA.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual: Woodcock–Johnson III Normative Update.* Itasca, IL: Riverside.

McGrew, K. S., & Wendling, B. J. (2010). CHC cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675.

McKeown, M. G., & Beck, I. L. (2004). Direct and rich vocabulary instruction. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 13–27). New York: Guilford Press.

Medina, J. (2008). *Brain rules.* Seattle, WA: Pear Press.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49,* 283–306.

Michas, I. C., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology, 12*(2), 147–164.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research and Practice, 22,* 47–57.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Moats, L. C. (2004). *Speech to print: Language essentials for teachers.* Baltimore: Brookes.

Moats, L. C. (2009). *Spellography for teachers: How English spelling works* (2nd ed., Vol. 3). Boston: Sopris West.

Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L., Carrillo, R., & Collazo, T., (2004). Working towards third space in content area literacy: An examination of everyday funds of knowledge and discourse. *Reading Research Quarterly, 39*(1), 38–69.

Morin, V. A., & Miller, S. P. (1998). Teaching multiplication to middle school students with mental retardation. *Education and Treatment of Children, 21,* 22–36.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8,* 161–168.

Mueller, S. T. (2002). *The roles of cognitive architecture and recall strategies in performance of the immediate serial recall task.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Nettelbeck, T. (1994). Speediness. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 1014–1019). New York: Macmillan.

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). Spatial development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1. Body and mind.* Oxford, UK: Oxford University Press.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell–Horn–Carroll-based assessment. *Psychology in the Schools, 47*(7), 621–634.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 51, pp. 45–100). San Diego, CA: Academic Press.

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*(3), 164–169.

Oberauer, K., & Kliegel, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601–626.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittman, W. W. (2000). Working memory capacity: Facets of a cognitive ability construct. *Personality and Individual Differences, 29,* 1017–1045.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory function predicts intelligence? *Intelligence, 6,* 641–652.

O'Brien, B. A., Wolf, M., Miller, L. T., Lovett, M. W., & Morris, R. (2011). Orthographic processing efficiency in developmental dyslexia: An investigation of age and treatment factors at the sublexical level. *Annals of Dyslexia, 61*(1), 111–135.

Ofiesh, N. S. (2000). Using processing speed tests to predict the benefit of extended test time for university students with learning disabilities. *Journal of Postsecondary Education and Disability, 14,* 39–56.

Pammer, K., Lavis, R., Hansen, P., & Cornelissen, P. L. (2004). Symbol-string sensitivity and children's reading. *Brain and Language, 89,* 601–610.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition.* San Antonio, TX: Author.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Pittelman, S. D., Heimlich, J. E., Berglund, R. L., & French, M. P. (1991). *Semantic feature analysis: Classroom applications.* Newark, DE: International Reading Association.

Posner, M. I., & DiGirolamo, G. J. (2000). Attention in cognitive neuroscience: An overview. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 623–631). Cambridge, MA: MIT Press.

Race, E. A., Kuhl, B. A., Badre, D., & Wagner, A. D. (2009). In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 705–723). Cambridge, MA: MIT Press.

Rack, J. P., Snowling, M., & Olson, R. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 28–53.

Ramani, G. R., & Siegler, R. S. (2005). *It's more than just a game: Effects of children's board games play on the development of numerical estimation.* Poster presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.

Rathvon, N. (1999). *Effective school interventions.* New York: Guilford Press.

Reynolds, M., & Niileksela, C. (2015), Test review: Woodcock–Johnson IV Tests of Cognitive Ability, *Journal of Psychoeducational Assessment, 33(4)* 381–390.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schrank, F. A., & Dailey, D. (2014) *Woodcock–Johnson online scoring and reporting system* [Online format]. Rolling Meadows, IL: Riverside.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV Cognitive Abilities assessment.* Hoboken, NJ: Wiley.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock–Johnson IV Tests of Oral Language.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2015a). *The WJ IV Gf-Gc composite and its use in the identification of specific learning disabilities* (Woodcock–Johnson IV Assessment Service Bulletin No. 3). Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2015b). *Woodcock–Johnson IV Tests of Early Cognitive and Academic Development.* Rolling Meadows, IL: Riverside.

Schrank, F. A., & Wendling, B. J. (2015a). *WJ IV interpretation and instructional interventions program.* Rolling Meadows, IL: Riverside.

Schrank, F. A., & Wendling, B. J. (2015b). *Manual and checklists: WJ IV interpretation and instructional interventions program.* Rolling Meadows, IL: Riverside.

Seidel, K. (2007). Social intelligence and auditory intelligence: Useful constructs? (Doctoral dissertation, Otto von Guericke Universität Magdenburg, Germany). Retrieved from *http://edoc2.bibliothek.uni-halle.de/hs/content/titleinfo/1229.*

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Sengpiel, F., & Hubener, M. (1999). Visual attention: Spotlight on the primary visual cortex. *Current Biology, 9,* 318–321.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and* *complete science-based program for overcoming reading problems at any level.* New York: Knopf.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72,* 116–141.

Short, D., & Fitzsimmons, D. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners—A report to the Carnegie Corporation of New York.* New York: Teachers College Press.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60,* 973–980.

Slowiaczek, M. L., & Clifton, C., Jr. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior, 19,* 573–582.

Squire, L. R., & Schacter, D. L. (Eds.). (2003). *Neuropsychology of memory* (3rd ed.). New York: Guilford Press.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology, 75*(4), 471–490.

Stanovich, K. E. (1991a). Conceptual and empirical problems with discrepancy definitions of reading disability. *Learning Disability Quarterly, 14,* 269–280.

Stanovich, K. E. (1991b). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26,* 7–29.

Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities, 27*(1), 34–50.

Swanson, H. L. (2001). Searching for the best model for instructing students with learning disabilities. *Focus on Exceptional Children, 34,* 1–15.

Swanson, H. L., Ashbaker, H. J., & Lee, C. (1996). Learning disabled readers' working memory as a function of processing demands. *Journal of Experimental Child Psychology, 61,* 242–275.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*(5245), 81–84.

Taylor, H. G., Lean, D., & Schwartz, S. (1989). Pseudoword repetition ability in learning-disabled children. *Applied Psycholinguistics, 10,* 203–219.

Temple, C. M. (1991). Procedural dyscalculia and number fact dyscalculia: Double dissociation in developmental dyscalculia. *Cognitive Neuropsychology, 8,* 155–176.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Oxford, UK: Oxford University Press.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385–398.

Unsworth, N. (2016). The many facets of individual differences in working memory capacity. *Psychology of Learning and Motivation, 65,* 1–46.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68–80.

Unsworth, N., & Engle, R. W. (2007a). Individual differences in working memory capacity and retrieval: A cue-dependent search approach. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in Honor of Henry L. Roediger, III* (pp. 241–258). New York: Psychology Press.

Unsworth, N., & Engle, R. W. (2007b). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both?: A direct test of the dual component model. *Journal of Memory and Language, 62,* 392–406.

van den Broek, P. (1989). Causal reasoning and inference making in judging the importance of story statements. *Child Development, 60,* 286–297.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition, 15,* 181–198.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269–1291.

Van Orden, G. C., & Goldinger, S. D. (1996). Phonological mediation in skilled and dyslexic reading. In C. H. Chase, G. D. Rosen, & G. F. Sherman (Eds.), *Developmental dyslexia: Neural, cognitive and genetic mechanisms* (pp. 185–223). Timonium, MD: York Press.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*(1), 53–70.

Vitevitch, M. S., & Goldstein, R. (2014). Keywords in the mental lexicon. *Journal of Memory and Language, 73,* 131–147.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 1–20.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Wallach, G. P., & Butler, K. G. (1994). Creating communication, literacy, and academic success. In G. P. Wallach & K. G. Butler (Eds.), *Language learning disabilities in school-age children and adolescents: Some principles and applications* (pp. 2–26). Boston: Allyn & Bacon.

Wang, P. L. (1987). Concept formation and frontal lobe function: The search for a clinical frontal lobe test. In E. Perecman (Ed.), *The frontal lobes revisited* (pp. 189–205). New York: IRBN Press.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition.* San Antonio, TX: Pearson.

Wendling, B. J., & Mather, N. (2009). *Essentials of evidence-based academic interventions.* Hoboken, NJ: Wiley.

Willis, J. (2008). *Teaching the brain to read.* Alexandria, VA: Association for Supervision and Curriculum Development.

Windsor, J., & Kohnert, K. (2008). Processing measures of cognitive–linguistic interactions for children with language impairment and reading disabilities. In M. Mody & E. R. Silliman (Eds.), *Brain, behavior, and learning in language and reading disorders* (pp. 135–160). New York: Guilford Press.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 387–407.

Wolfe, P. (2001). *Brain matters.* Alexandria, VA: Association for Supervision and Curriculum Development.

Wolff, P. H., Michel, G. F., Ovrut, M., & Drake, C. (1990). Rate and timing precision of motor coordination in developmental dyslexia. *Developmental Psychology, 26,* 349–359.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery—Revised.* Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *Journal of Special Education, 39,* 181–192.

Zhou, R., & Black, I. B. (2000). Development of neural maps: Molecular mechanisms. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 213–221). Cambridge, MA: MIT Press.

CHAPTER 15

The Cognitive Assessment System—Second Edition

From Theory to Practice

Jack A. Naglieri
Tulio M. Otero

In this chapter, we describe a suite of instruments to measure *planning, attention, simultaneous, and successive* (PASS) neurocognitive processes. This suite includes the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014a), the CAS2: Brief (Naglieri, Das, & Goldstein, 2014b), the CAS2: Rating Scale (Naglieri, Das, & Goldstein, 2014c) and the CAS2: Español (Naglieri, Moreno, & Otero, 2017). These measures extend the approach taken with the original CAS (Naglieri & Das, 1997a, 1997b) to accommodate a wider range of professionals of different qualification levels, as well as a wider array of contexts. These measures provide multiple ways to evaluate the four PASS neurocognitive processes, which we have most recently described in Chapter 6 of this book, and more thoroughly elsewhere (Naglieri & Otero, 2017). The strength of this theory as measured by these instruments lies in the fact that this approach dramatically changes the very concept of intelligence from that developed by the U.S. military in the early 1900s, and by others such as Alfred Binet, to a neurocognitive conceptualization. In practical terms, this means that the CAS2 battery, like the CAS before it, provides a way to do the following:

- Detect the neurocognitive variability that improves our understanding of individuals with, for example, specific learning disabilities (SLD), autism, and attention-deficit/hyperactivity disorder (ADHD) (Naglieri, 2012; Naglieri & Goldstein, 2011).
- Fairly assess neurocognitive abilities across races (Naglieri, Rojahn, Matto, & Aquilino, 2005) and ethnicities (Naglieri, Rojahn, & Matto, 2007).
- Equitably assess PASS across languages (Naglieri, Otero, DeLauder, & Matto, 2007; Naglieri, Taddei, & Williams, 2013; Naglieri et al., 2017; Otero, Gonzales, & Naglieri, 2013).
- Use PASS scores reflecting patterns of strengths and weaknesses for instructional planning and intervention (Naglieri & Otero, 2017; Naglieri & Pickering, 2010).

The strength of the CAS2 suite of measures is that it provides a way to evaluate neurocognitive abilities (i.e., basic psychological processes), using a specific theory (PASS theory) based on Luria's (1966a, 1966b, 1973, 1980, 1982) conceptualization of brain function (see Chapter 6, this volume, for further information). The subtests were explicitly designed to measure the four PASS constructs. In other words, the theory of human neurocognitive function came first; then ways to measure each construct were developed, tested, and validated for specific purposes. Importantly, this approach to test development provides clarity about interpretation of scores and what they mean. The theory defines the constructs that are represented by the test's scores. This approach is a unique way to develop a measure of ability because of its reliance on a specific neurocognitive theory that (1) measures a wider range of abilities than traditional IQ; (2) does not rely on verbal and quantitative test questions, which require knowledge rather than thinking; (3) provides a fair way to assess diverse populations; and (4) has instructional implications. These key differences have a considerable impact on the validity of the CAS2, as well as the utility of the information it yields.

DESCRIPTION OF THE MEASURES

The CAS and CAS2 represent ongoing efforts to operationalize the PASS theory. These efforts began on February 11, 1984, when J. P. Das and Jack A. Naglieri first decided to develop a different kind of intellectual ability test that would be based on Luria's conceptualization of the brain, easy to administer and score, straightforward to interpret, capable of detecting learning strengths and weaknesses, accessible to a wide variety of students, and helpful for teachers to maximize learning. At the initial meeting, it was determined that each subtest would have a clear correspondence to the theoretical framework now known as PASS. Development of the subtests was accomplished by following a carefully prescribed sequence of item generation, experimental research, test revision, and reexamination until the instructions, items, and other dimensions were refined over a series of pilot tests, research studies, national tryouts, and national standardizations. This process resulted in subtests that provided an efficient way to measure each of the processes (Das, Naglieri, & Kirby, 1994; Naglieri & Das, 1997b). Descriptions of the tasks that became the CAS and now the CAS2 subtests, and efforts to evaluate their practical utility and validity, are available in many published papers and several books (see Das et al., 1994; Naglieri & Das, 1997a; Naglieri & Otero, 2012, 2017). The most recent advancement in the assessment of PASS neurocognitive processes and related behaviors is represented by the various versions of CAS2 (see Figure 15.1), which can be used in different contexts and by different professionals.

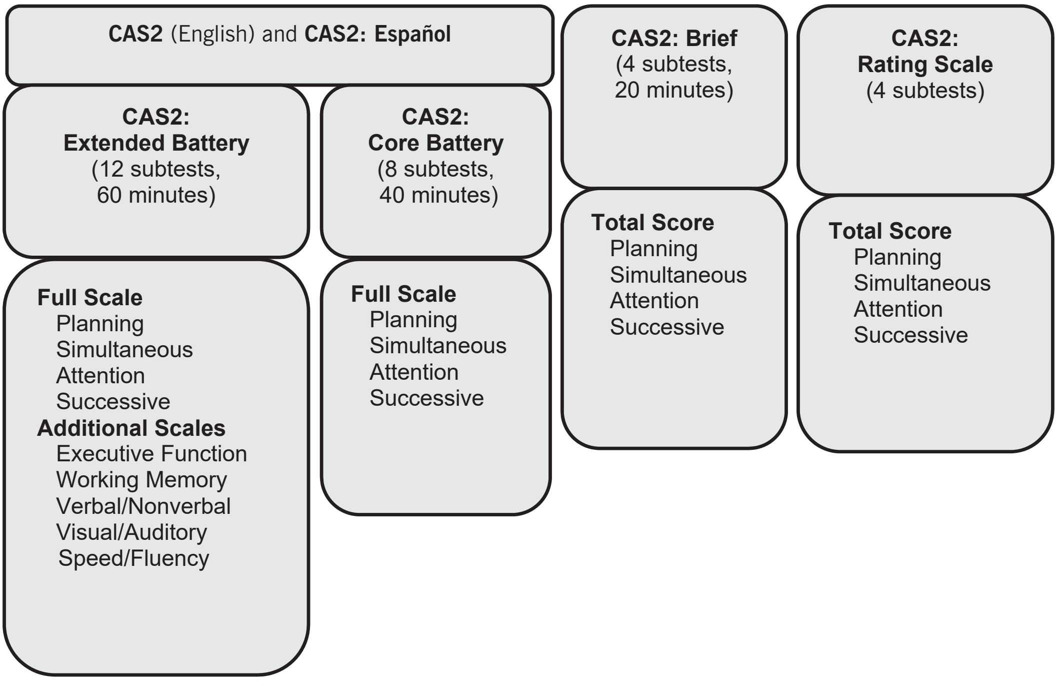


FIGURE 15.1. The CAS2 comprehensive system for assessment of PASS neurocognitive abilities and behaviors.

The CAS2 Suite as a Comprehensive System of Assessment

The CAS2, the CAS2: Brief, the CAS2: Español, and the CAS2: Rating Scale all provide scores for the four PASS neurocognitive processes. The PASS scale scores are combined into a Full Scale score (the CAS2 in English and the CAS2: Español) or Total Score (the CAS2: Brief and the CAS2: Rating Scale), both expressed as standard scores with a mean of 100 and standard deviation of 15. The PASS scores are computed on the basis of the subtest scaled scores included in each respective scale or the rating scale items for each neurocognitive ability. These scales represent a child's cognitive processing in specific areas and are used to examine cognitive strengths and/or weaknesses. It is important to note that the subtests vary in content (some are verbal, some involve memory, etc.), but each is an effective measure of a specific PASS process. In addition, the CAS2, CAS2: Brief, and CAS2: Rating Scale subtests are sometimes similar across the tests, but not identical, and the CAS2: Rating Scale's items focus on behaviors a teacher can observe. The subtests or items for these CAS2 measures of PASS are shown in Table 15.1 and more fully described in the sections that follow.

The CAS2

The CAS2 was designed to meet the need for a comprehensive evaluation of basic neurocognitive processes for children and adolescents ages 5 years, 0 months through 18 years, 11 months. Both the 8-subtest Core Battery (40-minute administration time) and the 12-subtest Extended Battery (60-minute administration time) provide a thorough evaluation of the four PASS constructs, using measures that vary by content, modality, and other demands. This individually administered test was normed on a sample of 1,342 cases representative of the U.S. population on several essential demographic variables (see the CAS2 interpretive and technical manual for more details). The two versions yield standard scores for the Full Scale and the Planning, Attention, Simultaneous, and Successive (PASS) cognitive processing scales. Subtest scores are reported as scaled scores, with a mean of 10 and a standard deviation of 3. The 12-subtest Extended Battery also includes scales to measure Working Memory, Executive Function (with and without Working Memory), Verbal Content, Nonverbal Content, Visual/Auditory, and Speed/Fluency. The CAS2 scales and their subtests are described more fully below.

Planning Scale

What the Scale Measures

The purpose of the Planning scale is to measure the student's ability to devise and apply strategies to solve novel problems (Naglieri & Otero, 2017). Planning is essential to all tasks where the examinee intentionally and independently decides to use some method to solve a problem. Better scores are obtained on these tasks if there is awareness of the need for a solution, monitoring for effectiveness of the chosen solution, consideration of alternative methods that might be appropriate, and evaluation of the relative value between continuing with the method or using a different one (Shadmehr, Smith, & Krakauer, 2010). Planning processing is also important when there is self-reflection on the results of a completed task, recognition of what worked and what did not work, and consideration of other possible solutions in the future. These unique human functions recruit activity within frontal lobes of the brain (Goldberg, 2009), as well as several neural networks (see Chapter 6, this volume).

How Planning Is Measured

To measure Planning, a test score *must* reflect how well a child has solved a novel problem for which there is no previously known strategy. To achieve this essential goal, the instructions for administration focus on *what* the student is to do, but the examiner provides no direction about *how* to complete the task. All of the Planning subtests on the CAS2, CAS2: Brief, and CAS2: Español (Naglieri et al., 2014a, 2014b, 2017) are solved via strategies selected by the examinee. The instructions explain the demands of the task, and the student is told to complete the task *in any way* that seems best. For this reason, the test scores reflect efficiency, measured by how long it takes to complete the task with the largest number of correct choices.

Planning Subtests (in Order of Administration)

- *Planned Codes*. The items on the Planned Codes subtest require the examinee to write specific codes (e.g., A, B, C, D to OX, XX, OO, XO) that correspond to letters (e.g., A, B, C, D) in empty spaces arranged in four rows and eight col-

TABLE 15.1. Subtests or Items Included in the Various CAS2 Measures

CAS2 Extended Battery (12 subtests)	CAS2 Core Battery (8 subtests)	CAS2: Español (12 subtests)	CAS2: Brief (4 subtests)	CAS2: Rating Scale (40 items)
		Planning		
Planned Codes	Planned Codes	Códigos planificados	Planned Codes	Planning items
Planned Connections	Planned Connections	Conexiones planificadas		
Planned Number Matching		Planificación de números pareados		
		Simultaneous		
Matrices	Matrices	Matrices	Simultaneous Matrices	Simultaneous items
Verbal–Spatial Relations	Verbal–Spatial Relations	Relaciónes verbales–espaciales		
Figure Memory		Memoria de figuras		
		Attention		
Expressive Attention	Expressive Attention	Attention expresiva	Expressive Attention	Attention items
Number Detection	Number Detection	Detección de numeros		
Receptive Attention		Atención receptiva		
		Successive		
Word Series	Word Series	Serie de palabras	Successive Digits	Successive items
Sentence Repetition/ Sentence Questions	Sentence Repetition/ Sentence Questions	Repetición de oraciones Preguntas a oraciónes		
Visual Digit Span		Retención visual de dígitos		

umns of letters without the codes. Students have 60 seconds per item to complete as many empty code boxes as possible. Each item is organized in a manner that requires a new strategy on each page.

- *Planned Connections.* The Planned Connections subtest requires the examinee to connect a series of numbers, alternating numbers, and letters of increasing length and difficulty in a specified order as quickly as possible within a time limit.
- *Planned Number Matching.* Each Planned Number Matching item presents the student with a page of eight rows with six numbers on each row. The student is required to find and underline the two numbers in each row that are the same within a time limit. This subtest is not included in the eight-subtest Core Battery.

Simultaneous Scale

What the Scale Measures

The Simultaneous scale measures how well a child or adolescent can combine separate stimuli into a conceptual whole (Naglieri & Otero, 2017). The spatial aspect of simultaneous ability involves perceiving stimuli as a group, seeing patterns in diagrams, recognizing a whole shape, and the formation of visual images. Simultaneous processing is also measured with items that require the comprehension of word relationships, prepositions, and spatial orientation. This ability is associated with the parietal, occipital, and temporal brain regions.

How Simultaneous Processing Is Measured

The distinguishing characteristic of the CAS2, CAS2: Brief, and CAS2: Español (Naglieri et al., 2014a, 2014b, 2017) subtests designed to measure simultaneous processing is that the correct answer to each item can only be obtained if the information that is provided is organized into a coherent whole. There is considerable variability in the content of the subtests, which illustrates the varied contexts in which simultaneous processing is used. For example, one of the subtests contains nonverbal visual–spatial content, while another is highly verbal; still another requires memory.

Simultaneous Subtests

- *Matrices.* The Matrices subtest utilizes shapes and geometric placements that are interrelated through spatial or logical organization. Students analyze the relationship among the parts of each item and solve for the missing part to choose the best of six options. This subtest is similar to other tests involving progressive matrices (e.g., the Naglieri Nonverbal Ability Test—Third Edition [NNAT3]; Naglieri, 2017).
- *Verbal–Spatial Relations.* Verbal–Spatial Relations is a six-item multiple-choice subtest in which each item consists of six drawings and a printed question at the bottom of each page. The examiner reads the question aloud, and the child is required to select the option that matches the verbal description. For example, the student may be asked, "Which picture shows a circle above a triangle to the left of a square under a triangle?"
- *Figure Memory.* This subtest requires a student to identify a familiar geometric shape (e.g., a triangle, a three-dimensional box) within a more complex design. The items are presented in a specific sequence. First, the examiner shows the student a geometric figure for 5 seconds. Then the stimulus is removed, and the student is asked to identify the original figure embedded in a larger, more complex geometric pattern. This subtest is not included in the eight-subtest Core Battery.

Attention Scale

What the Scale Measures

The purpose of the Attention scale is to determine how well a child or adolescent can focus on a specific stimulus and inhibit responding to distractions. The goal is to measure complex forms of attention, which involve "selective recognition of a particular stimulus and inhibition of responses to irrelevant stimuli" (Luria, 1973, p. 271). The ability to attend is demonstrated when a student can remain focused, selective, and effortful over time. Focused attention involves directed concentration toward a particular activity; selective attention provides the inhibition of responses to distracting stimuli; and sustained attention refers to the variation of performance over time, which can be influenced by the different amount of effort required to solve the test. Several brain structures, such as the reticular formation and the attention networks (see Chapter 6, this volume), allow an individual to focus selective attention toward a stimulus over time without the loss of attention to other competing stimuli.

How Attention Is Measured

Attention is measured in the CAS2, CAS2: Brief, and CAS2: Español (Naglieri et al., 2014a, 2014b, 2017) by subtests that demand focused, selective, sustained, and effortful activity. The subtest demands require the examinee to focus and resist distractions at different points. For example, in the Expressive Attention subtest, the most difficulty is experienced at the point of speaking. In contrast, the attentional demands of the Receptive Attention subtest are greatest when the stimuli are compared and a decision (match or no match) is made. The variety of task demands and content illustrates how attention can be measured in different ways.

Attention Subtests

- *Expressive Attention.* This subtest has two different versions. Children ages 5–7 years are given three pages consisting of seven rows that each contain six pictures of common animals, with each picture depicted as either big (1 inch × 1 inch) or small (½ inch × ½ inch). On each of the three pages, the student is required to identify whether the animal depicted is big or small in real life, ignoring the relative size of the picture on the page. On the first page, the pictures are all the same size. On the second page, big animals are depicted with big pictures, and small animals are depicted with small pictures. On the third page, the realistic size of the animal often differs from its printed size. For students ages 8–18 years the task is as follows: On page 1, students are asked to read the words BLUE, YELLOW, GREEN, and RED printed in black ink. On page 2, students are asked to name the colors of four colored rectangles (printed in blue, yellow, green, and red). On page 3, the four color words from page 1 are printed in the colors from page 2, but the word names and the colors do not match. In this item, students are required to name the color of the ink in which the word is printed, rather than to read the word.

- *Number Detection*. Each item presents the student with a page of 18 rows and 12 numbers per row. The task is to identify by underlining specific numbers on a page (ages 5–7 years) or specific numbers in a particular font (ages 8–18 years). The ratio of targets to distractors is carefully controlled by row and maintained across pages.
- *Receptive Attention*. This subtest consists of four sets of items, each containing 60 picture pairs (ages 5–7 years) or 180 letter pairs (8–18 years). The task is to underline pairs of objects or letters that either are identical in appearance or are the same from a lexical perspective (i.e., they have the same name). This subtest is not included in the eight-subtest Core Battery.

Successive Scale

What the Scale Measures

Successive processing is a neurocognitive ability that is used to work with information arranged in a specific serial order. This ability is required for recognition, recall, and reasoning when success on any task demands appreciation of a sequence. For example, successive processing is involved in the decoding of unfamiliar words, production of syntactic aspects of language, sequencing of non-automatized motor movements, and speech articulation. For this reason, successive processing is involved in activities such as phonological skills and the syntax of language (Das et al., 1994). This ability is associated with the temporo-parietal junctions of both the right and left hemispheres.

How Successive Processing Is Measured

Successive processing is measured on the CAS2, CAS2: Brief, and CAS2: Español (Naglieri et al., 2014a, 2014b, 2017) with tasks that vary in their content and complexity. Recall of words in order is a simple task, but repeating sentences made up only of color words or answering questions about the meaning of these statements demands appreciation of the syntactic relationships. The Successive scale's content also varies by modality; both auditory and visual stimuli are used. The use of various modalities illustrates that the content of the test is secondary to the basic cognitive processing demand.

Successive Subtests

- *Word Series*. This subtest utilizes nine single-syllable, high-frequency words: *book, car, cow, dog, girl, key, man, shoe,* and *wall*. The examiner reads aloud a series of two to nine of these words, at the rate of one word per second. The student is required to repeat the words in the same order as stated by the examiner.
- *Sentence Repetition or Sentence Questions*. The Sentence Repetition subtest, which is only administered to examinees ages 5–7 years, requires a child to repeat syntactically correct sentences containing little meaning, such as "The blue is yellowing." The Sentence Questions subtest is administered to examinees ages 8–18 years. For the items on this version, the child or adolescent answers a question by using words that are syntactically correct but contain little meaning. For example, the student is read the sentence "The blue is yellowing," and is then asked the following question: "Who is yellowing?"
- *Visual Digit Span*. The student is asked to recall a series of numbers in the order in which they were shown in the stimulus book. Items that are two to five digits in length are exposed for the same number of seconds as there are digits. Items with six digits or more are all exposed for a maximum of 5 seconds. This subtest is not included in the eight-subtest Core Battery.

CAS2: Español

The CAS2: Español is made up of the same 8- or 12-subtest versions as the CAS2 in English for students ages 5 years, 0 months through 18 years, 11 months. The test yields PASS and Full Scale scores, as well as additional scale scores to measure Working Memory, Executive Function, Verbal Content, and Nonverbal Content. We describe the translation and adaptation of the CAS2 into Spanish, as well as the rationale for this version, below.

Spanish Translation and Adaptation

The translation and cultural adaptation process was conducted by a team of psychologists from diverse geographical regions where Spanish is spoken and was organized by Mary Moreno, with Tulio M. Otero as consultant. The goal was to obtain equivalence between the original CAS2 and the Spanish version, based on a comprehensive cultural equivalence methodology for the translation and cultural adaptation proposed by Chávez and Canino (2005). This approach integrates the *emic–etic* perspective in an overall research methodology that is both culturally valid and generalizable. The *emic* perspective involves the evaluation of the phenomenon under study from within the culture and its context, while the *etic* perspective is basically comparative.

The back-translation method was used for the translation and adaptation of the CAS2: Español. In this method, a test is translated from English to Spanish, and then it is translated back from Spanish to English. This method was also used to translate the administration and scoring manual, the stimulus book, and the record form. The 12 CAS2 subtests were divided into two equal groups, and each group was assigned to a pair of translators. Each translator in each pair independently completed work on six subtests, and when they were finished, the two translators compared their results. They discussed the disagreements and, when necessary, consulted a translator on the other team. When they reached an agreement on the translation of their six subtests, one translator from each team also determined the consistency of the vocabulary used in the whole test. Once completed, the product was presented to two psychologists with broad experience in instrument translation, and they in turn checked for coherence between the English and Spanish versions.

The Spanish version of the CAS2 is more than a literal English–Spanish translation because it represents an adaptation of the original instrument. For example, the Word Series subtest in English is composed of one-syllable words that, when literally translated into Spanish, often contained more than one syllable. To ensure that the Spanish version of the subtest was not more difficult, one-syllable Spanish words familiar to 5-year-old participants were selected for this subtest. Similarly, modifications were made to the Sentence Repetition and Sentence Questions subtests to maintain the original intention of each subtest, so that the syntactic structure of the items was maintained. Finally, the vocabulary used in the Spanish version of the CAS2 took into account the diversity of concepts used by different Spanish-speaking cultural populations. For more information on the translation and adaptation of the CAS2 to the CAS2: Español, please see the CAS2: Español manual (Naglieri et al., 2017).

Rationale for the Spanish Edition

The CAS2: Español was developed to meet the need for assessment of children and adolescents with limited or no knowledge of English, and of English-language learners with limited cognitive academic language proficiency in both English and Spanish. Our goal was to provide a tool that can be effectively used with students who are learning English or who have limited educational opportunity. This goal is consistent with the position of Ceci (2000), Fagan (2000), Naglieri (2015), and Suzuki and Valencia (1997), all of whom believe that neurocognitive tests relying less on knowledge (vocabulary, information, etc.) are more appropriate for assessment of culturally and linguistically diverse populations.

There is evidence that the PASS neurocognitive processes can be assessed across language and cultures. Naglieri, Rojahn, and Matto (2007) examined the utility of the PASS theory with Hispanic children by comparing performance on the CAS of Hispanic and white non-Hispanic children from the standardization sample (N = 2,200). The study showed that the two groups differed by 4.8 standard score points when demographic differences were statistically controlled for. They also found that the correlations between achievement and the CAS scores did not differ significantly for the Hispanic and white non-Hispanic samples (Naglieri, Rojahn, & Matto, 2007). Similar results were reported for the CAS2 Full Scale scores in the test manual (Naglieri et al., 2014a). Without controls for demographic differences, Hispanics and non-Hispanics differed on the CAS2 Full Scale scores by 4.5 points; with such controls, the difference was 1.8 points. Very similar results are reported in the CAS2: Brief manual (Naglieri et al., 2014b). These findings suggest that measuring neurocognitive abilities rather than traditional IQ results in smaller differences between Hispanic and white non-Hispanic groups.

Other research involving Hispanics provided additional insights into the value of PASS theory as measured by the CAS. Naglieri, Otero, and colleagues (2007) compared PASS standard scores obtained on the CAS when administered in English and Spanish to bilingual children referred for reading problems. The children earned similar Full Scale scores on the English and Spanish versions of the CAS, and these scores were highly correlated (r = .96). As a group, this sample had a very low score on the Successive scale when given in English or Spanish. This finding was consistent with other research suggesting that children with reading decoding problems perform poorly in successive processing. Importantly, more than 90% of children who had a weakness on the English version of the CAS also had the same weakness on the Spanish version of the CAS. This study was followed by another (Otero et al., 2013) with a group of students referred for reading problems. Otero and colleagues (2013) found CAS Full Scale scores that differed by less than 1 point and, again, a high correlation between the scores (r = .94).

These findings are further supported by researchers who have used the PASS scales in other countries. Natur (2009) compared Arabic-speaking Palestinian students using the Arabic version of the CAS to a matched sample of children from the United States using the CAS in English. He found a very small difference between the Palestinian students' Full Scale mean of 101.0 and the U.S. Full Scale mean of 102.7 when the U.S. norms were used. Similarly, Naglieri and colleagues (2013) found that Italian children's (N = 809) Full Scale standard score of 100.9 on the Italian version of the CAS (Naglieri & Das, 2006) was very similar to the Full Scale standard score of 100.5 for a matched sample of U.S. children (N = 1,174) from the original standardization sample. Both samples' CAS standard scores were based on the U.S. norms. Importantly, multigroup confirmatory factor analysis (CFA) results supported the configural invariance of the CAS factor structure between Italians and Americans for the 5- to 7-year-old and 8- to 18-year-old age groups.

These research findings led us to conclude that although it would be optimal if scores for the CAS2: Español were based on separate norms collected on Spanish-speaking children in various countries and in the United States, the test can be used if appropriate actions are taken by the clinician. That is, the clinician should explicitly state that the scores obtained on the CAS: Español were obtained by comparing the child's performance to that of English-speaking students in the United States. We do expect, however, that this version of the CAS2 is likely to provide a good option for assessment of PASS processes. Of course, we anticipate more research on the utility of this approach to measuring these four neurocognitive processes.

There is, however, another important reason to use the CAS2: Español. The barrier posed by language-based measures of ability was the focus of findings in *McFadden v. Board of Education for Illinois School District U-46* (2013). This case involved Hispanic parents' concerns that their students were significantly underrepresented in gifted education classes. Over 40% of the students in the Elgin School District (U-46) were Hispanic, but only 2% of the students in the district's mainstream elementary school gifted program were Hispanic. The district required all students, including those who were learning the English language, to obtain high scores on a measure of ability that demands considerable verbal knowledge and quantitative skills (e.g., the Cognitive Abilities Test, Form 6 [CogAT6]; Lohman & Hagen, 2001), even if they had a very high score on a nonverbal measure of ability (the NNAT-2; Naglieri, 2008). The Court ruled that the use of tests requiring knowledge of English contributed to the underrepresentation of Hispanic students in gifted education. Judge Gettleman ruled that gifted children for whom English is a second language would be likely to score lower on a verbally demanding test than on the nonverbal, culturally neutral NNAT-2, which the plaintiffs' expert testified identified gifted students without a bias toward those students with higher English verbal skills. The judge further decided that the civil rights of the Hispanic students were violated in large part by the use of test questions demanding knowledge (traditional subtests like Vocabulary, Similarities, Information, and Arithmetic) to measure ability. The implications are clear: Using test questions that require knowledge should be avoided for those with limited verbal and quantitative skills. Because the language demands (particularly expressive language demands) made by the CAS2 are far less than those made by traditional tests of ability, and because the test is less reliant on previously acquired knowledge, it can be used in a manner consistent with fair and equitable assessment practices.

The CAS2: Brief

The CAS2: Brief is intended to be used as a short measure (administration time is 20 minutes) of PASS processes for children and adolescents. This version is appropriate for use in a prereferral evaluation of an initial concern, in a reevaluation, or in the decision-making process regarding educational programming.

The CAS2: Brief is a four-subtest, individually administered measure of PASS neurocognitive processing for students ages 4 years, 0 months through 18 years, 11 months. It was normed on a sample of 1,417 students representative of the U.S. population on a number of important demographic variables (see the examiner's manual for more details). The CAS2: Brief yields standard scores with a mean of 100 and a standard deviation of 15 for the four PASS scores and the Total Score. The subtests used in the CAS2: Brief are similar to, but not the same as, those used in the CAS2. Each PASS scale includes just one subtest, as described below.

Planning

The Planned Codes subtest is used on the CAS2: Brief to measure planning. This subtest, like its counterpart in the CAS2, contains four items, each with its own set of codes and particular arrangements of rows and columns. A legend at the top of each page shows which numbers (1, 2, 3, 4) correspond to different codes (a combination of O's and X's). The items differ from those in the CAS2 in the correspondence of numbers rather than letters and in the position of the numbers on the page. The test was designed to provide an examinee with the opportunity to complete the task more efficiently by using any strategy he or she may devise.

Simultaneous

The CAS2: Brief's Simultaneous Matrices subtest is a 44-item multiple-choice subtest composed of items different from those used in the CAS2 Matrices subtest. Simultaneous processing is measured with items composed of geometric shapes and designs that are interrelated through spatial or logical organization. The CAS2: Brief items have a different color scheme (yellow, blue, teal, and black) from those in the CAS2 (yellow, blue, and black). Like the CAS2, the items are composed like those in the NNAT-2 (Naglieri, 2008).

Attention

A modified version of the Expressive Attention subtest on the CAS2 is used to measure Attention in the CAS2: Brief. For children ages 5–7 years, the task is to identify the pictures of animals as either large or small, regardless of their relative size on each page. On the first page, all the animals are the same size. On the second page, the animals are sized relative to actual size. The item on the third page is the most sensitive to attention because the animals are usually sized opposite to their actual size. The orientation of the pictures is different from that in the CAS2, but the same pictures are used. The second set of items is administered to children ages 8–18. There are also three items in this version. First, the child reads words such as *blue* and *yellow;* second, the child identifies colors of a group of rectangles; and third, the child identifies the ink color in which words are printed. It is the last page that is sensitive to attention because the child must focus on one variable (specifically, the color), while inhibiting the more automatic response of word reading. All the stimuli are arranged differently for the CAS2: Brief than for the CAS2; the same images are used, but the color green is excluded in this version.

Successive

In the CAS2: Brief measure of successive processing, recall of numbers in order is used instead of recall of words (as on the CAS2). The Successive Digits subtest requires that the examinee repeat a series of numbers in the same order presented by the examiner. The numbers, which include numbers one through nine, are presented orally. There is a condition requiring the examinee to repeat the numbers in reverse order because, as found by Schofield and Ashman (1986), reversing the order of digits was equally correlated with both successive and planning processing.

CAS2: Brief and Identification of Gifted and Talented Students

Identification of gifted and talented students, like special education classification, has its own controversies that center on definitions and criteria for eligibility. These controversies are mostly attributable to the fact that there is no universally accepted definition of *gifted and talented,* which explains why nearly every U.S. state has its own definition. The National Association for Gifted Children (NAGC) defines *gifted* individuals as those students who demonstrate outstanding levels of aptitude, which can be explained as an exceptional ability to reason and learn. The NAGC defines *talented* individuals as those students with documented performance or achievement levels (e.g., mathematics, music, language) and/or sensory–motor skill levels (e.g., painting, dance, sports) in the top 10% of their age group. These definitions are similar to those proposed by Gagné (1985). According to Gagné, *giftedness* refers to natural abilities called *aptitudes* or *gifts* in at least one aspect of ability that fall in the top 10% of an individual's peer group, whereas *talent* indicates the superior mastery of knowledge and skills, as indicated by performance in the upper 10% of age peers. What is apparent in all these definitions is the view that gifted and talented students are those with a natural ability to perform above the 89th percentile, or those with the *potential* to perform at such a high level.

Given that identification of giftedness requires comparing a child's performance to that of a same-age norm group, the critical question becomes this: performance on which IQ test? A traditional IQ test actually assesses a mixture of ability (in those sections that *do not* require knowledge) and talents (in those sections that *do* require knowledge). Naglieri, Brulles, and Lansdowne (2008) and Naglieri and Ford (2015) proposed that gifted students with high ability can be identified with nonverbal tests that do not include verbal and quantitative measures, whereas talented students can be identified by measures of academic achievement. They also stressed that nonverbal tests help identify *all* gifted students, especially members of diverse populations that have been traditionally underrepresented (Ford, 2010, 2013). The importance of measuring ability without using verbal and quantitative tests found on IQ tests like the Wechsler scales or the CogAT6 (Lohman & Hagen, 2001) was recently tested in the *McFadden* legal case, as discussed above. We have taken the same position in our work on PASS theory. That is, measurement of intelligence is achieved more purely when the test questions have minimal knowledge demands.

Using PASS theory as measured by the CAS2: Brief to identify gifted students provides a way to measure cognitive ability rather than knowledge or achievement. For this reason, the CAS2: Brief meets the need for a short measure of ability that requires minimal verbal and math skills, but instead emphasizes neurocognitive abilities. Importantly, measuring cognitive constructs as defined in PASS theory would also result in the measurement of *more* abilities than would be assessed with a typical nonverbal test. For this reason, we suggest that the CAS2: Brief is a good tool for identification of gifted students, especially because it is appropriate for diverse populations of students. Of course, if a more comprehensive approach is warranted or desired, the CAS2 Core or Expanded Battery may be used.

Fair Assessment Using the CAS2, CAS2: Espanol, and CAS2: Brief

According to the Individuals with Disabilities Education Improvement Act of 2004 (IDEA, 2004), evaluators must use measures of intelligence that are considered nondiscriminatory (see Chapter 6, this volume), and fair assessment should also be viewed as a social justice issue. Padilla and Borsato (2008) explained that although tests can be technically sound, their content may privilege one group over another because knowledge is culturally embedded and the content of assessment tools may reflect the cultural values of the test developers. For this reason, Stobart (2005) argued that fairness or equity in assessment is fundamentally a sociocultural issue, and therefore equitable assessment is also a social justice issue (Gipps, 1999). Although test developers are now using more diverse norming samples that reflect the true diversity of the United States, obtaining more diverse samples is insufficient if the items on a measure are not appropriate for those diverse groups.

It is important that educators and parents understand the relationship of fair assessment to social justice in an often unjust educational system (as reflected in the *McFadden* case, described earlier). Equitable assessment is a form of social justice because a fair, appropriate assessment can prevent social problems and/or ensure access to resources (Mpofu & Ortiz, 2009). According to the *Standards for Educational and Psychological Testing* (American Educational Research Education [AERA], American Psychological Association, & National Council on Measurement in Education, 1999, as quoted in Mpofu & Ortiz, 2009, p. 57), "equitable assessment provides examinees an equal opportunity to display the requisite assessment processes, skills, and expectancies, as well as a fair chance to achieve the same level as others with equal ability on a given construct under measurement." The *Standards* also remind us that if a person has had limited opportunities to learn the content in, for example, a test of intelligence, that test may be considered unfair if it penalizes students for not learning content that is included in a test, even if the norming data do not demonstrate psychometric bias. Test developers are encouraged to use universal test design to ensure that "tests are as usable as possible for all test takers regardless of gender, age, language background, socioeconomic status or disability" (quoted in Mpofu & Ortiz, 2009, p. 57).

Poor test performance can have significant consequences because it is often used to make high-stakes decisions such as placement in special education, assignment to educational tracks, admission to college, and employment. The typical focus on finding a deficit is balanced when the emphasis is on understanding an examinee's current resources and assets in order to improve his or her well-being. In this way, assessment can embody a social justice perspective. We believe that the various CAS2 measures provide ways to assess children in a manner consistent with social justice tenets. These measures allow children's neurocognitive (PASS) assets as well as challenges to be determined, for the purpose of developing intervention plans that ultimately increase their access to the curriculum and to their own resources.

CAS2: Rating Scale

The CAS2: Rating Scale measures behaviors associated with the four PASS abilities for students ages 4 years, 0 months through 18 years, 11 months. The scale contains 40 items (10 for each PASS scale), which are rated by a teacher. It was normed on a sample of 1,383 students representative of the U.S. population on a number of key variables (see the examiner's manual for more details). Like the CAS2 and the CAS2: Brief, the CAS2: Rating Scale yields standard scores for the four PASS scales and a Total Score (mean of 100 and standard deviation of 15). The administration directions for the CAS2: Rating Scale are included in the record form. The CAS2: Rating Scale can be used whenever a teacher's evaluation of classroom behaviors related to PASS processes is desired. The information can be used for instructional planning, for screening, or (when paired with the CAS2: Brief or CAS2) as part of a larger evaluation.

Planning

The Planning scale is composed of items that ask about how well the student makes decisions about how to do things. These include questions that ask whether the student thinks before acting, controls impulses, devises new solutions when necessary, and determines the extent to which actions matched intentions. An example of a Planning scale item is "During the past month, how often did the child or adolescent . . . have a good idea about how to complete a task?"

Simultaneous

The Simultaneous scale includes items that describe how well the student understands interrelationships among ideas and objects. The questions determine whether a child can understand relationships among physical objects as well as verbal concepts. An example of a Simultaneous scale item is "During the past month, how often did the child or adolescent figure out . . . how parts of a design go together?"

Attention

The Attention scale items evaluate how well a student can focus attention and resist distractions, concentrate in noisy settings, and sustain attention over time. An example of an Attention scale item is "During the past month, how often did the child or adolescent work without getting distracted?"

Successive

The Successive scale evaluates the student's ability to work with information arranged in a specific order that includes remembering words, numbers, letters, and ideas in order, as well as blending sounds of words in the correct sequence. An example of a Successive scale item is "During the past month, how often did the child or adolescent follow three to four directions given in order?"

THEORETICAL BASIS OF THE CAS2 MEASURES

It is historically true that intelligence tests were not explicitly based on a specific theory of intelligence, which is why the definition "Intelligence is what intelligence tests measure" is so revealing. Neither Alfred Binet's (1890) nor David Wechsler's (1939) approach was theory-based. Instead, their tests were collections of tasks thought to have some relevance to the goal of categorizing people on a scale of ability, typically defined by the content of the tests (verbal, quantitative, and performance). The first test authors who intentionally developed an intelligence test with consideration of theory were Alan and Nadeen Kaufman when they published the Kaufman Assessment Battery for Children (K-ABC) in 1983. The first test authors to base a different kind of intelligence test on a specific neurocognitive theory were Naglieri and Das when they published the original CAS in 1997.

The CAS, like the CAS2, was unique in that it contained scales corresponding to Luria's (1973) view of neurocognitive abilities instead of the traditional verbal, quantitative, and nonverbal configuration. The goal was to provide a new way of defining ability based on a cognitive and neuropsychological theory, and to develop a test to measure these basic psychological processing abilities. The authors of the K-ABC and the CAS shared the intention to emphasize theory in the construction of the tests, and the view that this emphasis would stimulate an evolutionary step in the field of intelligence testing. Building a test on theory, rather than attaching theories or models to previously published tests, defines these "second-generation" approaches, and the research summarized in Naglieri and Otero (2017) illustrates the advantages.

In Chapter 6 of this volume, we summarize what we have learned about the validity of PASS theory, which leads us to suggest that the theory has a very strong research foundation that continues to grow. Consider these examples:

- Individuals with distinct disabilities such as SLD and ADHD have specific PASS profiles (Naglieri, 2012; Naglieri & Goldstein, 2011).
- Researchers found only small differences between PASS scores across races (Naglieri, Rojahn, et al., 2005) and ethnicities (Naglieri, Rojahn, & Matto, 2007).
- Only small differences between PASS scores have been found when the tests are administered in English or Spanish (Naglieri, Otero, et al., 2007; Otero et al., 2013).
- Only small differences between PASS scores were found when the Italian and English versions of the CAS were compared (Naglieri et al., 2013).
- PASS scores have shown to be strongly correlated with academic achievement scores (Naglieri & Rojahn, 2004).
- PASS scores have been successfully used for instructional planning and intervention (Naglieri & Conway, 2009; Naglieri & Feifer, in press).

ADMINISTRATION AND SCORING

Administration

The directions for administration are provided in the administration and scoring manual for the CAS2 (Naglieri et al., 2014a) and the CAS2: Español (Naglieri et al., 2017), and in the examiner's record form for the CAS2: Brief and the CAS2: Rating Scale. The instructions include general administration guidelines and scoring rules, as well as verbal statements and nonverbal actions to be used by the examiner. The combination of verbal and nonverbal communication is designed to ensure that all children understand each task.

The subtests on the CAS2, CAS2: Brief, and CAS2: Español are carefully sequenced to ensure the integrity of the subtests and to reduce the influence of extraneous variables on a student's performance. The Planning subtests are administered first because the child is given flexibility to solve the subtest in any manner. Attention subtests are presented after the Planning subtests because they must be completed in the prescribed order (i.e., left to right, top to bottom).

All Planning subtests include strategy assessment, which is conducted during and after the administration of each Planning subtest in two ways. First, observed strategies are those seen by the examiner through carefully watching the child completing the items; reported strategies are obtained following completion of an item. The examiner obtains this information by asking, "How did you find what you were looking for?" or "Tell me how you did these," or the like. The child can communicate the strategies by either verbal or nonverbal (gesturing) means. Strategies are recorded in the "Observed" and "Reported" sections of the Strategy Assessment Checklist included in the record form.

The administration directions were carefully written to ensure that each student fully understands the demands of the subtests. For example, sample and demonstration items are included, as well as opportunities for the examiner to clarify the requirements of the task. If, however, a child does not understand the demands of the subtest or appears in any way confused or uncertain, the examiner is instructed to provide a brief explanation if necessary. This instruction gives the examiner the freedom to explain what the child must do, in whatever terms are considered necessary to ensure that the child understands the task. The content of the help is determined by the examiner and can be given in any form, including gestures, as well as the use of a language other than English. However, it is important to remember that these alternative instructions are meant to ensure that the child understands what to do; they are not intended to teach the child how to complete the test.

Hand Scoring

The CAS2, CAS2: Brief, and CAS2: Español are scored via a standard conversion method of using the sum of the subtest raw scores to convert to subtest scaled scores; the subtest scaled scores are summed and converted to standard scores for the four PASS scales, the CAS2 Full Scale or Total Score (CAS2: Brief), or the Supplemental scales (CAS2). The CAS2 and CAS2: Español subtest raw scores are calculated in various ways: the total number correct (Nonverbal Matrices, Verbal–Spatial Relations, Figure Memory, Word Series, Sentence Repetition, Sentence Questions, and Visual Digit Span); time in seconds (Planned Connections); or ratio scores that combine time and number correct (Matching Numbers, Planned Codes, and Expressive Attention) or time, number correct, and number of false detections (Number Detection and Receptive Attention). *False detections* are defined as the number of times a child underlines a stimulus that is not a target. The examiner converts the raw scores to subtest scaled scores, using the appropriate table for the student's chronological age in years, months, and days. The PASS and Full Scale or Total Score standard scores are obtained from the sum of the subtests on either the Extended (12-subtest) or Core (8-subtest) Battery.

Online Scoring and Report Writing

Conversions of raw to standard scores and generation of a narrative report are made easier by use of the CAS2: Online Scoring and Report System (Naglieri, 2014). This online portal provides conversion of item level raw scores on the CAS2 and CAS2: Español to raw scores (and ratio scores) for each subtest; sums the subtest scores; and calculates all subtest scaled scores, as well as the PASS, Full Scale, and Supplemental scale standard scores. In addition, the system computes all the values needed to make comparisons among CAS2 scores for determining a pattern of strengths and weaknesses.

The CAS2: Online Scoring and Report System is a record-form-based program designed on the premise that computer scoring, interpretation, and report generation should be managed in a clear and easily understood environment. When the online program is opened, the first page that appears is one that looks very much like the first page of the CAS2 record form. Data entry requires the examiner to enter actual item data. For example, when time scores and number correct are entered for each Matching Numbers page, the program computes the ratio score, and then the sum of the ratio scores is calculated automatically. Alternatively, subtest raw scores can be entered directly on the front of the record form. Once the subtest raw scores are entered, the subtest scaled scores and PASS and Supplemental Scales' standard scores are computed, and analysis of the scores is completed. A text description of the results is provided, as are handouts for teachers and parents that describe the four PASS scales. The program performs all the interpretive tasks described by Naglieri and Otero (2017), and it also provides some of the intervention handouts included in the book *Helping Children Learn: Intervention Handouts for Use in School and at Home* (2nd ed.; Naglieri & Pickering, 2010).

PSYCHOMETRIC PROPERTIES

Standardization

The CAS2 scales were standardized, and norming was based on a total sample of 4,142 children and adolescents ages 4–18 years. The samples for the CAS2 (N = 1,342), CAS2: Brief (N = 1,417), and CAS2: Rating Scale (N = 1,383) were carefully chosen to match the 2013 U.S. census data on the variables of age, gender, geographic region, and ethnicity. Specific procedures used for data collection and norming, as well as details regarding representation by geographic region, gender, race, ethnicity, family income, educational level of parents, and educational classification of the three samples, are fully described in the respective manuals. Some important psychometric qualities are summarized below.

Reliability

The CAS subtests and scales have high reliability and meet or exceed minimum values suggested by Bracken (1987). The average reliability for the Extended Battery Full Scale is .97, and average reliabilities for the PASS scales are .88 (Planning and Attention) and .93 (Simultaneous and Successive). The Core Battery reliabilities are as follows: Full Scale, .87; Planning, .85; Simultaneous, .90; Attention, .84; and Successive, .90.

The average reliability coefficients for all age groups for the CAS2 measures are provided in Table 15.2 for the CAS2 subtests, Table 15.3 for all the scales on the CAS2, and Table 15.4 for the CAS2: Brief and CAS2: Rating Scale. The CAS2 subtest reliabilities range from .80 (Planned Connections) to .91 (Verbal–Spatial Relations) and illustrate that these subtests have good reliability overall. Importantly, when these subtests were combined into PASS scales, the reliability coefficients were very high, ranging from .84 (Attention) to .93 (Simultaneous) for the 8-subtest Core Battery and from .90 (Attention) to .94 (Simultaneous) for the 12-subtest Extended Battery. The Full Scale reliabilities were .95 and .97 for the CAS2 Core and Extended Batteries, respectively. The PASS scale and Total Score reliability coefficients for the CAS2: Brief were also high, ranging from .88 (Simultaneous) to .93 (Planning) with a .94 on the total score. The CAS2: Rating Scale reliability coefficients were high as well, ranging from .93 (Simultaneous) to .96 (Attention), with .98 on the Total Score. These reliability results were also similar across gender, race, ethnicity, and special education categories (see the respective CAS2 manuals).

TABLE 15.2. Coefficient Alpha Reliabilities for the CAS2 Subtests

Subtest	Alpha
Planned Codes	.88
Planned Connections	.80
Planned Number Matching	.82
Matrices	.88
Verbal–Spatial Relations	.91
Figure Memory	.85
Expressive Attention	.82
Number Detection	.80
Receptive Attention	.83
Word Series	.83
Sentence Repetition	.83
Sentence Questions	.85
Visual Digit Span	.86

TABLE 15.3. Reliability Coefficients for the CAS2 PASS Scales, Full Scale, and Supplemental Scales

CAS2 Scales	Alpha
Core Battery (8 subtests)	
Planning	.90
Simultaneous	.93
Attention	.86
Successive	.89
Full Scale	.95
Extended Battery (12 subtests)	
Planning	.92
Simultaneous	.94
Attention	.90
Successive	.92
Full Scale	.97
Supplemental scales	
Executive Function without Working Memory	.86
Executive Function with Working Memory	.91
Working Memory	.92
Verbal Content	.91
Nonverbal Content	.92

INTERPRETATION

Two Critical Questions

As emphasized throughout this chapter, the interpretation of the CAS2, CAS2: Brief, and CAS2: Rating Scale scores is based on the view that PASS scores are measures of neurocognitive processes or "thinking" processes. These processes are inferred through the performance of a child or adolescent

TABLE 15.4. Coefficient Alpha Reliabilities for the CAS2: Brief and CAS2: Rating Scale

Scale	Alpha for CAS2: Brief	Alpha for CAS2: Rating Scale
Planning	.93	.95
Simultaneous	.88	.93
Attention	.89	.96
Successive	.86	.94
Full Scale	.94	.98

on a particular test or behaviors observed by a parent. Later in this section, we explain how the scores from the CAS2 for the four types of thinking according to PASS theory can be understood when compared to achievement test scores, behavior ratings, and other measures of ability. A key requirement of this analysis, however, is understanding the PASS processing demand(s) of a test score from any measure. To do so, we need to ask two critical questions about a question from any test. First, what does the student have to know to answer the question (i.e., achievement)? Second, how does the student have to think to answer the question (i.e., PASS)?

Consider the test question provided in Figure 15.2. Solving the problem requires that the relationships among the shapes in the 2 × 2 matrix are understood. The examinee must understand that the difference between the left and right columns is shading, and that between the top and bottom rows is shape. The same kind of thinking is needed to solve a verbal analogy like "Man is to boy as woman is to ____________." That is, the relationships between the people must be understood, but in addition, the item requires *knowledge* of verbal concepts (*girl, woman, boy,* etc.). Furthermore, the examinee needs to understand that a girl becomes a woman and, similarly, a boy becomes a man. The relationships between the younger and older persons need to be comprehended to arrive at the correct answer. The same kind of thinking and knowledge components are involved when a student is asked, "What number comes next in this series: 1, 4, 7, __?" The relationships between the first and second numbers, as well as the second and third numbers, must be determined. In these last two examples, the student must know certain facts in order to understand the relationships among the words or numbers. In the example in Figure 15.2, in which only shapes are provided, the relationships among the shapes must be understood to answer the question, but knowledge of the names of the shapes is not needed. What is needed is *thinking about the relationships* between the shapes in the top and bottom rows. What distinguishes the different tasks is the fact that answering the verbal and math questions requires knowledge, whereas responding to the item presented with shapes requires very little knowledge. This distinction is the basis of interpretation of the CAS2 scores as well as scores from any other tests, especially tests of academic skills.

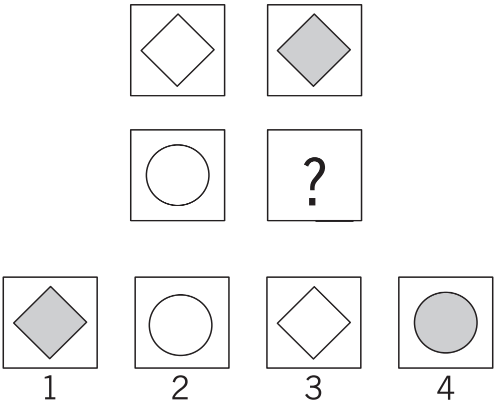


FIGURE 15.2. Illustration of a test item that demands simultaneous processing.

The cognitive demands of each CAS2 subtest on its respective PASS scale measure thinking (the PASS processes) with a minimum amount of knowledge. A person's CAS2 scores will often be compared, for example, to some test scores that may require a considerable amount of knowledge (e.g., reading, math, and writing). It is essential that examiners understand the relative contributions of knowledge and thinking to ability test scores when they are relating PASS scores to those from achievement tests and/or to reports of academic or behavioral problems in school or at home.

Overview of CAS2 Interpretation

In the remainder of this section on interpretation, we describe how to translate the PASS scores into theory-based explanation of a child or adolescent's neurocognitive processing abilities. We explain how to examine the scores for the CAS2, CAS2: Español, CAS2: Brief, and CAS2: Rating Scale from a normative basis (comparing the student to the standardization sample) and from an ipsative perspective (i.e., making comparisons within the individual's own set of scores). This analysis includes a method for determining whether there is a pattern of strengths and weaknesses that has relevance to educational programming, eligibility determination, and differential diagnosis, as well as intervention planning.

The CAS2 is unique among ability tests in that it was conceived in terms of and based upon the PASS neurocognitive theory, which provides the foundation for interpretation of the test scores. The focus of interpretation is on the child's ability to use the PASS neurocognitive concepts when thinking. Emphasis is placed on the four PASS scales' scores, which provide the most important information about the examinee. Indeed, we argue that the scores corresponding to the four PASS scores are *more* important than the Full Scale or the subtest scores. The CAS2 Full Scale (like the CAS2: Brief and CAS2: Rating Scale Total Scores) provides a very reliable overall description of a child's neurocognitive processing abilities, but it does not provide information regarding strengths or weaknesses according to PASS theory.

Interpretation of PASS profiles is the key to understanding a student's successes and difficulties in school and social/interpersonal situations, determining that student's eligibility for services, and planning appropriate interventions. Similarly, individual subtest score variability has little value, and efforts to reinterpret subtests from informally derived perspectives should be avoided. The purpose of having multiple subtests in the CAS2 is not to measure different abilities, but rather to provide highly reliable PASS scale scores by using subtests that measure each of the neurocognitive processes in a slightly different manner. For example, the Simultaneous subtests involve memory, verbal content, and reasoning with diagrams. The varying content of these complementary measures provides a broad evaluation of Simultaneous processing. When subtest scores are significantly different, they should only be interpreted when the additional data are compelling. Having one subtest score different from the others in that scale does not necessarily mean that the score for that scale is not interpretable; this phenomenon is more fully discussed later in this chapter. Even in such cases, the focus of interpretation of the CAS2 is on the PASS scale level.

Interpretation of PASS Scales

To be clear, the CAS2, CAS2: Español, and CAS2: Brief all measure planning, attention, simultaneous, and successive neurocognitive processing abilities in slightly different ways. These are briefly described as follows:

- *Planning.* This neurocognitive ability is used to create, apply, self-monitor, and self-correct thoughts and actions so that effective solutions to problems can be achieved. Planning provides the means to solve novel problems when no solution is immediately apparent, and often involves retrieval of information and use of the other PASS abilities to process information. All the CAS2 Planning subtests involve the use of strategies for efficient performance and the application of these strategies to novel tasks of relatively reduced complexity.
- *Attention.* This neurocognitive ability is used to focus selectively on a specific aspect of a complex stimulus while inhibiting responses to competing aspects. Successful performance on the Attention subtests requires attention to be focused, selective, sustained, and sometimes quite effortful. All CAS2 Attention subtests present tasks that require focus on one stimulus and resistance to responding to distractions over time.
- *Simultaneous.* This neurocognitive ability is used to understand how separate elements fit together into a conceptual whole. This ability is often applied to visual–spatial content, but can also be used to comprehend logical–grammatical verbal statements. Simultaneous processing subtests require perception of parts as a single whole, understanding of relationships, and synthesis of parts into integrated groups; all this occurs either through examination of the stimuli during the activity or through recall of the stimuli.
- *Successive.* This neurocognitive ability is used to integrate information into a specific order, where each element is only related to those that precede it. Successive processing involves working with stimuli in sequence, ordering thoughts and ideas, and forming sounds and movements in order. Some Successive processing subtests require recognition and reproduction of the serial order of information; others require comprehension of complex linear relationships.

The scores obtained with the CAS2: Rating Scale reflect behaviors a teacher can observe in the classroom that are associated with the PASS neurocognitive constructs. For planning, the goal is to estimate the extent to which behaviors associated with planning can be detected through observations of the strategies, organizational skills, and problem solving that a student may exhibit in the classroom. Attention is inferred from how well the student can focus on a task, resist distractions, and pay attention when dealing with everyday school activities. A student's ability to see relationships among concepts or physical objects, work with visual materials, and recognize faces can be informative about simultaneous processing. Finally, how well the student works with information in a sequence, remembers words or numbers in order, and can spell long words tells us about successive processing. It is important to note, however, that environmental factors can influence the PASS-related behaviors a student shows in the classroom. Instruction that discourages or encourages strategy use, for example, can influence classroom behaviors that would be related to planning.

Analytical Support for Interpretation Based on the PASS Scales

Interpretation of all the CAS2 measures thus begins with examination of the four PASS scores. This emphasis on the multidimensional theoretical level of analysis was supported by Naglieri and Rojahn's (2004) finding that the four PASS scale scores cumulatively accounted for more of the variance on an achievement test than the CAS Full Scale score in a sample of 1,559 students between the ages of 5 and 17. This study, in conjunction with those showing distinct PASS profiles for students with different disabilities (see Chapter 6, this volume), strengthens this emphasis on the separate neurocognitive abilities. Interpretation of PASS scores was also supported by Canivez (2011), who found that compared to subtests in other tests of intelligence, CAS subtests measured less general factor variance; the Planning, Attention, Simultaneous, and Successive factors had sufficient specific variance.

Recent CFA results are provided in the CAS2 and CAS2: Rating Scale manuals and provide support for focusing on the four PASS scales. The subtest-to-PASS-scale configuration was tested with individuals in the normative sample at four age intervals (ages 5–7, 8–10, 11–13, and 14–18 years), using maximum-likelihood CFA. Several models were tested, including models with one factor (PASS scales), two factors (Planning/Attention and Simultaneous/Successive), three factors (Planning/Attention, Simultaneous, and Successive), and four factors (Planning, Attention, Simultaneous, and Successive). The findings reported in the CAS2 manual indicated that for the four age groups, the four-factor PASS model was the best fit, thereby supporting the relative importance of the PASS scales versus the subtests (see the manual for more details).

Similar results were obtained from the analysis of the CAS2: Rating Scale. The correspondence of items to PASS scales on the CAS2: Rating Scale was studied by using CFA to test one-factor (no PASS scales), two-factor (Planning/Attention and Simultaneous/Successive), three-factor (Planning/Attention, Simultaneous, and Successive), and four-factor (Planning, Attention, Simultaneous, and Successive) models. The various fit statistics across the four different models improved as the number of factors increased, indicating that the four-factor PASS model was the best fit. These findings support the assignment of items to the four PASS scales. Perhaps most importantly, the results from these analyses of the assignment of *behavioral* items observed by teachers from the CAS2: Rating Scale and the correspondence of *subtests* from the CAS2 both support the structure of these measures based on the PASS theory. These results lend still more support to the interpretation of these tests at the PASS level, as well as the interpretation of the overall scores.

Steps for Interpreting the CAS2, CAS2: Brief, and CAS2: Rating Scale

Step 1: Interpreting the PASS Profile

The interpretation of the CAS2, CAS2: Brief, and CAS2: Rating Scale should begin with an examination of the four PASS scales and their associated confidence intervals, percentile ranks, and categorical labels. These derived scores are found in the standard score conversion tables in the CAS2, CAS2: Brief, and CAS2 Rating Scale manuals. Determining whether the PASS scores differ significantly is accomplished by comparing each PASS scale score with the average of the student's four PASS scores. This method is known as an *ipsative* comparison, which has been used often in intelligence testing (see Kaufman, 1994; Naglieri, 1999, 2011) because it can be used to determine whether any of the four PASS scores deviate from the student's level of functioning. This method thus provides a way to determine if the student's profile of neurocognitive processes is reliable. The values needed to use this approach for the CAS2, CAS2: Brief, and CAS2: Rating Scale are provided in the respective manuals and in Naglieri and Otero (2017).

However, the ipsative approach to determining whether any PASS scores differ significantly from the student's average is not sufficient to define a weakness or strength that can be used for diagnostic purposes (Naglieri, 1999; Naglieri & Otero, 2017). A second rule is needed for a better understanding of any high or low scores that may be found. For instance, a PASS score that is significantly lower than the person's average *must* *also* fall below the national average (at least below a standard score of 90) to be considered a weakness appropriate for eligibility determination or diagnosis. Table 15.5 provides several examples that are useful for using this approach.

Example 1 in Table 15.5 provides a scenario in which the Planning score of 84 is significantly lower than the student's average PASS score of 96.8, *and* that score falls below the average range. This meets our definition of a weakness because (1) it is low for this individual and (2) it is low in relation to the norm (i.e., the standardization sample). Similarly, the Simultaneous score is interpreted as a strength because it is significantly *above* the student's average and *above* the 84th percentile rank. This profile is often found for individuals who have been diagnosed with ADHD (Naglieri & Otero, 2012) and who lack control of their behavior and thinking. We suggest that scores above or below the average (90–109), or, to be stricter, scores corresponding to plus or minus one standard deviation from the average (standard score of 85/16th percentile and standard score of 115/84th percentile), be considered scores that are sufficiently unusual. There is no perfect cutoff score, so ultimately it is the clinician's decision.

Interventions for a student such as the one described in Example 1 should focus on having the student using the strength in simultaneous processing when learning, as well as on encouraging the use of strategies. The student can be taught that learning is most efficient when the big pic-

TABLE 15.5. Illustrations of PASS Profile Interpretations

PASS scale	Score	Difference from average score	Significant?	Weakness or strength?
		Example 1		
Planning	84	–12.8	Yes	Weakness
Simultaneous	116	19.3	Yes	Strength
Attention	95	–1.8	No	
Successive	92	–4.8	No	
Average	96.8			
		Example 2		
Planning	90	–1.5	No	
Simultaneous	103	11.5	Yes	
Attention	95	3.5	No	
Successive	78	–13.5	Yes	Weakness
Average	91.5			
		Example 3		
Planning	93	–13.5	Yes	
Simultaneous	104	–2.5	No	
Attention	111	4.5	No	
Successive	118	11.5	Yes	Strength
Average	106.5			
		Example 4		
Planning	86	–4.5	No	(Weakness)
Simultaneous	102	11.5	Yes	
Attention	95	4.5	No	
Successive	79	–11.5	Yes	Weakness
Average	90.5			

Note. Significance of the difference between each PASS score and the average PASS score for each case was tested at $p = .05$. A cutoff score of 85 was used.

ture is clear. The teacher can use handouts from Naglieri and Pickering (2010) that encourage the use of manipulatives, such as Cuisenaire Rods (pp. 114–115) for math, and that rely on simultaneous processing, such as the Summarization Strategy (p. 83) for reading comprehension. To encourage the use of planning processes, the Planning Facilitation (pp. 111–112) method for math and Plans for Reading Comprehension (p. 85) are good resources. The ultimate goal is to help the student use plans more frequently and to develop a repertoire of strategies that can be skillfully applied whenever needed.

Example 2 illustrates how a weakness can give the impression of a strength. In this case, the Simultaneous score of 103 is above the child's average PASS score by 11.5 points, and the difference is statistically significant—but because this score is within the average range (90–109), it would not be described as a strength. However, it would be appropriate to describe the Simultaneous score as the student's strongest area of ability, and to emphasize this neurocognitive ability when identifying or developing instructional methods that emphasize a big-picture perspective on information (e.g., Story Maps and Webbing, from Naglieri & Pickering, 2010). The Planning scale score of 90 is close to the average of 91.5, and the Attention score of 95 is 3.5 points above the child's average, indicating that neither of these is a strength or weakness. The Successive score of 78 is 13.5 points below the child's average, which is statistically significant *and* would be considered a weakness because the score is well below the average range. This finding indicates that the student has considerable difficulty working with information (words, numbers, motor movements, thoughts) arranged in a specific order. This profile is often found for students with SLD in reading decoding; they struggle in working with letters, sounds, and words arranged in a specific order (Naglieri & Otero, 2012).

Interventions for a student with a good score on the Simultaneous scale and an identified weakness on the Successive scale should focus on improving sequencing by organizing the information in groups—a strategy that brings in simultaneous processing. Naglieri and Pickering (2010) suggest using methods they describe, such as Chunking for Reading Decoding (p. 86) and Word Sorts for Improving Spelling (pp. 98–99), because they help a student learn to use simultaneous processing (the focus is on the bigger picture) to manage the sequence of information more easily. They also urge the student to recognize when a task requires successive processing (see the handout entitled Successive Processing Explained, p. 61) as a cue to use these intervention methods.

Example 3 illustrates that a relatively low score on the Planning scale (93), which is 13.5 points below the child's average but still within the average range, would not be considered a weakness. This does *not* mean that the score is unimportant, however: Because the Planning score is the lowest of the four scores, it has implications for the child's educational planning and self-esteem. Parents, teachers, and the student should be mindful that limited planning ability may be the reason the student does not perform his or her best when tasks demand developing and using strategies, organizing ambiguous tasks, and figuring out how to get things done. This suboptimal performance will be particularly apparent in contrast to tasks that demand getting the big picture (simultaneous), sequencing (successive), or sustaining focus and resisting distractions (attention). In fact, the Successive score of 118 is 11 points above the child's average and a strength, indicating that the student is likely to do well working in situations where the solutions are clearly spelled out in a logical, linear manner.

Interventions for a student with a strength on the Successive scale (above the PASS mean and the average range), and a score on the Planning scale that is at the bottom of the average range, should focus on the strength and encourage the student to be mindful about the value of strategy use. This is a student who needs to know that he or she may be prone to act impulsively at times, without careful consideration of the consequences. Teaching the student about the importance of being thoughtful and strategic will be very important, and handouts from Naglieri and Pickering (2010) such as Planning Explained (p. 55) and How to Be Smart: Planning (p. 63) should be utilized. The handout Successive Processing Explained (p. 61) should be used to inform the student of his or her strongest way to think, and to explain that seeing the whole (simultaneous) and focusing while resisting distractions (attention) are also important.

Example 4 presents a situation in which analysis of the PASS profile demands that rules be flexibly applied. The Simultaneous scale score of 102 is significantly above the child's average PASS score by 11.5 points—but as in Example 2, it would not be described as a strength because this score is within the average range. Nevertheless, as in Example 2, it would be reasonable to describe the Simultaneous score as reflecting the student's strongest ability and to recommend instructional methods that emphasize a big-picture perspective on information (e.g., Story Maps and Webbing, from Naglieri & Pickering, 2010). The Successive scale score of 79 is 11.5 points below the child's average, which is statistically significant; importantly, because the score is below the average range, this score is also considered a weakness. This weakness has educational and diagnostic implications. The Planning score of 86 is 4.5 points below the child's PASS average, and although not significantly different from the child's average PASS score, it is below the average range. In this case, it would be reasonable to interpret the Planning score as a weakness. The detection of Successive and Planning weaknesses like these has considerable implications for a student, especially in the early elementary grades, in which tasks that require sequencing are emphasized; when the student is unsure how to do something, he or she will have limited ability to use strategies to be successful. The intervention considerations for these processes described in the previous cases would also apply.

Step 2: Examining the Full Scale or Total Score

The next step is to interpret the Full Scale (CAS2) or Total Score (CAS2: Brief and CAS2: Rating Scale). The important question to consider with every case is this: Does the Full Scale or Total Score represent all four PASS scores? When significant variability in the four PASS scales is *not* found, the overall score can be considered an adequate description of the child or adolescent across the four scales. In contrast, when significant PASS score variability is found, it should be clearly stated that the Full Scale or Total Score will not be representative of all of the four PASS scales. In these instances, it is necessary simply to deemphasize the Full Scale or Total Score.

Step 3: Comparing the CAS2 or CAS2: Brief Scores with CAS2: Rating Scale Scores

Comparing scores from the CAS2 or CAS2: Brief with those obtained from the CAS2: Rating Scale allows for an examination of how PASS neurocognitive abilities can be contrasted to behaviors observed in the classroom. The values needed for significance when comparing the CAS2 and CAS2: Brief with the CAS2: Rating Scale are provided in Naglieri and Otero (2017). It is important to recognize that like any observational measure, the CAS2: Rating Scale is more influenced by the environment than PASS neurocognitive abilities measured by student performance on the CAS2 subtests. For this reason, PASS scores from the CAS2 and those from the CAS2: Rating Scale may differ because of the impact of environment, especially schooling. For example, a student may have good planning ability, yet may not demonstrate behaviors related to planning in the classroom if he or she has been taught to follow a specific method of solving a math question rather than to devise several ways of completing a problem. In this case, instruction reduces the role of planning and increases the need to remember the exact solution taught by the teacher. The underlying message to the student is to think less about possible ways of doing things and remember more of what is being taught (Meltzer, 2010). In such instances, the CAS2 or CAS2: Brief score may be different from the CAS2: Rating Scale score. This information is valuable, especially in devising interventions.

Step 4: Comparing PASS Scores to Achievement Test Scores

A key component of a comprehensive assessment is the comparison of scores on tests of cognitive ability and academic achievement. Comparing PASS scores to achievement test scores helps us understand if a cognitive processing weakness corresponds to an academic weakness, and if a cognitive processing strength corresponds to an academic strength. The statistical significance of the difference between scores on the CAS2 and a variety of achievement tests can be determined by using the same method as that for comparing CAS2 with CAS2: Rating Scale scores (Anastasi & Urbina, 1997). The values needed for significance provided in Naglieri and Otero (2017) were computed by using the standard errors of measurement reported in the technical manuals of the Kaufman Test of Educational Achievement, Third Edition (Kaufman & Kaufman, 2014), the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), the Woodcock–Johnson IV Tests of Achievement (McGrew, LaForte, & Schrank, 2014), the Feifer Assessment of Reading (Feifer, 2015), the Feifer Assessment of Mathematics (Feifer, 2016), and the Bateria-III Woodcock–Muñoz (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005). This method provides an efficient way to determine whether PASS strengths and weaknesses correspond to academic strengths and weaknesses.

Analyzing the Scores

The comparison of the CAS2 scores with achievement test scores should focus on the four PASS scales. The Full Scale should only be included if the four PASS scales are not significantly different and there is some requirement to contrast achievement with a total score. The emphasis, however, should be on comparison of the PASS scales to specific areas of achievement. To illustrate, if a child has earned scores of 90 on Planning, 102 on Simultaneous, 96 on Attention, and 77 on Successive on the CAS2 Extended Battery, these scores can be compared to subtest and composite scores from an achievement test. If the student has earned scores of 80 and 77 on the WIAT-III Pseudoword Decoding subtest and Basic Reading composite, respectively, these scores would be significantly lower (at $p = .10$) than the Simultaneous and Attention scores. However, the Pseudoword Decoding and Basic Reading scores would *not* be significantly different from the Successive score. The differences among the PASS scales (i.e., Successive score significantly below the child's average PASS score); the differences between the high ability scores (Simultaneous and Attention) in contrast to low achievement; and the similarity between the low Successive score and low achievement form the basis of the discrepancy–consistency model for SLD determination, which is described next. In this illustration, the CAS2 Full Scale score is not needed, nor does it help explain the relationship between PASS neurocognitive and achievement scores.

Determining Patterns of Strengths and Weaknesses: The Discrepancy–Consistency Method for SLD Identification

Naglieri (1999) first introduced the idea that a pattern of strengths and weaknesses in neurocognitive scores could be used as part of the process of identifying SLD. Similar approaches have also been suggested (Flanagan, Ortiz, & Alfonso, 2007; Hale & Fiorello, 2004). These authors' shared view that a pattern of strengths and weaknesses in cognitive ability scores can be used to identify SLD is sometimes referred to as a "third option" (Zirkel & Thomas, 2010). We propose that the discrepancy–consistency method should be used for the identification of SLD, based on a systematic examination of the variability of PASS scores from the CAS2 and academic achievement test scores. This method is based on evidence of one or more PASS weaknesses (as described in Step 1 above) and of variability in achievement test scores corresponding to PASS strengths and weaknesses. The results are two discrepancies and one consistency:

- A discrepancy between high and low PASS scale scores.
- A discrepancy between high PASS scores and low achievement test scores.
- A consistency between the PASS weakness(es) and low achievement test scores.

When these two discrepancies and a consistency are found, there is evidence that the child has what IDEA calls "a disorder in one or more of the basic psychological processes" involved in learning, and thereby that this student can be identified as having an SLD (Naglieri, 1999, 2005, 2011). An example of a decision tree for identifying a pattern of strengths and weaknesses, and making an SLD eligibility determination via the discrepancy–consistency method, is provided in Figure 15.3; a graphic representation of the findings is found in Figure 15.4.

Evidence of "a disorder in one or more of the basic psychological processes" referred to in IDEA is found in the successive processing weakness; when this evidence is combined with similarly low reading scores and with academic failure, eligibility for services is supported. This kind of a weakness in sequencing information underlies the inability to decode words successfully, as especially noted in the Pseudoword Decoding task. The consistency between the low Successive scale score and reading scores for this illustration provides evidence of a cause of the academic failure (assuming adequate instruction, motivation to learn, etc.). The significant difference between the high PASS scores and the two low reading scores further suggests that the student's achievement is below the ability to work with information that forms a whole, as well as to attend, shift focus, and resist distraction. This evidence, in conjunction with other relevant data and when other inclusionary–exclusionary conditions are also met, supports the student's eligibility for SLD services.

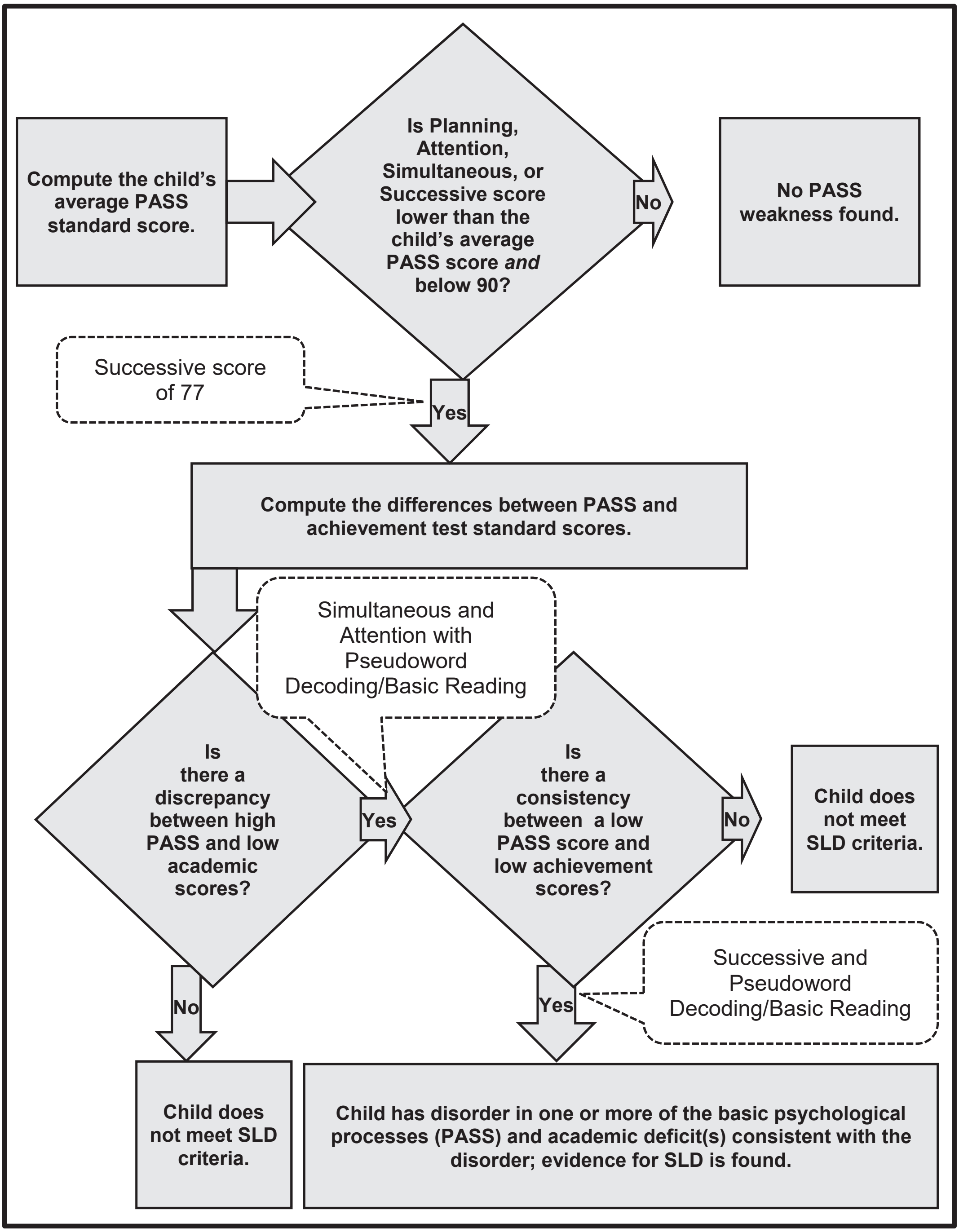


FIGURE 15.3. An example of using the discrepancy–consistency method for SLD eligibility determination.

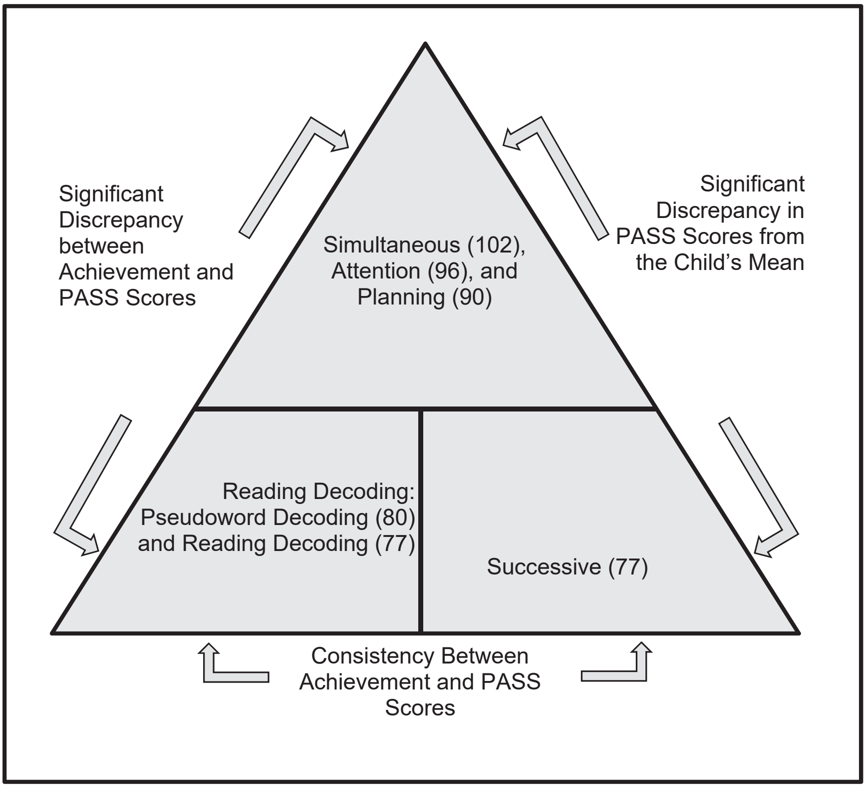


FIGURE 15.4. PASS and achievement scores for the discrepancy–consistency method.

Step 5: Examining the CAS2 Supplemental Scales

The CAS2 Extended Battery of 12 subtests includes Supplemental scales. These scales provide a way to integrate CAS2 results with measures of executive function, working memory, and speed/fluency, or of other concepts such as verbal–nonverbal and visual–auditory skills. These scores are thus useful for comparing results on the CAS2 to results obtained from other tests included in a comprehensive assessment battery. Even though our view is that PASS scores constitute the most valid way to interpret the CAS2, connecting this test's scores with other commonly used concepts is important. When doing so, practitioners should follow a few guidelines.

PASS scales represent the best way to understand the CAS2 subtests, which is why the theory drives interpretation. However, the Supplemental scales provide information that assists the clinician in incorporating other tests' data with CAS2 data during the interpretation. For example, it may be useful to relate findings from a rating scale of executive function, such as the Comprehensive Executive Function Inventory (CEFI; Naglieri & Goldstein, 2013), to some of the CAS2 subtests that also measure this construct. There are two different ways of looking at executive function, based on different combinations of CAS2 subtests; one CAS2 Supplemental scale includes, while another excludes, a measure of working memory.

Executive Function Scales

There is little agreement about the definition of executive function(s), the number of components it may have, and the role of working memory within it (Goldstein, Naglieri, Princiotta, & Otero, 2014). The two Supplemental scales on the CAS2 for assessing executive function are titled Executive Function without Working Memory and Executive Function with Working Memory. The Executive Function without Working Memory score is based on the scores obtained on the Planned Connections and Expressive Attention subtests, which are most closely aligned with the research on the executive functioning concept. The selection of these subtests is supported by Weyandt and colleagues (2014), who reported that the trail-making test (our Planned Connections) and the Stroop test (our Expressive Attention) are the most widely used measures of executive functioning. These tests measure two important components of executive function: shifting and inhibition (Georgiou, Das, & Hayward, 2008), respectively.

The Executive Function with Working Memory composite includes the Planned Connections, Expressive Attention, Verbal–Spatial Relations, and Sentence Repetition (for ages 5–7 years) or Sentence Questions (for ages 8–18 years) subtests. This combination of subtests adds the working memory concept, which is described by Baddeley and Hitch (1974) and is central to some theories of executive functioning.

The CAS2 Executive Function scales can be easily compared to the scores on the CEFI (Naglieri & Goldstein, 2013) because both measures use the same metric (i.e., mean of 100 and standard deviation of 15), where higher scores indicate better performance. The values needed for significance when comparing scores on CEFI and the CAS2 are provided in Naglieri and Otero (2017). This comparison will help determine whether a student's ability to think, devise a plan, apply the solution, evaluate its effectiveness, modify the plan as needed, and successfully complete the tasks on the CAS2 is consistent with behaviors observed by a parent or teacher or reported by the student. It is important to note that it is not assumed that these scores should be similar because behaviors will be greatly influenced by the environment within which the student functions. For example, if the student's educational experiences have consisted largely of explicit instructions in how to complete assignments, then the student's opportunity to think about different ways to solve problems may be inhibited. This possibility may explain why a child or adolescent with good scores on the Executive Function scales of the CAS2 has low scores on the CEFI.

Working Memory Scale

Georgiou and colleagues (2008) described *working memory* as the capacity to store information for a short period of time and manipulate that information using the *phonological loop* and *visuospatial sketchpad,* which are two concepts introduced by Baddeley and Hitch (1974). The visuospatial sketchpad refers to the person's use of mental images with visual and spatial features, and the phonological loop refers to retention of information from speech-based systems that are particularly important when order of information is required (Engle & Conway, 1998). The Verbal–Spatial Relations and Sentence Repetition or Sentence Questions subtests have cognitive demands similar to those of the visuospatial sketchpad and phonological loop, respectively, which is why these subtests were selected to comprise the Working Memory scale on the CAS2.

The CAS2 Working Memory scale can be compared to the Working Memory Index in the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014). The values needed for significance for testing any score differences are provided by Naglieri and Otero (2017). This comparison between scales that have the same name but are part of different tests must take into account the differences between the subtests used to measure this concept. For example, the CAS2 subtests that make up the Working Memory scale closely represent the concepts of the visuospatial sketchpad and phonological loop (Baddeley & Hitch, 1974; Engle & Conway, 1998). In contrast, the WISC-V Working Memory Index consists of the Digit Span and Picture Span subtests. We can assume that Digit Span Forward (like the CAS2 Sentence Repetition and Sentence Questions subtests) has a strong successive processing demand, but that Digit Span Backward involves successive and planning processes (Schofield & Ashman, 1986). The contribution of PASS processes to digit sequencing (reproducing the string of numbers in ascending order) was not evaluated by Schofield and Ashman, as this task postdates their original work. Because the task does involve mental manipulation and resequencing, we believe that this too would recruit successive and planning processes. It is not clear what processes are involved in the WISC-V Picture Span subtest, although we suspect that this test may involve some successive processing.

Verbal and Nonverbal Scales

Verbal Scores. The use of *verbal* and *performance* (i.e., nonverbal) test content as a way to define scales on measures of ability was initiated by Wechsler (1939) and based on the 1917 U.S. Army Alpha and Beta tests (Naglieri, 2015). The verbal (Army Alpha) tests were intended for those who had been educated, and the nonverbal (Army Beta) tests were intended for those who were illiterate. In fact, Yoakum and Yerkes (1920) stated that the purpose of the Beta (nonverbal) tests was "to avoid injustice by reason of unfamiliarity in English" (p. 19) and that the Alpha (verbal) tests would *not* measure intelligence reliably for those with a limited knowledge of English (p. 51). Not only did the Alpha and Beta tests set the stage for the Verbal and Performance scales on the Wechsler–Bellevue scale (Wechsler, 1939), but the U.S. Army's tests became embedded in both individual and group-administered tests of intelligence.

The WISC-V (Wechsler, 2014) still includes verbal and nonverbal subtests, although the publisher has created new scales and different names. The Verbal Comprehension Index is made up of subtests (Vocabulary and Similarities) intended to measure "verbal concept formation and abstract reasoning" and "word knowledge and verbal concept formation" (p. 7); "a child's ability to acquire, retain, and retrieve general factual knowledge" (p. 8); and "verbal reasoning and conceptualization, verbal comprehension and expression, the ability to evaluate and use past experience, and the ability to demonstrate practical knowledge and judgment" (p. 8). These subtests clearly require verbal skills and especially knowledge. It is important to recognize that verbal tests that contain considerable knowledge will be influenced by the student's opportunity to learn, just as Yoakum and Yerkes (1920) suggested.

The CAS Verbal scale differs from scales with similar names in traditional IQ tests. The Verbal scale on the CAS2 was created to measure thinking with verbal content, but with questions requiring as little factual knowledge as possible and very limited expressive language skills. The intent of the Verbal scale on the CAS2 is thus to measure verbal reasoning, but with less of a demand on language and on general knowledge. The CAS2 Verbal scale consists of Verbal–Spatial Relations, Receptive Attention, and Sentence Repetition (ages 5–7 years) or Sentence Questions (ages 8–18 years), which are derived from the Simultaneous, Attention, and Successive scales, respectively. This means that scores obtained on the CAS2 Verbal scale are likely to differ from those obtained on Verbal Comprehension measures found on the WISC-V and other traditional cognitive tests, especially for individuals with limited opportunity to acquire verbal knowledge.

The CAS2 Verbal scale demands the use of language across three of the four PASS scales and meets the objective described by Suzuki and Valencia (1997), as well as Naglieri and Bornstein (2003): to create a measure of ability with greatly reduced English demands and general factual knowledge demands, but one that still assesses thinking with language. Comparing the WISC-V Verbal Comprehension Index and the CAS2 Verbal scale, using the values presented by Naglieri and Otero (2017), allows the practitioner a way to determine if the child can work with the English language; by contrast, interpreting the WISC-V Verbal Comprehension Index score in isolation will reflect how much the child *knows,* as well as his or her ability to work with English.

Nonverbal Scores. Wechsler (1939) adapted the Army Beta tests and called his version the Performance Scale. That scale has been split into the Visual Spatial Index and the Fluid Reasoning Index in the WISC-V. These scales have also been described as nonverbal measures, which have gained popularity as the U.S. population has become more diverse (see Bracken & McCallum, 2009; Naglieri & Brunnert, 2009). These tests typically demand that the student solve problems using shapes that are arranged in a 2 × 2 or 3 × 3 matrix. The solution to each problem demands that the logic is understood in terms of changes in shape, color, and rotation.

The CAS2 Matrices, Figure Memory, and Planned Codes subtests constitute a Nonverbal scale that requires reasoning, has minimal language requirements, has a visual–spatial memory component, and provides the examinee with the opportunity to use strategies in solving a task. This scale involves simultaneous and planning processes. In addition to the two index scores mentioned above, the WISC-V has an ancillary Nonverbal Index that consists of six subtests: Block Design, Matrix Reasoning, Coding, Figure Weights, Visual Puzzles, and Picture Span. We suggest that this Nonverbal Index involves simultaneous processing (Block Design, Visual Puzzles and Matrix Reasoning), perhaps some successive processing (Picture Span), and some quantitative knowledge (Figure Weights). Practitioners should recognize that differences between scores on the CAS2 Nonverbal scale and the WISC-V Nonverbal Index may reflect the different cognitive and academic demands of the subtests in these scales.

Visual versus Auditory Subtest Comparisons

The idea that visual or auditory demands may have a significant impact on a student's performance has been studied since the early 1970s. Despite the limited support for interventions based on sensory modality (Dehn, 2014), interest in this area persists. Assessment of the difference in scores obtained from visual and auditory tests is typically confounded by differences in the intellectual requirements of the tasks. For example, WISC-V Block Design and Digit Span scores might be used when comparing visual and auditory stimuli, respectively, but this comparison is confounded by differences in the processing demands. That is, Block Design can be interpreted as a simultaneous processing task, while Digit Span has a strong successive processing requirement (Naglieri, Kamphaus, & Kaufman, 1983). This problem is circumvented in the CAS2 by comparing two subtests on the Successive scale.

Specifically, the CAS2 provides a means of directly comparing performance on tasks with auditory versus visual demands through evaluation of scores on the Word Series and Visual Digit Span subtests. Both these subtests demand the same neurocognitive ability (i.e., successive processing), but use visual (Visual Digit Span) versus auditory (Word Series) information. If the two subtests' scores differ by 3 points (one standard deviation), then the difference is statistically significant (see the CAS2 interpretive manual, Table B.3), and the difference in the modalities of stimuli may be an important factor to consider.

Speed/Fluency Scale

The concept of speed of processing, or how fast a person responds to highly learned stimuli, can be understood in terms of PASS theory by referring to Goldberg's (2009) description of the right and left hemispheres of the brain. In his book *The New Executive Brain,* Goldberg states that when a task is new, brain activity is maximized—especially in the frontal lobes, where planning takes place. When the task has been well learned, the functioning of the brain shifts from thinking about how to perform the task to solving the task with automaticity. Fluency is the result of the interaction of many factors, such as instruction, motivation, intention, and opportunity—but especially PASS abilities because they provide a foundation for learning. The transition from putting forth greater effort to putting forth less effort represents not only a change in hemispheric dominance from right to left (Goldberg, 2009), but also greater vertical organization of the brain. That is, as any task is learned, there is a shift to subcortical dominance, with greater involvement of the cerebellum. The cerebellum drives the speed, force, and accuracy of the expression of what is learned.

Speed/fluency can be measured with the first two pages of the Expressive Attention subtest on the CAS2. These initial two pages of Expressive Attention items require that the student respond to very well-known stimuli (either identifying the size of well-known animals, or reading the same words or naming the same set of basic colors) as quickly as possible. These items are used to prime the examinee for the final page that measures attention, but the scores on these pages are *not* used to measure attention. Naming sizes of animals, colors, or words requires speedy and fluent retrieval of knowledge, but little attention. Thus the first two pages of the Expressive Attention subtest are used to provide the Speed/Fluency scale score.

Step 6: Examining the Subtest Scores within Each of the Four PASS Scales

Determining whether there is variation among the CAS2 subtests is different for the 12-subtest Extended Battery and the 8-subtest Core Battery. The subtest scores from the Extended Battery on each of the four PASS scales are compared to the mean of the three subtest scores for each PASS scale. The differences are tested for significance via the ipsative comparison method (see Naglieri & Otero, 2017, for the values needed) *and* compared to the average subtest score range (8–12) via the same approach as that used for the PASS scales. That is, these differences need to be significant, *and* the subtest score must be below 8 (weakness) or above 12 (strength). These subtest score variations should be interpreted within the context of the PASS theory, with consideration of strategy use and other relevant variables.

Subtest variability, when interpreted within the context of PASS theory, may have potential value. For example, a child may have a low score on the Planned Connections subtest because an ineffective plan or no plan has been used. If the child has used strategies on the other two Planning subtests, the low score and observation of no plan would indicate that inconsistent strategy use is the problem. Similarly, a low score on Planned Codes may reflect a lack of self-correction if the child completes a page using a strategy that was appropriate on a previous page, without recognizing that the arrangement has changed. Similarly, a child with a very low Attention score may have problems shifting focus from letters to numbers as required on the Planned Connections subtest, resulting in a very low score on that subtest, despite average scores on the other two Planning subtests. In this example in which one subtest score on a PASS scale (Planning) is influenced by a problem with a different process (Attention), that low subtest score should not be included in the calculation of the overall Planning scale score, and a prorating approach should be applied.

A pairwise comparison method is used to examine subtest variation within each PASS scale for the eight-subtest Core Battery. The examiner should calculate the difference between the two subtests on each PASS scale (ignoring the sign) and compare the result to the values in Naglieri and Otero (2017) to determine when two subtests within each of the four PASS scales differ significantly. If the difference is equal to or greater than that found in the table, then the difference is significant. If a significant difference is found and a student has obtained a higher score on the Word Series subtest than the Sentence Repetition subtest, for example, the examiner look to see if the higher score is associated with a chunking strategy. If so, then it is likely that the use of a plan has influenced the results on the Successive scale. In such an instance, the difference between the two subtest scores illustrates the potential improvement on tasks requiring successive processing when a strategy is used. The best estimate of successive processing is the lower score, however, because it is less confounded by other processing abilities. This information is very useful for intervention: A teacher can simply encourage the use of plans when the student is completing work, especially work that demands successive processing (see Naglieri & Pickering, 2010).

Using the CAS2: Online Scoring and Report System

The CAS2: Online Scoring and Report System (Naglieri, 2014) provides an efficient way to convert all raw scores to derived scores; complete all scoring and comparisons of scales within the CAS2; and provide a narrative report describing each scale, the scores obtained, and what they mean. Item scores or subtest raw scores can be entered in the online program, using an interface that looks like the CAS2 record form. The report is provided as either a .pdf or a Word document. Two report formats are available: the Score Summary and the Scoring and Interpretive Report. The Scoring and Interpretive Report provides all the results that can be included within a comprehensive report; the user can modify and/or add information as needed.

INTERVENTION

One of the greatest strengths of the CAS2 is that this assessment approach gives a practitioner a way to understand how a student learns best (i.e., a PASS strength), and whether there are any obstacles to learning (i.e., a PASS weakness), so that a path to maximize learning is apparent. This information is most valuable when it is shared with the student, teachers, and parents. Fortunately, the four PASS neurocognitive abilities are easy to explain. One way to do this is to describe when PASS abilities are used, as follows:

"Planning is used when you think about how to do something before you act."

"Attention is used when you focus your thinking on something and resist distractions."

"Simultaneous processing is used when you think about how ideas or things go together."

"Successive processing is used when you must think of things in a specific order."

These four PASS neurocognitive processes underlie everything we do, and are most apparent in relation to academic success and difficulties. A PASS weakness can pose a substantial obstacle to learning (Naglieri, 2000). Conversely, if someone has a strength in one of the PASS processes, that strength can form the basis of success. It is essential that we communicate information about strengths and weaknesses to the students we assess, to maximize their likelihood of success in school and in life. This is an important initial step in the intervention process, but we begin here by explaining our view of intervention, and of how it differs from instruction.

We use the term *intervention* to indicate a specific way of teaching that is selected or developed with consideration of a student's PASS cognitive processing profile. *Instruction* is the application of any method of teaching, such as a phonics or whole-language curriculum, regardless of the learner's PASS profile. The use of an instructional method without consideration of the student's PASS and academic profiles is *not* an intervention. Ordinary instruction becomes an intervention when it is based on the results of an assessment that includes information about PASS processes and other relevant factors, such as mental health, previous educational history, home environment, and so on. The more informed we are about the characteristics of the student, the more efficient the selection of an instructional method will be, and the more likely it becomes that the intervention will be successful.

Informing the Student

The student who has been evaluated with the CAS2 is the most important person to inform about the results, in terminology appropriate for the student's age and ability to understand. When a comprehensive evaluation is conducted, we can expect that this student's difficulties at school have affected his or her self-concept. Just as success in school is often associated with being "smart," the lack of success can lead a student to doubt his or her ability. Thoughts such as "I am not very capable of learning" can lead a student to have limited expectations for success. Therefore, the first step in the intervention process is to inform the student of his or her PASS strengths as well as weaknesses, in a manner that is age-appropriate. Importantly, the examiner should clearly describe observed strengths, while also emphasizing that observed weaknesses can be managed with thoughtful effort. The goal is to change the student's attitude toward learning by helping him or her understand that the areas of strength can be used to overcome the observed weaknesses. This objective can be facilitated by handouts for students (as well as teachers and parents) provided by Naglieri and Pickering (2010), which explain each PASS processing ability as well as academic interventions. This strategy can also help change the student's mindset about the future, which is a key to acceptance of any interventions.

Teaching a student with a PASS processing disorder about his or her *strengths* and *needs,* and indicating which tools to use to address the learning needs, will empower him or her. Shifting the mindset from "I can't do this work," to "If I think smart, I know I can do better," takes time and requires a concerted effort on the part of all those working with the student. The evaluator should explain to the student that *mindset* is a description of the way a person thinks about his or her abilities, especially when tasks are demanding (Dweck, 2006). Students with a fixed mindset believe they cannot improve with effort, and they give up easily. Those students with a growth mindset believe they *can* achieve, with effort, persistence, and hard work. Helping not only the student, but his or her parents and teachers, adopt a growth mindset is important. When the evaluator is informing parents and teachers about the child's cognitive strengths and weaknesses, it is critical that the conversation encourage a growth mindset perspective. The goal is to ensure that the student and the adults adopt language like this: "I can't do it . . . *yet.* So I am going to keep trying until I succeed." For more information, and for two informal scales for rating mindset, see Naglieri and Otero (2017).

As noted above, the PASS handouts provided by Naglieri and Pickering (2010) can be used to inform a student about a strength or weakness in a PASS area. Each PASS "thinking ability" is described in simple terms, and the student is encouraged to "work smarter" by using a specific PASS way of thinking. For example, the theme of the handout on planning is "You can be smarter if you PLAN before doing things" (Naglieri & Pickering, 2010, p. 63). This handout is very important because it teaches that academic problems can be overcome if strategies are used, the success of strategies is evaluated, and modifications are made as needed. The message in the handout entitled Think Smart and Use a Plan! is that the student can achieve more than in the past by taking a strategic approach. This requires that the student learn to recognize when the demands of a task are particularly hard, and whether the obstacle is related to his or her PASS weakness. This suggestion is well supported by previous PASS research, which shows that encouraging students to use plans when doing math and reading has been very effective (Iseman & Naglieri, 2011)

When students with SLD in reading decoding have a weakness in successive processing (Naglieri & Otero, 2011), they especially need to use planning strategies to succeed. A student who has a weakness in successive processing should be informed that *any* task that demands sequencing will be difficult, but that using a strategy will make the task easier. Examples of such tasks include sequencing letters or sounds to make and spell words, remembering information in order, doing things in a specific order (e.g., tying shoelaces or remembering the combination to a lock), and so forth. One way to meet the demands of any sequencing task that requires reading or spelling is to put sounds and/or letters in groups. This chunking, or grouping, strategy is very helpful. These students should be given the handout Chunking for Reading Decoding (Naglieri & Pickering, 2010, p. 86), which teaches students how to use a chunking strategy for reading decoding instead of trying to sound out and blend sounds to make a word. Using strategies allows the student to arrive at the correct answer by *thinking* about how to solve the problem (using a plan), rather than by trying to decode words sequentially, which demands much successive processing. The change in the thinking reduces the successive processing demands of the task because seeing letters in groups reduces the length of the sequence and involves planning and simultaneous processes. The result is that by shifting the cognitive processing demands of a task, the child can read in a way that does not rely on his or her cognitive weakness (i.e., successive processing), and this shift improves the chances for success. More success means more confidence and the belief that a student can achieve by "thinking smarter."

Informing Teachers and Parents

The ultimate goal of the CAS2, CAS2: Brief, or CAS2 Rating Scale is to provide results that explain how children and adolescents learn and how to maximize their learning. An analysis of strength and/or weakness in any of the PASS neurocognitive abilities will provide information about any challenges to learning and about which instructional strategies or interventions are likely to be the most beneficial. It is important that the selection of teaching methods is based on thoughtful consideration of PASS demands, and that those PASS demands are aligned with the each student's profile. This task means that the teacher must understand the PASS processes required by the instructional lesson plan. For example, a child with a weakness in successive processing is likely to have problems learning from a reading program that demands blending sounds to read or spell words. Knowing this, the teacher can select methods that more efficiently match the learner's characteristics to the method of instruction. A critical part of this process is to examine the academic and PASS demands of any learning environment.

One way for teachers to understand the cognitive demands of any instructional method is to ensure that they understand the four PASS neurocognitive abilities. Naglieri and Pickering (2010) provide teacher handouts that describe the PASS neurocognitive abilities and their relationships to learning. Additionally, a teacher who has completed the CAS2: Rating Scale has already learned something about PASS through the questions included on the scale. When this information is combined with information obtained from the CAS2 or CAS2: Brief, the teacher can gain a greater understanding of the relationships between a student's PASS profile and the PASS demands of the academic tasks. Two important issues should be considered: (1) Most academic tasks involve more than one PASS ability, and (2) the roles of PASS processes may change as the task is learned.

Academic tasks like reading and math clearly require more than one PASS process (Naglieri & Rojahn, 2004), so the identification of a student's neurocognitive weakness helps teachers, parents, and the clinician understand where the learning breaks down. For example, a student with a weakness in planning ability may do poorly in reading because of a failure to consider all the possible meanings of the text. A student with a weakness in successive processing may not remember the order of events described in a paragraph and may arrive at an incorrect understanding of the text. Similarly, a student with a weakness in simultaneous processing may do poorly because of a difficulty in integrating all the relevant information into a cohesive whole to get the overall meaning. Finally, a student with a weakness in attention is likely to miss the details and get distracted by less relevant information, leading to a faulty understanding of the text. If both the teacher and the learner are aware of the student's PASS strengths and weaknesses vis-à-vis the demands of the academic task, then these obstacles can be anticipated and overcome. All this illustrates the importance of encouraging the student, teacher, and parents to work together to help the student succeed.

Intervention Options

There are several resources for using PASS theory to identify interventions that can augment learning for students with academic needs. These include books such as Naglieri and Pickering's (2010) *Helping Children Learn: Intervention Handouts for Use in School and Home*; book chapters by Naglieri and Feifer (2017) and Feifer and Naglieri (2017); and the PASS Reading Enhancement Program (PREP; Das, 1999). Other resources include, for example, *Learning Problems: A Cognitive Approach* (Kirby & Williams, 1991), *Cognitive Strategy Instruction That Really Improves Children's Academic Performance* (2nd ed., Pressley & Woloshyn, 1995), and *Helping Students Become Strategic Learners* (Scheid, 1993). Some of the options provided in these resources are now discussed in more detail.

Planning Strategy Instruction

The connection between planning as described in PASS theory and interventions to improve the use of strategies has been examined in a series of studies. These investigations have involved both math and reading; they have focused on the concept that children can be taught to be more strategic when they complete academic tasks, and that the facilitation of plans has a positive impact on academic performance. While it is true that some students need to be explicitly taught to use strategies, encouraging students to approach their work strategically and devise their own plans has the added advantage of near and far transfer (Iseman & Naglieri, 2010). For this reason, the *planning facilitation* method is designed so that children discover the value of strategy use without being specifically instructed to do so. The students are encouraged to examine the demands of a task in a strategic and organized manner. Research on this intervention method and its relationship to PASS scores on the CAS has been carefully examined in several important research studies.

The first two studies using planning strategy instruction showed that children's performance in math calculation improved substantially (Naglieri & Gottling, 1995, 1997). The children in these two studies attended a special school for those with learning disabilities. Students completed mathematics worksheets in sessions over about a 2-month period. The method designed to teach planning was applied in individual one-on-one tutoring sessions (Naglieri & Gottling, 1995) or in the classroom by the teacher (Naglieri & Gottling, 1997), two to three times per week in 30-minute blocks of time. Students were encouraged to recognize the need to plan and use strategies when completing mathematics problems during the intervention periods. The teachers provided probes that facilitated discussion and encouraged the students to consider various ways to be more successful. More details about the method are provided by Naglieri and Gottling (1995, 1997).

The relationship between planning strategy instruction and the PASS profiles for children with learning disabilities and mild mental impairments was studied by Naglieri and Johnson (2000). The purpose of their study was to determine whether children with cognitive weaknesses in each of the four PASS processes, and children with no cognitive weaknesses, would show different rates of improvement in math when given the same group-based planning strategy instruction. The findings from this study showed that children with a cognitive weakness in planning improved considerably over baseline rates, while those with no cognitive weakness improved only marginally. Similarly, children with cognitive weaknesses in simultaneous, successive, and attention processing showed substantially lower rates of improvement. The importance of this study was that the five groups of children responded very differently to the same intervention (see Table 15.6). Those with a weakness in planning on the CAS benefited the most from the intervention, while those with weaknesses in the other PASS processing domains or no weakness did not benefit as much. Thus the PASS processing scores were predictive of the children's response to this math intervention (Naglieri & Johnson, 2000).

The effects of planning strategy instruction on reading comprehension were reported by Haddad and colleagues (2003). Their study assessed whether an instruction designed to facilitate planning would have differential benefits on reading comprehension, and whether improvement would be related to each child's PASS scores. The researchers used a sample of children in general education, sorted into groups based on their CAS PASS scale profiles. Even though the groups did not differ on CAS Full Scale scores or pretest reading comprehension scores, children with a planning weakness benefited substantially (effect size of 1.4) from the instruction designed to encourage the use of strategies and plans. In contrast, children with no PASS weakness or a successive weakness did not benefit as much (effect sizes of 0.52 and 0.06, respectively). These results further support the research suggesting that the PASS profiles are relevant to instruction.

TABLE 15.6. PASS Standard Scores: Intervention Effect Sizes for Five Groups of Students with Learning Disabilities

Planning	Simultaneous	Attention	Successive	Effect size
69.7	89.0	86.3	91.7	1.4
86.5	92.5	**71.0**	101.5	0.3
88.0	88.0	103.0	**72.0**	0.4
93.0	**70.0**	90.3	89.7	–0.2
83.3	85.0	89.2	83.1	2.0

Note. Cognitive weaknesses are noted in boldface.

Iseman and Naglieri (2011) examined the effectiveness of the strategy instruction for students with learning disabilities and ADHD, who were randomly assigned to an experimental group or a control group that received standard math instruction. They found large pre–post effect sizes for students in the experimental group (0.85), but not those in the control group (0.26), on classroom math worksheets; they also found standardized test score effect sizes on Math Fluency (1.17 and 0.09, respectively) and Numerical Operations (0.40 and –0.14, respectively). One year later, the experimental group continued to outperform the control group. These findings strongly suggested that the students in the experimental group evidenced not only greater improvement than the control students on the math worksheets, but far transfer to standardized tests of math, and that they sustained these improvements at the 1-year follow-up. The findings also illustrate the effectiveness of planning strategy instruction, especially for those with low planning scores on the CAS.

PASS Reading Enhancement Program

PREP was developed as a cognitive remedial program based on the PASS theory. These researchers summarized studies showing that students could be trained to use successive and simultaneous processes more efficiently, with subsequent improvement in reading. PREP aims to improve the use of the PASS cognitive processing strategies that underlie reading. The tasks in the program teach children to focus their attention on the sequential nature of many tasks, including reading, which helps the children better utilize successive processing (a very important cognitive process in reading decoding). PREP is also founded on the premise that the transfer of principles is best facilitated through inductive, rather than deductive, approaches. The program is structured so that tacitly acquired strategies are likely to be used in appropriate ways. For example, the tasks teach children to focus on the sequences of information included in a variety of tasks, including reading.

Support for PREP, which has been summarized elsewhere (Naglieri, 2015), has demonstrated the effectiveness of the instructional method. Children (including children with learning disabilities) who received PREP, in comparison to a regular reading program, improved significantly on measures of nonsense-word decoding and word recognition. When PREP was compared to a meaning-based reading program in a study using two carefully matched groups of first-grade children, the results showed a significant improvement in reading scores for the PREP group over the control group. Specific relevance to the children's CAS profiles was also demonstrated by the fact that those children with a higher level of successive processing at the beginning of the program benefited the most from the PREP instruction, but those with the most improvement in the meaning-based program had higher levels of planning. Taken as a whole, these studies support the effectiveness of PREP in remediating deficient reading skills during the elementary school years. Additionally, they illustrate the connection between the PASS theory and intervention.

The interventions described here and earlier in this chapter (see "Step 1: Interpreting the PASS Profile," above) range from handouts for students, teachers, and parents to guidance in selecting various published reading and math programs. The correspondence of PASS strengths and weaknesses with published reading and math programs is discussed by Naglieri and Feifer (2017). For example, they suggest that programs such as the Wilson Reading System, Read 180, Orton–Gillingham, Ladders to Literacy, and Read to the Code can be used to improve the sequential processing of sounds. Programs such as Academy of Reading, RAVE-O, Reading Recovery, and Great Leaps Reading can be used to improve processing of sounds as groups. Teachers of students who need to make better use of planning and attention processing while reading should also consider approaches such as narrative retelling, active participation, and creating questions. These more formal programs and approaches offer the advantage that teachers may already be familiar with them or have the necessary materials. Even if formal programs are used, the handouts from Naglieri and Pickering (2010) should also be used to inform all parties of the meaning of a student's PASS scores and to build confidence that cognitive processing strengths can be used to maximize learning. For more on interventions in reading and math, see Naglieri and Feifer (2017) and Feifer and Naglieri (2017).

CONCLUDING THOUGHTS

This chapter has described the CAS2, CAS2: Español, CAS2: Brief, and CAS2: Rating Scale, which provide four ways to measure PASS neurocognitive theory (see Chapter 6, this volume). The greatest advantage of the various versions of the CAS2 is that they are grounded in the well-supported PASS theory. This theory helps us understand two key questions: Why does a student fail? And what neurocognitive strengths can be used to help the student succeed? This collection of measures also provides ways to apply PASS theory within a multi-tiered service delivery system—that is, in tier 1 (CAS2: Rating Scale), tier 2 (CAS2: Rating Scale and CAS2: Brief), and tier 3 (CAS2 or CAS2: Español and CAS2: Rating Scale). These four ways to evaluate basic psychological processes have been designed to address the need in the field to assess ability more broadly, more fairly, and more effectively, with a minimum of complexity. Which brings us to one final reminder: The test an examiner select has a profound impact on what the examiner learns about a student and on what can be done to help that student. We suggest that the decision be made with full knowledge of the strengths and weaknesses of all the options.

REFERENCES

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Binet, A. (1890). *On double consciousness: Experimental psychological studies* (No. 8). Chicago: Open Court.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*(4), 313–326.

Bracken, B. A., & McCallum R. S. (2009). Universal nonverbal intelligence test. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 291–314). Hoboken, NJ: Wiley.

Canivez, G. L. (2011). Hierarchical factor structure of the Cognitive Assessment System: Variance partitions from the Schmid–Leiman (1957) procedure. *School Psychology Quarterly, 23*(4), 305–317.

Ceci, S. J. (2000). So near and yet so far: Lingering questions about the use of measures of general intelligence for college admission and employment screening. *Psychology, Public Policy, and Law, 6,* 233–252.

Chávez, L. M., & Canino, G. (2005). Toolkit on translating and adapting instruments. Retrieved from *www.hsri.org/files/uploads/publications/PN54_Translating_and_Adapting.pdf.*

Das, J. P. (1999). A neo-Lurian approach to assessment and remediation. *Neuropsychology Review, 9*(2), 107–116.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *The assessment of cognitive processes: The PASS theory of intelligence.* Boston: Allyn & Bacon.

Dehn, M. J. (2014). *Essentials of processing assessment* (2nd ed.). Hoboken, NJ: Wiley.

Dweck, C. (2006). *Mindset: The new psychology of success.* New York: Random House.

Engle, R. W., & Conway, A. R. (1998). Working memory and comprehension. In R. H. Logie & K. J. Gilhooly (Eds.), *Working memory and thinking* (pp. 67–89). Hove, UK: Psychology Press.

Fagan, J. R. (2000). A theory of intelligence as processing: Implications for society. *Psychology, Public Policy, and Law,* 6(1), 168–179.

Feifer, S. G. (2015). *Feifer Assessment of Reading.* Lutz, FL: Psychological Assessment Resources.

Feifer, S. G. (2016). *Feifer Assessment of Mathematics.* Lutz, FL: Psychological Assessment Resources.

Feifer, S., & Naglieri, J. A. (2017). A neurodevelopmental approach to assessment. In S. G. Feifer, *The neuropsychology of mathematics: An introduction to the FAM* (pp. 105–132). Middletown, MD: School Neuropsych Press.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Ford, D. Y. (2010). Underrepresentation of culturally different students in gifted education: Reflections about current problems and recommendations for the future. *Gifted Child Today, 33*(3), 31–35.

Ford, D. Y. (2013). Multicultural issues: Gifted underrepresentation and prejudice—Learning from Allport and Merton. *Gifted Child Today, 36*(1), 62–67.

Gagné, F. (1985). Giftedness and talent: Reexamining a reexamination of the definitions. *Gifted Child Quarterly, 29,* 103–112.

Georgiou, G. K., Das, J. P., & Hayward, D. V. (2008). Comparing the contribution of two tests of working memory to reading in relation to phonological awareness and rapid naming speed. *Journal of Research in Reading, 31,* 302–318.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, 24,* 355–392.

Goldberg, E. (2009). *The new executive brain: Frontal lobes in a complex world.* New York: Oxford University Press.

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical clinical construct. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 3–12). New York: Springer

Haddad, F. A., Garcia, Y. E., Naglieri, J. A., Grimditch, M., McAndrews, A., & Eubanks, J. (2003). Planning facilitation and reading comprehension: Instructional relevance of the PASS theory. *Journal of Psychoeducational Assessment, 21*(3), 282–289.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Individuals with Disabilities Education Improvement Act (IDEA) of 2004, Pub. L. 108-446, 20 U.S.C. 1400 (2004).

Iseman, J., & Naglieri, J. A. (2011). A cognitive strategy instruction to improve math calculation for children with ADHD: A randomized controlled study. *Journal of Learning Disabilities, 44*(2), 184–195.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement, Third Edition*. Bloomington, MN: Pearson.

Kirby, J. R., & Williams, N. H. (1991). *Learning problems: A cognitive approach*. Toronto: Kagan & Woo.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test*. Itasca, IL: Riverside.

Luria, A. R. (1966a). *Higher cortical functions in man*. New York: Basic Books.

Luria, A. R. (1966b). *Human brain and psychological processes*. New York: Harper & Row.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

Luria, A. R. (1982). *Language and cognition*. New York: Wiley.

McFadden v. Board of Education for Illinois School District U-46, 984 F. Supp. 2d 882 (2013).

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV*. Rolling Meadows, IL: Riverside.

Meltzer, L. (2010). *Promoting executive function in the classroom*. New York: Guilford Press.

Mpofu, E., & Ortiz, S. O. (2009). Equitable assessment practices in diverse contexts. In E. L. Grigorenko (Ed.), *Multicultural psychoeducational assessment* (pp. 41–76). New York: Springer.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock–Muñoz*. Itasca, IL: Riverside.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.

Naglieri, J. A. (2000). Can profile analysis of ability test scores work?: An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly, 15*, 419–433.

Naglieri, J. A. (2005). The Cognitive Assessment System. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 441–460). New York: Guilford Press.

Naglieri, J. A. (2008). *Naglieri Nonverbal Ability Test—Second Edition*. San Antonio, TX: Pearson.

Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 145–172). Hoboken, NJ: Wiley.

Naglieri, J. A. (2012). Psychological assessment by school psychologists: Opportunities and challenges of a changing landscape. In K. Geisinger & B. A. Bracken (Eds.), *APA handbook of testing and assessment in psychology* (pp. 1–17). Washington, DC: American Psychological Association.

Naglieri, J. A. (2014). *CAS2: Online scoring and report system*. Austin, TX: PRO-ED.

Naglieri, J. A. (2015). One hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 295–316). New York: Springer.

Naglieri, J. A. (2017). *Naglieri Nonverbal Ability Test—Third Edition*. Bloomington, MN: Pearson.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psycheducational Assessment, 21*, 244–260.

Naglieri, J. A., Brulles, D., & Lansdowne, K. (2008). *Helping all gifted students learn: A teacher's guide to using Naglieri nonverbal assessments*. San Antonio, TX: Pearson.

Naglieri, J. A., & Brunnert, K. (2009). Wechsler Nonverbal Scale of Ability. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 291–314). Hoboken, NJ: Wiley.

Naglieri, J. A., & Conway, C. (2009). The Cognitive Assessment System. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 3–10). Hoboken, NJ: Wiley.

Naglieri, J. A., & Das, J. P. (1997a). *Cognitive Assessment System*. Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (1997b). *Cognitive Assessment System: Administration and scoring manual*. Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (2006). Are intellectual processes important in the diagnosis and treatment of ADHD? *The ADHD Report, 14*(1), 1–6.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014a). *Cognitive Assessment System—Second Edition*. Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014b). *Cognitive Assessment System—Second Edition: Brief*. Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014c). *Cognitive Assessment System—Second Edition: Rating Scale*. Austin, TX: PRO-ED.

Naglieri, J. A., & Feifer, S. (2017). Intervention. In J. A. Naglieri & T. M. Otero, *Essentials of CAS2 assessment*. Hoboken, NJ: Wiley.

Naglieri, J. A., & Feifer, S. (in press). Pattern of strengths and weaknesses made easy: The discrepancy/consistency method. In D. Flanagan & V. Alfonso, *Essentials of specific learning disabilities identification* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A., & Ford, D. Y. (2015). Misconceptions about the Naglieri Nonverbal Ability Test: A commentary of concerns and disagreements. *Roper Review, 37,* 234–240.

Naglieri, J. A., & Goldstein, S. (2011). Assessment of cognitive and neuropsychological processes. In S. Goldstein & J. A. Naglieri (Eds.), *Understanding and managing learning disabilities and ADHD in late adolescence and adulthood* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A., & Goldstein, S. (2013). *Comprehensive Executive Function Inventory.* Toronto: Multi-Health Systems.

Naglieri, J. A., & Gottling, S. H. (1995). A cognitive education approach to math instruction for the learning disabled: An individual study. *Psychological Reports, 76*(3), 1343–1354.

Naglieri, J. A., & Gottling, S. H. (1997). Mathematics instruction and PASS cognitive processes: An intervention study. *Journal of Learning Disabilities, 30*(5), 513–520.

Naglieri, J. A., & Johnson, D. (2000). Effectiveness of a cognitive strategy intervention in improving arithmetic computation based on the PASS theory. *Journal of Learning Disabilities, 33*(6), 591–597.

Naglieri, J. A., Kamphaus, R. W., & Kaufman, A. S. (1983). The Luria–Das simultaneous– successive model applied to the WISC-R. *Journal of Psychoeducational Assessment, 1,* 25–34.

Naglieri, J. A., Moreno, M. A., & Otero, T. (2017). *Cognitive Assessment System—Second Edition: Spanish.* Austin, TX: PRO-ED.

Naglieri, J. A., & Otero, T. (2011). Cognitive Assessment System: Redefining intelligence from a neuropsychological perspective. In A. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 320–333). New York: Springer.

Naglieri, J. A., & Otero, T. M. (2012). The Cognitive Assessment System: From theory to practice. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 376–399). New York: Guilford Press.

Naglieri, J. A., & Otero, T. M. (2017). *Essentials of CAS2 assessment.* Hoboken, NJ: Wiley.

Naglieri, J. A., Otero, T., DeLauder, B., & Matto, H. (2007). Bilingual Hispanic children's performance on the English and Spanish versions of the Cognitive Assessment System. *School Psychology Quarterly, 22,* 432–448.

Naglieri, J. A., & Pickering, E. (2010). *Helping children learn: Intervention handouts for use in school and at home* (2nd ed.). Baltimore: Brookes.

Naglieri, J. A., & Rojahn, J. (2004). Validity of the PASS theory and CAS: Correlations with achievement. *Journal of Educational Psychology, 96,* 174–181.

Naglieri, J. A., Rojahn, J., & Matto, H. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence, 35,* 568–579.

Naglieri, J. A., Rojahn, J., Matto, H., & Aquilino, S. A. (2005). Black–white differences in intelligence: A study of the PASS theory and Cognitive Assessment System. *Journal of Psychoeducational Assessment, 23,* 146–160.

Naglieri, J. A., Taddei, S., & Williams, K. (2013). US and Italian children's performance on the Cognitive Assessment System: A cross cultural equivalence study. *Psychological Assessment, 25,* 157–166.

Natur, N. H. (2009). An analysis of the validity and reliability of the Das–Naglieri Cognitive Assessment System (CAS)—Arabic Edition (Doctoral dissertation, Howard University). Retrieved from *https://search.proquest.com/openview/c9fccb2c6da6c11ff48d85340b2834f8/1?pq-origsite=gscholar&cbl=18750&diss=y* (UMI 3342065).

Otero, T. M., Gonzales, L., & Naglieri, J. A. (2013). The neurocognitive assessment of Hispanic English-language learners with reading failure. *Applied Neuropsychology: Child, 2*(1), 24–32.

Padilla, A. M., & Borsato, G. (2008). Issues in culturally appropriate psychoeducational assessment. In L. A. Suzuki & J. Ponterotto (Eds.), *Handbook of multicultural assessment* (pp. 5–21). Hoboken, NJ: Wiley.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition.* San Antonio, TX: Author.

Pressley, M., & Woloshyn, V. (1995). *Cognitive strategy instruction that really improves children's academic performance* (2nd ed.). Cambridge, MA: Brookline Books.

Scheid, K. (1993). *Helping students become strategic learners: Guidelines for teaching.* Cambridge, MA: Brookline Books.

Schofield, N. J., & Ashman, A. F. (1986). The relationship between digit span and cognitive processing across ability groups. *Intelligence, 10,* 59–73.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33,* 89–108.

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education, 12*(3), 275–287.

Suzuki, L. A., & Valencia, R. R. (1997). Race–ethnicity and measured intelligence. *American Psychologist, 52,* 1103–1114.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Weyandt, L. L., Willis, W. G., Swentosky, A., Wilson, K., Janusis, G. M., Chung, H. J., et al. (2014). A review of the use of executive function tasks in externalizing and internalizing disorders. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 69–88). New York: Springer.

Yoakum, C., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

Zirkel, P. A., & Thomas, L. B. (2010). State laws for RTI: An updated snapshot. *Teaching Exceptional Children, 42,* 56–63.

CHAPTER 16

The Wechsler Adult Intelligence Scale—Fourth Edition and the Wechsler Memory Scale—Fourth Edition

Lisa Whipple Drozdick
Susan Engi Raiford
Dustin Wahlstrom
Lawrence G. Weiss

The Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV; Wechsler, 2008) is the most recent edition of the WAIS. It incorporates numerous changes from previous editions to ensure that the instrument continues to reflect advances in modern test theory and research. It is used to assess intellectual ability in adults and adolescents ages 16–90 and provides information about specific cognitive abilities across various domains. It is frequently used with other instruments in comprehensive evaluations.

Memory is frequently evaluated along with intellectual functioning. The WAIS-IV and the Wechsler Memory Scale—Fourth Edition (WMS-IV; Wechsler, 2009) are co-normed to facilitate direct comparison of performance across the two measures. A common referral suggesting the combined use of the WAIS-IV and WMS-IV involves poor or questionable cognitive or memory performance. When the two batteries are used together, the WAIS-IV provides a measure of general cognitive ability that serves as a backdrop for WMS-IV interpretation.

This chapter presents an overview of the WAIS-IV and the WMS-IV, followed by guidelines on the use and interpretation of the instruments separately and in combination. Each overview presents a brief description of the theory and structure of the instrument, descriptions of the subtests and composites, and information on psychometric properties. The section on interpretation provides specific interpretive information for each instrument and on joint use of the two instruments. Finally, a chapter appendix presents a case study to illustrate the combined use of the WAIS-IV and WMS-IV.

WECHSLER ADULT INTELLIGENCE SCALE–FOURTH EDITION

Theoretical Framework

The Wechsler intelligence scales utilize a *theoretical framework of intelligence* to guide development of each edition. Within this framework, multiple lines of inquiry are simultaneously considered to align each new edition of a Wechsler intelligence scale with contemporary thought and knowledge. This framework draws upon current theories of intellect and specific cognitive abilities, and upon active progress in developmental psychology, neuropsychology, cognitive neuroscience, and clinical research and utility. This ensures that the Wechsler scales are highly innovative and modern. Tying an instrument to a single theory limits the ability of the test to evolve dynamically, based on all relevant research and in response to other information from related fields. Moreover, if the theory changes to accommodate new information, the test becomes outdated. The modern Wechsler theoretical framework is described in greater detail elsewhere in this volume (see Wahlstrom, Raiford, Breaux, Zhu, & Weiss, Chapter 9).

David Wechsler, in his time, set the tone of responsiveness to contemporary theory and research as well as clinical utility when developing the Wechsler intelligence scales. He was best known for his clinical acumen and developed his tests to be clinical instruments, although he was trained in statistics by Charles Spearman and Karl Pearson (A. S. Kaufman, Raiford, & Coalson, 2016). Wechsler was influenced by two key theorists of intelligence at the time: the aforementioned Charles Spearman, as well as Edward Thorndike. Spearman's concept of *g* and general intelligence was an obvious influence (A. S. Kaufman et al., 2016; Tulsky, Zhu, & Prifitera, 2000; Weiss, Saklofske, Coalson, & Raiford, 2010), as Wechsler viewed intelligence as a global entity. However, Thorndike's influence was also evident, as Wechsler conceived of this global entity as consisting of qualitatively different elements (A. S. Kaufman et al., 2016; Wechsler, 1939, 1950, 1975). He articulated this view best in the Wechsler–Bellevue Intelligence Scale manual, where he described intelligence as

> the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his [or her] environment. It is global because it characterizes the individual's behavior as a whole; it is aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. (Wechsler, 1939, p. 3)

Wechsler also postulated that a number of abilities not assessed by intelligence tests affect an individual's ability to navigate his or her environment effectively. These include personality and conative factors, such as drive, persistence, curiosity, and temperament (Wechsler, 1950). Wechsler was unsuccessful in developing a measure of these noncognitive intellective skills, but his practical and clinically based overarching theory of intelligence resulted in a number of strengths that have made the Wechsler intelligence scales the most widely used in the world today (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Rabin, Paolillo, & Barr, 2016). Perhaps most importantly, the Wechsler scales are considered valid and clinically useful instruments, providing a clinician with an accurate snapshot of an individual's functioning that is related to the person's success in real-world settings—including job performance (Hunt & Madhyastha, 2012; Kuncel, Ones, & Sackett, 2010; Schmidt, 2014); mental and physical health and health behaviors (Johnson, Corley, Starr, & Deary, 2011; Rindermann & Meisengerg, 2009); and academic achievement and educational attainment (Deary & Johnson, 2010; S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012).

Many of the features discussed later in this chapter were incorporated into the WAIS-IV to address recent advances in the lines of inquiry considered within the Wechsler theoretical framework of intelligence. For example, a growing body of literature emphasizes the importance of fluid reasoning in general cognitive functioning and age-related cognitive decline (e.g., Keage et al., 2015; Pineda-Pardo, Martinez, Román, & Colom, 2016), and fluid reasoning is commonly found to be highly associated with general intelligence (Reynolds, Keith, Flanagan, & Alfonso, 2013). A measure of quantitative fluid reasoning, Figure Weights, was included in the WAIS-IV to expand coverage of this construct. Similarly, working memory has been found to be an important predictor of individual differences in learning and fluid reasoning (Chuderski, 2014; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Fry & Hale, 1996; Perlow, Jattuso, & Moore, 1997), and a very close association has also been observed between working memory capacity and *g* (Colom, Rebollo, Palacios, Espinosa, & Kyllonen, 2004; Conway, Kane, & Engle, 2003). In addition, working memory training has been linked to improvements in fluid reasoning ability, suggesting a dynamic interplay between the two in the control of cognitive functioning (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Furthermore, working memory appears to play a role in age-related cognitive decline (Pineda-Pardo et al., 2016). To address the growing importance of working memory in cognitive ability, the Digit Span Sequencing task was added to Digit Span. This change increases the role of mental manipulation and results in greater demands on working memory, relative to previous versions of Digit Span. Arithmetic was also altered to reduce the role of verbal comprehension skills or mathematical knowledge relative to working memory. For example, difficult items require several simple math computation steps that have to be represented in working memory in favor of the complex calculations included in the prior edition.

Processing speed is another construct important to the expression of intelligence, as research suggests that it interacts with working memory and fluid intelligence; processing speed mediates the relationship between working memory and reasoning (Dang, Braeken, Colom, Ferrer, & Liu, 2015; Fry & Hale, 1996, 2000). Specifically, it has been proposed that rapid information processing may reduce working memory demands, and that this reduction in turn releases cognitive resources for more complex forms of reasoning (Weiss et al., 2010). Furthermore, factor-analytic research has identified processing speed as an important component of the Cattell–Horn–Carroll (CHC) model of intelligence (Schneider & McGrew, 2012 and Chapter 3, this volume), and it has been shown to be sensitive to a number of clinical syndromes, including attention-deficit/hyperactivity disorder (ADHD) (Nielsen & Wiig, 2011), traumatic brain injury (TBI) (Donders & Strong, 2015), epilepsy (Baxendale, McGrath, & Thompson, 2014), and multiple sclerosis (Ryan, Gontkovsky, Kreiner, & Tree, 2012). Processing speed is additionally implicated in cognitive decline related to aging (Manard, Carabin, Jaspar, & Collette, 2014). The WAIS-IV includes both the Coding and Symbol Search subtests in the calculation of the Full Scale IQ (FSIQ). As a result, the contribution of processing speed subtests to the FSIQ increased from 9% in the WAIS-III to 20% in the WAIS-IV, reflecting the growing importance of this construct in intellectual assessment.

Overall, confounding variables within cognitive domains have been reduced in the WAIS-IV, thereby increasing the instrument's ability to tap cognitive functions more purely. For example, to reduce the influence of declining processing speed on the scores of older adults (Lee, Gorsuch, Saklofske, & Patterson, 2008), time-bonus points were removed from Arithmetic and reduced on Block Design. Similarly, Cancellation was added as an alternative processing speed subtest with fewer fine motor demands than Coding, and Visual Puzzles demands less fine motor control than Object Assembly, which it replaces on the Perceptual Reasoning Index (PRI). Despite these changes, care has been taken to ensure that the test retains its validity and clinical utility. A century of cognitive research has demonstrated that human cognitive functions form a dynamically unified entity, as evidenced by the discussion above of fluid reasoning, working memory, and processing speed. Wechsler himself noted the dynamic nature of intellectual abilities, noting that they "appear to behave differently when alone from what they do when operating in concert" (1975, p. 138). Therefore, while the measurement of pure cognitive functions is ideal from theoretical and psychometric standpoints, it does not necessarily result in information that is clinically rich or practically useful in real-world applications (Weiss et al., 2010; Zachary, 1990). Thus the WAIS-IV reflects a balance between theoretically sound cognitive constructs and the need to maintain the predictive value and clinical utility that are central to Wechsler's framework of intelligence.

Criticisms of the WAIS-IV Framework

"It's Atheoretical"

The Wechsler intelligence scales have been criticized for what some perceive as a lack of theoretical orientation (Beres, Kaufman, & Perlman, 2000; A. S. Kaufman & Lichtenberger, 1999; McGrew & Flanagan, 1997). Wechsler selected cognitive tasks for his intelligence battery that were clinically grounded rather than tied to a single theory of intelligence, thus allowing subsequent revisions to incorporate modern advances from a wider array of ongoing clinical and neuropsychological research. As a result, the current editions align and correlate well with tests created more recently and based on specific theories of intelligence, such as the CHC and Luria models. For instance, after analyzing the factor structure of more than 450 datasets, Carroll (1993) revealed the presence of a general intelligence factor, and several studies have provided evidence suggesting that intelligence is composed of specific narrow abilities that appear to cluster into higher-order ability domains (Carroll, 1993; Cohen, 1952, 1957; Horn, 1994). Numerous independent analyses of the WAIS-IV normative sample suggest that the cognitive domains measured by the test align closely with those specified by recent models of intelligence (Benson, Hulac, & Kranzler, 2010; Keith, as cited in Lichtenberger & Kaufman, 2012; Weiss, Keith, Zhu, & Chen, 2013), which include crystallized ability, visual processing, fluid reasoning, working memory, and processing speed. These results suggest that the cognitive constructs measured by the WAIS-IV are similar to those of other tests that are designed around specific theories. This alignment is also supported by the fact that the Wechsler scales correlate highly with theory-based intelligence measures, such as the Kaufman Adolescent and Adult Intelligence Test

(KAIT; A. S. Kaufman & Kaufman, 1993) and the Differential Ability Scales—Second Edition (Elliott, 2007). The totality of this evidence has caused some to rethink their positions regarding the theoretical nature of the Wechsler scales. For instance, Alan S. Kaufman has stated that "I have since reconsidered my stance on the lack of a theoretical framework for Wechsler's scales" (2010, p. xvi), as "the WAIS-IV also was developed with specific theoretical foundations in mind. In fact, revisions were made purposely to reflect the latest knowledge from literature in the areas of intelligence theory, adult cognitive development, and cognitive neuroscience" (Lichtenberger & Kaufman, 2012, p. 20).

"The Index Scores Don't Matter"

Assumptions make a difference in an investigator's approach to factor-analytic studies and to interpretation of those studies' results. One group of researchers takes an approach that is different from the intended test model, assuming that only general intelligence (*g*) is relevant, and not broad cognitive abilities like crystallized intelligence, fluid reasoning, and working memory. Consistent with their assumption, they statistically remove *g* from the index scores and then examine the index scores' residual validity (the *bifactor* model). They then conclude that variance attributed to the first-order factors (the index scores) is too small to be of importance, and that the FSIQ is the only score worth interpreting.

This approach is problematic for a number of reasons. First, individual routing of *g* resources to specific cognitive domains results in greater development of those abilities over time. Thus removing the influence of *g* from the index scores attenuates their power, and creates a rather artificial situation. As Schneider (2013) observed, "the independent portion is not the 'real Gc'. We care about a sprinter's ability to run quickly, not residual sprinting speed after accounting for general athleticism. So it is with Gc: g is a part of the mix" (p. 188).

Second, as noted by Flanagan and Alfonso (2017), even these researchers have published studies demonstrating that these factors have incremental validity in predicting academic outcomes (e.g., Canivez, 2013; Nelson, Canivez, & Watkins, 2013), with up to 2–30% of additional variance explained. Thus the broad factors are worth interpreting.

Finally, in examinees with neurological disorders or brain injuries, discrepancies often characterize the index scores. For example, while the FSIQ may be 100, the Verbal Comprehension Index (VCI) may be much higher than 100, and the Processing Speed Index (PSI) may be much lower. This situation renders the FSIQ insufficient to describe a person's abilities and provide guidance for treatment planning, and it could do the person the disservice of ignoring a real decline in cognitive functioning that should be acknowledged and accommodated. Thus the approach taken by these researchers excludes important information and is too one-sided. Both *g* and the factor index scores are important, and each construct has a place in the practice of assessment.

Test Model and Scores

All Wechsler intelligence scales provide factor-based index scores that measure major cognitive domains identified in contemporary theories of intelligence. The primary advantage of the index scores is their measurement of relatively discrete cognitive domains, which allows a clinician to evaluate specific aspects of cognitive functioning more clearly. For example, individuals with learning disorders may be expected to perform poorly on working memory tasks (Cohen-Mimran & Sapir, 2007; Proctor, 2012; Wechsler, 2008), and individuals with TBI may exhibit poor processing speed skills (Donders & Strong, 2015; Mathias & Wheaton, 2007). The WAIS-IV provides four index scores: the VCI, the PRI, the PSI, and the Working Memory Index (WMI). In addition to these four index scores, it provides the optional General Ability Index (GAI), which is derived from the six FSIQ subtests that contribute to the VCI and the PRI.

The WAIS-IV consists of 15 subtests (10 core and 5 supplemental), three of which are new: Visual Puzzles, Figure Weights, and Cancellation. Figure 16.1 shows the WAIS-IV subtests, as well as the composite scores to which each contributes. Despite the changes made to the WAIS-IV, the correlations with the WAIS-III are high, suggesting that they measure closely related constructs. For example, the two versions' FSIQs correlate .94, and the index scale correlations range from .84 for the PRI (correlated with the Perceptual Organization Index [POI] on the WAIS-III) to .91 for the VCI (Wechsler, 2008).

The WAIS-IV is normed for individuals between the ages of 16 years, 0 months (16:0) and 90:11. The following sections describe the WAIS-IV scales and each composite score that is derived

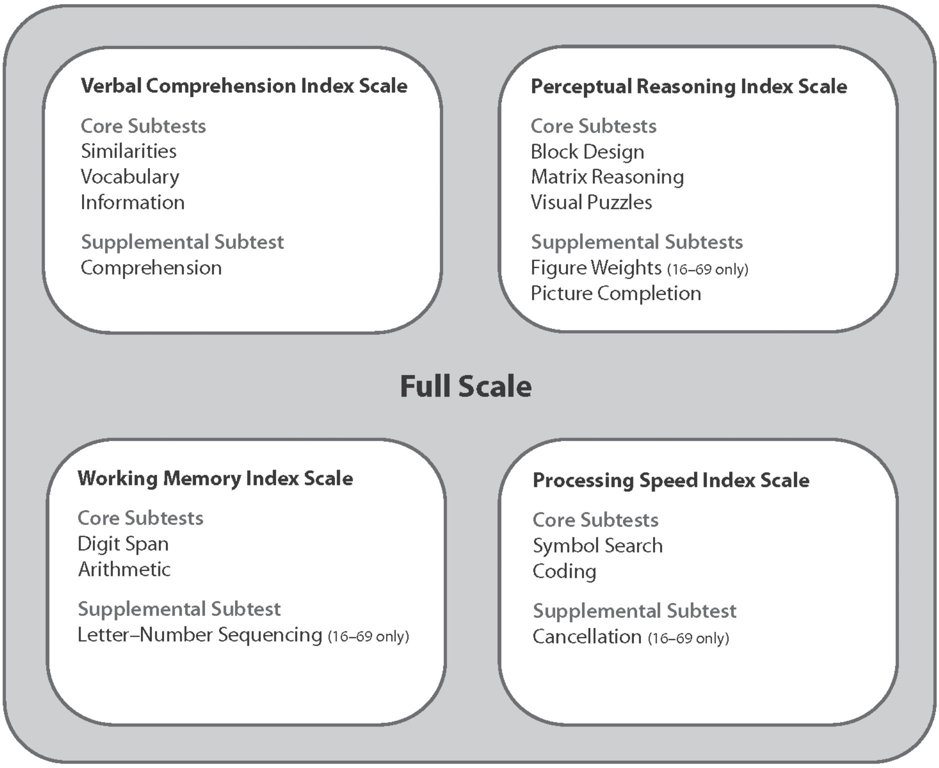


FIGURE 16.1. Test framework of the WAIS-IV. Figures found in the manual for the *Wechsler Adult Intelligence Scale®, Fourth Edition (WAIS®-IV)*. Copyright © 2008 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

from the core subtests on that scale. General descriptions of the FSIQ, index scores, and subtests are provided elsewhere in this volume (Wahlstrom et al., Chapter 9); thus, rather than focusing on the general nature of the subtests, scales, and index scores that constitute the WAIS-IV, this chapter instead highlights information specific to the WAIS-IV that does not appear in Chapter 9. For example, some subtests that are not described in Chapter 9 are described in the sections that follow.

Full Scale

The Full Scale contains all 15 of the WAIS-IV subtests. The FSIQ is a composite score that estimates an individual's general level of intellectual functioning. It is derived from the sum of 10 core subtest scaled scores. As shown in Figure 16.1, 10 subtests contribute to the WAIS-IV FSIQ: three from the Verbal Comprehension scale, three from the Perceptual Reasoning scale, two from the Working Memory scale, and two from the Processing Speed scale.

As discussed in Chapter 9 (this volume), the FSIQ is considered the score most representative of general intellectual functioning (i.e., *g*), and is a robust predictor of an array of important life outcomes (Deary & Johnson, 2010; Gottfredson, 1997; Gottfredson & Deary, 2004; Hunt & Madhyastha, 2012; Johnson et al., 2011; S. B. Kaufman et al., 2012; Kuncel et al., 2010; Rindermann & Meisengerg, 2009, Schmidt, 2014). Furthermore, *g* consistently emerges in factor-analytic studies—a finding that has been replicated in research using international adaptations of the Wechsler scales (Bowden, Lange, Weiss, & Saklofske, 2008; Georgas, Weiss, van de Vijver, & Saklofske, 2003; Golay & Lecerf, 2011; Weiss et al., 2013).

Verbal Comprehension Scale

There are four subtests on the Verbal Comprehension scale. As shown in Figure 16.1, the VCI is derived from the sum of the three core subtest scaled scores: Similarities, Vocabulary, and Information. If necessary, Comprehension may be used as a substitute for one of the three core Verbal Comprehension subtests in deriving the VCI. Relative to the WISC-V VCI described in Chapter 9, which primarily emphasizes verbal concept formation and verbal reasoning, the WAIS-IV VCI places slightly more emphasis on acquired knowledge.

Perceptual Reasoning Scale

There are five subtests on the Perceptual Reasoning scale: Block Design, Matrix Reasoning, Visual Puzzles, Figure Weights (for ages 16–69 only), and Picture Completion. As depicted in Figure 16.1, the PRI is derived from the sum of the three core Perceptual Reasoning subtest scaled scores: Block Design, Matrix Reasoning, and Visual Puzzles. If necessary, Picture Completion may be used as a substitute for one of the three core subtest scaled scores to derive the PRI. For ages 16–69, Figure Weights may also be used as a substitute.

The PRI is a measure of fluid reasoning, spatial processing, and visual–motor integration. It also reflects working memory and processing speed skills, based on evidence that these abilities are intertwined with fluid reasoning.

Picture Completion

Picture Completion requires the examinee to view a picture and then point to or name the important part missing within a specified time limit. This subtest measures visual perception and organization. It also requires visual discrimination, visual recognition of essential details of objects, crystallized knowledge, reasoning, and visual long-term memory (Lichtenberger & Kaufman, 2012; Sattler & Ryan, 2009).

Working Memory Scale

Three subtests are on the Working Memory scale: Digit Span, Arithmetic, and Letter-Number Sequencing (for ages 16–69 only). As depicted in Figure 16.1, the WMI is derived from the sum of the two core Working Memory subtest scaled scores. If necessary, Letter–Number Sequencing may be used for ages 16–69 as a substitute for one of the core subtest scaled scores to derive the WMI.

In contrast to the WISC-V WMI described in Chapter 9 (this volume), which emphasizes both auditory and visual working memory, the WAIS-IV WMI solely involves auditory working memory. It also draws on quantitative reasoning due to the Arithmetic subtest.

Processing Speed Scale

The Processing Speed scale has three subtests: Symbol Search, Coding, and Cancellation (for ages 16–69 only). As shown in Figure 16.1, the PSI is derived from the sum of the two core subtest scaled scores. If necessary, Cancellation may be used for ages 16–69 as a substitute for one of the core subtest scaled scores to derive the PSI. The PSI is of specific interest on the WAIS-IV, as it is especially sensitive to aging (A. S. Kaufman, 2001; Lichtenberger & Kaufman, 2012), and because processing speed has been hypothesized to underlie age-related declines in other cognitive domains (Salthouse, 2004).

General Ability Index

The GAI is described in Chapter 9 (this volume). It is utilized for all ability–memory comparisons with the WMS-IV to increase the ability to detect discrepancies in individuals with processing speed and working memory deficits, which are commonly observed in individuals with memory deficits (see Glass, Bartels, & Ryan, 2009; Lange & Chelune, 2006; Lange, Chelune, & Tulsky, 2006). However, the GAI should not be used as a substitute for overall ability simply because the WMI and PSI are significantly lower than the VCI and PRI, as working memory and processing speed are important contributors to intelligence (Prifitera, Saklofske, & Weiss, 2005; Weiss et al., 2010). It is best practice to derive and interpret the FSIQ alongside the GAI.

Psychometric Properties

Normative Sample

The WAIS-IV has an excellent normative sample (Sattler & Ryan, 2009). It included 2,200 individuals divided into 13 age bands (see Wechsler, 2008, for detailed descriptions of the age bands). Each age band below 70 years contained 200 examinees, whereas each age band from the ages of 70 to 90 years included 100 examinees. The stratification of the normative sample matched 2005 U.S. census data closely on five key demographic variables: age, gender, race/ethnicity, educational level, and geographic region.

Reliability

The WAIS-IV scores have strong reliability (Wechsler, 2008). First, the overall internal-consistency reliability coefficients for the normative sample are in the .90s for all index scores and .98 for the FSIQ. At the subtest level, the overall internal-consistency reliability coefficients of the normative sample are in the .80s or .90s for all core subtests. In addition, Cancellation is the only supplemental subtest with a reliability below .80, although it is still within the acceptable range (.78). The internal-consistency coefficients of the WAIS-IV subtests calculated from special group samples are very consistent with those obtained from the normative sample, with all subtest coefficients in the .80–.90 range. Second, the test–retest stability coefficients for the WAIS-IV index scores vary from .87 to .96, and the subtest stability coefficients range from .74 to .90. The test–retest stability of the FSIQ is .96. Finally, the interscorer agreement for most WAIS-IV subtests is .98–.99. Even the Verbal Comprehension subtests, which require greater judgment in scoring, have interscorer agreement above .90.

Validity

There is ample evidence to support the validity of the WAIS-IV. The confirmatory factor-analytic studies reported in the test manual provide strong evidence of construct validity and clearly demonstrate that in addition to measuring general intellectual ability, the Wechsler scales measure four cognitive domains: Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed (Wechsler, 2008). Further evidence of construct validity is provided by independent examinations of the WAIS-IV data. Independent factor-analytic studies verify the factor structure (Bowden, Saklofske, & Weiss, 2011a, 2011b; Ward, Bergman, & Hebert, 2012) and also indicate that the basic factor structure of the WAIS-IV holds for individuals with clinical syndromes such as schizophrenia and TBI (Goldstein & Saklofske, 2010). Similarly, factor analyses in samples of individuals with autism spectrum disorder have revealed factors of verbal comprehension, perceptual reasoning, and freedom from distractibility, as well as a social cognition factor (Goldstein et al., 2008; Goldstein & Saklofske, 2010).

Since the publication of the WAIS-IV, investigators have identified an alternative five-factor structure of the WAIS-IV that also has merit (Benson et al., 2010; Weiss et al., 2013). The factor names proposed by different teams of investigators vary according to the taxonomic systems used, but the names appear to align with Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed. In the five-factor model, the multidimensionality inherent in the WAIS-IV PRI is disentangled to reveal unique factors for Visual Spatial (Block Design and Visual Puzzles) and Fluid Reasoning (Matrix Reasoning, Figure Weights, and Arithmetic), similar to the five factor structure of the WISC-V, and consistent with CHC theory (Weiss et al., 2013).

In addition, the WAIS-IV and its predecessors correlate highly with other measures of intelligence. The correlation of the WAIS-IV FSIQ with the WAIS-III FSIQ is .94, and the correlations between index scores are all above .83 (Wechsler, 2008). Moreover, previous versions of the WAIS correlate highly with other measures of intelligence, such as the KAIT and the Stanford–Binet Intelligence Scale: Fourth Edition (Wechsler, 1997). Finally, the WAIS-IV has good concurrent validity, as it correlates highly with composites from the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), with correlation coefficients ranging from .42 to .81. In studies with the Delis–Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001), the California Verbal Learning Test—Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998), high correlations are found between similar constructs, such as the WAIS-IV PRI and RBANS Visuospatial/Constructional scale, and lower correlations between dissimilar constructs, such as WAIS-IV index scores and delayed memory scores on the CVLT-II and RBANS (Wechsler, 2008).

WECHSLER MEMORY SCALE—FOURTH EDITION

Theoretical Underpinnings

The assessment of memory is a key component in many evaluations of cognitive functioning. The measurement of memory ability in evaluations for dementia or mild cognitive impairment is required to make an accurate diagnosis. However, memory is also affected in many other neurocognitive and learning disorders, including TBI, stroke, epilepsy, intellectual disabilities, learning disorders, and autism spectrum disorder (Bauer, 2008; Carlozzi, Grech, & Tulsky, 2014; Eichenbaum, 2008; Squire & Schacter, 2002). Also, many medical and neurological disorders influence memory performance, including multiple sclerosis, HIV/AIDS, lupus, epilepsy, and cancer (Fama, Rosenbloom, Sassoon, Pfefferbaum, & Sullivan, 2012; Lafosse, Mitchell, Corboy, & Filley, 2013; Lindner et al., 2014; Schucard, Lee, Safford, & Schucard, 2011; Soble et al., 2014). Finally, individuals with psychiatric disorders, such as depression, schizophrenia, or bipolar disorder, also exhibit memory difficulties (Grillon, Krebs, Gourevitch, Giersch, & Huron, 2010; Porter, Robinson, Malhi, & Gallagher, 2015; Rock, Roiser, Riedel, & Blackwell, 2014). In addition to memory difficulties known to be associated with specific disorders, knowledge of an examinee's memory ability can be helpful in assessments to help tailor interventions or training programs to a student's learning ability.

The WMS is one of the most popular memory assessment instruments (Rabin et al., 2016). It was first published in 1945 (Wechsler, 1945) and has undergone several revisions since the initial publication (e.g., Russell's WMS [Russell, 1975, 1988]; the WMS-III [Wechsler, 1997]). The changes in the WMS across revisions reflect the growing research on and theories of learning and memory. *Learning* is the process through which new information is acquired, and *memory* is the persistence of learning so it can be recalled at a later time (Squire, 1987). The WMS-IV measures both learning and memory ability. Several key concepts are used across measures of learning and memory and are defined here. For more detailed descriptions of learning and memory processes, theories, and measurement, see Lampinen and Beike (2014), Lezak, Howieson, Bigler, and Tranel (2012), or Squire and Schacter (2002).

While other skills and abilities are involved in memory, the WMS-IV measures the learning and memory constructs of encoding, storage or consolidation, and retrieval of information. *Encoding* is the transformation of external information into mental representations or memories; it reflects the entry of information into the memory system. *Consolidation* is the process through which information in immediate memory is transferred and solidified into or stored in long-term memory stores, and *retrieval* involves retrieving or recalling information from long-term memory stores into active conscious awareness.

Many theories divide memory into short-term memory and long-term memory (e.g., Atkinson & Shiffrin, 1968; Strauss, Sherman, & Spreen, 2006). *Short-term memory* refers to brief, temporary storage of information, lasting from a few seconds to a few minutes. More permanent or long-term memories, lasting from hours to years, are considered *long-term memory*. The WMS-IV measures both short- and long-term memory with the immediate and delayed conditions of Logical Memory, Verbal Paired Associates, Designs, and Visual Reproduction.

Working memory has recently been included as a component of short-term memory. Working memory involves temporary storage and manipulation of information, and the amount of information that can be held in it is very limited. In Baddeley's (2000, 2003) model, the working memory system has multiple components. The *central executive* is a regulatory system that oversees two information activation–storage systems, the *phonological loop* and the *visuospatial sketchpad*. Auditory information is processed and temporarily stored in the phonological loop, and visual information is processed and temporarily stored in the visuospatial sketchpad. In addition to these processes, the *episodic buffer*, regulated by the central executive, transfers information into long-term memory and holds interrelated information in working memory. The central executive controls the flow of information and the attention system, and engages long-term memory as needed. This coordination of cognitive processes by the central executive facilitates learning and other complex cognitive tasks.

Long-term memory is often described as *implicit* (procedural) or *explicit* (declarative) memory. *Implicit* or procedural memory involves learning from experiences without being consciously aware of learning, such as learning to ride a bike or drive a car. *Explicit* or declarative memory involves the conscious storage and retrieval of information, such as personal or factual knowledge. Explicit memory consists of *semantic* and *episodic* memory. *Semantic memory* is the memory for facts and concepts, whereas *episodic memory* is the recollection of personal events and the contexts in which they occur. None of the information required for performance on the WMS-IV is learned prior to the testing session. Therefore, the WMS-IV is primarily a measure of explicit episodic memory, as the "information presented is novel and contextually bound by the testing situation and requires the examinee to learn and retrieve information" (Wechsler, 2009, p. 2).

Test Framework and Scores

Two batteries were developed for the WMS-IV: the Adult and Older Adult batteries. The Older Adult battery contains fewer subtests than the Adult battery and provides a shorter administration time. In addition, the content of the auditory memory subtests differs across the two batteries, with fewer stimuli in the Older Adult battery subtests. These changes have reduced the testing time and improved the subtest floors for older adults. The Adult battery can be used for ages 16–69, and the Older Adult battery is used for individuals ages 65–90. For examinees in the overlapping ages of 65–69, the examiner may select the battery that is more appropriate for the individual being tested.

The WMS-IV Adult battery contains seven subtests, including six primary subtests and one optional subtest. Four of the primary subtests have immediate-recall (I), delayed-recall (II), and delayed-recognition conditions. Scores from the primary subtests combine to create five index scores: the Auditory Memory Index (AMI), the Visual Memory Index (VMI), the Visual Working Memory Index (VWMI), the Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). The WMS-IV Older Adult battery contains five subtests, including four primary subtests and one optional subtest. Three of the primary subtests have both immediate- and delayed-recall conditions and a delayed-recognition condition, which are combined to form four index scores. The VWMI is not available in the Older Adult battery. Unlike the overall FSIQ in the WAIS-IV, there is not an overall memory ability score; index scores are related to specific domains of memory.

Table 16.1 lists the WMS-IV subtests, the conditions within each subtest, and (where applicable) the index scores to which each subtest contributes. For subtests that have both immediate and delayed conditions, the separate conditions are listed in this table. Note that a single subtest may contribute to more than one index. Some subtest conditions produce scores that are not used to derive index scores, but are considered process scores. These are also optional and denoted with parentheses in Table 16.1. Process scores assist

TABLE 16.1. Subtests and Related Composite Scores of the WMS-IV

	Contribution to index scales				
Subtest and condition	AMI	VMI	VWMI*	IMI	DMI
(BCSE)					
Logical Memory I	✓			✓	
Logical Memory II	✓				✓
(Logical Memory Recognition)					
Verbal Paired Associates I	✓			✓	
Verbal Paired Associates II	✓				✓
(Verbal Paired Associates Recognition)					
(Verbal Paired Associates Word Recall)					
Designs I*		✓		✓	
Designs II*		✓			✓
(Designs Recognition)*					
Visual Reproduction I		✓		✓	
Visual Reproduction II		✓			✓
(Visual Reproduction Recognition)					
(Visual Reproduction Copy)					
Spatial Addition*			✓		
Symbol Span			✓		

Note. AMI, Auditory Memory Index; VMI, Visual Memory Index; VWMI, Visual Working Memory Index; IMI, Immediate Memory Index; DMI, Delayed Memory Index. Asterisks indicate subtests/conditions not included in the Older Adult battery. Check marks indicate primary subtests contributing to an index score. Optional subtests or conditions are in parentheses.

in the interpretation of performance by describing specific skills or abilities required to perform the tasks on the WMS-IV. For example, the Visual Reproduction subtest requires the examinee to draw responses. In order to assess the impact of poor motor ability, a copy condition provides a measure of motor skills. The WMS-IV test models for the Adult battery and Older Adult battery are depicted in Figures 16.2 and 16.3, respectively.

Subtest Descriptions

It is important to note that the factor analyses conducted with the WMS-IV standardization do not include the Brief Cognitive Status Exam (BCSE). In addition, they do not separate the immediate and delayed memory factors, due to the high correlations between these factors (Hoelzle, Nelson, & Smith, 2011; Wechsler, 2009). Although the immediate and delayed factors cannot be separated in factor analyses, they are clinically meaningful and useful (Millis, Malina, Bowers, & Ricker, 1999; Tulsky, Ivnik, Price, & Wilkins, 2003).

Brief Cognitive Status Exam

The BCSE consists of multiple types of items designed to quickly assess various areas of cognitive functioning. The items included in the BCSE were derived from subtests included in the WMS-III (e.g., mental control) or from common mental status and neuropsychological measures (e.g., inhibition). Items include orientation, mental control, incidental memory, clock drawing, inhibition control, and verbal production. Each section of items is used to create a weighted score representing that cognitive ability. The weighted scores are summed to provide an overall BCSE Total Raw Score, which is converted into a classification level indicating the examinee's general cognitive ability.

The BCSE is designed to measure an individual's basic cognitive ability across a variety of tasks. Unlike the comprehensive assessment of cognitive ability provided by the WAIS-IV, the BCSE is intended to provide a quick snapshot of an individual's cognitive status. When the WAIS-IV and WMS-IV are used together, it may not be necessary to administer the BCSE. However, the BCSE provides an opportunity for success early in an assessment session, as most examinees can answer some of the items on the BCSE. Individual items measure orientation to time, mental manipulation of information, incidental memory, planning and organization, the ability to inhibit an overlearned response, and verbal fluency. Other abilities that may be used during this task include confrontation naming, visual perception and attention, processing speed, working memory, planning, cognitive flexibility, auditory comprehension, and verbal expression (Drozdick, Holdnack, & Hilsabeck, 2011; Wechsler, 2009).

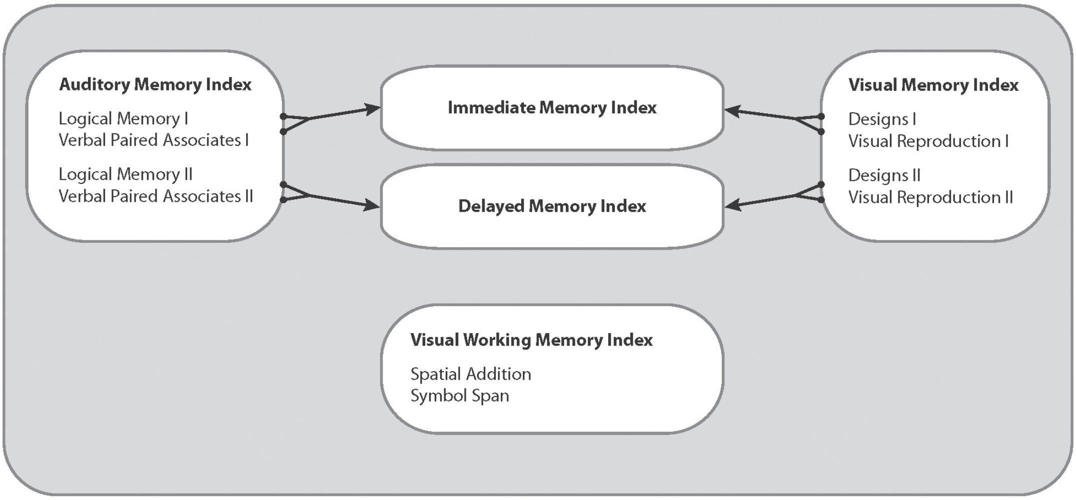


FIGURE 16.2. Test framework of the WMS-IV: Adult Battery. Figures found in the manual for the *Wechsler Memory Scale®, Fourth Edition (WMS®-IV)*. Copyright © 2009 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

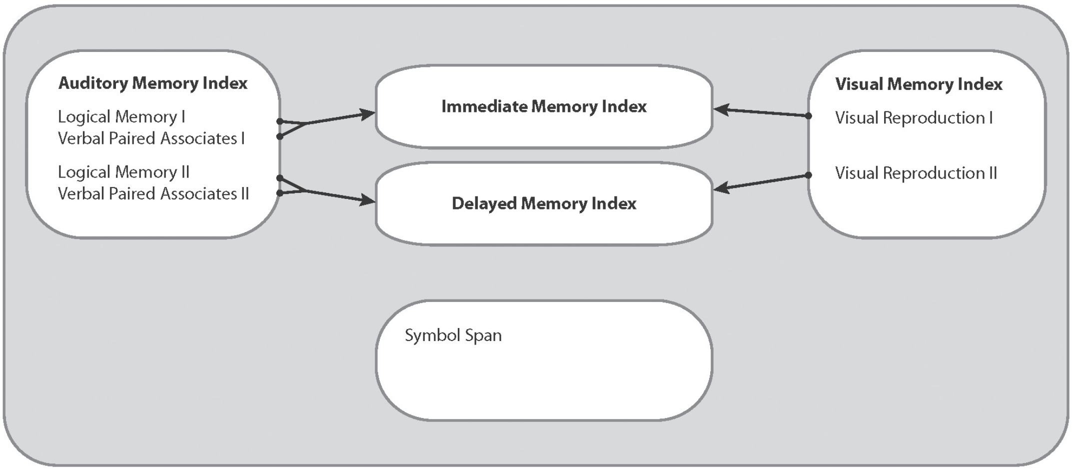


FIGURE 16.3. Test framework of the WMS-IV: Older Adult Battery. Figures found in the manual for the *Wechsler Memory Scale®, Fourth Edition (WMS®-IV)*. Copyright © 2009 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

Logical Memory

In the Logical Memory subtest, the examinee is read two stories and asked to recall them immediately after presentation and after a 20- to 30-minute delay. Different stories are presented in the Adult and Older Adult batteries. One story in the Older Adult battery is presented twice, creating a multitrial learning task in this age range. Neither story is repeated in the Adult battery. In the delayed recognition task, the examinee answers a series of yes–no questions about details from the stories. Logical Memory measures memory for structured information presented orally. It measures encoding, immediate and long-term retrieval, and recognition of organized, sequentially related auditory information. It also involves auditory comprehension, receptive and auditory language, auditory working memory, auditory attention and concentration, and hearing acuity (Drozdick et al., 2011; Groth-Marnat & Wright, 2016; Lichtenberger, Kaufman, & Lai, 2002). There is a long history of clinical and concurrent validity for Logical Memory (see Lezak et al., 2012, for a review). This subtest contributes to the AMI, IMI, and DMI.

Verbal Paired Associates

In Verbal Paired Associates, the examinee is read a series of paired words. The pairs are either related (e.g., *cat–dog*) or unrelated (*rain–pencil*). After being presented the list, the examinee is read the first word of each pair and asked to recall the second word. If the examinee misses an item, the correct word is given. Four learning trials are given during the immediate condition. Following a 20- to 30-minute delay, the examinee completes three delayed-memory tasks: delayed recall, word recall, and recognition. In the delayed-recall condition, the examinee recalls the second word in each pair again. In the recognition condition, the examinee recalls the second word pair and asked to indicate whether it is a pair from the list. In the word recall condition, the examinee is asked to recall the individual words from the word pairs; no cues are provided, and the examinee does not have to recall the words in pairs.

Verbal Paired Associates measures the ability to learn and retrieve semantically associated and unassociated words presented orally. It involves short-term auditory learning, long-term cued and recognition auditory memory, and associative memory. Other skills include auditory perception and comprehension, auditory working memory or rehearsal, expressive and receptive language, and hearing acuity (Groth-Marnat & Wright, 2016; Holdnack & Drozdick, 2010; Lichtenberger et al., 2002). Like Logical Memory, Verbal Paired Associates has a long history of clinical and concurrent validity studies (see Lezak et al., 2012, for a review). This subtest contributes to the AMI, IMI, and DMI.

Designs

Designs requires the examinee to view a 4 × 4 grid with designs in some of the cells in the stimulus book. Following exposure, the grid is removed from view, and the examinee uses a blank grid and design cards to reproduce the image from memory. Four grids are shown and recalled both immediately and following a 20- to 30-minute delay. A delayed-recognition condition is included in which the examinee selects the two cards displayed in a pictured grid that match the correct design and location as shown in the immediate condition. This subtest measures the ability to recall visual details and spatial information for abstract designs. It involves memory for visual detail and spatial memory; encoding and immediate and long-term retrieval of visual and spatial information; visual perception and organization; self-monitoring; and auditory comprehension. It also involves visual acuity, visual working memory, and gross motor ability (Drozdick et al., 2011; Groth-Marnat & Wright, 2016; Holdnack & Drozdick, 2010). This subtest contributes to the VMI, IMI, and DMI in the Adult battery. It is not included in the Older Adult battery.

Visual Reproduction

For each Visual Reproduction item, the examinee is shown a design for 10 seconds and then asked to draw the design from memory. A total of five designs are shown sequentially, with recall both immediately after presentation and following a 20- to 30-minute delay. Performance on Visual Reproduction is affected by motor constructional abilities (Gfeller, Meldrum, & Jacobi, 1995; Larrabee & Curtiss, 1995). An optional copy condition is available to assess the influence of motor difficulties. The scoring reflects the memory aspects of recall instead of motor accuracy. Visual Reproduction is designed to assess immediate and long-term visual memory for abstract designs. It also involves visual–motor coordination, visual–spatial construction, visual perception and organization, processing speed, planning, self-monitoring, and language ability (Drozdick et al., 2011; Groth-Marnat & Wright, 2016; Lezak et al., 2012). This subtest contributes to the VMI, IMI, and DMI.

Spatial Addition

Spatial Addition requires the examinee to look at two grids, shown sequentially, which contain blue and/or red circles. The examinee is instructed to remember the locations of the blue circles and to ignore the red circles. After being shown both grids, the examinee creates a new grid by placing cards in a blank grid based on a set of rules. This subtest measures visual–spatial working memory. It also involves visual perception and organization, executive functions, and gross motor ability (Drozdick et al., 2011; Groth-Marnat & Wright, 2016). This subtest contributes to the VWMI in the Adult battery. It is not included in the Older Adult battery.

Symbol Span

For each Symbol Span item, the examinee is shown an array of abstract symbols. The array is then removed, and the examinee is shown a second array of symbols containing both the symbols the examinee was shown previously and distractor symbols. The examinee must select the correct symbols in the order presented in the original array. Partial credit is awarded for selecting the correct symbols in an incorrect order. The task involves visual-sequencing working memory. It also involves mental manipulation of visual material, attention, mental flexibility, visual–spatial imaging, and visual perception and attention (Drozdick et al., 2011; Groth-Marnat & Wright, 2016). This subtest is included in both the Older Adult and Adult batteries, but only contributes to the VWMI in the Adult battery.

Index Descriptions

The WMS-IV index scores measure major domains of memory, whereas the subtest scores measure specific aspects of memory. For example, the AMI score summarizes an individual's performance across subtests designed to measure auditory memory, while Logical Memory and Verbal Paired Associates measure more specific aspects of auditory memory. An individual can obtain an average AMI with multiple ranges of subtest scores—for example, average scores on all contributing subtests, or above-average scores on Logical Memory and below-average scores on Verbal Paired Associates. Therefore, it is important to evaluate performance on index scores in the context of the performance on the contributing subtests.

Auditory Memory Index

The AMI is a measure of memory for orally presented information, both immediately and following a delay. It requires auditory attention, comprehension, and retrieval, and receptive and expressive language ability (Groth-Marnat & Wright, 2016; Wechsler, 2009). It consists of the Logical Memory and Verbal Paired Associates immediate- and delayed-recall conditions. It is composed of the same subtests across both batteries.

Visual Memory Index

The VMI is a measure of memory for information presented visually—specifically, abstract visual designs and spatial information. It measures visual perception and organization, as well as immediate and delayed recall and recognition of visual and spatial details (Groth-Marnat & Wright, 2016; Wechsler, 2009). It consists of the Designs and Visual Reproduction immediate and delayed conditions in the Adult battery. For the Older Adult battery, it consists of the Visual Reproduction immediate and delayed conditions.

Visual Working Memory Index

The VWMI is a visual analogue to the WAIS-IV WMI, which measures auditory working memory. It involves attending to, encoding, mentally manipulating, organizing, and recalling visual and spatial information (Groth-Marnat & Wright, 2016; Wechsler, 2009). It consists of the Spatial Addition and Symbol Span subtests and is only available in the Adult battery. The WAIS-IV WMI contains auditory working memory tasks, whereas the WMS-IV VWMI contains visual and spatial working memory tasks. The primary benefit of this structure is interpretive clarity, as some theories of working memory posit distinct systems underlying visual and auditory information processing (Baddeley, 2003).

Immediate Memory Index

Performance on the IMI is an indication of an examinee's ability to recall presented information immediately. It is a good indication of the examinee's ability to encode auditory and visual information. It consists of the immediate-recall conditions of Logical Memory, Verbal Paired Associates, Designs, and Visual Reproduction in the Adult battery, and the immediate-recall conditions of Logical Memory, Verbal Paired Associates, and Visual Reproduction in the Older Adult battery. Since it includes measures of auditory and visual memory, discrepancies between performances on these modalities will affect performance on this index.

Delayed Memory Index

Performance on the DMI is an indication of the examinee's ability to recall presented information following a delay in which other tasks are being completed. It is a good indication of an examinee's ability to consolidate and retrieve auditory and visual information. Although the Designs subtest includes aspects of recognition memory, recognition tasks are not included in the DMI. The DMI consists of the delayed-recall conditions of Logical Memory, Verbal Paired Associates, Designs, and Visual Reproduction in the Adult battery, and the delayed-recall conditions of Logical Memory, Verbal Paired Associates, and Visual Reproduction in the Older Adult battery. Since it includes measures of auditory and visual memory, discrepancies between performances on these modalities will affect performance on this index.

Psychometric Properties

Normative Sample

The WMS-IV has excellent normative samples, with 900 individuals included in the Adult battery sample and 500 included in the Older Adult battery sample. The samples were divided into 14 age bands (9 in the Adult battery and 5 in the Older Adult battery), with 100 individuals in each age band (see Wechsler, 2009, for detailed descriptions of the age bands). The stratification of the normative sample matched 2005 U.S. census data closely on five key demographic variables: age, gender, race/ethnicity, educational level, and geographic region.

The co-collection of the WAIS-IV and WMS-IV normative samples provided the opportunity for the mean ability level for each WMS-IV age band to be set at a GAI of 100, without requiring weighting of the normative sample. This ensured that the norms for the WMS-IV were not biased due to a high- or low-ability sample. In addition, the normative sample was screened for cognitive impairment and suboptimal effort.

Reliability

The reliabilities for the WMS-IV are good to excellent (Groth-Marnat & Wright, 2016; Wechsler, 2009) as indicated by internal-consistency and test–retest reliability estimates and interscorer agreement rates. The overall internal-consistency reliability coefficients for the normative samples are in the .90s for all index scores. At the subtest level, the overall internal-consistency reliability coefficients of the normative samples are in the .80s or .90s for all of the primary subtest scores, with the exception of Verbal Paired Associates II in the Older Adult battery. The lower reliability (.74) in this subtest is due to the small range of scores available in this measure. The internal-reliability coefficients for the process scores are lower, ranging from .74 to .77. The internal-consistency coefficients calculated using clinical samples are generally higher but fairly consistent with those obtained from the normative group, with all subtest coefficients in the .86–.97 range. The test–retest stability coefficients for the WMS-IV tend to be lower than those observed in the WAIS-IV, due to ceiling effects, fluctuations in motivation, and practice effects, which are observed in most memory measures (Strauss et al., 2006). Stability coefficients for the index scores range from .80 to .87, and the subtest stability coefficients range from .64 to .79. The test–retest stability of the process scores ranges from .59 to .76. All WMS-IV subtests have interscorer agreement above .90, with most WMS-IV subtests at .98–.99.

Validity

There is strong evidence to support the validity of the WMS-IV (Drozdick et al., 2011; Groth-Marnat & Wright, 2016; Wechsler, 2009). The confirmatory factor-analytic studies reported in the WMS-IV technical and interpretive manual provide strong evidence of construct validity for the AMI, VMI, and VWMI. As described earlier, the IMI and DMI are not supported by factor-analytic studies, due to the high correlation of these indexes. Further evidence of construct validity has also been provided by independent examinations of the Adult battery normative sample. A series of exploratory principal-component analyses conducted by Hoelzle and colleagues (2011) on each of the Adult battery normative age bands supported a two-factor structure for the WMS-IV, differentiating auditory and visual factors. This study failed to differentiate the VMI and VWMI, but confirmed the visual factor. Factor analyses examining the WAIS-IV and WMS-IV together also support the construct validity of the instruments (Drozdick, Holdnack, Weiss, & Zhou, 2013; Miller, Davidson, Schindler, & Messier, 2013).

In addition to factor-analytic support, the WMS-IV correlates highly with its predecessor and with other memory measures. The WMS-IV technical and interpretive manual (Wechsler, 2009) reports the correlations of the WMS-IV with the WMS-III, WMS-III Abbreviated, CVLT-II, and Children's Memory Scale (CMS; Cohen, 1997). In general, the correlations with the WMS-III are high, although they are lower in the visual memory domain because of the significant changes to visual memory measures in the WMS-IV. With regard to the other memory measures, the correlations are low to moderate, with higher correlations in measures assessing similar constructs (e.g., the AMI with Verbal Immediate Memory in the CMS). Finally, the WMS-IV has good predictive validity, as it correlates highly with related composites from the WIAT-II (Harcourt Assessment, 2005), with correlation coefficients ranging from .29 to .77. Correlations with the D-KEFS, RBANS, Texas Functional Living Scales (Cullum, Saines, & Weiner, 2009), and Independent Living Scales (Loeb, 1996) are also reported; these support the concurrent validity and structure of the WMS-IV (Cullum et al., 2009; Wechsler, 2009).

ADMINISTRATION OPTIONS

The WAIS-IV and WMS-IV can be administered digitally via the Q-interactive™ system. Briefly, the Q-interactive system allows an examiner to administer tests by using two tablets connected via Bluetooth. The examiner uses one to read instructions, record and score examinee responses, and send visual stimuli to the examinee's tablet. The examinee uses the other to view and respond to visual stimuli. Q-interactive is described in more detail elsewhere in this volume (Wahlstrom et al., Chapter 9). Furthermore, an in-depth review of the WAIS-IV on Q-interactive can be found in Cayton, Wahlstrom, and Daniel (2013).

The WAIS-IV and WMS-IV in digital format are scored by using the same normative data as those for the original paper-based tests. As a result, the approach to "digitizing" the tests has been conservative, in order to ensure construct and total raw score equivalence across the two formats (e.g., whether in paper or in digital format, WAIS-IV Block Design requires the examinee to use the traditional blocks; WAIS-IV processing speed tasks and WMS-IV Visual Reproduction require paper response booklets; and WMS-IV Designs requires the physical grid and cards). The conservative design approach and continued use of manipulatives ensure that the response processes used to complete the tasks are identical regardless of whether the tests are given in digital or paper format, and it allows for similar interpretation of test results in both modalities.

Although the design approach to the WAIS-IV and WMS-IV was guided by the requirement of construct equivalence, it is possible that scores could differ between the digital and paper formats for reasons the development team did not anticipate, which could negate the use of the paper norms to score the tests administered in digital format. In order to test this assumption of equivalence, studies were conducted that compared the raw scores obtained between paper and digital administrations of the WAIS-IV and WMS-IV (Daniel, 2012, 2013). Different research designs were utilized across the studies, but for each it was determined a priori that equivalence would be defined by an effect size of less than or equal to 0.20 (i.e., slightly over 0.5 scaled score points).

In the WAIS-IV study, each examiner administered half of the subtests in paper format and the other half of the subtests in digital format, with the order counterbalanced so that every subtest was administered in both paper and digital format. For each subtest, the format effect was calculated by comparing the score from the digital format to a predicted score obtained from the other subtests administered in paper format, scores from the WAIS-IV normative sample, and demographic variables. Scores obtained from the digital format that were significantly higher or lower than the predicted score would indicate a significant format effect. The results of this initial study suggested that 12 of 15 subtests were equivalent, with Coding, Information, and Picture Completion all yielding effect sizes greater than 0.20. Picture Completion was found to be more difficult in the digital format. This was determined to be due to subtle differences in the clarity of the stimuli. When the clarity of the art for the digital format was adjusted to match more closely what was seen on paper, a follow-up study confirmed that there was not a significant format effect. Coding was easier in the initial study—which is a curious finding, considering that the examinee completes the task in a response booklet in both the paper and digital formats, and that the only digital adaptation of this task involved replacing a hand-held stopwatch with a timer on the Practitioner Device. Investigation revealed that an issue with the timer calibration was allowing additional time in the digital format. When this issue was corrected, the format effect dropped below 0.20. Finally, Information was determined to be more difficult in the digital format, with no apparent explanation of the reason why, since the questions are asked and answered verbally in the same manner as in the paper format. Follow-up analysis of the paper Information cases found that when the same prediction analysis was applied to them, scores were also lower than what the predicted score suggested. The reason for this sample effect is unknown, but when accounted for the digital format effect was within a priori standards.

The WMS-IV study utilized three different study designs: (1) an equivalent-groups method, in which individuals were randomly assigned to take the tests in either paper or digital format; (2) a test–retest design, in which examinees took the subtests in both paper and digital format in a counterbalanced order; and (3) a dual-capture method, in which video-recorded administrations of tests were scored by examiners using either the paper or digital format. More detail about why these study designs were selected can be found in Daniel (2013), but note that that the dual-capture design was deemed appropriate only for subtests where the stimulus presentation and the examinee's response do not require a tablet (e.g., Logical Memory). In these subtests, the only threat to equivalence is the examiner's interaction with the tablet. Overall, the studies indicate no format effects between paper and digital administrations of WMS-IV subtests.

One question of particular interest is whether older individuals, who have less digital experience than their younger counterparts (Hart, Chaparro, & Halcomb, 2008), might struggle with the digital format of the WAIS-IV and WMS-IV. The WAIS-IV equivalence study found no correlation between format effect and age, which suggests that this is not the case. Perhaps this result is not surprising. Because of the conservative design approach mentioned above, there are few requirements for the examinee to interact with the tablet. Many subtests do not require a tablet at all, and even those that use the tablet to log an examinee's responses (e.g., Matrix Reasoning) require only a simple tap. In addition, the responses are numbered so that if need be, the examinee can provide the response verbally.

These findings, in conjunction with the overall lack of format effects, suggest that the WAIS-IV and WMS-IV can be administered in either format, and that the reliability and validity information that has been established with the original normative samples can be applied to both. Furthermore, interpretation of the subtests is unchanged by format.

CLINICAL APPLICATIONS

Because the Wechsler intelligence scales are reliable and valid instruments for comprehensive assessment of general cognitive functioning, clinicians have found many clinical applications of these instruments. The WAIS-IV is one of several key instruments for the assessment or diagnosis of (1) psychoeducational and developmental disorders, such as developmental delay, developmental risk, intellectual disabilities, learning disabilities, ADHD, language disorders, motor impairment, and autism; (2) intellectual giftedness; (3) neuropsychological disorders, such as TBI, Alzheimer disease, Huntington disease, Parkinson disease, multiple sclerosis, and temporal lobe epilepsy; and (4) alcohol-related disorders, such as chronic alcohol abuse and Korsakoff syndrome. In addition to the many clinical validation studies reported in the WAIS-IV technical and interpretive manual (Wechsler, 2008), please refer to Lichtenberger and Kaufman (2012), Sattler and Ryan (2009), and Weiss and colleagues (2010) for more detailed discussion of the clinical utility of the WAIS-IV.

The various editions of the WMS are among the mostly widely used assessments of memory functioning (Rabin et al., 2016). The WMS-IV can be an instrumental component of any evaluation involving memory functioning, including evaluations of (1) memory complaints, such as those seen with dementia, depression, and mild cognitive impairment; (2) neuropsychological disorders, such as brain injury or insult, temporal lobe epilepsy, or multiple sclerosis; (3) medical disorders involving memory difficulties, such as lupus, systemic illnesses, or central nervous system infections; (4) psychiatric disorders that influence memory functioning, such as depression, anxiety, or schizophrenia; (5) learning disorders; and (6) substance use disorders or exposure to toxic substances, such as Korsakoff syndrome or heavy-metal exposure. See the WMS-IV technical and interpretive manual (Wechsler, 2009) for the clinical validation studies of the WMS-IV. In addition, see Drozdick and colleagues (2011), Groth-Marnat and Wright (2016), Lichtenberger and colleagues (2002), and Strauss and colleagues (2006) for more information on clinical utility of the WMS-IV and WMS-III.

A number of issues need to be considered when clinicians are administering the WAIS-IV and WMS-IV in conjunction with each other. First, the WMS-IV should be administered prior to the WAIS-IV if they are given during the same testing session. Zhu and Tulsky (2000) found small order effects on the WMS-III when it was administered following the WAIS-III, but not vice versa. Second, as described previously, the GAI is used instead of the FSIQ for ability–memory comparisons. Finally, the WAIS-IV and WMS-IV both assess working memory. Although they measure different modalities, only one of the index scores needs to be obtained if modality-specific weaknesses are not observed.

INTERPRETATION

In addition to the basic interpretation steps and procedures suggested in the technical and interpretive manuals of the WAIS-IV and the WMS-IV (Wechsler, 2008, 2009), many interpretation strategies, methods, and procedures developed by experienced clinicians and researchers for the previous and current versions of the Wechsler intelligence scales are valid and useful (Lichtenberger & Kaufman, 2012; Sattler & Ryan, 2009; Weiss et al., 2010). Although a detailed discussion of interpretation strategies, methods, and procedures is beyond the scope of this chapter, the following interpretive considerations may help readers understand the nature of clinical interpretation. However, these suggestions should not be used as a "cookbook" or comprehensive guideline for interpretation. Clinical interpretation is a very complicated hypothesis-testing process that varies across evaluations. Therefore, no single approach will work for all scenarios, and WAIS-IV and WMS-IV data should always be interpreted within the context of such information as medical and psychosocial history, behavioral observations, and referral questions.

Basic Interpretation of Wechsler Scores

Scores

The Wechsler scales utilize two types of standard scores: scaled scores and composite scores (i.e., FSIQ and index scores). Converting raw scores into standard scores allows clinicians to interpret scores within the Wechsler scales and between the Wechsler scales and other related measures. The scaled scores and composite scores are age-corrected standard scores that allow comparison of each individual's cognitive functioning with other individuals in the same age group.

Scaled scores are derived from the total raw scores on each subtest. They are scaled to a metric with a mean of 10 and a standard deviation (*SD*) of 3. A subtest scaled score of 10 reflects the average performance of a given age group. Scores of 7 and 13 correspond to 1 *SD* below and above the mean, respectively, and scaled scores of 4 and 16 deviate 2 *SD*s from the mean.

Composite scores (e.g., the FSIQ, PRI, and AMI) are standard scores derived from various combinations of subtest scaled scores. They are scaled to a metric with a mean of 100 and an *SD* of 15. A score of 100 on any composite defines an average performance. Scores of 85 and 115 correspond to 1 *SD* below and above the mean, respectively, and scores of 70 and 130 deviate 2 *SD*s from the mean.

In general, standard scores provide the most accurate descriptions of test data. However, for individuals unfamiliar with test interpretation, standard scores are often difficult to understand. Other methods, such as percentile ranks and verbal descriptive classifications, are often used in conjunction with standard scores to describe an examinee's performance. Composite scores should be reported with confidence intervals, so that each score is evaluated in light of the score's reliability. Confidence intervals delineate a range of scores in which the examinee's true score is most likely to fall and remind the examiner that the observed score contains measurement error.

For several process scores on the WMS-IV, the majority of individuals in the normative sample obtained perfect or near-perfect scores, resulting in a skewed distribution of raw scores. Cumulative percentage ranges are used to describe performance for these scores (i.e., >2%, 3–9%, 10–16%, 17–25%, 26–50%, 51–75%, and >75%). Cumulative percentages describe the percentage of individuals who obtained the same or lower score on a task as the examinee. For example, a cumulative percentage range of 51–75 means that the examinee scored as well as 51–75% of the normative sample.

Level of Performance

The composite scores can be characterized as falling within a certain level of performance (e.g., superior, high average, average, low average). The *level of performance* refers to the rank obtained by an individual on a given test, compared to the performance of an appropriate normative group. The descriptive classifications corresponding to the WAIS-IV and WMS-IV composite scores are presented in Table 16.2.

TABLE 16.2. Descriptive Classifications of WAIS-IV and WMS-IV Composite and Index Scores

Score	Classification	Percent included in theoretical normal curve
130 and above	Very superior	2.2
120–129	Superior	6.7
110–119	High average	16.1
90–109	Average	50.0
80–89	Low average	16.1
70–79	Borderline	6.7
69 and below	Extremely low	2.2

Descriptive classifications allow communication of results in terms most individuals can comprehend. Test results can be described in a manner similar to the following example: "Compared to individuals of similar age, the examinee performed in the low average [descriptive classification] range on a measure of general intelligence [domain content]."

For clinical evaluations, the level of performance provides an estimate of the presence and severity of relative strengths or weaknesses in an individual's performance. This can help with clinical decisions for individuals whose level of performance is significantly lower than that of the normative group, either overall or within specific cognitive or memory domains. Alternatively, clinical decisions can be based on relative strengths and weaknesses within an individual's scores (e.g., a specific score is significantly lower than the individual's other scores, representing an intraindividual weakness).

Description of Composite Scores

For both the WAIS-IV and the WMS-IV, the composite scores are more reliable than the subtest scaled scores, and in general they are the first scores examined by the practitioner. On the WAIS-IV, the FSIQ is typically the first score reported and described, as it is the most reliable score and describes the examinee's general intellectual functioning. The standard score is typically provided, as well as the percentile rank and the confidence interval surrounding the score.

Although it is the best single-point predictor of cognitive ability, there are many cases in which additional scores are required to portray a person's cognitive ability accurately, especially cases characterized by extreme discrepancies between index scores. See other recently released volumes for guidance on interpreting the FSIQ in cases with extreme index score variability (Lichtenberger & Kaufman, 2012; Weiss et al., 2010). However, note that extreme index score discrepancies do not invalidate the FSIQ; rather, interpretation of the FSIQ should account for the index discrepancies. Regardless of the presence of index discrepancies, the four index scores should always be included with the FSIQ in the first level of interpretation, as differences in index score profiles may yield important clinical information. For example, research using previous editions of the WAIS has demonstrated that several clinical syndromes are characterized by unique index score profiles, including schizophrenia, autism, and TBI (Goldstein & Saklofske, 2010). It is important to note, however, that index score patterns are not diagnostic of such clinical disorders because the index scores tap broad cognitive abilities that may be impaired in several different clinical conditions. For example, working memory is typically impaired in patients with TBI as well as individuals with specific learning disabilities. Thus index score patterns are best thought of as indicating deficits in specific cognitive abilities, but not diagnostic of a specific clinical disorder. Patterns of index scores may be consistent or inconsistent with specific clinical disorders, but they cannot be diagnostic of them because such patterns overlap substantially across different disorders.

Analysis of Discrepancies among Composite Scores

Given the clinical significance of index score differences on the WAIS and WMS (Dori & Chelune, 2004; Goldstein & Saklofske, 2010; Lange & Chelune, 2006; Lange et al., 2006; Wilde et al., 2001) and the impact of extreme index score discrepancies on the interpretation of the FSIQ, composite score discrepancy analyses are an important aspect of WAIS-IV and WMS-IV interpretation. The statistical significance of the discrepancy between a pair of composite scores can be examined within and between the WAIS-IV and WMS-IV.

Three different methods can be used for score comparisons: simple difference, predicted difference, and contrast scaled scores. The simple-difference method can be used to compare scores within the WAIS-IV or WMS-IV, and the contrast scaled score approach can be used for comparing scores within the WMS-IV. In addition, all three methods can be used to evaluate ability–memory discrepancies between the WAIS-IV and WMS-IV.

For the simple-difference and predicted-difference methods, statistical significance and clinical significance of score differences are important concepts to understand. A *statistically significant difference* between scores (e.g., between the VCI and the PRI scores) means that the likelihood of obtaining a similar difference by chance is very low (e.g., $p < .05$) if the true difference between the scores is zero. The level of significance reflects the level of confidence a clinician can have that the difference between the scores, called the *difference score,* is a true difference. Along with statistical significance, it is important to consider the base rate of various difference scores in the general population. Even a statistically significant difference may occur frequently in normally developing and aging individuals. Often the difference between an individual's composite scores is significant in the statistical sense, but occurs frequently among individuals in the general population.

The statistical significance of discrepancies between scores, and the frequency of the difference in the normative population, are two separate issues and have different implications for test interpretation (Holdnack, Drozdick, Weiss, & Iverson, 2013; Sattler & Ryan, 2009). In order for a score difference to be considered clinically meaningful, it should be relatively rare. There are no strict guidelines to determine whether a significant difference is rare, and clinicians are advised to take into account medical history, cultural context, and other factors when making that decision. That being said, Sattler and Ryan (2009) advise that score differences occurring in fewer than 15% of the normative sample should be considered unusual. However, such discrepancies do not necessarily suggest a clinical condition, as at least one index score discrepancy of 23 points is observed in approximately 50% of examinees from both nonclinical and clinical populations (Kaufman, Raiford, & Coalson, 2016; Raiford & Coalson, 2014).

Simple- and Predicted-Difference Methods

The simple-difference method involves the subtraction of an obtained score from a second obtained score. In the predicted-difference method, the ability score is used in a regression equation to calculate a predicted score. The examinee's actual performance is then compared to the predicted score. These methods are easy to compute and to interpret. The score difference is identified as either a relative strength or a relative weakness, or no difference is observed between the variables. To determine whether the scores are statistically different, the difference score for each comparison is compared to the established cutoff required for statistical significance. The WAIS-IV and WMS-IV administration and scoring manuals provide clinicians with the minimum differences between index scores required for statistical significance at the .15 and .05 levels by age group. The manuals also provide base rates of the various index score discrepancies in the normative sample. The WMS-IV technical and interpretive manual also provides this information for the ability-memory discrepancies.

Contrast Scaled Score Method

The contrast scaled score methodology applies standard norming procedures to adjust a dependent measure by a control variable. These scores answer specific clinical questions, such as "Is the DMI above or below average, given the examinee's IMI score?" Tables for deriving the contrast scaled scores for the WMS-IV indexes are found in the WMS-IV administration and scoring manual. The tables for deriving the ability–memory contrast scaled scores are found in the WMS-IV technical and interpretive manual. A contrast scaled score is interpreted in the same manner as all scaled scores. For example, a 53-year-old examinee with a GAI of 120 and a DMI of 95 would obtain a contrast scaled score of 6, indicating borderline delayed-memory ability given the examinee's general ability.

Evaluation of Subtest Strengths and Weaknesses

Cognitive profiles describe an examinee's relative strengths and weaknesses. It is very common for examinees to perform better on some subtests than on others, and for practitioners to want to interpret these differences as meaningful. Given the frequency of subtest variation, the risk of overinterpretation is high. Practitioners can minimize this risk by assessing subtest strengths and weaknesses in the context of referral questions, as well as all sources of corroborating and contradicting evidence.

However, there are many situations in which subtest-level differences are important (e.g., neuropsychological evaluation of individuals with focal brain damage, interpretation of composite scores with only two subtests contributing), and the Wechsler scales provide three types of subtest score comparisons for these situations: comparison to the mean subtest score, difference scores, and contrast scaled scores.

First, in order to identify subtest strengths or weaknesses, the WAIS-IV provides the minimum difference between a single subtest and the mean of all 10 core subtests required for statistical significance at the .15 and .05 levels, as well as the base rates of these differences. Second, for cases in which it is more appropriate to compare two subtest scores directly (e.g., Digit Span and Arithmetic within the context of interpreting the WMI), the WAIS-IV provides the difference between all possible pairs of subtests required for statistical significance at the .15 and .05 levels, as well as the base rate information for the differences. The WMS-IV provides the same information for specified subtest comparisons. Finally, contrast scaled scores are provided for several WMS-IV subtest comparisons. For more advanced approaches to clinical interpretation of the WAIS-IV and WMS-IV, see Holdnack and colleagues (2013).

Appendix 16.1 provides a case study that demonstrates how the WAIS-IV and WMS-IV can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

APPENDIX 16.1

Brief Case Study

Examinee: Elizabeth H.
Report date: 6/13/2020
Age: 72 years
Date of birth: 2/18/1948
Gender: Female
Years of education: 14
Tests administered: Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV), Wechsler Memory Scale—Fourth Edition (WMS-IV)

REASON FOR REFERRAL

Mrs. H. was referred for an evaluation by her primary care physician, Dr. Jordan Lane, due to concerns about her memory and ability to manage her medications.

HISTORY

Mrs. H. is a 72-year-old widowed female who was born in the United States to parents who were first-generation immigrants from Mexico. She speaks English as her primary language and has 14 years of education in the United States. She attended college for 2 years before marrying and raising five children. She has good relationships with her children and grandchildren, and is active in several church and charity groups. She currently lives alone with some assistance from her children. Recently her children have noticed that she frequently misplaces items, repeats stories, and has trouble organizing her bills and calendar. Mrs. H. denies significant problems, but reports feeling more tired and admits forgetting where she places things. She also reports that it is harder for her to learn new things, and that she is sometimes overwhelmed by preparing for meetings and appointments.

BEHAVIORAL OBSERVATIONS

Mrs. H. was well groomed and polite. She exhibited appropriate behavior throughout the assessment. She was alert and oriented to time and place. She appeared to give forth good effort and did not require redirection to the testing. She appeared to fatigue easily, however, and several breaks were given to allow her to rest and refocus on testing.

TEST RESULTS

Mrs. H. completed the 10 core subtests of the WAIS-IV (see Tables 16.A.1 and 16.A.2 for her composite and subtest scores, respectively). She obtained a Full Scale IQ (FSIQ) of 88 (84–92 at the 95% confidence interval), indicating low average overall intellectual functioning in comparison to her same-age peers in the normative sample. Her performance across the index scores shows variability among the skills measured by the FSIQ. Specifically, she demonstrated average verbal comprehension (Verbal Comprehension Index [VCI] = 91 [85–98]) and perceptual reasoning (Perceptual Reasoning Index [PRI] = 100 [94–106]) abilities. However, her auditory working memory (Working Memory Index [WMI] = 80 [74–88]) and processing speed (Processing Speed Index [PSI] = 84 [77–94]) are in the low average range. The General Ability Index (GAI) was also computed in order to examine ability-memory discrepancies. Her GAI of 96 (91–101) indicates that her general verbal and nonverbal problem-solving ability is in the average range.

The WAIS-IV index comparisons for Mrs. H. show that verbal and perceptual reasoning abilities are significantly better than working memory. Similarly, processing speed is significantly lower than perceptual reasoning abilities. FSIQ is significantly lower than GAI. The score profile indicates relative weaknesses in both working memory and processing speed. Subsequently, FSIQ is lower

TABLE 16.A.2. Mrs. H.'s WAIS-IV Subtest Scores

Subtests	Scaled score	Percentile rank
Similarities	8	25
Vocabulary	8	25
Information	10	50
Block Design	8	25
Matrix Reasoning	14	91
Visual Puzzles	8	25
Digit Span	6	9
Arithmetic	7	16
Coding	8	25
Symbol Search	6	9

TABLE 16.A.1. Mrs. H.'s WAIS-IV Composite Scores

Scale	Composite score	Percentile rank	Confidence interval (95%)	Qualitative description
Verbal Comprehension Index (VCI)	93	32	88–99	Average
Perceptual Reasoning Index (PRI)	100	50	94–106	Average
Working Memory Index (WMI)	80	9	74–88	Low average
Processing Speed Index (PSI)	84	14	77–94	Low average
General Ability Index (GAI)	96	39	91–101	Average
Full Scale IQ (FSIQ)	88	21	84–92	Low average

than GAI, due to the impact of significantly lower processing speed and working memory. Although significant index-level variability in performance was observed, significant variability in subtest performance was only observed in the perceptual reasoning domain.

Mrs. H.'s performance on the measures within the PRI was variable. On a task of nonverbal visual–spatial reasoning, she performed above average (Matrix Reasoning = 14); however, scores on measures of nonverbal concept formation and nonverbal reasoning were in the average range (Block Design = 8, Visual Puzzles = 8). It is interesting to note that there were no time constraints on the subtest on which Mrs. H. performed in the above-average range. Her lower performance on the other tests may have been influenced by time constraints. Outside of this relative strength, Mrs. H.'s performance was relatively consistent across subtests and domains.

Mrs. H. also completed the WMS-IV (see Tables 16.A.3 and 16.A.4 for her index and subtest scores, respectively) to examine her memory abilities. Her low average score on the Brief Cognitive Status Exam (BCSE) is consistent with her performance on the WAIS-IV. Her performance across the WMS-IV index scores shows variability. Specifically, she has borderline auditory memory ability (Auditory Memory Index [AMI] = 72 [67–80]) and low average visual and immediate memory abilities (Visual Memory Index [VMI] = 82 [78–87]; Immediate Memory Index [IMI] = 80 [75–87], respectively). However, her delayed memory is extremely low (Delayed Memory Index [DMI] = 67 [62–77]). Mrs. H. appears to be having fairly significant problems with her memory—in particular, her ability to recall information following a delay. On a measure of visual working memory, her performance was in the average range, suggesting relatively intact ability to encode and mentally manipulate visual information.

Mrs. H.'s DMI is significantly lower than her IMI (note that neither the IMI nor the DMI is compared to the AMI or VMI, because they share common subtests). Contrast scores also indicate that delayed memory is extremely low, given her immediate recall ability. The AMI and VMI are not significantly different, suggesting that the memory problems Mrs. H. is experiencing are not specific to either auditory or visual memory.

TABLE 16.A.4. Mrs. H.'s WMS-IV Subtest Scores

Subtests	Scaled score	Percentile rank
Logical Memory I	6	9
Logical Memory II	3	1
Verbal Paired Associates I	6	9
Verbal Paired Associates II	6	9
Visual Reproduction I	8	25
Visual Reproduction II	5	5
Symbol Span	8	25
BCSE	—	Low average

Within the AMI, IMI, and DMI, none of Mrs. H.'s subtest scores were significantly different from the mean of the subtest scaled scores within the domain. However, a significant pairwise difference between Visual Reproduction I and II is evident. This suggests relatively stable performance within domains. The contrast scores for the primary subtest scores revealed relatively consistent performance on Verbal Paired Associates delayed recall when immediate-recall abilities were considered. However, delayed-recall scores on both Logical Memory and Visual Reproduction were unexpectedly low, given her immediate recall ability. Overall, Mrs. H. appears to be having difficulties with long-term memory. In addition, delayed memory appears to be a significant weakness compared to immediate memory.

In addition to the recall conditions, Mrs. H. also completed optional WMS-IV conditions, including the recognition trials for all three delayed-memory conditions—those in Logical Memory, Verbal Paired Associates, and Visual Reproduction; the Word Recall condition of Verbal Paired Associates; and the Copy condition of Visual Re-

TABLE 16.A.3. Mrs. H.'s WMS-IV Index Scores

Scale	Composite score	Percentile rank	Confidence interval (95%)	Qualitative description
Auditory Memory Index (AMI)	72	3	67–80	Borderline
Visual Memory Index (VMI)	82	12	78–87	Low average
Immediate Memory Index (IMI)	80	9	75–87	Low average
Delayed Memory Index (DMI)	67	1	62–77	Extremely low

production. The optional conditions provide additional information about Mrs. H.'s performance on specific subtests. For both verbal memory subtests, recall memory was slightly lower than expected, given the examinee's ability to recognize information. Her performance on Logical Memory II Recognition was extremely low, suggesting that her memory does not improve with cueing and that her problems with delayed memory are due to an inability to encode and retain verbal information. On Verbal Paired Associates Recognition, she performed in the low average range (17–25%) for delayed recognition and word recall (scaled score = 6). The Visual Reproduction optional conditions indicate that Mrs. H. had average recognition (26–50%) for the visual designs. Her ability to copy the designs was also in the average range (26–50%). The recognition and immediate-recall versus delayed-recall contrast scaled scores indicate that delayed free recall is unexpectedly low compared to both recognition and immediate encoding ability, indicating difficulties with encoding and retrieval of visual details and spatial relations among the details. The Visual Reproduction Copy versus Immediate Recall scores show that her immediate-memory performance was lower than expected, given her ability to directly copy the design. Thus the low scores on this subtest are not likely to be attributable to poor motor control.

The WAIS-IV and WMS-IV index comparisons were completed to determine whether Mrs. H.'s memory functioning is consistent with her general cognitive functioning. Table 16.A.5 displays Mrs. H.'s WAIS-IV versus WMS-IV index comparisons, using the predicted-difference method and contrast scores. When the predicted-difference discrepancy model was applied with GAI as the predictor, significant differences were observed between GAI and AMI (base rate = 2–3%), VMI (base rate = 10%), IMI (base rate = 5–10%), and DMI (base rate >1%). These results indicate that memory functioning is significantly lower than expected, considering Mrs. H.'s general problem-solving ability. Her low memory scores are not likely to be attributable to low general ability. The contrast scaled scores also suggest memory problems that are not consistent with her general cognitive abilities. When the GAI was controlled for, her AMI, VMI, IMI, and DMI were all lower than expected.

Mrs. H. was referred for this evaluation due to reports of memory difficulties. Her profile supports the presence of significant memory problems, in addition to problems with processing speed and verbal working memory. Although her performance in most domains was relatively stable, a strength was noted on an untimed measure of nonverbal fluid reasoning.

SUMMARY

Mrs. H. is a 72-year-old widowed female with a history of memory problems. She was administered the WAIS-IV and WMS-IV to determine her general cognitive and memory abilities. Her general intellectual ability is in the low average range, with weaknesses observed in verbal working memory and processing speed. Her memory performance was lower than expected given her general problem-solving ability, with particular problems with delayed memory. Differences were not observed between auditory and verbal memory. It is very likely that Mrs. H. is experiencing mild cognitive impairment.

TABLE 16.A.5. Mrs. H.'s Ability–Memory Discrepancy Analyses and Contrast Scores

	Predicted-difference method				
	Predicted WMS-IV index score	Actual WMS-IV index score	Difference	Critical value	Base rate
AMI	98	72	26	10.41	2–3
VMI	98	82	16	7.89	10
IMI	97	80	17	9.97	5–10
DMI	98	67	31	12.33	<1
Contrast scaled score method					
	WAIS-IV composite	WMS-IV index score	Contrast score		
GAI vs. AMI	96	72	4		
GAI vs. VMI	96	82	6		
GAI vs. IMI	96	80	5		
GAI vs. DMI	96	67	3		
VCI vs. AMI	93	72	4		
PRI vs. VMI	100	82	5		
WMI vs. AMI	80	72	6		

REFERENCES

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87,* 84–94.

Atkinson, R. C., & Shiffrin, R. M. (1968). A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 82–90). New York: Academic Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839.

Bauer, R. M. (2008). The three amnesias. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (pp. 729–742). New York: Taylor & Francis.

Baxendale, S., McGrath, K., & Thompson, P. J. (2014). Epilepsy and IQ: The clinical utility of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) indices in the neuropsychological assessment of people with epilepsy. *Journal of Clinical and Experimental Neuropsychology, 36*(2), 137–143.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22,* 121–130.

Beres, K. A., Kaufman, A. S., & Perlman, M. D. (2000). Assessment of child intelligence. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (3rd ed., pp. 65–96). New York: Pergamon Press.

Bowden, S. C., Lange, R. T., Weiss, L. G., & Saklofske, D. H. (2008). Invariance of the measurement model underlying the WAIS-III in the United States and Canada. *Educational and Psychological Measurement, 68,* 1024–1040.

Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011a). Augmenting the core battery with supplementary subtests: Wechsler Adult Intelligence Scale—IV measurement invariance across the United States and Canada. *Assessment, 18*(2), 133–140.

Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011b). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale—IV in the United States and Canada. *Educational and Psychological Measurement, 71*(1), 186–199.

Canivez, G. L. (2013). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment, 25,* 484–495.

Carlozzi, N. E., Grech, J., & Tulsky, D. S. (2014). Memory functioning in individuals with traumatic brain injury: An examination of the Wechsler Memory Scale—Fourth Edition (WMS-IV). *Journal of Clinical and Experimental Neuropsychology, 35,* 906–914.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cayton, T. G., Wahlstrom, D., & Daniel, M. H. (2013). The initial digital adaptation of the WAIS-IV. In E. O. Lichtenberger & A. S. Kaufman, *Essentials of WAIS-IV assessment* (2nd ed., pp. 389–426). Hoboken, NJ: Wiley.

Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory and Cognition, 42*(3), 448–463.

Cohen, J. (1952). A factor-analytically based rationale for the Wechsler–Bellevue. *Journal of Consulting Psychology, 16,* 272–277.

Cohen, J. (1957). A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology, 21,* 451–457.

Cohen, M. (1997). *Children's Memory Scale.* New York: Psychological Corporation.

Cohen-Mimran, R., & Sapir, S. (2007). Deficits in working memory in young adults with reading disabilities. *Journal of Communication Disorders, 40*(2), 168–183.

Colom, R., Rebollo, I., Palacios, A., Espinosa, M. J., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by *g. Intelligence, 32*(3), 277–296.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547–552.

Cullum, M., Saines, K., & Weiner, M. F. (2009). *Texas Functional Living Scales.* San Antonio, TX: Pearson.

Dang, C., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2015). Do processing speed and short-term storage exhaust the relation between working memory capacity and intelligence? *Personality and Individual Differences, 74,* 241–247.

Daniel, M. H. (2012). *Equivalence of Q-interactive-administered cognitive tasks: WAIS-IV* (Q-interactive Technical Report No. 1). Bloomington, MN: Pearson.

Daniel, M. H. (2013). *Equivalence of Q-interactive and paper administration of WMS-IV cognitive tasks* (Q-interactive Technical Report No. 6). Bloomington, MN: Pearson.

Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology, 39*(5), 1362–1369.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System.* San Antonio, TX: Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test—Second Edition.* San Antonio, TX: Psychological Corporation.

Donders, J., & Strong, C. H. (2015). Clinical utility of the Wechsler Adult Intelligence Scale—Fourth Edition after traumatic brain injury. *Assessment, 22*(1), 17–22.

Dori, G. A., & Chelune, G. J. (2004). Education-stratified base-rate information on discrepancy scores within and between the Wechsler Adult Intelligence Scale—Third Edition and the Wechsler Memory Scale—Third Edition. *Psychological Assessment, 16,* 146–154.

Drozdick, L. W., Holdnack, J. A., & Hilsabeck, R. (2011). *Essentials of WMS-IV assessment.* Hoboken, NJ: Wiley.

Drozdick, L. W., Holdnack, J. A., Weiss, L. G., & Zhou, X. (2013). Overview of the WAIS-IV/WMS-IV/ACS. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation* (pp. 1–73). Amsterdam: Elsevier.

Eichenbaum, H. (2008). *Learning and memory.* New York: Norton.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Fama, R., Rosenbloom, M. J., Sassoon, S. A., Pfefferbaum, A., & Sullivan, E. V. (2012). Differential effect of alcoholism and HIV infection on visuomotor procedural learning and retention. *Alcoholism: Clinical and Experimental Research, 36,* 1738–1747.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment.* Hoboken, NJ: Wiley.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237–241.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54,* 1–34.

Georgas, J., Weiss, L. G., van de Vijver, F. J. R., & Saklofske, D. H. (Eds.). (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III.* San Diego, CA: Academic Press.

Gfeller, J. D., Meldrum, D. L., & Jacobi, K. A. (1995). The impact of constructional impairment on the WMS-R visual reproduction subtests. *Journal of Clinical Psychology, 51,* 58–63.

Glass, L. A., Bartels, J. M., & Ryan, J. J. (2009). WAIS-III FSIQ and GAI in ability-memory discrepancy analysis. *Applied Neuropsychology, 16,* 19–22.

Golay, P., & Lacerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment, 23,* 143–152.

Goldstein, G., Allen, D. N., Minshew, N. J., Williams, D. L., Volkmar, F., Klin, A., et al. (2008). The structure of intelligence in children and adults with high functioning autism. *Neuropsychology, 22*(3), 301–312.

Goldstein, G., & Saklofske, D. H. (2010). The Wechsler intelligence scales in the assessment of psychopathology. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 189–216). London: Elsevier.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13,* 1–4.

Grillon, M.-L., Krebs, M.-O., Gourevitch, R., Giersch, A., & Huron, C. (2010). Episodic memory and impairment of an early encoding process in schizophrenia. *Neuropsychology, 24,* 101–108.

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Hoboken, NJ: Wiley.

Harcourt Assessment. (2005). *Wechsler Individual Achievement Test—Second Edition: Update 2005.* San Antonio, TX: Author.

Hart, T., Chaparro, B., & Halcomb, C. (2008). Evaluating websites for older adults: Adherence to senior-friendly guidelines and end-user performance. *Behaviour and Information Technology, 27*(3), 191–199.

Hoelzle, J. B., Nelson, N. W., & Smith, C. A. (2011). Comparison of Wechsler Memory Scale—Fourth Edition (WMS-IV) and Third Edition (WMS-III) dimensional structures: Improved ability to evaluate auditory and visual constructs. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 283–291.

Holdnack, J. A., & Drozdick, L. W. (2010). Using WAIS-IV with WMS-IV. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 237–283). London: Elsevier.

Holdnack, J. A., Drozdick, L. W., Weiss, L. G., & Iverson, G. L. (Eds.). (2013). *WAIS–IV, WMS–IV, and ACS: Advanced clinical interpretation.* Amsterdam: Elsevier.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: Macmillan.

Hunt, E., & Madhyastha, T. M. (2012). Cognitive demands of the workplace. *Journal of Neuroscience, Psychology, and Economics, 5,* 18–37.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences USA, 105*(19), 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory and Cognition, 42*(3), 464–480.

Johnson, W., Corley, J., Starr, J. M., & Deary, I. J. (2011). Psychological and physical health at age 70 in the Lothian birth cohort 1936: Links with early life IQ, SES, and current cognitive function and neighborhood environment. *Health Psychology, 30,* 1–11.

Kaufman, A. S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence, 29,* 131–167.

Kaufman, A. S. (2010). Foreword. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. xiii–xxii). London: Elsevier.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman*

Adolescent and Adult Intelligence Test. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Hoboken, NJ: Wiley.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic *g* one and the same *g*?: An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123–138.

Keage, H. A. D., Kurylowicz, L., Lavrencic, L. M., Churches, O. F., Hofmann, J., Kohler, M., et al. (2015). Cerebrovascular function associated with fluid, not crystallized, abilities in older adults: A transcranial Doppler study. *Psychology and Aging, 30,* 613–623.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences, 49,* 331–336.

Lafosse, J. M., Mitchell, S. M., Corboy, J. R., & Filley, C. M. (2013). The nature of verbal memory impairment in multiple sclerosis: A list-learning and meta-analytic study. *Journal of the International Neuropsychological Society, 19,* 995–1008.

Lampinen, J. M., & Beike, D. R. (2014). *Memory 101*. New York: Springer.

Lange, R. T., & Chelune, G. J. (2006). Application of new WAIS-III/WMS-III discrepancy scores for evaluating memory functioning: Relationship between intellectual and memory ability. *Journal of Clinical and Experimental Neuropsychology, 28,* 592–604.

Lange, R. T., Chelune, G. J., & Tulsky, D. S. (2006). Development of WAIS-III General Ability Index minus WMS-III memory discrepancy scores. *The Clinical Neuropsychologist, 20,* 382–395.

Larrabee, G. J., & Curtiss, J. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Neuropsychology, 17,* 536–547.

Lee, H. F., Gorsuch, R., Saklofske, D. H., & Patterson, C. (2008). Cognitive differences for ages 16 to 89 (Canadian WAIS-III): Curvilinear with Flynn and processing speed corrections. *Journal of Psychoeducational Assessment, 26*(4), 382–394.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Lichtenberger, E. O., Kaufman, A. S., & Lai, Z. C. (2002). *Essentials of WMS-III assessment*. New York: Wiley.

Lindner, O. C., Phillips, B., McCabe, M. G., Mayes, A., Wearden, A., Varese, F., et al. (2014). A meta-analysis of cognitive impairment following adult cancer chemotherapy. *Neuropsychology, 28,* 726–740.

Loeb, P. A. (1996). *Independent Living Scales*. San Antonio, TX: Psychological Corporation.

Manard, M., Carabin, D., Jaspar, M., & Collette, F. (2014). Age-related decline in cognitive control: The role of fluid intelligence and processing speed. *BMC Neuroscience, 15,* 7.

Mathias, J. L., & Wheaton, P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: A meta-analytic review. *Neuropsychology, 21*(2), 212–223.

McGrew, K. S., & Flanagan, D. P. (1997). Beyond *g*: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26*(2), 189–210.

Miller, D. I., Davidson, P. S. R., Schindler, D., & Messier, C. (2013). Confirmatory factor analysis of the WAIS-IV and WMS-IV in older adults. *Journal of Psychoeducational Assessment, 31,* 375–390.

Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the WMS-III. *Journal of Clinical and Experimental Neuropsychology, 21,* 87–93.

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale—Fourth Edition with a clinical sample. *Psychological Assessment, 25*(2), 618–630.

Nielsen, N. P., & Wiig, E. H. (2011). AQT cognitive speed and processing efficiency differentiate adults with and without ADHD: A preliminary study. *International Journal of Psychiatry in Clinical Practice, 15*(3), 219–227.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition*. San Antonio, TX: Author.

Perlow, R., Jattuso, M., & Moore, D. D. (1997). Role of verbal working memory in complex skill acquisition. *Human Performance, 10*(3), 283–302.

Pineda-Pardo, J. A., Martínez, K., Román, F. J., & Colom, R. (2016). Structural efficiency within a parieto-frontal network and cognitive differences. *Intelligence, 54,* 105–116.

Porter, R. J., Robinson, L. J., Malhi, G. S., & Gallagher, P. (2015). The neurocognitive profile of mood disorders: A review of the evidence and methodological issues. *Bipolar Disorders, 17,* 21–40.

Prifitera, A., Saklofske, D. H., & Weiss, L. G. (Eds.). (2005). *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives*. Amsterdam: Elsevier Academic Press.

Proctor, B. (2012). Relationships between Cattell–Horn–Carroll (CHC) cognitive abilities and math achievement within a sample of college students with learning disabilities. *Journal of Learning Disabilities, 45*(3), 278–287.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31,* 206–230.

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV assessment*. Hoboken, NJ: Wiley.

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: Psychological Corporation.

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51*(4), 535–555.

Rindermann, H., & Meisengerg, G. (2009). Relevance of education and intelligence at the national level for health: The case of HIV and AIDS. *Intelligence, 37,* 383–395.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44,* 2029–2040.

Russell, E. W. (1975). A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology, 43,* 800–809.

Russell, E. W. (1988). Renorming Russell's version of the Wechsler Memory Scale. *Journal of Clinical and Experimental Neuropsychology, 10,* 235–249.

Ryan, J. J., Gontkovsky, S. T., Kreiner, D. S., & Tree, H. A. (2012). Wechsler Adult Intelligence Scale—Fourth Edition performance in relapsing–remitting multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology, 34*(6), 571–579.

Salthouse, T. A. (2004). Localizing age-related individual differences in a hierarchical structure. *Intelligence, 32,* 541–561.

Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. San Diego, CA: Jerome M. Sattler.

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A commentary on four recent articles. *Perspectives on Psychological Science,* 9(2), 211–218.

Schneider, W. J. (2013). What if we took our models seriously?: Estimating latent scores in individuals. *Journal of Psychoeducational Assessment, 31,* 186–201.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schucard, J. L., Lee, W. H., Safford, A. S., & Schucard, D. W. (2011). The relationship between processing speed and working memory demand in systemic lupus erythematosus: Evidence from a visual *n*-back task. *Neuropsychology, 25,* 45–52.

Soble, J. R., Eichstaedt, K. E., Waseem, H., Mattingly, M. L., Benbadis, S. R., Bozorg, A. M., et al. (2014). Clinical utility of the Wechsler Memory Scale—Fourth Edition (WMS-IV) in predicting laterality of temporal lobe epilepsy among surgical candidates. *Epilepsy and Behavior, 41,* 232–237.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R., & Schacter, D. L. (2002). *Neuropsychology of memory* (3rd ed.). New York: Guilford Press.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Tulsky, D. S., Ivnik, R. J., Price, L. R., & Wilkins, C. (2003). Assessment of cognitive functioning with the WAIS-III and WMS-III: Development of a 6-factor model. In D. S. Tulsky, D. H. Saklofsky, G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. Bornstein, et al. (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 147–179). San Diego, CA: Academic Press.

Tulsky, D. S., Zhu, J., & Prifitera, A. (2000). Assessment of adult intelligence with the WAIS-III. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (3rd ed., pp. 97–130). New York: Pergamon Press.

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*(2), 328–340.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87–95.

Wechsler, D. (1950). Cognition, conative, and nonintellective intelligence. *American Psychologist, 5,* 78–83.

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist, 30,* 135–139.

Wechsler, D. (1997). *WAIS-III–WMS-III technical manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition*. San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale—Fourth Edition*. San Antonio, TX: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV clinical validation of the four- and five factor interpretive approaches [Special edition]. *Journal of Psychoeducational Assessment, 31*(2), 94–113.

Weiss, L. G., Saklofske, D. H., Coalson, D., & Raiford, S. E. (Eds.). (2010). *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives*. London: Elsevier.

Wilde, N., Strauss, E., Chelune, G. J., Loring, D. W., Martin, R. C., Hermann, B. P., et al. (2001). WMS-III performance in patients with temporal lobe epilepsy: Group differences and individual classification. *Journal of the International Neuropsychological Society,* 7(7), 881–891.

Zachary, R. A. (1990). Wechsler's intelligence scales: Theoretical and practical considerations. *Journal of Psychoeducational Assessment, 8,* 276–289.

Zhu, J., & Tulsky, D. S. (2000). Co-norming of the WAIS-III and WMS-III: Is there a test order effect on IQ and memory scores? *Clinical Neuropsychologist, 14*(4), 461–467.

CHAPTER 17

The Wechsler Nonverbal Scale of Ability

Assessment of Culturally and Linguistically Diverse Populations

Jack A. Naglieri
Tulio M. Otero

MEASUREMENT OF GENERAL ABILITY

As our environment becomes increasingly diverse in terms of ethnicities, languages, and disabilities/medical fragilities, nonverbal tests of general ability now meet the growing need for cognitive assessment, particularly in North America. Nonverbal tests help address the increased diversity and complexity of a society in which the evaluation of intellectual ability has rapidly become a common practice. Some examinees simply cannot demonstrate their ability through verbal means, and for these persons, nonverbal assessment offers a more effective alternative for obtaining estimates of ability. Researchers have found that nonverbal tests can be used for detecting strengths and weaknesses in children with disabilities (Mayes & Calhoun, 2003), for screening (Webster, Plante, & Couvillion, 1997), for predicting recovery and treatment outcomes (Matson, 2007), and for defining high- versus low-functioning subcategories within disability groups (Bishop, Richler, & Lord, 2006). Nonverbal tests are useful for persons with autism; deafness; speech impairment; traumatic brain injury; limited English proficiency; Intellectual disability; neurocognitive disorders such as dementia and stroke; selective mutism; and other disorders affecting communication (McCallum, 2017).

The assessment of intellectual ability extends as far back as the 1800s. Edouard Séguin, a French psychiatrist developed a test that became known as the *Séguin form board,* which measures a child's ability nonverbally (Pichot, 1948). The test requires placement of various geometric shapes to inserts of the same shape. The form board has since been modified and adapted for use in several other tests and cultures (see Venkatesan & Basavarajappa, 2007, for examples).

The use of nonverbal test questions to assess ability was part of the approach used by the U.S. Army in 1917. Robert Yerkes, then the president of the American Psychological Association, assembled a team of psychologists to design tests that would be used to determine whether Army recruits were fit for military service and advanced training (see Wasserman, Chapter 1, this volume, for a full account). The Army intelligence tests were called the Alpha and Beta, and were designed to be administered to groups of recruits by trained psychological examiners. The Army Alpha was designed to evaluate the general ability of literate English speakers, and the Army Beta was intended for recruits who were illiterate or nonproficient in English. These tests evolved into the Wechsler scales (Naglieri, 2015).

The concept of *general ability,* as measured by traditional IQ tests such as the Wechsler scales, has had a long and successful history in psychology and education—so much so that the tests have been used to define intelligence. There is strong support for a view of the Wechsler scales as good measures of general ability, using both verbal and nonverbal (i.e., visual–spatial) test questions. However, the visual–spatial tests (e.g., Block Design and Visual Puzzles) are often described as measures of *nonverbal ability,* and the verbal tests (e.g., Vocabulary and Similarities) as measures of *verbal ability,* even though Wechsler did not have any intention to measure verbal and nonverbal abilities. In fact, Wechsler did not view verbal and nonverbal tests as measures of two types of intelligence, even though his tests yielded Verbal and nonverbal (Performance) IQs. Wechsler stated that "the subtests are different measures of intelligence, not measures of different kinds of intelligence" (1958, p. 64), and he viewed verbal and performance tests as "equally valid measures of intelligence" (Wechsler & Naglieri, 2006a, p. 1). Furthermore, Bracken and Naglieri (2003) and Naglieri (2003a, 2003b, 2008a, 2008b) have clarified that the term *nonverbal* refers to the *content* of the test, not a type of ability, and that the goal of this type of test is to measure general ability.

Wechsler believed that nonverbal tests would help to "minimize the over-diagnosing of feeblemindedness that was, he believed, caused by intelligence tests that were verbal in content . . . and he viewed verbal and performance tests as equally valid measures of intelligence and criticized the labeling of performance [nonverbal] tests as measures of special abilities" (Boake, 2002, p. 396). Those individuals who have not had the chance to develop proficiency in English do poorly on verbal tests. As a result, their scores on such tests are not good reflections of their general ability to learn, despite their having been exposed to seemingly appropriate instruction. This problem is complicated by the fact that the content of traditional IQ test questions is often similar to that of questions found on tests of achievement.

Traditional IQ tests often include tests of word knowledge—for example, tests that ask the examinee to define vocabulary words or determine how two words are alike. We have described some examples of the similarity of test questions across ability and achievement tests in Chapter 6 of this volume, and similar examples can still be found in the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014) and the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014). For example, the WJ IV COG (like its predecessor in the WJ III) contains a Verbal Comprehension subtest item comparable to "Tell me another word for *small,*" and the WJ IV ACH contains a Reading Vocabulary question similar to "Tell me another word for *little.*" This type of overlap in content artificially increases the correlations between these tests of ability and achievement (thereby distorting the validity of the tests) and raises important questions about the utility of measuring intellectual ability with questions that are clearly achievement-laden. More importantly, however, these test questions pose particular problems for individuals with limited English-language proficiency.

One way to assess ability without the confounding variables of language and knowledge is to use a nonverbal test of ability. This type of test provides a way to assess individuals from diverse linguistic groups, especially those who have limited language skills, as well as children with language impairments. In addition, children who cannot tolerate a lengthy test battery—such as those with autism, those who are inattentive and/or hyperactive, or those who fatigue easily because of traumatic brain injury—are evaluated more easily with nonverbal tests, especially brief ones, than with verbal tests (see Stepler & Brown, 2016).

MEASUREMENT OF GENERAL ABILITY WITH NONVERBAL TESTS

The greatest advantage of a nonverbal test of general ability is that it measures intelligence without using test questions that are unduly reliant on verbal skills. The test questions evaluate *general* ability *nonverbally*—not *nonverbal ability*—via subtests with strong spatial requirements (e.g., blocks to make a design or progressive matrices). The specific administration format and subtests may vary, but the basic concept remains the same. For example, Bracken and McCallum (1998) suggest that an entire ability test must be administered nonverbally and that pantomimed instructions must be used. Others suggest that test directions for administration may be spoken, but that the content of the test questions should not require knowledge of a specific culture (Naglieri, 1997). Another method is to use pictorial directions, as found in the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006a) and the Naglieri Nonverbal Ability Test—Second and Third Editions (NNAT-2 and NNAT3; Naglieri, 2008a, 2016). Nonverbal tests of general ability also differ in the diversity of their subtests. For example, some nonverbal tests consist of one type of item such as the progressive matrix (e.g., the NNAT3), given in either a group or an individual format (e.g., the Naglieri Nonverbal Ability Test—Individual Form [NNAT-I]; Naglieri, 2003b). Another method is to use several different types of nonverbal subtests, as found in the WNV as well as the Universal Nonverbal Intelligence Test—Second Edition (UNIT2; Bracken & McCallum, 2016). Despite their differences in administration methods and subtest content, these nonverbal tests of general ability provide ways to assess a wide variety of individuals fairly, regardless of their educational or linguistic backgrounds and/or disabilities. The goal of fair and equitable assessment is as critical today as it was when the concepts of verbal and nonverbal assessment were introduced in the early 1900s.

Assessment of ability for populations of individuals who vary in educational and linguistic skills has been and continues to be important in the fields of both psychology and education. The initial conceptualization of verbal and nonverbal assessment was stated clearly by Yoakum and Yerkes (1920) when they wrote that the Army Beta (nonverbal) test was used during World War I instead of the Army Alpha (verbal) test to assess soldiers with limited skills in English, to avoid "injustice by reason of relative unfamiliarity with English" (p. 19). Today, as indicated above, there are several different types of nonverbal tests of general ability (for reviews, see Naglieri & Goldstein, 2009; McCallum, 2017). In the remainder of this chapter, we describe an individually administered measure of general ability, the WNV (Wechsler & Naglieri, 2006a).

STRUCTURE AND ADMINISTRATION OF THE WNV

The WNV consists of six subtests that are organized into two forms (one for ages 4–7 years and the other for ages 8–21 years) and two versions (a two-subtest and a four-subtest version). The subtests measure general ability with tasks that vary in specific requirements. Some of the subtests have a strong visual–spatial requirement; others demand recall of spatial information or recall of the sequence of information; and still others involve paper-and-pencil skills. The multidimensionality of these tasks distinguishes the WNV from tests such as the NNAT3 (Naglieri, 2016), which exclusively uses progressive matrices.

Subtests

Most of the WNV subtests have appeared in previous editions of the Wechsler scales and have an established record of reliability and validity for the nonverbal measurement of general ability. The origins and descriptions of the WNV subtests are provided in Table 17.1. Adaptation of the subtests was necessary to accommodate the new pictorial-directions format, identify items that were most appropriate for specific ages, and provide directions in several languages. WNV subtest norms (mean of 50 and standard deviation [*SD*] of 10) and Full Scale score norms (mean of 100 and *SD* of 15) are based on standardization samples collected in the United States and Canada. The subtests are briefly described below.

- *Matrices.* The Matrices subtest requires the examinee to discover how different geometric shapes are spatially or logically interrelated. The multiple-choice items are constructed of geometric figures such as squares, circles, and triangles, in some combination of the colors black, white, yellow, blue, and green. Matrices is always administered (i.e., it is given to examinees in both age bands and is included in both the four- and two-subtest batteries).

- *Coding.* The Coding subtest requires the examinee to copy symbols (e.g., two vertical lines, a dash) that are paired with simple geometric shapes or numbers according to a key provided at the top of the page. Form A is used in the four-subtest battery for ages 4 years, 0 months to 7 years, 11 months (4:0–7:11), and Form B is used in the four-subtest battery for ages 8:0–21:11.

- *Object Assembly.* The Object Assembly subtest consists of items that require the examinee to complete pieces of a puzzle to form a recognizable object, such as a ball or a car. Object Assembly is included in the four-subtest battery of the WNV for examinees ages 4:0–7:11.

- *Recognition.* The Recognition subtest was created for use in the WNV and is included in both the four- and two-subtest batteries for examinees ages 4:0–7:11. It requires the examinee to look at a stimulus (e.g., a square with a small circle in the center) for 3 seconds, and then to choose which option is identical to the stimulus that was just seen. The figures are colored black, white, yel-

TABLE 17.1. Origins and Description of WNV Subtests

Subtest	Origin and description
Matrices	This subtest was adapted from the Naglieri Nonverbal Ability Test—Individual Form (NNAT-I; Naglieri, 2003b). The examinee chooses an option that solves a progressive matrix.
Coding	This subtest was adapted from the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003). The examinee copies symbols that are paired with simple geometric shapes or numbers using a key that appears at the top of the page.
Object Assembly	This subtest was adapted from the Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III; Wechsler, 2002) and the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991). The child assembles puzzle pieces that make an object (e.g., a car).
Recognition	This is a new match-to-stimulus subtest. The child looks at a page with a design containing geometric patterns for 3 seconds. Then the child chooses which of several options matches the previously viewed stimulus.
Spatial Span	This subtest was adapted from the Wechsler Memory Scale—Third Edition (WMS-III; Wechsler, 1997a). The examinee repeats the examiner's touching of a series of blocks, either in the same order as the examiner or in the reverse order.
Picture Arrangement	This subtest is adapted from the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; Wechsler, 1997b) and a research version of the Wechsler Intelligence Scale for Children—Fourth Edition—Integrated (WISC-IV Integrated; Wechsler et al., 2004). The examinee puts picture cards in the correct order to tell a logical story within a specified time limit.

low, blue, and/or green, to maintain interest and minimize the likelihood that impaired color vision will influence the scores.

- *Spatial Span.* The Spatial Span subtest requires the examinee to touch a group of blocks arranged in an irregular pattern on an 8½-inch by 11-inch board in the same and reverse order demonstrated by the examiner. Spatial Span is included in both the four- and two-subtest batteries for ages 8:0–21:11.
- *Picture Arrangement.* The Picture Arrangement subtest involves cartoon-like illustrations, which must be put into a sequence that is logical and makes sense. Picture Arrangement is included in the four-subtest battery for examinees ages 8:0–21:11.

Administration

Administration of the WNV starts with short standardized introductions that tell examinees (1) to look at the pictorial directions to understand what to do, and (2) to ask the examiner questions if necessary. These introductions can be provided in English, French, Spanish, Chinese, German, or Dutch. The three steps below provide guidance on administration of the WNV.

1. The first step is to administer the test following standardized directions. These directions include gestures that correspond to the pictorial directions. Pictorial directions are used at step 1 to provide a standardized method of communicating the demands of each task. The pictorial directions show a scene like the one the examinee is in currently. The pictures show an examinee being presented with the question, then thinking about the item, and finally choosing the correct solution. These instructions include actions by the examiner that must be followed carefully, to ensure that the examinee understands the correspondence between the materials and the task. Gestures are used to direct the examinee's attention to specific portions of the pictorial directions and to the stimulus materials, and sometimes to demonstrate the task itself. Sometimes simple statements are also included to convey the importance of both time and accuracy to the examinee.

2. The second step involves using additional directions *as needed* after the standard directions are used. These instructions must be followed exactly and are only given when an examinee is unclear about what he or she is being asked to do. These directions include simple sentences and gestures for further communicating the requirements of

the task. The verbal instructions provide another way to ensure that the examinee understands the demands of the tasks; again, they can be given in English, Spanish, German, French, Chinese, or Dutch. These instructions are only used when two conditions are met: (a) The WNV is administered to an examinee who speaks one of the languages just mentioned; and (b) the examiner or a professional interpreter speaks the same language as the examinee.

3. The third step gives an examiner flexibility to communicate the demands of the task in any form. In general, examiners are given the opportunity to communicate in whatever manner they think will best clarify the demands of the subtest, based on their judgment of an examinee's needs. This may include providing further explanation or demonstration of the task, restating or revising the verbal directions, or using additional words (in a combination of languages, including sign language) to describe the requirements of the task. At no time, however, should an examiner teach an examinee how to solve the items. Instead, the goal is to provide additional help to ensure that the examinee understands the demands of the task. The amount of help provided and the determination about when to stop should both be based on professional judgment.

Using an Interpreter

When an interpreter is used to facilitate communication prior to and during administration, it is important for the interpreter to understand what is and what is not permitted. The interpreter should translate a general explanation of the testing situation for the examinee. It is especially important to translate the introductory paragraph at the beginning of Chapter 3 in the WNV administration and scoring manual (Wechsler & Naglieri, 2006b) before the administration begins. The interpreter must also recognize the boundaries of his or her role in administration. For example, although it is appropriate for the interpreter to translate the examiner's responses to an examinee's response to a sample item, it is not acceptable for the interpreter to make additional statements unless instructed to do so. Importantly, at no time should the interpreter communicate any information that could influence the examinee's scores. See Brunnert, Naglieri, and Hardy-Braz (2009) for more information about working with translators, especially in testing examinees who are deaf or hard of hearing.

SUBTEST ADMINISTRATION TIPS

Administration of the WNV subtests was designed to be straightforward. The WNV administration and scoring manual describes each subtest; the materials needed; start, stop, and reverse rules; scoring; and issues to consider that are unique to each subtest. Below, we highlight the most salient points about individual subtest administration.

- *Matrices.* On this subtest, examiners should be aware of possible responses that may suggest concern. For example, some students who are impulsive may select the option that is mostly but not completely correct. Because the response options were written with varying degrees of accuracy, an examinee who is impulsive may not carefully consider all options. If an examinee is not looking at the options closely, an option that is *almost* correct may be selected. Also important to note is that if an examinee takes a long time to respond (i.e., more than 30 seconds), the examiner may prompt a response.
- *Coding.* On this subtest, the examiner should ensure that the examinee works from left to right and from top to bottom without skipping any items or rows, by providing the appropriate instruction when needed. The examinee can correct mistakes by crossing out the incorrect symbol and writing the new response next to it. The examiner also briefly instructs the examinee to work as quickly as possible. For this reason, examinees should not be allowed to spend too much time making corrections.
- *Object Assembly.* On this subtest, it is important to remember that the examiner should set up the puzzle pieces on the same side of the stimulus book as the examinee's dominant hand, and the stimulus book should be removed before administering the sample item. The examiner should also ensure that the examinee works as quickly as possible. If the examinee is still completing a puzzle when the time limit expires, the examiner should place his or her hand over the puzzle to stop the examinee's progress and then record the examinee's answer. If the examinee seems upset at that point, the examiner should allow the examinee to finish, but should not consider any additional work for scoring purposes. It is also important to remember to begin timing after the last word of the instruction is provided. Note that the puzzle pieces for each item on this subtest are put before the examinee in a specific format *face down*. Once

all the pieces are in front of the examinee, they are turned over in the order indicated by the number on the back of each piece.

- *Recognition.* This subtest requires that the examinee expose each stimulus page for 3 seconds. Examinees are not permitted to turn the pages.
- *Spatial Span.* On this subtest, it is important to arrange the Spatial Span board so that an examinee can reach all cubes on the board easily. Also, the arrangement of the board should allow only the examiner to see the numbers on the back of each blue block. The Spatial Span board should always be placed on the same side of the stimulus book as the examinee's dominant hand. The blocks must be tapped at a rate of one per second, and the examiner should raise his or her hand approximately 1 foot above the Spatial Span board between each tap. If the examinee does not respond after the examiner taps a sequence, the examiner can say, "It's your turn." The examiner should always administer Spatial Span Backward, regardless of the examinee's performance on the initial Spatial Span Forward test.
- *Picture Arrangement.* On this subtest, the examiner should always place the Picture Arrangement cards on the same side of the stimulus book as the examinee's dominant hand. To facilitate administration, the cards for each item of this subtest should be arranged in the order in which they are to be exposed to the examinee. When the examinee completes the item, the examiner should record the sequence and then re-sequence the cards in the presentation order for the next administration. If the examinee is working too slowly, it is permissible to inform the examinee that he or she should work as quickly as possible. If the examinee orders the cards from right to left instead of left to right, the examiner should ask, "Where does it start?" If the examinee is still working when the time limit expires, the examiner should place his or her hand over the story to stop the examinee's progress, and record the examinee's answer. If the examinee seems upset at being stopped while completing the story, the examiner should allow the examinee to finish. No credit can be given for any work completed after the time has expired.

SCORING

Five of the six WNV subtests (i.e., Matrices, Coding, Recognition, Spatial Span, and Picture Arrangement) are scored by summing the number of points earned during administration. The sixth subtest (i.e., Object Assembly) has time bonuses for some items that may be part of the raw score. The raw scores are converted to *T* scores. The sum of *T* scores is converted to a Full Scale score. Percentile ranks and confidence intervals are included in the conversion table. The WNV Scoring Assistant is a computer scoring program that provides scores based on U.S. and Canadian samples. The report-writing feature of the scoring program provides reports for clinicians and parents. The parent report is available in English, French, and Spanish. The scoring program is also linked to the Wechsler Individual Achievement Test—Second Edition (Wechsler, 2001), which allows for comparisons between ability and achievement.

PSYCHOMETRIC PROPERTIES

Standardization

The WNV was standardized in the United States and Canada. The U.S. sample consisted of 1,323 children and adolescents stratified across five demographic variables: age (4:0–21:11); sex; race/ethnicity (described as black, white, Hispanic, Asian, and other); education level (8 years or less of school, 9–11 years of school, 12 years of school [high school diploma or equivalent], 13–15 years of school [some college or associate's degree], and 16 or more years of school [college or graduate degree]); and geographic region (Northeast, North Central, South, and West). Educational level was determined by parental education for examinees ages 4:0–17:11 and by the examinees' own education for ages 18:0–21:11. Approximately 4% of the U.S. normative sample consisted of individuals with limited English skills.

The Canadian sample consisted of 875 examinees stratified across five demographic variables: age (4:0–21:11); sex; race/ethnicity (described as Caucasian, Asian, First Nations, and other); education level (less than a high school diploma; high school diploma or equivalent; college/vocational diploma or some university, but no degree obtained; and a university degree); and geographic region (West, Central, and East). In addition, the Canadian sample consisted of 70% English speakers, 18% French speakers, and 12% speakers of other languages. See the WNV technical and interpretive manual (Wechsler & Naglieri, 2006c) for more details.

Reliability

WNV reliability coefficients are provided for subtest and Full Scale scores by age for the U.S. and Canadian normative samples and for the special groups described in the WNV technical and interpretive manual (Wechsler & Naglieri, 2006b). The reliability coefficients for the U.S. normative sample were .91 for both the two-subtest and four-subtest versions' Full Scale scores across ages, and .74–.91 for the subtests. The reliability estimates for the Canadian normative sample were .73–.90 for the subtests, .90 for the Full Scale score (four-subtest version), and .91 for the Full Scale score (two-subtest version). The reliability estimates for special-group studies (i.e., individuals with giftedness, mild and moderate intellectual disabilities, reading and written expression disorders, or language disorders; English-language learners; and individuals who were deaf or hard of hearing) are provided in the manual. Other information, such as the standard error of measurements (*SEM*), confidence intervals, and test–retest stability estimates for the U.S. and the Canadian normative samples, is also provided in the WNV technical and interpretive manual (Wechsler & Naglieri, 2006c) and administration and scoring manual (Wechsler & Naglieri, 2006b).

Validity

English as a Second Language

As the United States continues to become more diverse, the number of individuals whose primary language is not English has also increased substantially. The largest of these groups is the Hispanic population. Pew research studies indicate this population was dominated by individuals of Mexican origin (66.9%), who resided in the Western (44.2%) and Southern (34.8%) regions of the United States (Motel & Patten, 2015). A report on educational attainment from the U.S. Census in 2015 (Ryan & Bauman, 2016) showed that although dropout rates are lower than in previous years, Hispanics are less likely to have a high school diploma than white non-Hispanics (66.7% and 88.7%, respectively), and less likely to attain a college degree (Krogstad, 2016). These facts make clear the need for psychological tests that are appropriate for examinees from working-class homes with parents who have limited academic and English-language skills. The WNV is useful for assessment of minority children because it yields smaller race and ethnic differences than do intelligence tests with substantially more verbal content, while retaining good correlations with achievement.

Wechsler and Naglieri (2006a) studied the utility of the WNV for a sample of individuals speaking English as a second language (ESL), in comparison to a matched group from the WNV standardization sample. The ESL sample included 55 examinees ages 8–21 years whose native language was not English, who spoke a language other than English at home, and whose parents had resided in the United States for less than 6 years. There were 27 Hispanics and 28 examinees whose primary language was either Cantonese, Chinese, Korean, Russian, or Urdu. This sample earned scores very similar to those of their matched counterparts from the normative sample, with negligible effect sizes for the Full Scale scores from both WNV test batteries, as shown in Table 17.2. Additional information about this sample is available in the WNV technical and interpretive manual (Wechsler & Naglieri, 2006c).

Giftedness

The fact that minority children are underrepresented in classes for the gifted has been and con-

TABLE 17.2. WNV Means, *SD*s, and Effect Sizes for Diverse Populations and Matched Control Groups

	Diverse sample			Control group			
	Mean	*SD*	*n*	Mean	*SD*	*n*	Effect size
Gifted students	123.7	13.4	41	104.2	12.3	41	1.5
English-language learners	101.7	13.4	55	102.1	13.4	55	0.0
Hard-of-hearing students	96.7	15.9	48	100.5	14.2	48	0.3
Profoundly deaf students	102.5	9.0	37	100.8	14.3	37	0.1

Note. Effect size = (X1 – X2)/SQRT [(n1 * SD12 + n2 * SD22)/(n1 + n2)] tinues to be an important educational problem (Ford, 1998; Naglieri & Ford, 2003, 2005). Discussions of this issue have often focused on the types of tests used to evaluate children who may be eligible for gifted programming. Some researchers have argued that the verbal and quantitative contents of some of the ability tests used and procedures followed for identifying students for gifted programming are inconsistent with the characteristics of culturally, ethnically, and linguistically diverse populations (Naglieri & Ford, 2005). Additionally, since IQ has traditionally been measured with verbal, quantitative, and nonverbal questions, students with limited English proficiency and math skills earn lower scores on the verbal and quantitative scales of these tests because they do not have sufficient knowledge of the language or training in math, not necessarily because they are of low ability (Bracken & Naglieri, 2003; Naglieri, 2008a).

The use of measures of ability that demand knowledge of English and math skills was the focus of *McFadden v. Board of Education for Illinois School District U-46* (2013). As we describe in more detail in Chapter 15, this litigation was brought by Hispanic parents who were concerned that their students were significantly underrepresented in gifted education classes. About 40% of the students in the Elgin School District (U-46) were Hispanic, but only 2% of the students in the district's school gifted program were Hispanic. The Court ruled that the use of tests demanding knowledge of English (e.g., the Cognitive Abilities Test, Form 6 [CogAT6]; Lohman & Hagen, 2001), rather than a nonverbal measure of ability (e.g., the NNAT-2), contributed to the underrepresentation of Hispanic students in gifted education. This case illustrates that placing emphasis on a nonverbal test could help to circumvent the problem of underrepresentation of Hispanic students in gifted programs (see Card & Giuliano, 2017; Ford, 2013). To test this hypothesis, a large-scale study using the NNAT-2 was conducted.

Card and Giuliano (2017) conducted an 8-year longitudinal study of rates of gifted identification in a district of approximately 40,000 students. Low-income and minority students were substantially underrepresented in gifted education programs in years 1 and 2, when identification was based on referrals by parents and teachers. During years 3 and 4, the district used the NNAT-2. Without any changes in the standards for gifted eligibility, use of the NNAT-2 led to large increases in the numbers of economically disadvantaged and minority students placed in gifted programs. The newly identified gifted students included blacks, Hispanics, those receiving free/reduced-price lunch, and English-language learners (ELLs)—all members of groups identified at much lower rates during the previous 2 years, when the parent–teacher referral system was used. Importantly, these findings suggest that parents and teachers often fail to recognize the potential of poor and minority students and those with limited English skills. During the following 4 years, when the NNAT-2 was no longer used due to budgetary restraints, the numbers of students in these groups who were found to qualify for gifted programming declined and reverted to numbers consistent with those in years 1 and 2.

The WNV manual reports a study involving gifted children who were matched to control subjects from the standardization sample on the basis of age, race/ethnicity, and education level. The differences between the means were calculated by using Cohen's (1988) formula (i.e., the difference between the means of the two groups divided by the square root of the pooled variance). The study included 41 examinees, all of whom had already been identified as gifted via a standardized ability measure on which they performed at 2 *SD* above the mean or more. The students in the gifted programs performed significantly better than their matched counterparts from the normative sample, with large effect sizes for the Full Scale score on both the two- and four-subtest WNV batteries. See Table 17.2 for more details.

Deafness/Hearing Impairments

The issues of limited spoken language and educational attainment are also relevant to those with deafness or hearing impairments. Because the directions are given pictorially and can be augmented with additional statements and/or communication in sign language, the WNV offers considerable advantages for appropriate evaluation of individuals who are deaf or hard of hearing, as the research studies described below illustrate (see also Brunnert et al., 2009).

Wechsler and Naglieri (2006c) reported two studies involving individuals who were deaf or hard of hearing. The first study involved a sample of profoundly deaf examinees who were compared with a group from the standardization sample of the WNV matched on many important demographic variables. The deaf sample consisted of 37 examinees who "must not have been able to hear tones to interpret spoken language after the age of 18 months, must not lip read, must not be trained in the oral or auditory–verbal approach, and must not use cued speech (i.e., they must have routine discourse by some means of communicating other than spoken language). They must have had severe to profound deafness (hearing loss measured with dB, Pure Tone Average greater than or equal to 55)" (Wechsler & Naglieri, 2006c, p. 65). These examinees performed very similarly to their matched counterparts from the normative sample, with negligible effect sizes (Cohen's *d*) for the Full Scale score for both the two- and four-subtest batteries, as shown in Table 17.2.

Wechsler and Naglieri (2006c) also described a study of individuals who were hard of hearing and compared them to a demographically matched group from the standardization sample. This study included 48 examinees who "could have a unilateral or bilateral hearing loss or deafness, and the age of onset of their inability to hear could be any age and [they] could have cochlear implants" (pp. 65–66). This group also performed similarly to the matched counterparts from the normative sample, again with negligible effect sizes (Cohen's *d*) for both batteries, as shown in Table 17.2.

INTERPRETATION

Like all test results, the WNV scores should be interpreted within the full context of the examinee and the administration setting. Issues such as the behaviors observed during testing, relevant educational and environmental backgrounds, physical and emotional status, and reason for referral must be considered when the results are examined. To obtain the greatest amount of information from the WNV, some interpretive methods are the same for the four- and two-subtest batteries, whereas others are unique to each version. These are discussed next.

Interpretation of Both Versions

The WNV subtests are set at a mean of 50 and *SD* of 10. These scores are combined to yield the Full Scale score. The WNV is the first of the Wechsler tests to express the subtest scores on the *T*-score metric (as opposed to a traditional scaled score with a mean of 10 and *SD* of 3). This format was selected because the individual subtests had a sufficient range of raw scores to allow for the use of *T* scores, which have a greater range and precision than scaled scores. The use of *T* scores also provides greater precision on each subtest, allowing for higher reliability coefficients of the Full Scale score.

The WNV Full Scale is set to have a mean of 100 and *SD* of 15, regardless of whether the four- or two-subtest battery is used. This score provides a nonverbal estimate of general ability that has excellent reliability and validity. It is important to recognize that even though the WNV subtests have different demands—that is, some are spatial (e.g., Matrices or Object Assembly), whereas others involve sequencing (Picture Arrangement and Spatial Span), require memory (e.g., Recognition and Spatial Span), or use symbol associations (Coding)—they all measure general ability. General ability, as represented by the Full Scale standard score, provides an estimate for predicting how well a person, for example, will be able to understand spatial as well as verbal and mathematical concepts, remember visual relationships as well as quantitative or verbal facts, and work with sequences of information of all kinds (Naglieri, Brulles, & Lansdowne, 2009). The content of the questions may be visual or verbal, and may require memory or recognition, but general ability (sometimes referred to as *g*) underlies performance on all these kinds of tasks.

WNV Interpretation—Level 1

In the first interpretive step, the Full Scale score should be reported with its associated percentile score, categorical description (average, above average, etc.), and confidence interval. The following illustrates how this information could be included in a written document:

> Annie obtained a WNV Full Scale score of 91, which is ranked at the 27th percentile and falls within the average classification. This means that she performed as well as or better than 27% of examinees her age in the normative sample. There is a 90% chance that her true Full Scale score falls within the range of 85–99.

The second step in interpretation of the four-subtest version of the WNV is to examine the *T* scores the examinee earned on the subtests, taking into consideration the lower reliability of these scores. Examination of the four WNV subtests should also take into consideration that even though the subtests are all nonverbal measures of general ability, they do have unique attributes, as noted above. In addition, statistical guidelines should be followed to ensure that any differences interpreted are beyond those that could be expected by chance. The values needed for significance when comparing a WNV subtest for an examinee to that examinee's mean *T* score are provided in the WNV administration and scoring manual (Wechsler & Naglieri, 2006b, Table B.1) and in more detail by Brunnert and colleagues (2009), and should be used in examining subtest variability. The following steps should be used to compare each of the four WNV subtest *T* scores to the examinee's mean subtest *T* score:

1. Calculate the mean of the four subtest *T* scores.
2. Calculate the difference between each subtest *T* score and the mean.
3. Subtract the mean from each of the subtest *T* scores (retain the sign).
4. Find the value needed for significance, using the examinee's age group and the desired significance level in Table 12.3 of the WNV manual (Wechsler & Naglieri, 2006b).
5. If the absolute value of the difference is equal to or greater than the value in the table, the result is statistically significant.
6. If the subtest difference from the mean is lower than the mean, then the difference is a weakness; if the subtest difference from the mean is greater than the mean, then the difference is a strength.

When there is significant variability in the WNV subtests, it is also important to determine whether a weakness relative to the examinee's overall mean is also sufficiently below the average range. Determining whether an examinee has significant variability relative to his or her own average score is a valuable way to determine strengths and weaknesses, but Naglieri (1999) has cautioned that a relative weakness could also be significantly below the normative mean. He recommends that any subtest score that is low relative to the examinee's mean score should also fall below the average range to be considered a noteworthy weakness (i.e., at least 1 *SD* below the normative mean).

WNV Interpretation—Level 2

Spatial Span Forward and Backward

The WNV Spatial Span Forward and Backward scores can be interpreted separately. The sizes of the differences required for statistical significance by age and for the U.S. and Canadian standardization samples are 11 and 13 for the .10 and .05 levels for the U.S. sample, and 10 and 13 for these levels for the Canadian sample, respectively, for the combined ages of 8:0–21:11. Table C.1 of the WNV administration and scoring manual (Wechsler & Naglieri, 2006b) provides the information necessary to allow for the comparison between Spatial Span Forward and Backward. A difference of 9 *T*-score points is needed at the .15 level (13 at the .05 level) for significance. (Note that base rate data by the direction of the difference are provided in the WNV manual.)

Information about Spatial Span Forward and Backward *T* scores may be useful, but it should be integrated within the greater context of a comprehensive assessment. For example, if a difference between Spatial Span Forward and Backward is found, it may be expected that similar results will be found on similar tests (e.g., Digit Span Forward vs. Digit Span Backward on the Wechsler Intelligence Scale for Children—Fifth Edition [WISC-V; Wechsler, 2014]). The Backward scores may also be related to the Planning scale of the Cognitive Assessment System (CAS; see Naglieri, 1999) and the CAS2 (see our discussion of the CAS2 in Chapter 15, this volume), and may suggest that the examinee has difficulty with development and utilization of strategies for reversing the order of serial information. In addition, Digit Span and Spatial Span Forward may be considered measures of sequencing, whereas Digit Span and Spatial Span Backward may be considered measures of sustained concentration and visual–spatial working memory, respectively (Miller, 2007, 2010).

Spatial Span can be considered a nonverbal version of the Digit Span subtest of the WISC-V, even though the task has a distinct spatial component. For Spatial Span Forward, the examinee touches a sequence of blocks randomly arranged on an 8½- by 11-inch board in the same order as demonstrated by the examiner. For Spatial Span Backward, the examinee repeats a sequence in the reverse order of that demonstrated by the examiner. Again, then, Spatial Span Backward can be viewed as a task that requires visual–spatial working memory. Goldberg (2009) defines *working memory* as "the selection of task-relevant information" (p. 94), and it is the selection process incorporated into the task that demands strategy use. Observing the examinee's performance on Spatial Span Forward can reveal information about how well he or she initially commits sequenced visual–spatial information to memory. Spatial Span Backward allows the examiner to observe visual working memory capacity and efficiency in the selection of the sequence executed. Normative information for comparisons of Spatial Span Forward and Backward, as well as normative sample base rates found in the WNV manual (Wechsler & Naglieri, 2006b), should be used to understand the reliability of any differences found.

Coding

The WNV Coding subtest also provides the opportunity for finer clinical interpretation. For example, Koziol and Budding (2009) hypothesized that a task such as Coding places demands on working memory because there are numbers and symbols, and that quick performance may be facilitated by "holding this information online in working memory in the course of performing the task" (p. 261). The associations between symbols and numbers are maintained within working memory, and a short-term plan of action is activated. If the number–symbol associations are made quickly, it is assumed that less conscious effort is required for the task. This observation is consistent with Gabrieli, Stebbins, Singh, Willingham, and Goetz's (1997) formulation that working memory capacity facilitates fast performance and the attainment of procedural learning.

An examinee who completes the Coding subtest accurately but very slowly is approaching the task differently than an examinee who completes the task quickly but with many errors (making the wrong number–symbol association, skipping), who in turn is different from an examinee who completes the task quickly and accurately. Useful ways to interpret this kind of performance may be in terms of attention to the instructions (e.g., whether the instruction to "do the task as fast as you can" is ignored) or of conscious effort or concentration. The examinee who works faster for the 120 seconds and gives more responses has approached the task differently than the examinee who works more slowly and gives fewer responses. These examinees have completed the tasks at different rates over the same time interval. From this information, it can be hypothesized that the examinee who works more slowly has to put forth greater conscious control and effort, which may be related to recruitment of more brain areas (Saling & Philips, 2007). The examinee who works more slowly has to concentrate harder, and the examinee who works quickly has probably expended less effort. In this way, the subtest score for Coding may be viewed as a measure of efficiency of concentration.

Matrices

The WNV Matrices subtest is like others that have a long history of being good measures of general ability, as demonstrated by high *g* loadings (Jensen, 1998). Tests like Matrices can be viewed as measures of visual-perceptual reasoning. Matrices can also be considered a test of simultaneous processing, a mental activity by which a person integrates stimuli into interrelated groups or a whole (Naglieri, 1999). Simultaneous processing tests typically have strong visual–spatial aspects. The cognitive demands of the task require the integration of information (Naglieri & Otero, 2011).

SUMMARY AND CONCLUSIONS

The assessment of diverse populations of children and adults with verbal ability tests is particularly problematic for those with limited language skills and/or educational opportunities. The WNV was designed to provide a nonverbal measure of general ability that would be appropriate for a wide variety of culturally and linguistically diverse populations. This assessment should be useful in many settings, including the identification of gifted minority students. This chapter has summarized some of the validity evidence provided in the WNV technical and interpretive manual (Wechsler & Naglieri, 2006c), which supports the utility of the WNV for fair assessment of cognitive ability in several groups (students from culturally diverse backgrounds and/or with language differences, gifted students, and students who are deaf or hard of hearing). Within the context of a comprehensive evaluation, the WNV can provide important information about an examinee's level of general ability. It is important, however, to recognize that traditional tests based on the general ability model can be substantially augmented by measures of neurocognitive processes, which can detect underlying strengths and weaknesses that further explain academic performance.

Appendix 17.1 provides essential aspects of a case study that illustrates the use of the WNV in in the context of a multicomponent evaluation. All personally identifiable information has been altered in order to preserve the anonymity of the examinee.

APPENDIX 17.1

Case Study

REASON FOR REFERRAL

Lucia was seen by a multidisciplinary team to assess her educational needs. The author of this report (Tulio M. Otero, Ph.D.) was requested to evaluate Lucia in the areas of cognitive and academic functioning. Lucia is a third grader who is currently in a regular education class with English-language learner (ELL) services. A review of her files indicated that she suffered a closed head injury at age 3 in a motor vehicle accident. There was brief loss of consciousness (approximately 10 minutes). At the time of this evaluation, no further medical history was available. Despite interventions that included 8 weeks of Lexia Reading and Symphony Math, Lucia has been making only limited progress in both reading and math.

OBSERVATIONS

Lucia is a 9-year-old Hispanic female who speaks both English and Spanish. At home, Lucia speaks only Spanish, since neither of her parents speaks English; at school, she is in a regular classroom setting with ELL services. At school, she speaks mostly English with peers and teachers. Although she has conversational English skills, she is still developing cognitive academic language skills. Lucia preferred to converse with the examiner in English, but during the evaluation both languages were used in order allow maximum opportunities to comprehend and respond to the tests. Although both languages were used, her performance did not improve on verbal portions of the tests. She was alert, oriented, friendly, and cooperative. Her range of emotion was good, and she did not report any significant worries or concerns. Lucia reported that she had gone to bed late the night before because the family had gone to a party. She further indicated that she sometimes naps in the afternoon and then stays awake watching TV until very late in the evening. Although she put forth good effort on all tasks, it was obvious that she struggled from time to time because of fatigue, and therefore was provided with brief breaks and snacks as necessary. She did not have any negative reaction to failure and became quite animated when she did well on tasks.

Lucia struggled on academic achievement tests. During word-reading tasks, she read slowly. Her reading errors can be described as approximations to words based on what they looked like. For example, she read “ground” instead of *around*, “worm” for *wrong*, and “throw” for *threw*. Reading comprehension was deficient and very slow. She did not seem to benefit from the visual stimuli accompanying the passage she needed to read. On math reasoning tasks, Lucia took a long time to initiate a response, needed prompting to use paper and pencil to assist her in deriving the answer, and often used ineffective strategies. Relatively simple items took her several minutes to work through. On calculation tasks, she also did poorly because of either treating subtraction as addition or providing a result that did not make sense. During testing of the limits, Lucia was encouraged to notice her errors and the operation sign. She correctly reworked five items, although she could not receive credit for the reworkings. Her spelling skills seem typical of ELL students who use sounds in Spanish to spell English words. Yet she also missed sounds or sound clusters. Examples of some of her errors were “joump” for *jump*, “forrise” for *forest*, “unfar” for *unfair*, and “meniger” for *manager*.

TESTS ADMINISTERED

- Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003)
- Escalas de Inteligencia de Reynolds (RIAS-Spanish; Reynolds & Kamphaus, 2009)
- Wechsler Nonverbal Test of Ability (WNV; Wechsler & Naglieri, 2006b)
- Cognitive Assessment System—Second Edition: Spanish Edition (CAS2: Español; Naglieri, Moreno, & Otero, 2017)
- WJ IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014)
- Batería III Pruebas de Aprovechamiento (Batería III; Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005)

NONVERBAL TEST RESULTS

Lucia’s overall cognitive ability was assessed by eliminating or minimizing the impact of limited language proficiency through the use of the WNV. This test is used to assess the general cognitive ability of individuals ages 4–21 years. Lucia’s WNV Full Scale score was 95; she scored higher than approximately 37 out of 100 individuals her age. Her general cognitive ability, as assessed by the WNV, is in the average range. These results suggest that her general cognitive ability is somewhat higher than that measured by tests requiring a student to process language both receptively and expressively. This is an example of how nonverbal tests can be particularly important when Hispanic children are assessed, as these students are more likely to have varying histories of educational opportunity and levels of academic English-language proficiency. Lucia's WNV performance is presented in Table 17.A.1.

Lucia was administered the RIAS in English with Spanish support, using select items from the Spanish version of the RIAS. The RIAS is an individually administered measure of intellectual functioning normed for individuals between the ages of 3 and 94 years. It contains several individual tests of intellectual problem solving and reasoning ability, which are combined to form a Verbal Intelligence Index and a Nonverbal Intelligence Index. These two indexes of intellectual functioning are then combined to form an overall Composite Intelligence Index. Each of these indexes is expressed as an age-corrected standard score, scaled to a mean of 100 and an *SD* of 15. The RIAS also contains Verbal Memory and Nonverbal Memory subtests, which are combined to form a Composite Memory Index.

Lucia earned a Composite Intelligence Index score of 81 on the RIAS. This level of performance falls within the range of scores designated as below average and exceeds the performance of 10% of individuals Lucia's age. Lucia earned a Verbal Intelligence Index score of 79, which falls within the moderately below-average range of verbal intelligence skills and exceeds the performance of 8% of individuals Lucia's age. Lucia earned a Nonverbal Intelligence Index score of 88, which falls within the low average range of nonverbal intelligence skills and exceeds the performance of 21% of individuals Lucia's age. Lucia's Verbal and Nonverbal Intelligence Index scores were not significantly different; however, there was a significant difference between scores obtained on nonverbal versus verbal subtests comprising the Composite Memory Index score (86, low average range). This score exceeds the performance of 18% of individuals Lucia's age. Lucia's performance in the Nonverbal Memory domain significantly exceeded her performance within the Verbal Memory domain. The difference between Nonverbal Memory and Verbal Memory is reliable and indicates that Lucia functions at a significantly higher level when asked to perform visual memory tasks (*T* score = 51) as opposed to verbal memory tasks (*T* score = 33). Lucia's RIAS performance is presented in Table 17.A.2.

General Ability Summary

The results of the WNV and RIAS suggest that when general ability is measured by using questions that demand the use of verbal stimuli, the scores are lower than when general ability is measured via tests that are commonly described as nonverbal. These nonverbal test items typically involve spatial and visual stimuli. Given that Lucia is limited in her English-language skills, these results lead to two important conclusions: first, that the verbal measures underestimate her general ability; and second, that the nonverbal scores are likely to give a better indication of her general ability.

TABLE 17.A.1. Wechsler Nonverbal Scale of Ability (WNV) Results for Lucia

	Standard score	Percentile rank	90% confidence interval
Full Scale	95	37	89–102

Subtest	Subtest *T* score	Difference from mean	Critical value	Variability	Base rate
Matrices	51	3	8	NS	32.8
Coding	47	–1	10	NS	47.2
Spatial Span	42	–6	8	NS	18.6
Picture Arrangement	52	4	10	NS	27.7
Within-subtest analysis					
Spatial Span Forward–Backward	39	–8	13	NS	21.6

Note. NS, nonsignificant. The mean WNV subtest score is 48.0.

TABLE 17.A.2. Reynolds Intellectual Assessment Scales (RIAS) Results for Lucia

Subtest	*T* score
Guess What	34
Odd-Item Out	40
Verbal Reasoning	35
What's Missing	43
Verbal Memory	33
Nonverbal Memory	51

Scale	Standard score	90% confidence interval	Percentile rank
Verbal Intelligence Index	79	75–86	8
Nonverbal Intelligence Index	88	83–94	21
Composite Intelligence Index	81	77–87	10
Composite Memory Index	86	81–92	18

Note. Lucia's Verbal Memory score is significantly lower than her Nonverbal Memory score.

ACADEMIC SKILLS RESULTS

Lucia was administered select subtests from the WJ IV ACH and Batería III to evaluate her academic achievement. Relative strengths and weaknesses among her academic skills are described below.

Academic Assessment in English

The WJ IV ACH Broad Reading cluster includes tests of reading decoding, reading speed, and the ability to comprehend connected discourse while reading. Lucia's standard score on this cluster was 61, which falls within the very low range (percentile rank range of <1–1; standard score range of 58–63) for her age. This score suggests that Lucia's overall reading ability in English is very limited; reading tasks above the age 7:5 level will be quite difficult for her. Lucia is likely to require intensive instructional support and targeted interventions in reading. The Letter–Word Identification subtest measured Lucia's ability to identify words. Lucia seemed unable to apply phoneme–grapheme relationships. Passage Comprehension measured Lucia's ability to understand what she had read. The items required Lucia to read a short passage and identify a missing word that made sense in the context of the passage. Lucia appeared to read each passage very slowly and had difficulty identifying the missing word. Reading Fluency measured Lucia's ability to read simple sentences quickly. Lucia appeared to read and respond to the sentences slowly.

The WJ IV ACH Broad Math cluster includes tests of mathematics reasoning and problem solving, number facility, and automaticity. Although her standard score on this cluster was within the very low range (58), her performance varied on different types of math tasks. Lucia's performance was very limited on tasks requiring the ability to analyze and solve applied mathematics problems (the Applied Problems subtest; standard score = 80). Her performance was limited on tasks requiring knowledge of how to perform mathematical computations, either with or without time limits (the Math Calculation Skills cluster, including the Calculation and Math Fluency subtests; standard score = 39). These results indicate that intensive instructional support in math, including targeted interventions, is likely to be needed for Lucia. Calculation measured Lucia's ability to perform mathematical computations; she worked very slowly and relied on the use of strategies that appeared to be inefficient for her age level. To solve each item in Applied Problems, Lucia was required to listen to the problem, recognize the procedure to be followed, and then perform relatively simple calculations. Because many of the problems included extraneous information, Lucia needed to decide not only the appropriate mathematical operations to use, but also what information to include in the calculation. Lucia appeared to have limited understanding of age-appropriate math application tasks. Finally, Math Fluency measured Lucia's ability to solve a series of simple addition, subtraction, and multiplication problems quickly in a 3-minute time limit. Lucia appeared to take longer to work on such problems than is typical for her same-age peers.

In the area of Writing, Lucia attained a standard score within the low range (79), suggesting underdeveloped writing skills. Her Spelling score was significantly lower than her Writing Samples score. Spelling measured Lucia's ability to write orally presented words correctly; Lucia spelled words in a laborious manner. Writing Samples measured Lucia's skill in writing responses to meet a variety of demands. She was asked to produce written sentences that were evaluated with respect to the quality of expression. Lucia was not penalized for any errors in basic writing skills, such as spelling or punctuation. Many of her sentences were inadequate to meet the task demands. Yet she managed to score within the average range on this particular subtest.

Overall, Lucia's academic skills in English, including spelling, sight reading, and math calculation, appear very limited (see Table 17.A.3 for a summary). Lucia's overall ability to apply her academic skills via tests of writing ability, reading comprehension, and math problem solving also seems limited.

Academic Assessment in Spanish

On the Batería III, Breve Lectura includes tests of reading decoding and the ability to comprehend connected discourse while reading. Lucia's standard score of 66 is within the very low range (percentile rank range of 1–2; standard score range of

TABLE 17.A.3. Woodcock–Johnson IV Tests of Achievement (WJ IV ACH) Results for Lucia

CLUSTER/test	AE	Easy to difficult	SS (90% confidence interval)	GE
ACHIEVEMENT	7:4	7:0 to 7:8	67 (65–70)	2:0
BROAD READING	7:1	6:9 to 7:5	61 (58–63)	1:7
BROAD MATH	6:11	6:4 to 7:7	58 (55–62)	1:6
MATH CALC. SKILLS	6:4	5:11 to 7:0	39 (32–45)	1:1
BRIEF WRITING	7:7	7:2 to 8:2	79 (76–82)	2:2
ACADEMIC SKILLS	6:11	6:7 to 7:3	58 (55–61)	1:6
ACADEMIC APPLICATIONS	7:6	7:1 to 8:1	72 (69–76)	2:2
Letter–Word Identification	7:2	6:11 to 7:5	67 (65–70)	1:8
Applied Problems	7:10	7:4 to 8:4	80 (76–83)	2:5
Spelling	7:2	6:10 to 7:7	72 (68–76)	1:9
Passage Comprehension	6:11	6:7 to 7:2	67 (62–71)	1:6
Math Fluency	7:1	< 5:1 to 9:0	70 (66–73)	1:8
Calculation	6:3	6:1 to 6:6	38 (31–46)	1:0
Writing Samples	8:5	7:6 to 10:4	92 (87–97)	3:1
Reading Fluency	7:3	6:1 to 8:2	71 (65–76)	1:9

	SS			Variation		
Variations	Actual	Predicted	Difference	Discrepancy PR	Discrepancy *SD*	Significant at ±1.50 *SD* (*SEE*)
Intra-Achievement (Brief)						
BROAD MATH	58	79	–20	5	–1.63	Yes

Note. Norms based on ages 9–10. AE, age equivalent; SS, standard score; GE, grade equivalent; PR, percentile rank; *SD*, standard deviation; *SEE*, standard error of estimate. For information on how to interpret the discrepancy percentile rank and discrepancy standard deviation, see Jaffe (2009).\

64–69) when compared to those of others her age. This suggests that reading tasks above the age 7:6 level will be quite difficult for her. Similarly, Ortografía evaluates Lucia's ability to write orally presented words correctly in Spanish. Her standard score is within the low range (74; percentile rank range of 3–7; standard score range of 71–78) for her age. This indicates that Lucia's spelling ability is very limited; spelling above the age 7:9 level will be quite difficult for her. Lucia's Batería III performance is presented in Table 17.A.4.

Academic Assessment Summary

The academic test results suggest that Lucia has weaknesses in academic skills in both English and Spanish. Lucia's performance on nonverbal tests suggests that her performance on academic tasks could be expected to be higher. That is, there is a substantial difference between Lucia's nonverbal scores and her level of achievement in both English and Spanish. To clarify why her levels of academic skills seem low, a measure of neurocognitive processing was administered.

NEUROCOGNITIVE TEST RESULTS

The CAS2: Español was administered to Lucia, in an effort to understand the difference between her nonverbal general ability scores (which were in the average range; WNV Full Scale = 95) and her poor academic achievement scores in both English and Spanish. The CAS2: Español is an individually administered test designed to measure intelligence as a group of neurocognitive processes. It is based on the *planning, attention, simultaneous, and successive* (PASS) theory of intelligence.

The Planning scale of the CAS2: Español measures cognitive control, use of strategies, knowledge and skills, intentionality, and self-regulation. This scale measures a child's ability to determine how to solve a problem, execute that solution, monitor its effectiveness, and modify the approach as needed to achieve the goal. Tasks administered as part of this scale require impulse control, as well as generation, evaluation, and execution of a plan. The Planning scale measures how well the child can solve problems of varying complexity that may involve control of attention, simultaneous, and successive processes, as well as acquisition of knowledge and skill. This ability is associated with the brain area known as the prefrontal cortex. Lucia earned a CAS2: Español Planning standard score of 80, which is within the low average classification and is ranked at the 9th percentile. This percentile rank score means that Lucia did as well as or better than 9% of children her age in the standardization group. There is a 90% probability that Lucia's true Planning score is within the range of 75–88.

The Attention scale of the CAS2: Español measures a child's ability to demonstrate focused, selective cognition over time, with resistance to distraction. Attention can be described as focused, selective, sustained, and effortful activity. *Focused* attention involves concentration directed toward a particular activity, and *selective* attention is important for the inhibition of responses to distracting stimuli. *Sustained* attention refers to the variation of performance over time, which can be influenced by the varying amounts of *effort* required to solve the test. An effective measure of attention presents children with competing demands and requires sustained focus. The components of attention are subserved by neural networks spanning subcortical, dorsal–ventral steams and frontal brain regions in an interactive manner. *Executive attention* is defined as maintaining behavior goals and using these goals as a basis for choosing what aspects of the environment or tasks to attend to and which action to select. Paying attention is

TABLE 17.A.4. Batería III Normative Update Pruebas de Aprovechamiento Results for Lucia

CLUSTER/test	AE	Easy to difficult	SS (90% confidence interval)	GE
LECTURA (READING)	7:2	6:11 to 7:6	66 (64–69)	1:9
Ident. de Letras y Palabras (Letter–Word Identification)	7:3	7:1 to 7:6	70 (68–73)	2:0
Ortografía (Spelling)	7:4	7:0 to 7:9	74 (71–78)	2:0
Comprensión de Textos (Reading Comprehension)	7:1	6:9 to 7:5	71 (68–75)	1:8

Note. AE, age equivalent; GE, grade equivalent; SS, standard score.

the first step in the learning process. Lucia earned a CAS2: Español Attention standard score of 102, which is within the average classification and is ranked at the 55th percentile. This percentile ranking means that Lucia did as well as or better than 55% of children her age in the standardization group. There is a 90% probability that Lucia's true Attention score is within the range of 94–109.

The Simultaneous scale of the CAS2: Español measures a child's ability to integrate separate but interrelated stimuli into groups or into a whole. Simultaneous processing tests have strong visual–spatial aspects for this reason, but this ability is also used to solve tasks with verbal content (e.g., reading comprehension), if the tasks require integration of information into a coherent whole. Simultaneous processing underlies the use and comprehension of grammatical statements because they demand comprehension of word relationships, prepositions, and inflections, so that a person can obtain the meaning of a whole idea. Select aspects of verbal reasoning tasks, such as those in which an examinee is given two to four clues and asked to deduce the object or concept being described, also require simultaneous processing. There was variability in Lucia's scores on the subtests that make up this scale. Lucia had her lowest score on the Verbal–Spatial Relations subtest; however, overall, she earned a CAS2: Español Simultaneous standard score of 95, which is within the average classification and is ranked at the 34th percentile. This means that Lucia did as well as or better than 34% of children her age in the standardization group. There is a 90% probability that Lucia's true Simultaneous score is within the range of 89–101.

The Successive scale of the CAS2: Español evaluates how well a child works with stimuli in a specific serial order, where each element is only related to those that precede it and these stimuli are not interrelated. Successive processing ability involves both the perception of stimuli in sequence and the formation of sounds and movements in order. For example, successive processing is involved in the decoding of unfamiliar words, production of syntactic aspects of language, and speech articulation. This process is measured with tests that demand use, repetition, or comprehension of information based on order. Following a sequence such as the order of operations in a math problem is another example of successive processing. Lucia earned a CAS2: Español Successive standard score of 81, which is within the low average classification and is ranked at the 10th percentile. This means that Lucia did as well as or better than 10% of children her age in the standardization group. There is a 90% probability that Lucia's true Successive score is within the range of 76–89.

Lucia's overall CAS2: Español performance is presented in Table 17.A.5. Note that the Full Scale is omitted from this report and Table 17.A.5 because the significant variability among the four PASS scales renders the overall score misleading. In addition to the PASS scores, the test provides a Nonverbal score; Lucia obtained a Nonverbal score of 97, which is within the average classification and is ranked at the 42nd percentile. This means that Lucia did as well as or better than 42% of children her age in the standardization group. There is a 90% probability that Lucia's true Nonverbal score is within the range of 91–104.

SUMMARY

Lucia's RIAS Composite Intelligence Index (81) indicates mild deficits in general intelligence relative to others of her age when ability is measured with verbal and nonverbal tests. Although students earning a Composite Intelligence Index in the 80s frequently experience at least some difficulty acquiring information through traditional educational methods provided in the classroom setting, Lucia's limited verbal skills depressed her Verbal score on this test. Lucia's score on the RIAS Nonverbal Memory scale is higher than her Verbal Memory score. Importantly, the WNV Full Scale result (95) and the CAS2: Español Nonver-

TABLE 17.A.5. Cognitive Assessment System—Second Edition: Spanish Edition (CAS2: Español) Results for Lucia

Scale	Scaled score	90% confidence interval	Percentile rank
Planning	80	75–88	9
Simultaneous	95	89–101	37
Attention	102	94–109	55
Successive	81	76–89	10
Average PASS score	89.5		

bal score (97) suggest that Lucia's general ability when language-processing requirements are markedly reduced is within the average range.

Lucia's English achievement test scores suggest academic skills within the low range. When compared to those of others at her age level, Lucia's standard score is low in writing, and her standard scores are very low in broad reading, broad mathematics, and math calculation skills. Lucia demonstrated a significant weakness in mathematics, particularly because she confused the operation signs and made calculation errors when borrowing and regrouping. Similarly, the Spanish achievement test results included a very low score in reading. These results strongly suggest that the verbal test scores should be considered underestimates of overall ability, and that the neurocognitive results better explain her levels of intellectual functioning.

Lucia earned scores on the four PASS scales of the CAS2: Español that differed significantly. She earned significantly low scores on measures of planning and successive neurocognitive processes, which help to answer the question "Why is Lucia having academic problems?" These two cognitive weaknesses provide evidence of a disorder in one or more of the basic psychological processes described in the Individuals with Disabilities Education Improvement Act (IDEA) of 2004. When low scores on such measures are accompanied by similarly low achievement scores with adequate opportunity to learn, there is evidence that a child probably meets eligibility criteria for a specific learning disability, according to the discrepancy–consistency method (see Figure 17.A.1) as described by Naglieri and Otero (2017). That is, Lucia appears to have significant variability in basic neurocognitive processes (Planning and Successive weaknesses on the CAS2, with Simultaneous and Attention strengths), a discrepancy between low achievement scores and her CAS2 Simultaneous and Attention scores, and a consistency between low achievement and low CAS2 Planning and Successive scores. For more details on this method, see Naglieri and Otero (2017) and Naglieri and Feifer (2017a, 2017b).

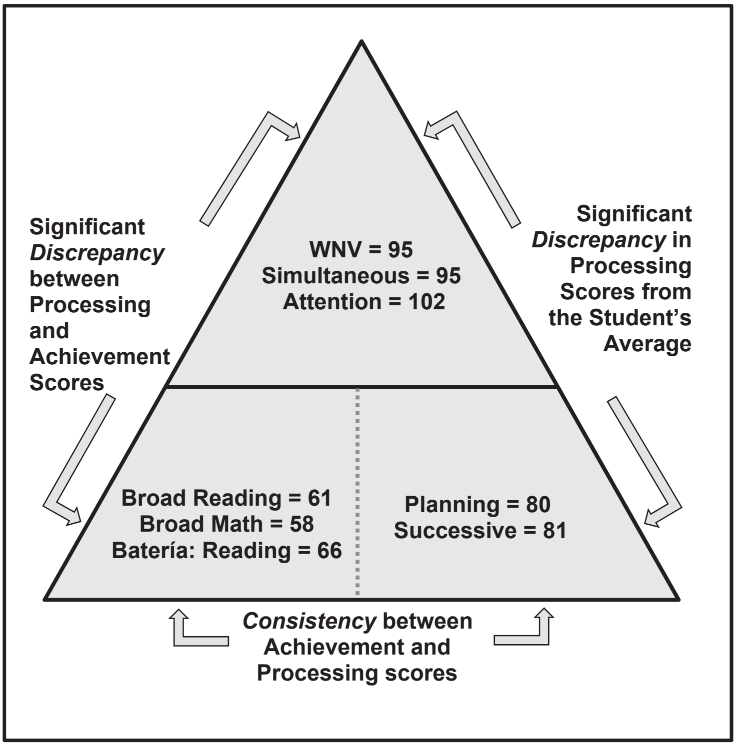


FIGURE 17A.1. Discrepancy–consistency method of SLD eligibility determination for the case of Lucia.

POSSIBLE INSTRUCTIONAL RECOMMENDATIONS AND INTERVENTIONS

- Lucia will probably gain the most from reading instruction presented within the middle- to late-first-grade range. The primary goal is to increase Lucia's exposure to printed words, in hopes of increasing the number of words that she can recognize orthographically.
- Daily reading is recommended for fluency-building intervention. Lucia should read a short passage several times until she can read the passage with ease. The instructor should select material that is at Lucia's instructional reading level. Then the instructor should have her read through the passage aloud; the instructor should record the number of errors, as well as the time it takes Lucia to read the passage. After Lucia completes the passage, the instructor should review the misread words and then have her read it again. This approach should be continued until Lucia has read the passage three to five times or has reached a preestablished goal for accuracy or rate.
- Lucia and her parents should be encouraged to spend time reading every day outside of school.
- To address Lucia's weaknesses in planning and successive neurocognitive processes, reading strategies described in *Helping Children Learn: Intervention Handouts for Use in School and at Home* (Naglieri & Pickering, 2010) should be used. Lucia's teachers and parents can help her follow the instructions in the handouts (e.g., Summarization Strategy for Reading Comprehension, Chunking for Reading/Decoding, Word Families for Reading/Decoding).
- Math instruction presented within the middle-first-grade to early-second-grade range will be likely to produce the greatest gains for Lucia.
- Use of a concrete–representational–abstract sequence will ensure that Lucia understands a computation or math fact: first by using manipulatives, then by drawing representations (pictures or tallies) of the problem, and finally by solving the problem with actual numbers.
- Teachers and parents should consider using math techniques described by Naglieri and Pickering (2010), such as Touch Math, the Part–Whole Strategy, and the Seven-Step Strategy. These methods will help Lucia to solve math problems in a variety of settings.
- The cover–copy–compare intervention requires teacher-made worksheets that provide correctly completed problems on the left side of the paper and the unsolved problems on the right side of the paper. The teacher will instruct Lucia to study each correctly completed problem, then cover it with an index card, complete the matching problem to the right, and check her work by comparing it to the model problem.
- Writing instruction presented within the early- to middle-second-grade level is appropriate for Lucia.
- The Write–Say method may be helpful in addressing Lucia's spelling skills. This intervention will require Lucia to study a spelling list on her own on Monday, and then to participate in an orally administered spelling test on Tuesday. The teacher will provide verbal feedback, and Lucia will then say and write the correct spelling of missed words, letter by letter, five times. The same procedure is to be followed Wednesday and Thursday; however, Lucia will then practice incorrectly spelled words 10 times and 15 times, respectively. Finally, the teacher will administer a summative spelling test on Friday.
- The Add-a-Word spelling program may assist Lucia in developing better spelling skills. To implement this intervention, Lucia's teacher will provide five daily spelling words. Each word must be correctly spelled 5 days in a row before an individual word is replaced with a new spelling word. If on subsequent spelling tests a previously learned word is missed, that word will be placed back onto the current spelling list until it is mastered again.
- Lucia needs to devote more time to writing. Daily writing practice at school and at home facilitates writing for different purposes and for different audiences. Making the connection between writing and real-world applications should be an important motivator in developing Lucia's writing skills.
- Because Lucia earned a low score on the Planning scale of the CAS2, she should be provided with strategies for writing. The following descriptive handouts of specific methods (Naglieri & Pickering, 2010) should be provided to the teachers and parents: Story Plans for Written Composition, Story Grammar for Writing, and Plans for Writing. These will help Lucia acquire strategies for communicating her ideas in writing. These and other handouts are available in Spanish for use by Lucia's parents.
- Lucia will probably gain the most from Spanish reading instruction presented at the late-first-grade to early-second-grade level.

REFERENCES

Bishop, S. L., Richler, J., & Lord, C. (2006). Association between restricted and repetitive behaviors and nonverbal IQ in children with autism spectrum disorders. *Child Neuropsychology, 12*(4–5), 247–267.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 383–405.

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test.* Itasca, IL: Riverside.

Bracken, B. A., & McCallum, R. S. (2016). *Universal Nonverbal Intelligence Test—Second Edition.* Austin, TX: PRO-ED.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 243–273). New York: Guilford Press.

Brunnert, K., Naglieri, J. A., & Hardy-Braz, S. (2009). *Essentials of WNV assessment.* Hoboken, NJ: Wiley.

Card, D., & Giuliano, L. (2017). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences USA, 113,* 13678–13683.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education, 32,* 4–77.

Ford, D. Y. (2013). *Recruiting and retaining culturally different students in gifted education.* Waco, TX: Prufrock Press.

Gabrieli, J. D., Stebbins, G. T., Singh, J., Willingham, D. B., & Goetz, C. G. (1997). Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington's disease: Evidence for dissociable memory systems in skill learning. *Neuropsychology, 11,* 272–281.

Goldberg, E. (2009). *The new executive brain: Frontal lobes in a complex world.* New York: Oxford University Press.

Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the Relative Proficiency Index* (Woodcock–Johnson III Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Koziol, L., & Budding, D. E. (2009). *Subcortical structures and cognition: Implications for neuropsychological assessment.* New York: Springer.

Krogstad, J. M. (2016, July 28). 5 facts about Latinos and education. Retrieved from *www.pewresearch.org/fact-tank/2016/07/28/5-facts-about-latinos-and-education.*

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.

Matson, J. L. (2007). Determining treatment outcome in early intervention programs for autism spectrum disorders: A critical analysis of measurement issues in learning based interventions. *Research in Developmental Disabilities, 28*(2), 207–218.

Mayes, S. D., & Calhoun, S. L. (2003). Ability profiles in children with autism: Influence of age and IQ. *Autism, 7*(1), 65–80.

McCallum, R. S. (2017). *Handbook of nonverbal assessment.* New York: Springer.

McFadden v. Board of Education for Illinois School District U-46, 984 F. Supp. 2d 882 (2013).

Miller, D. C. (2007). *Essentials of school neuropsychology assessment.* Hoboken, NJ: Wiley.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Motel, S., & Patten, E. (2015). The 10 largest Hispanic origin groups: Characteristics, rankings, top counties. *Religion, 5,* 12.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock–Muñoz: Pruebas de aprovechamiento.* Itasca, IL: Riverside.

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test—Multilevel Form.* San Antonio, TX: Psychological Corporation.

Naglieri, J. A. (1999). *Essentials of CAS assessment.* Hoboken, NJ: Wiley.

Naglieri, J. A. (2003a). *Naglieri Nonverbal Ability Test—Individual Form.* San Antonio, TX: Psychological Corporation.

Naglieri, J. A. (2003b). Naglieri Nonverbal Ability Tests: NNAT and MAT-EF. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 175–190). New York: Kluwer.

Naglieri, J. A. (2008a). *Naglieri Nonverbal Ability Test—Second Edition.* San Antonio, TX: Pearson.

Naglieri, J. A. (2008b). Traditional IQ: 100 years of misconception and its relationship to minority representation in gifted programs. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 67–88). Waco, TX: Prufrock Press.

Naglieri, J. A. (2015). One hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.) *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 295–316). New York: Springer.

Naglieri, J. A. (2016). *Naglieri Nonverbal Ability Test—Third Edition.* Bloomington, MN: Pearson.

Naglieri, J. A., Brulles, D., & Lansdowne, K. (2009). *Helping all gifted children learn: A teacher's guide to using the NNAT-2.* San Antonio, TX: Pearson.

Naglieri, J. A., & Feifer, S. (2017a). Interventions based on PASS. In J. A. Naglieri & T. M. Otero, *Essentials* *of CAS2 assessment* (pp. 147–180). Hoboken, NJ: Wiley.

Naglieri, J. A., & Feifer, S. (2017b). Pattern of strengths and weaknesses made easy: The discrepancy/consistency method. In D. Flanagan & V. Alfonso, *Essentials of specific learning disabilities identification* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47,* 155–160.

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly, 49,* 29–36.

Naglieri, J. A., & Goldstein, S. (Eds.). (2009). *Practitioner's guide to assessing intelligence and achievement.* Hoboken, NJ: Wiley.

Naglieri, J. A., Moreno, M. A., & Otero, T. M. (2017). *Cognitive Assessment System—Second Edition: Spanish Edition.* Austin, TX: PRO-ED.

Naglieri, J. A., & Otero, T. (2011). Cognitive Assessment System: Redefining intelligence from a neuropsychological perspective. In A. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 320–333). New York: Springer.

Naglieri, J. A., & Otero, T. M. (2017). *Essentials of CAS2 assessment.* Hoboken, NJ: Wiley.

Naglieri, J. A., & Pickering, E. (2010). *Helping children learn: Intervention handouts for use in school and at home* (2nd ed.). Baltimore: Brookes.

Pichot, P. (1948). French pioneers in the field of mental deficiency. *American Journal of Mental Deficiency, 53,* 128–137.

Ramirez, R. R., & de la Cruz, G. (2002). *The Hispanic population in the United States: March 2002* (Current Population Reports No. 20-545). Washington, DC: U.S. Bureau of the Census.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and Reynolds Intellectual Screening Test: Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Reynolds, C. R., & Kamphaus, R. W. (2009). *Escalas de Inteligencia de Reynolds.* Madrid: TEA Ediciones.

Ryan, C. L., & Bauman, K. (2016). Educational attainment in the United States: 2015. Population characteristics [Current Population Reports, P20-578]. Washington, DC: U.S. Census Bureau.

Saling, L. L., & Philips, J. G. (2007). Automatic behavior: Efficient not mindless. *Brain Research Bulletin, 73,* 1–20.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Stepler, R., & Brown, A. (2016, April 19). 2014, Hispanics in the United States statistical portrait. Retrieved from *www.pewhispanic.org/2016/04/19/2014-statistical-information-on-hispanics-in-united-states.*

Venkatesan, S., & Basavarajappa, D. M. (2007). Seguin Form Board Test: Field try out on a modified procedure of test administration. *Indian Journal of Applied Psychology, 44*(1), 1–5.

Webster, P. E., Plante, A. S., & Couvillion, L. M. (1997). Phonologic impairment and prereading: Update on a longitudinal study. *Journal of Learning Disabilities, 30*(4), 365–375.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1991). *Wechsler Intelligence Scale For Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2001). *Wechsler Individual Achievement Test—Second Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scales of Intelligence—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., et al. (2004). *Wechsler Intelligence Scale for Children Fourth Edition—Integrated.* San Antonio, TX: Harcourt Assessment.

Wechsler, D., & Naglieri, J. A. (2006a). *Wechsler Nonverbal Scale of Ability.* San Antonio, TX: Harcourt Assessment.

Wechsler, D., & Naglieri, J. A. (2006b). *Wechsler Nonverbal Scale of Ability: Administration and scoring manual.* San Antonio, TX: Harcourt Assessment.

Wechsler, D., & Naglieri, J. A. (2006c). *Wechsler Nonverbal Scale of Ability: Technical and interpretive manual.* San Antonio, TX: Harcourt Assessment.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

CHAPTER 18

The Reynolds Intellectual Assessment Scales, Second Edition, and the Reynolds Intellectual Screening Test, Second Edition

Tara C. Raines
Cecil R. Reynolds
Randy W. Kamphaus

This chapter provides the reader with an extensive introduction to the Reynolds Intellectual Assessment Scales, Second Edition (RIAS-2; Reynolds & Kamphaus, 2015), and the Reynolds Intellectual Screening Test, Second Edition (RIST-2; Kamphaus & Reynolds, 2015)—recently updated cognitive assessments for children and adults. Brief overviews of the instruments and of noteworthy updates to the second editions are provided. Furthermore, a review of the theory and structure of the RIAS-2 is offered, using a goal-oriented developmental framework. Additionally, more extensive descriptions of the subtests, their administration, and their scoring are provided. Psychometric characteristics of the RIAS-2 as well as guidelines for interpretation are presented. Clinical applications of the RIAS-2 and RIST-2 are briefly discussed. A chapter appendix provides a case study using the RIAS-2 as the featured measure of intelligence.

The RIAS-2 is an individually administered test of intelligence appropriate for ages 3 years through 94 years with co-normed, supplemental measures of processing speed and memory. The RIAS-2 consists of eight subtests. It also includes a two-subtest Verbal Intelligence Index (VIX) and a two-subtest Nonverbal Intelligence Index (NIX). The scaled sums of *T* scores for the four subtests are combined to form the Composite Intelligence Index (CIX), which is the suggested summary estimate of global intelligence. For individuals preferring to include measures of memory and processing speed in the global estimate of intelligence, composite scores that include these subtests are also available. Administration of the four intelligence scale subtests by a trained, experienced examiner requires approximately 20–25 minutes. A Composite Memory Index (CMX) is derived from the two supplementary memory subtests, which requires approximately 10–15 minutes of additional testing time. Additionally, a composite Speeded Processing Index (SPI) consists of two subtests, which take an additional 5–10 minutes to administer. Each index is representative of both verbal and nonverbal subtests. Table 18.1 provides an overview of the RIAS-2 indexes and subtests.

Using a population-proportionate stratified sampling plan, the test developers standardized the RIAS-2 on a sample of 2,154 individuals. The RIAS-2 was designed to eliminate or minimize any cultural biases in the assessment of intelligence for individuals acculturated to and educated in the United States. In essence, this test was designed for individuals with English language proficiency. The RIAS-2 provides standardized *T* scores, which have a mean of 50 and a standard deviation of 10. Each of the indexes is scaled to the common standard score metric used for assessment of intellectual ability, with a mean of 100 and a standard deviation of 15. Via a method known as *continu-*

TABLE 18.1. RIAS-2 Subtests and Indexes

Subtest or index	Description
Composite Intelligence Index (CIX)	CIX provides a summary estimate of global intelligence, designed to estimate *g*, the general intelligence factor.
Verbal Intelligence Index (VIX)	VIX provides a summary estimate of verbal intelligence as assessed by verbal reasoning and reflecting primarily crystallized intellectual functions.
Guess What (GWH)	Examinees are given a set of two to three clues, and are asked to deduce the object or concept being described. This subtest measures verbal reasoning in combination with vocabulary, language development, and overall fund of available information.
Verbal Reasoning (VRZ)	Examinees listen to a propositional statement that essentially forms a verbal analogy, and are asked to respond with one or two words that complete the idea or proposition. This subtest measures verbal-analytical reasoning ability, but with fewer vocabulary and general knowledge demands than Guess What.
Nonverbal Intelligence Index (NIX)	NIX provides a summary estimate of nonverbal intelligence as assessed by nonverbal reasoning and reflecting primarily fluid intellectual functions.
Odd-Item Out (OIO)	Examinees are presented with a picture card containing five to seven pictures or drawings, and are asked to designate which one does not belong or go with the others. This subtest measures nonverbal reasoning skills, but also requires the use of spatial ability, visual imagery, and other nonverbal skills on various items. It is a form of reverse nonverbal analogy.
What's Missing (WHM)	A redesign of a classic task present on various ability measures. Examinees are shown a picture with some key element or logically consistent component missing, and are asked to identify the missing essential element. This subtest assesses nonverbal reasoning: The examinee must conceptualize the picture, analyze its gestalt, and deduce what essential element is missing.
Composite Memory Index (CMX)	CMX provides a summary estimate of verbal and nonverbal memory functions in general, as its component parts correspond to the broad areas of memory skills.
Verbal Memory (VRM)	This scale consists of a single verbal memory subtest. Depending upon the examinee's age, a series of sentences or brief stories are read aloud by the examiner and then recalled by the examinee. This task assesses the ability to encode, store briefly, and recall verbal material in a meaningful context where associations are clear and evident.
Nonverbal Memory (NVM)	This scale consists of a single visual memory subtest. It contains a series of items in which a stimulus picture is presented for 5 seconds, following which an array of pictures is presented. The examinee must identify the target picture from the new array of six pictures. It assesses the ability to encode, store, and recognize pictorial stimuli that are both concrete and abstract or without meaningful referents.
Speeded Processing Index (SPI)	SPI provides a summary estimate of processing speed, primarily involving both decision speed and reaction time while minimizing the efforts of fine motor speed.
Speeded Naming Task (SNT)	Examinees are asked to name pictures of common objects presented in a grid format, and must recognize them quickly, access their names, and speak these aloud to the examiner.
Speeded Picture Search (SPS)	Examinees are asked to find three target pictures within a large array of pictures as quickly as they can.

Note. Index titles appear in boldface. From Reynolds and Kamphaus (2015, p. 10). Copyright © Psychological Assessment Resources (PAR), Inc. Adapted by permission.

ous norming, all indexes yield age-corrected deviation scaled scores (Angoff & Robertson, 1987; Evers, Sijtsma, Lucassen, & Meijer, 2010; Reynolds & Kamphaus, 2015; Zachary & Gorsuch, 1985). Percentile ranks, *T* scores, *z* scores, normal curve equivalents, and stanines are also offered. Age-equivalent subtest scores are only offered for ages 3–14 (Reynolds & Kamphaus, 2015).

The RIST-2 is a general intelligence screening measure derived from subtests of the RIAS-2. This instrument is not designed to provide diagnostic information. Instead, it is designed to provide information on risk for intellectual atypicality. It consists of two subtests of the RIAS-2 (Guess What and Odd-Item Out), and is to be used to determine whether intellectual functioning deviates enough from the mean to warrant additional assessment. This type of screening may be helpful in screening for giftedness and talents or developmental delays, screening for the impact of a stroke or cerebrovascular accident, or screening for the impact of dementia on cognitive functioning.

THEORY AND STRUCTURE

The RIAS-2 has combined practical and theoretical aspects of the assessment of intelligence. This instrument measures five valuable facets of intelligence: general intelligence (including fluid reasoning); verbal intelligence (crystallized abilities); nonverbal intelligence (visual and spatial abilities); memory (short-term memory and learning); and processing speed (decision speed or reaction time). The selection of these aspects was based on their sound history of scientific support (Carroll, 1993; Goldstein & Reynolds, 2011; Horn & Cattell, 1966; Kamphaus, 2001; Reynolds & Kamphaus, 2015). The RIAS-2 is rooted in the Cattell–Horn–Carroll (CHC; Horn & Cattell, 1966) model of intelligence, which is grounded in the assessment of fluid and crystallized intelligence. However, based on more current and practical acknowledgment of the value of both processing speed and memory in the assessment of cognitive ability, subtests measuring these abilities were added to the RIAS-2. It is the test authors' hope that with these additional measures, practitioners can glean comprehensive and useful information about individual cognitive abilities. Additional information about the theoretical underpinnings of the RIAS-2 is available in the RIAS-2 manual (Reynolds & Kamphaus, 2015).

Development Goals

Reynolds and Kamphaus (2015) describe a set of nine primary goals for development of the RIAS-2. These goals are based on their experiences of teaching courses on the assessment of cognitive abilities, and of administering and interpreting many different measures of intelligence; they are also based on the current literature surrounding theoretical models of intelligence and research on intelligence test interpretation. The goals of the instrument build upon the goals established in the development of its predecessor, the RIAS (Reynolds & Kamphaus, 2003). Table 18.2 summarizes these goals, which are reviewed and discussed only briefly below. Additional information on these goals can be found in the RIAS-2 manual (Reynolds & Kamphaus, 2015).

- *Goal 1: Provide a reliable and valid measurement of g and its two primary components, verbal and nonverbal intelligence, with close correspondence to*

TABLE 18.2. Summary of Nine Primary Goals for Development of the RIAS-2

1. Provide a reliable and valid measurement of *g* and its two major practical components, verbal and nonverbal intelligence, with close correspondence to crystallized and fluid intelligence.
2. Provide a practical measurement device in terms of efficacies of time, direct costs, and information needed from a measure of intelligence.
3. Allow continuity of measurement across all developmental levels for ages 3–94 years for both clinical and research purposes.
4. Substantially reduce or eliminate dependence upon motor coordination and visual–motor speed for measurement of intelligence.
5. Eliminate dependence upon reading for measurement of intelligence.
6. Provide accurate predictions of basic academic achievement, at levels at least comparable to those of intelligence tests twice the length of the RIAS-2.
7. Apply familiar, common concepts that are clear and easy to interpret, coupled with simple, easy administration and scoring.
8. Eliminate items that show differential item functioning associated with gender or ethnicity.
9. Provide for the verbal and nonverbal assessment of memory.

Note. From Reynolds and Kamphaus (2015, p. 11). Copyright © Psychological Assessment Resources (PAR), Inc. Adapted by permission.

crystallized and fluid intelligence. The general intelligence factor, *g*, is the most reliable and enduring aspect of any multifactorial view of intelligence (Deary, 2012; Jensen, 1998). In the Cattell–Horn model (Horn & Cattell, 1966; Kamphaus, 2001) of intelligence, *g* is the dominant factor in the hierarchy of multiple abilities, with the next two dominant facets being crystallized and fluid intelligence. These components bear a close association to the common view of verbal and nonverbal intelligence. The RIAS-2 includes subtests that match both constructs closely to share in the theoretical support of the Cattell–Horn model of crystallized and fluid intelligence, while taking advantage of the very practical division of intelligence into verbal and nonverbal components. Verbal and nonverbal components of intelligence also have strong support from factor-analytic work (e.g., Kaufman, 1994) and the brain sciences (e.g., Riccio & Hynd, 2000).

- *Goal 2: Provide a practical measurement device in terms of efficacies of time, direct costs, and information needed from a measure of intelligence.* Time, cost, and efficiency have always been necessary considerations in the delivery of effective psychological and psychoeducational services. With the growing demand to provide services to children with disabilities in schools, time has become a fundamental consideration for practitioners assessing cognitive abilities. A useful intelligence test needs to provide an objective, reliable, and valid assessment of the major constructs that underlie psychometric intelligence. Intellectual assessment can be conducted efficiently and at a significantly lower cost than it has been in the past. The RIAS-2 was designed to provide an efficient measure of intelligence that is aligned with the guidelines for eligibility, such as those put forth by the special education and Social Security laws in the United States.

- *Goal 3: Allow continuity of measurement across all developmental levels from ages 3 years through 94 years for both clinical and research purposes.* Individuals often require reevaluation of cognitive ability throughout their lives. As they age, different versions of intelligence tests may be required. These various tests have different subtests, measure different aspects of intelligence, were normed in different years, and may have sample stratifications that are not aligned. Thus scores obtained over time may not be comparable because of measurement artifact, and not because of any real changes in an individual's cognitive abilities or structure (Sattler, 2001). Thus there is clinical utility in having a common set of subtests and a common reference group for such comparisons, just as there is utility in longitudinal research that includes intelligence as a variable, whether dependent, independent, or moderator in nature (Andrews, 2007).

- *Goal 4: Substantially reduce or eliminate dependence on motor coordination and visual–motor speed in the measurement of intelligence.* The majority of current individually administered intelligence tests rely heavily on visual–motor coordination and motor speed for accurate assessment of intelligence (Dombrowski & Mrazik, 2008). However, many children referred for special education services struggle with visual–motor tasks or have other neurodevelopmental disorders that produce motor-related problems (Goldstein & Reynolds, 2011). Furthermore, in older populations, the incidence of tremor and related motor problems is quite high. To attempt to measure the intelligence of such individuals with tasks that require rapid manipulation of blocks, small cardboard pieces, or even pencil markings, where speed and accuracy of performance are substantial contributors to the resulting IQ or cognitive index, seems inappropriate. It is our view that intelligence tests should emphasize thinking, reasoning, and problem solving.

- *Goal 5: Eliminate dependence on reading ability in the measurement of intelligence.* Tasks where the ability to read the English language facilitates individual item performance confound the measurement of intelligence with academic ability. Certainly, intelligence cannot be assessed completely independently of prior knowledge, despite many failed attempts to do so (e.g., culture-free tests; see Anastasi & Urbina, 1997; Kamphaus, 2001). However, to confound intellectual assessment with academic skills such as reading and spelling makes such tests inappropriate for nonreaders, individuals with limited English-reading skills, and persons with visual impairments.

- *Goal 6: Provide accurate predictions of basic academic achievement, at levels at least comparable to those of intelligence tests twice the length of the RIAS-2.* Predicting academic achievement and acquired knowledge in such areas as reading, language, and math is an important function for intelligence tests. Predicting achievement should remain a function of any new intelligence test.

- *Goal 7: Apply familiar, common concepts that are clear and easy to interpret, coupled with simple administration and scoring.* Formal intelligence testing via Binet-type tasks is only a century old. During this time, innumerable tasks have been devised to measure intelligence and related abilities. Many of these tasks are quite good at the measurement of cognitive ability and possess long histories in psychology and education. The use of familiar, well-researched tasks has many advantages over the use of novel, less well-established tasks. Many of these tasks are simple and easy to administer, despite the complex mental functions required for deriving a correct solution. Avoiding tasks that require lengthy verbal responses or split-second timing for awarding bonus points can facilitate objective scoring. Tasks that are simple to administer and clearly objective to score reduce or nearly eliminate administration and scoring errors.

- *Goal 8: Eliminate items that show differential item functioning associated with gender or ethnicity.* The problem of bias, whether a function of gender, race, or culture, has long produced debate in psychology, education, and the assessment of intelligence (e.g., Brown, Reynolds, & Whitaker, 1999; Reynolds, Lowe, & Saenz, 1999). Following years of debate, a host of methods for detecting test items that function differentially across nominally defined groups has been devised (Reynolds, 2000). However, in view of the availability of sound statistical approaches for identifying such test items, all intelligence test items should be scrutinized during development and standardization to determine whether the functions are measuring the same constructs and skills across groups.
- *Goal 9: Provide for the verbal and nonverbal assessment of memory.* Brief assessments of memory have been included in intelligence measures since the onset of the practice (Binet & Simon, 1905). The outcomes of this portion of the assessment are typically included in the composite intelligence score or quotient. The RIAS-2 was designed to assess memory, with the understanding that memory is a pivotal part of the diagnostic assessment of many disorders throughout the lifespan. Via conorming, the RIAS-2 allows practitioners to compare IQ directly with key memory functions.

In addition to these nine primary goals, the developers of the RIAS-2 updated the item content from the original RIAS by removing controversial and dated items, adding new content, and extending the basal and ceiling rules. Additionally, all norms were updated with current U.S. census data. These updates were made at the requests of RIAS users to increase the value, acceptance, and applicability of the RIAS-2 as a measure of intellectual functioning. It was with these overarching goals in mind that the RIAS-2 was designed and its structure implemented.

Theory

As noted above, the RIAS-2 was designed to measure five important aspects of intelligence: (1) *general intelligence* (of which the major component is *fluid* or *reasoning abilities*); (2) *verbal intelligence* (sometimes referred to as *crystallized abilities,* which is a closely related though not identical concept); (3) *nonverbal intelligence* (referred to in some theories as *visualization* or *spatial abilities,* and closely allied with fluid intelligence); (4) *memory* (subtests covering this ability have been labeled variously as assessing *working memory, short-term memory,* or *learning*); and (5) *processing speed* (subtests assessing *rapid decision making* and *reaction time*). These five constructs are measured by combinations of the eight RIAS-2 subtests (see Table 18.1).

The RIAS-2 subtests were selected and designed to measure intelligence constructs that have a substantial history of scientific support. In addition, Carroll's (1993) seminal and often-cited three-stratum theory of intelligence informed the creation of the RIAS-2 by demonstrating that many of the latent traits tapped by intelligence tests were test-battery-independent. He clearly demonstrated, for example, that numerous tests measured the same crystallized, visual-perceptual, and memory abilities. However, Kamphaus (2001) concluded that these same test batteries did not measure fluid abilities well.

The RIAS-2 focuses on the assessment of stratum III and stratum II abilities from Carroll's (1993) three-stratum theory. Stratum III is composed of one construct only, *g.* Psychometric *g* accounts for the major portion of variance assessed by intelligence test batteries. More important, however, is the consistent finding that the correlations of intelligence tests with important outcomes, such as academic achievement and occupational attainment, are related to the amount of *g* measured by the test battery. In other words, so-called "*g*-saturated" tests are better predictors of important outcomes than are tests with low *g* saturation. Although the nature of *g* has yet to be fully understood, the scores from *g*-saturated tests have known utility, especially in terms of predictive validity. One theory posits that *g* is actually a measure of working memory capacity (Kyllonen, 1996), whereas another theory posits that it is reasoning ability (Gustafsson, 1999). Regardless of the theory that will eventually be supported, the utility of psychometric *g* remains, much in the same way that the usefulness of certain pharmaceutical drugs will continue before their mechanisms of action are fully understood. For these reasons, the RIAS-2 places great importance on the assessment of psychometric *g* and on the assessment of its theorized main components (i.e., verbal and nonverbal reasoning, working memory and processing speed). With this in mind, the developers of the RIAS-2 settled upon the inclusion of subtests that emphasize reasoning skills and the application of knowledge to problem solving on the intelligence subscales. This also guided the inclusion of two working memory and two processing speed tasks on the separate memory and processing speed subscales.

The second stratum in Carroll's (1993) hierarchy consists of traits that are assessed by combinations of subtests, or stratum I measures. A stratum I measure is typically a single subtest that measures the trait of interest. Combinations of stratum I subtests, such as those used to form the VIX and NIX, are considered stratum II measures and should result in enhanced measurement of complex traits such as verbal and nonverbal intelligence. Combining stratum II index measures into an overarching composite measure, such as the CIX, allows for the measurement of a complex stratum III trait, such as general intelligence.

There are, however, several stratum II traits to choose from. These traits include fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, broad cognitive speed, and processing speed (i.e., reaction time or decision speed). However, it is important to note that myriad investigations suggest that these abilities are ordered by their assessment of *g* (Kamphaus, 2001). Specifically, subtests that tap fluid abilities are excellent measures of *g*, whereas tests of psychomotor speed are weak. If one accepts the aforementioned finding that *g* saturation is related to predictive validity, then the first few stratum II factors become the best candidates for inclusion in an intelligence test battery like the RIAS-2.

Any test of *g* must measure so-called "higher-order" cognitive abilities and those associated with fluid abilities, such as general sequential reasoning, induction, deduction, syllogisms, series tasks, matrix reasoning, analogies, quantitative reasoning, and so on (Carroll, 1993). Kamphaus (2001) advocated the following definition of reasoning: "that which follows as a reasonable inference or natural consequence; deducible or defensible on the grounds of consistency; reasonably believed or done" (*New Shorter Oxford English Dictionary,* 1999). This definition emphasizes a central cognitive requirement to draw inferences from knowledge. This characteristic of general intelligence is measured best by two RIAS-2 subtests, Verbal Reasoning and Odd-Item Out, although all subtests have substantial *g* saturation (see Reynolds & Kamphaus, 2015).

First-order factors of crystallized ability typically have one central characteristic: They involve language abilities (Vernon, 1950). These language abilities range from vocabulary knowledge to spelling to reading comprehension. On the other hand, it is not possible to dismiss this type of intelligence as a general academic achievement factor (Kamphaus, 2001), for several reasons. First, indicators of a latent construct do not necessarily reflect the content of the abilities that they represent. For example, the Snellen chart is the standard measure of visual acuity, but it is essentially a letter recognition test. In other words, the contrived nature of this task is far removed from the typical daily activities involving vision; however, performance on the task is a good indicator of visual acuity in numerous activities. Second, Carroll (1993) identified a separate set of factors associated with school achievement and knowledge that included tests of specific academic subject areas (e.g., English, history), among other tests—a finding suggesting that the "reading" and "writing" tests of the crystallized intelligence factor were measuring an intelligence construct. Otherwise, these same tasks would have loaded on factors that are clearly recognizable as academic attainments.

Kamphaus (2001) has suggested that the term *crystallized* for this second-order factor does not fully capture the centrality of language processes involved in successful performance on subtests typically associated with this ability. He proposed the term *verbal* to describe the latent construct tapped by subtests like those selected for the RIAS-2 to assess this ability. Furthermore, Kamphaus suggested this label would better represent the centrality of language development to successful performance on these subtests, while recognizing the wide variety of tests that measure this latent trait (e.g., word reading and spelling). In addition, the term *verbal* has often been used by the general public when describing types of intelligence—a fact suggesting that this label may be helpful for communicating the results derived from such measures (Weinberg, 1989). Kamphaus proposed that the verbal factor be defined as "oral and written communication skills that follow the system of rules associated with a language" (p. 45), including comprehension skills. The RIAS-2 includes Guess What and Verbal Reasoning as measures of verbal ability. Verbal Memory is also a member of the verbal factor in terms of factor-analytic results.

Nonverbal tests have come to be recognized as important measures of spatial and visual-perceptual abilities—abilities that may need to be assessed for a variety of clients, including those with brain injuries. In the 1963 landmark Educational Testing Service Kit of Factor-Referenced Cognitive Tests, spatial ability was defined as "the ability to manipulate or transform the image of spatial patterns into other visual arrangements" (quoted in Carroll, 1993, p. 316). The RIAS-2 subtests What's Missing and Odd-Item Out follow in this long tradition of tasks designed to measure visual–spatial abilities.

Digit recall, sentence recall, geometric design recall, bead recall, and similar measures loaded consistently on a *general memory and learning* stratum II factor identified by Carroll (1993) in his numerous analyses. The RIAS-2 Verbal Memory and Nonverbal Memory subtests are like these tasks, although more complex than simple confrontational memory tasks such as pure digit recall. Carroll's findings suggest that the RIAS Verbal Memory and Nonverbal Memory subtests should be good measures of the memory construct that has been identified previously in many investigations of a diverse array of tests. Carroll described memory span as "attention to a temporally ordered stimulus, registration of the stimulus in immediate memory, and output of its repetition" (p. 259). This operational definition is an accurate description of the RIAS-2 memory subtests and composite. Memory is typically considered a complex trait with many permutations, including visual, verbal, long-term, and short-term. Carroll's analysis of hundreds of datasets supports the organization of the RIAS-2, in that he found ample evidence of a general memory trait that may be subdivided further for particular clinical purposes.

The rate at which an individual can perform a simple task quickly is typically considered a reflection of cognitive processing speed. This is aligned with Carroll's stratum II construct of *processing speed*. In alignment with the verbal–nonverbal approach of the RIAS-2, subtests that measure processing speed in both modalities are offered. More notably, the developers of this assessment actively sought tasks of processing speed that reduce reliance on motor skills to demonstrate ability in this area. However, performance on this task may still be confounded by other variables such as attention. In a three-factor solution, these processing speed tasks emerge as distinct from other RIAS-2 tasks and are thus valuable as a unique construct to be assessed. The Speeded Naming Task and Speeded Picture Search subtests are used to measure these abilities. Additional information about the theory that drives the RIAS-2 can be found in the test manual.

Description of Subtests

Subtests from the original RIAS were updated and expanded for the RIAS-2. These subtests have a familiar look and feel, with deep histories in the field of intellectual assessment. There are a total of four intelligence subtests, two memory subtests, and two processing speed subtests. The intelligence subtests were chosen due to their complex nature: Each assesses many intellectual functions and requires their integration for successful performance. The memory and processing speed subtests were chosen not only for complexity, but also due to their representation of the primary content domains of the constructs being measured. A brief description of the subtests is provided below.

- *Guess What.* This subtest measures vocabulary knowledge in combination with reasoning skills that are predicated on language development and funds of information. For each item, the examinee is asked to listen to a question that contains clues presented orally by the examiner, and then to give a verbal response (typically one or two words) that is consistent with the clues. The questions pertain to physical objects, abstract concepts, and well-known places and historical figures from a variety of cultures and geographic locations. Many disciplines are represented among the test questions.
- *Verbal Reasoning.* The second verbal subtest measures analytical reasoning abilities. More difficult items, of necessity, also require advanced vocabulary knowledge. For each item, the exam-

inee is asked to listen to an incomplete sentence presented orally by the examiner, and then to give a verbal response (typically one or two words) that completes the sentence—most commonly completing a complex analogy. Completion of the sentences requires the examinee to evaluate the various conceptual relationships that exist between the physical objects or abstract ideas contained in the sentences. Many different types of relationships from a broad content range are represented in these items.

- *Odd-Item Out.* This subtest measures general reasoning skills, emphasizing nonverbal ability. For each item, the examinee is presented with a picture card containing five to seven figures or drawings. One of the figures or drawings on the picture card has a distinguishing characteristic, making it different from the others. For each item, the examinee is given two chances to identify the figure or drawing that is different from the others. Two points are awarded for a correct response given on the first attempt. One point is awarded for a correct response given on the second attempt (i.e., if the first response was incorrect).

- *What's Missing.* This subtest measures nonverbal reasoning skills through the presentation of pictures in which some important component of the pictured object is missing. Examinees must conceptualize the pictured object, assess its gestalt, and distinguish essential from nonessential components. For each item the examinee is shown a picture card, asked to examine the picture, and then to indicate (in words or by pointing) what is missing. Verbally naming the missing part correctly is not required, so long as the examinee can indicate the missing component correctly. For each item, the examinee is given two chances to identify what is missing from the picture. Two points are awarded for a correct response provided on the first attempt. One point is awarded for a correct response provided on the second attempt. Two incorrect responses are awarded a score of 0.

- *Verbal Memory.* This subtest measures the ability to encode, briefly store, and recall verbal material in a meaningful context. Young children (ages 3–4 years) are asked to listen to sentences of progressively greater length as each is read aloud by the examiner; they are then asked to repeat each sentence back to the examiner, word for word, immediately after it is read aloud. Older children and adults are asked to listen to two stories read aloud by the examiner and then to repeat each story back to the examiner, word for word, immediately after it is read aloud. The sentences and stories were written to provide developmentally appropriate content and material of interest to the targeted age group. Specific stories are designated for various age groups.

- *Nonverbal Memory.* This subtest measures the ability to encode, briefly store, and recall visually presented material, whether the stimuli represent concrete objects or abstract concepts. For each item, the examinee is presented with a target picture for 5 seconds, and then with a picture card containing the target picture and an array of similar pictures. The examinee is asked to identify the target picture among the array of pictures presented on the picture card. For each item, the examinee is given two chances to identify the target picture. Two points are awarded for a correct response given on the first attempt. One point is awarded for a correct response given on the second attempt. Two incorrect responses are awarded a score of 0. The pictures are, primarily abstract at the upper age levels, and pictures of common objects at the lower age levels. The use of naming and related language strategies, however, is not helpful, due to the design of the distractors. For example, one early item presents as a target stimulus a picture of a cat. On the recall page, six cats are presented, each characteristically different (except one) from the target stimulus.

- *Speeded Naming Task.* This subtest measures the ability to differentiate stimuli and verbally name them under time constraints. Examinees are presented with pictures of common objects in a grid format. They are then asked to recognize them quickly and speak the name of each object aloud to the examiner. The subtest provides an accurate report of verbal processing speed (aligned with CHC definitions) for individuals with intact attentional, visual-perceptual, and speech abilities.

- *Speeded Picture Search.* This subtest assesses the ability to nonverbally differentiate stimuli under time constraints. Examinees are shown an array of pictures and asked to identify three target pictures as quickly as possible. This, like the other processing speed subtest, is aligned with CHC theory, as the stimuli are simple and could be identified with ease without the time constraints. It is the coupling of the task with the time limits that makes this a measure of processing speed. This measure provides an accurate account of processing speed ability for individuals who do not have reported attentional, visual-perceptual, or speech impairments.

ADMINISTRATION AND SCORING

The RIAS-2 was specifically designed to be easy to administer and objective to score. Examiners giving the RIAS-2 should be experienced with general test administration procedures and properly credentialed in their state or province. They should also review all materials associated with the test, including the manual, stimulus books, and record form, before attempting to administer this test. To promote efficient administration, the RIAS-2 and RIST-2 record forms contain the necessary instructions and examiner guides necessary to administer the tests. Experienced examiners as well as graduate students have consistently reported that these tests are surprisingly easy to administer and score.

Once an examiner has established rapport with an examinee, the protocols outlined in the manual for standardized administration should be followed. Instructions for the examinee are short and simple. Basal and ceiling rules as well as start points are clearly labeled and outlined in both the record form and the test manual, and should be followed to maintain fidelity of administration. For all subtests except Verbal Memory, there are clear, objective lists of correct responses for each test item, and seldom are any judgment calls required. Studies of the interscorer reliability of these five subtests with trained examiners produced reliability coefficients of .99 (Reynolds & Kamphaus, 2015). On Verbal Memory, some judgment is requited when examinees do not give verbatim responses; however, the scoring criteria provide clear examples and guidelines for such circumstances, making the Verbal Memory subtest only slightly more difficult to score than the others. The interscorer reliability study of this subtest produced a coefficient of .95.

The time required to administer the entire RIAS-2 (including the intelligence, memory, and processing speed subtests) averages 40–45 minutes once the examiner has practiced giving the RIAS-2 and has become fluent in its administration. The administration time for the RIST-2 (the screening subtests of the RIAS-2) averages about 12–15 minutes. As with most tests, the first few administrations are likely to take longer. The four intelligence subtests alone (i.e., Guess What, Odd-Item Out, Verbal Reasoning, and What's Missing), can be administered to most examinees in about 20–25 minutes. The two memory and two processing speed subtests can typically be administered in about 10 minutes for each set. However, significant time variations can occur as a function of special circumstances (e.g., very-low-functioning individuals will likely take much less time to complete the battery, and very-high-functioning individuals may take longer). A detailed description of the methods used for setting these administration parameters is given in the test manual.

PSYCHOMETRIC PROPERTIES

The psychometric characteristics of any measurement device and its scores are crucial in determining its utility. In this section, we summarize the reliability of the scores derived from the RIAS-2 and RIST-2, as well as evidence related to the validity of their interpretations. Due to the length restrictions in a single book chapter, a discussion of the developmental process of the tests simply cannot adequately be provided; such a discussion can be found in the test manual. However, the RIAS-2 underwent years of development, including tryout and review of the items on multiple occasions by school psychologists, clinical psychologists, and neuropsychologists. Items were written to conform to clear specifications, consistent with the goals for development of the test as given previously in this chapter. Items were reviewed by panels of expert psychologists for content and construct consistency to ascertain the cultural saliency of the items and to eliminate any potential problems of ambiguity or offensiveness in various settings. The developmental process speaks directly to the psychometric characteristics of the tests and is described in greater detail in Reynolds and Kamphaus (2015). These should be considered carefully in any full evaluation of the instrument.

Standardization

The RIAS-2 was normed on a sample of 2,154 participants residing in 32 states between the years of 2013 and 2014. U.S. Census Bureau projected characteristics of the U.S. population for 2012 were used to select a population-proportionate sample. Age, gender, ethnicity, educational level (parental educational level was used for ages 3 years through 16 years, and the participants' actual educational level was used at all other ages), and region of residence were used as stratification variables. Participants in the norming were screened for many low-incidence disorders that might interfere with their performance on the test, such as color blindness, uncorrected hearing and vision difficulties, or a history of traumatic brain injury (TBI) or attention-deficit/hyperactivity disorder (ADHD). The resulting norms for the RIAS-2 and the RIST-2 were calculated on a weighted sampling that provided a virtually perfect match to the census data. The overall sample was a close match to the population statistics. Tables with specific demographic information are available in the test manual.

Score Reliability

Reliability is a psychometric concept referring to the accuracy and consistency of scores obtained on a measurement device, and not that of the test itself. To the degree to which test scores contain error, the reliability of the score is reduced. Errors in test scores may come from a variety of sources, such as errors associated with content sampling, time sampling, administrative and scoring errors, and the like. Traditionally in psychology, these error sources are evaluated by using specialized correlation procedures. The largest source of error in test scores is typically errors due to content sampling. To assess this source of error, measures of the internal consistency of the items are used, the most widely known and applied being Cronbach's (1951) alpha.

Since the RIAS-2 is a power test (items are presented in order of difficulty, from least to most difficult, and individuals' scores depend entirely on how many items they respond to correctly), the internal-consistency reliability of the items on the RIAS-2 subtests was investigated by using Cronbach's coefficient alpha. Internal-consistency estimates for the RIAS-2 indexes (i.e., the VIX, NIX, CIX, CMX, and SPI) were derived by using a simplification of Guilford's (1954, p. 393) formula. This formula was designed for application to the special case present in the RIAS-2, where both the VIX and the NIX have only two components, each scaled to a common metric, and where the CIX represents a linear, equally weighted composite of the sums of *T* scores for the four subtests (see Nunnally, 1978, p. 249, formula 7-15). Alpha reliability coefficients for the RIAS-2 subtest scores and the Nunnally reliability estimates for the index scores are presented in the test manual. It is noteworthy that 100% of the alpha coefficients for the RIAS subtest scores reach .84, or higher, for every age group. Further, the median alpha reliability estimate for each RIAS-2 subtest across age equals or exceeds .90. This point is important because many measurement experts recommend that reliability estimates above .80 are necessary and those above .90 are highly desirable for tests used to make decisions about individuals. All RIAS-2 subtests meet these recommended levels.

One cannot always assume that because a test is reliable for a general population, it will be equally reliable for every subgroup within that population. Therefore, test developers and researchers should demonstrate, when possible, that the instrument demonstrates measurement equivalence and reliability for subgroups. This is particularly important for groups that, because of gender, racial, cultural or linguistic differences, might experience test bias (Reynolds, 2000). Thus it is instructive to view the various reliability estimates for the RIAS-2 (or any test) for smaller, meaningful subgroups of a population. As noted in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1999, 2014), these values may also provide information relevant to the consequences of test use. Reliability estimates reported in the RIAS-2 manual suggest uniform internal consistency across age (3–94 years), as well as across gender and racial groups.

Test score stability (sometimes referred to as *error due to time sampling*) refers to the extent to which an individual's test performance is constant over time and is usually estimated by the test–retest method. In this procedure, the test is given to a group of individuals, a period of time is allowed to pass, and the same individuals are tested again. Then the results of the two administrations are compared. The stability of RIAS-2 scores over time was investigated by using the test–retest method with 97 individuals ages 3 years through 72 years. The intervals between the two test administrations ranged from 7 to 43 days, with a median test–retest interval of 18 days. The correlations for the two administrations, along with mean scores and standard deviations, are reported in detail in the RIAS-2 manual (Reynolds & Kamphaus, 2015).

The obtained coefficients are of sufficient magnitude to allow confidence in the stability of RIAS-2 test scores over time. The values are quite good for all of the subtests, but especially for the index scores. Both the uncorrected coefficients and the corrected, or disattenuated, coefficients are reported (the corrected coefficients have been corrected for the alpha for each subtest). When viewed across age groups, the values are generally consistent with the values obtained for the total test–retest sample. The test–retest stability coefficients for scores on the RIAS-2 subtests and indexes are quite strong and provide evidence of more than sufficient short-term temporal stability of the scores to allow examiners to be confident in the obtained results.

Validity of RIAS-2/RIST-2 Test Scores as Measures of intelligence

According to the 1999 edition of the *Standards, validity,* in this context, refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed users of tests" (AERA et al., 1999, p. 9). Reynolds (1998) has defined *validity* similarly, arguing that it refers to the appropriateness and accuracy of the interpretation of performance on a test (see also Montgomery, Torres, & Eiseman, Chapter 30, this volume). Validation of the meaning of test scores is also a process—one that involves an ongoing, dynamic effort to accumulate evidence for a sound scientific basis for proposed test score interpretations (AERA et al., 1999; Reynolds, 1998). Validity will always be a relative concept because the validity of an interpretation will vary according to the purpose for which test scores are being used, the types of individuals or populations being examined, and the specific interpretations being made.

As with any measure of intelligence, many other basic but subsidiary cognitive processes, such as auditory and visual perception, logical reasoning, language processing, spatial skills, visual imagery, attention, and the like, play a role in performance on the RIAS-2. These skills are the building blocks of the primary intellectual functions assessed by the RIAS-2. As previously described, the cognitive abilities measured by the RIAS-2 are grounded in Carroll's (1993) work. Initial validity evidence associated with the RIAS-2 indexes is highly consistent with the long history of intelligence testing research. A comprehensive review of the validity evidence is presented in the test manual.

Correlations with the Wechsler Scales

Measures of intelligence should generally correlate well with one another, if they are measuring *g* and related constructs. Thus, to understand a new measure and its appropriate interpretations, it is instructive to assess the relationship of the new measure to other measures of intelligence. For children (ages 6 years through 16 years), the best-known and most widely researched scale over the years and one that has maintained a reasonably consistent structure is the Wechsler Intelligence Scale for Children, which at the time the RIAS-2 was developed was in its fourth edition (WISC-IV; Wechsler, 2003). Tables available in the test manual demonstrate the correlations between the WISC-IV and the RIAS-2. The RIAS-2 indexes all correlated highly with the WISC-IV Full Scale IQ (FSIQ), with correlations ranging from a low of .58 (SPI to FSIQ) to a high of .77 (CIX to FSIQ). This pattern is predictable, as the highest correlations were between aspects of the tests measuring similar elements of *g*.

A group of 72 adults were administered the RIAS-2 and the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV; Wechsler, 2008) in a counterbalanced design. Most of the RIAS-2 indexes correlated positively with WAIS-IV composites, ranging from a low of .23 (SPI with Perceptual Reasoning) to a high of .77 (VIX with Verbal Comprehension). Again, the strongest relationships were demonstrated between indexes measuring similar constructs (e.g., VIX with Verbal Reasoning).

Correlations with Measures of Academic Achievement

One of the major reasons for the development of the early, individually administered intelligence tests was to predict academic achievement levels. Intelligence tests have done well as predictors of school learning, with typical correlations in the mid-.50s and .60s (for summaries, see Kamphaus, 2001; Sattler, 2001). To evaluate the relationship between the RIAS-2 and academic achievement, 253 individuals across the lifespan (ages 4–85 years) were administered the RIAS-2 and the Academic Achievement Battery (AAB; Messer, 2014). Correlations between the RIAS-2 and the AAB indexes were moderately strong and positive, ranging from .16 (SPI with Expressive Communication and Mathematical Reasoning) to .56 (CIX with Total Achievement). Learning is school is heavily dependent on language, which is confirmed by the positive correlations between the VIX and all AAB indexes. A more in-depth look at these relationships is provided by the relevant tables in the RIAS-2 manual.

Performance of Clinical Groups

Examination of performance on the RIAS-2 by groups of individuals who meet formal criteria for various diagnoses can also be instructive. For example, individuals with intellectual and developmental disabilities (IDD), dementia, and other cognitive problems associated with intellectual impairment should earn lower scores on the RIAS-2 than their typically developing peers in the RIAS-2 standardization sample. In interpreting such scores of preselected samples when those samples are selected on the basis of extreme scores on a cognitive measure, one must always consider the problem of regression to the mean on a second testing. For example, scores obtained from a sample with IDD will typically be higher on a second testing, but the scores should still be well below the population mean.

During the standardization of the RIAS-2, 12 different clinical groups were identified, and their scores on the RIAS-2 were analyzed to supplement the validation of the interpretation of RIAS scores. In each instance, the primary diagnosis given by the agency serving the examinee was accepted, and no independent review or diagnosis was undertaken. The samples included individuals with TBI, dementia, stroke/cerebrovascular accident, IDD, deafness and hearing impairments, specific learning disabilities (SLD), and a variety of others. The various impairments represented many organic deficits within each category, along with diffuse brain lesions. There were samples of children with SLD and ADHD that came from more than one source. In reviewing the outcomes of testing with the RIAS-2 with these 12 clinical samples, Reynolds and Kamphaus (2015) concluded that all of the various clinical groups in these studies demonstrated some levels of deviation from the population mean on the RIAS-2. Although most deviations are small, as is commonly found in the literature, the samples with more severe disorders performed more poorly. Again, however, the purpose for presenting these data is not to make definitive statements about these clinical groups, but to describe how the RIAS-2 scores function for each group. Moreover, these data are preliminary and not definitive as to any score patterns on the RIAS-2 that may emerge for clinical groups. Replication with larger and more carefully defined samples will be necessary before firm conclusions can be drawn. The evidence thus far is quite supportive because the score patterns do conform well.

APPLICATIONS OF THE RIAS-2 AND RIST-2

As a measure of intelligence, the RIAS-2 is appropriate for a wide array of purposes and should be useful when assessment of an examinee's intellectual level is needed. The RIAS-2 can be used across school and clinical settings with preschool and school-age children for purposes of diagnosis and intervention plan development; it is useful for students with disabilities, as well as gifted and talented students. Diagnosis of specific disorders, such as IDD, SLD, and the various dementias, often require the use of an intellectual test as a part of evaluation. The RIAS-2 is appropriate for such applications. Clinicians who perform general clinical and neuropsychological evaluations will find the RIAS-2 very useful when a measure of intelligence is needed. Practitioners will also find the RIAS-2 useful in disability determinations under various state and federal programs, such as the Social Security Administration's disability program and Section 504 regulations. Employers may find the RIAS-2 or RIST-2 helpful as a tool to predict success in job training programs and, in other instances, to predict possible difficulties when lower limits are set on IQ levels for specific jobs. Finally, the RIAS-2 and RIST-2 are strong predictors of academic performance. When intelligence level is a question in such situations, the RIAS and the RIST are appropriate choices.

Appendix 18.1 presents a case report that demonstrates how the RIAS-2 can be integrated into a comprehensive evaluation. It is an amalgamation of several authentic cases, and therefore does not resemble any one individual.

CONCLUSIONS

The RIAS-2 and RIST-2 are standardized assessments of intellectual functioning for individuals ages 3–94 years. The RIAS-2 is a quick and efficient comprehensive measure of cognitive ability. The RIST-2 is a screening instrument designed to guide identification of risk for atypical cognitive functioning. These instruments, standardized on a large and diverse sample, are firmly rooted in widely accepted intelligence test theory. The RIAS-2/RIST-2 indexes capture the most empirically sound aspects of intelligence, supporting their use across populations and affirming the utility of the results. These instruments were developed with special attention to well-reasoned modern standards for psychological and educational assessment (AERA et al., 1999, 2014). Therefore, the RIAS-2 and RIST-2 are optimal for use in making determinations for educational or clinical purposes, as well as for benefits eligibility.

APPENDIX 18.1

Sample Case Report

Child's name: Tamik Parris
Parent's name: Tanya Parris
School: One Great Elementary School
Address: 123 Awesome Street, Excellent City, Fine State
Grade: 1
Test date: November 10, 2016
Examiner: Karen Schuyler
DOB: 10/14/2010 Age: 6:0

CONSULTATIONS

Teacher, Nicole Talapatra; parent, Tanya Parris

INSTRUMENTS ADMINISTERED

Reynolds Intellectual Assessment Scales, Second Edition (RIAS-2)
Kaufman Test of Educational Achievement—Second Edition (KTEA-II)
Behavior Assessment System for Children, Third Edition (BASC-III), Teacher Rating Scale (TRS) and Parent Rating Scale (PRS)
Children's Self-Report and Projective Inventory (CSRPI), selected subtests
Records review
Interviews with parent, teacher, and child

REFERRAL REASON

Tamik was referred for evaluation by the Student Support Team at One Great Elementary School. Teachers report concerns regarding Tamik's social-emotional and academic performance. He is significantly below grade level in all academic areas. In addition, Tamik has had several major disruptive episodes. These episodes have included threatening to harm himself, running from his designated area, appearing disoriented, and requiring administrative time outs. Tamik has a current certification of special education eligibility from another state and requires reevaluation to determine eligibility for special education services in Fine State.

BACKGROUND INFORMATION

Information for this section was gathered from the parent information form and parent interview. Tamik currently lives with his biological mother and his older sister. The family recently relocated from another state following his parents' divorce. Tamik's family is African American. It was reported that he has a good relationship with his family members, despite behavioral difficulties. Tamik's mother disclosed that he did witness domestic violence as a toddler, and that his behavior has escalated since she separated from his father 2 years ago. Ms. Parris reports that Tamik and his father were very close, and adds that he still often screams for his father to come home. Currently, Tamik's father has no custodial rights, and he has not seen Tamik in over 2 years.

It was revealed during the parent interview that Tamik was the product of a typical pregnancy and delivery. Developmental milestones were reported as met within normal limits. No significant medical history was revealed.

School history indicates that Tamik attended school in the previous state of residence before enrolling at One Great. According to Ms. Parris, his previous schools noted behavioral and academic concerns. Tamik was found eligible for special education services under the category of "emotional disturbance" at the beginning of his kindergarten year. This information was also obtained from the interview with Ms. Parris. She reports that Tamik was in a classroom with seven other children, a teacher, and a paraprofessional. At the time of the evaluation, Tamik's records from his previous school were not available. Multiple unsuccessful requests have been made to obtain this information.

Since beginning school at One Great, Tamik has had notable difficulty adhering to classroom rituals and routines. Additionally, he has had several major disruptive episodes that include running from the school building, threatening to kill himself, and being verbally aggressive with staff. Please review the functional behavior assessment data for detailed accounts of Tamik's behaviors.

The observed behaviors have had a negative impact on Tamik's academic functioning. He is on a specific behavior management plan to address problem behaviors.

On November 11, 2016, Tamik passed his hearing and vision screenings.

TEST BEHAVIOR AND OBSERVATIONS

At the time of the evaluation, Tamik separated from his class reluctantly. He appeared well groomed and well nourished. He was dressed in the style of his peers. Upon reaching the assessment area, Tamik appeared to be more comfortable, as demonstrated by his relaxed posture and the ease with which he completed tasks asked by the examiner. Tamik was asked to draw a picture of a person to build rapport. His conversation in response to the drawing was morose. He indicated that he drew a "grim reaper" that "kills real demons." His response to questions about his picture included repeated themes of death. At times, he struggled to answer the questions appropriately, and his conversation appeared disjointed. Despite these observations, it is believed that rapport was adequately established and maintained.

Tamik fidgeted throughout the assessment session and asked numerous questions about various objects in the assessment room. He required frequent redirection to attend to the tasks presented. Despite this, Tamik enthusiastically attempted all items presented. He appeared to enjoy verbal praise as he worked through the tasks.

It is believed that Tamik's distractibility may have had a negative impact on his performance during this assessment. However, this behavior is consistent with teacher reports of his classroom behavior. Therefore, due to optimal participation and test setting results, this evaluation is considered an accurate estimate of Tamik's current functioning.

TEST RESULTS AND INTERPRETATION

Cognitive Processes

Intelligence refers to a person's ability to receive information through various perceptual modalities, to retain information in memory, and to organize it meaningfully through various cognitive processes (i.e., concept formation, comprehension, reasoning, judgment, planning, information processing, working memory, and problem solving).

Reynolds Intellectual Assessment Scales, Second Edition

The Reynolds Intellectual Assessment Scales, Second Edition (RIAS-2), is an individually administered measure of intellectual functioning normed for individuals between the ages of 3 and 94 years. The RIAS-2 contains several individual tests of intellectual problem-solving and reasoning ability that are combined to form a Verbal Intelligence Index (VIX) and a Nonverbal Intelligence Index (NIX). These two indexes of intellectual functioning are then combined to form an overall Composite Intelligence Index (CIX). When the VIX and the NIX are combined into the CIX, a strong, reliable assessment of general intelligence is obtained. Scores for this portion of this instrument are presented in terms of standard scores (SS) with a mean of 100 and a standard deviation of 15. The index scores are categorized within the following descriptive ranges:

69 or below: Significantly below average
70–79: Moderately below average
80–89: Below average
90–109: Average
110–119: Above average
120–129: Moderately above average
130 or above: Significantly above average

Tamik's CIX, VIX, and NIX scores were as follows:

- *Composite Intelligence Index.* Tamik obtained a CIX score of 103. This level of performance is considered within the average range, falls within expected limits, and falls into the 58th percentile. The results obtained with this and the other indexes of the RIAS-2 impress the examiner as being accurate reflections of Tamik's current intellectual functioning.
- *Verbal Intelligence Index.* Skills measured by the VIX include the ability to deduce or infer relationships and the ability to apply knowledge to problem solving that involves using words and language skills. The VIX provides a measure of verbal reasoning ability, with primary emphasis on crystallized intelligence functions. Tamik performed in the average range on this index, with a score of 108, which falls into the 47th percentile. Subtests within this index measure verbal reasoning

in combination with vocabulary, language development, and an overall fund of available information.

- *Nonverbal Intelligence Index.* Skills measured by the NIX include the ability to perceive, manipulate, or transform accurately the image of spatial patterns into other visual arrangements. The NIX provides a measure of nonverbal reasoning ability with primary emphasis on fluid intelligence functions. Tamik performed in the average range on this index as well. His performance yielded a score in the 66th percentile. Subtests within this index measure nonverbal reasoning, spatial ability, visual imagery, visual discrimination and deductive reasoning. It appears that Tamik demonstrated a relative strength in nonverbal tasks.

Table 18.A.1 presents the *T* scores that reflect Tamik's performances on the individual subtests used to derive the CIX, VIX, and NIX. Each *T* score has a mean of 50 and a standard deviation of 10. Approximately two-thirds of the population earns *T* scores between 40 and 60.

In addition to the four subtests that yield the CIX, VIX, and NIX as measures of intellectual functioning, the RIAS-2 offers two further index scores. These are the Composite Memory Index (CMX) and the Speeded Processing Index (SPI). Tamik's performances on these, as well as his performances on the subtests used to derive these scores, are reported below.

- *Composite Memory Index.* The CMX provides a summary estimate of verbal and nonverbal memory functions in general as its component parts correspond to the broad areas of memory skills. Tamik obtained a CMX score of 96, exhibiting memory capacity in the average range. Tamik's performance within this cluster revealed fairly even development across verbal and nonverbal short-term memory. His performance fell into the 39th percentile.
- *Speeded Processing Index.* The SPI provides a summary estimate of processing speed, primarily involving both decision speed and reaction while minimizing the efforts of motor speed. Tamik obtained an SPI score of 82 (12th percentile), which is in the low average range for processing speed. Tamik's performance within this cluster revealed better developed visual short-term memory than verbal short-term memory. It is important to note that Tamik required redirection during these tasks multiple times. This may have affected his score on the verbal speeded processing activity. Therefore, while this may be an accurate portrayal of his current functioning, it is likely to be an underestimate of his ability.

Table 18.A.2 presents the *T* scores that reflect Tamik's performances on the individual subtests used to derive the CMX and SPI. Each *T* score has a mean of 50 and a standard deviation of 10. Approximately two-thirds of the population earns *T* scores between 40 and 60.

Academic Performance

Kaufman Tests of Educational Achievement, Second Edition

Tamik's academic skills were measured with the Kaufman Test of Educational Achievement—Second Edition (KTEA-II). The KTEA-II is an individually administered, standardized test instrument designed to measure skills in reading, mathematics, written language, and oral language. The KTEA-II contains a series of tasks arranged in eight subtests, with six additional reading-related subtests. For purposes of interpretation, Tamik's performance on the questions within each subtest is compared to that of a normative sample of students at the same age, and converted to a standard score. Standard subtest scores are combined to yield scores for Reading, Mathematics, Written

TABLE 18.A.1. Tamik's RIAS-2 Subtest *T* Scores for the CIX, VIX, and NIX

Subtest	Age-adjusted *T* score	Percentile	Description
Guess What (GWH)	48	45	Vocabulary, language development, overall knowledge
Odd-Item Out (OIO)	59	82	Nonverbal reasoning, spatial ability
Verbal Reasoning (VRZ)	51	53	Verbal-analytical reasoning
What's Missing (WHM)	52	55	Nonverbal deduction, attention to visual detail

TABLE 18.A.2. Tamik's RIAS-2 Subtest *T* Scores for the CMX and SPI

Subtest	Age-adjusted *T* score	Percentile	Description
Verbal Memory (VRM)	45	32	Verbal recall in a meaningful context
Nonverbal Memory (NVM)	51	53	Nonverbal recognition without meaningful context
Speeded Naming Task (SNT)	35	7	Verbal differentiation/recognition of simple stimuli under time constraints
Speeded Picture Search (SPS)	46	34	Visual differentiation/recognition of simple stimuli under time constraints

Language, and Oral Language. Each of the standard scores has a mean of 100 and a standard deviation of 15. A selection of Tamik's KTEA-II scores is presented in Table 18.A.3.

These results are inconsistent with classroom reports of Tamik's academic ability. In class, he has demonstrated limited abilities in all academic areas. He appeared to enjoy this portion of the assessment. He was allowed to stand and take frequent breaks when completing tasks. Tamik does not complete assignments in the classroom and requires one-on-one assistance to begin tasks in the classroom.

Social-Emotional Functioning

Social-emotional assessment is the evaluation of an individual's social relations, coping strategies, and self-perceptions. The assessment is based upon information obtained from a variety of sources, which can include evaluations by teachers and parents, as well as individual personality testing. Current reports indicate that Tamik is having difficulty attending to tasks, is demonstrating a great deal of overactivity in the classroom as well as at home, and is having major disruptive episodes in the classroom and other school settings.

TABLE 18.A.3. Selected KTEA-II Scores for Tamik

Subtest or **composite**	SS
Letter and Word Recognition	92
Reading Comprehension	92
Reading Composite	**91**
Math Concepts and Applications	97
Math Computation	104
Math Composite	**101**
Spelling	90
Phonological Awareness	98

Note. Composites are given in boldface.

Children's Self-Report and Projective Inventory

Tamik was administered selected sections of the Children's Self-Report and Projective Inventory (CSRPI). Due to limited attention to tasks and increasing agitation as the assessment progressed, the entire instrument was not attempted. Tamik was asked to draw a picture of his family with everybody doing something. His explanation of the picture was disjointed. However, he supplied a drawing of himself on the right side of the picture, holding a long straight object with a point like an arrow. He stated that in the picture, he was exercising. He drew his mother and sister on the other side, much smaller, and said they were talking. He then stated that his aunt was also exercising until his mother left. In the center of the picture very large, he drew Giovanni (a family friend). He reported that Giovanni "knows how to walk." Tamik was additionally asked to draw a picture of a child in the rain. Tamik growled when the examiner requested this picture. He hastily drew one large black cloud at the top of the page. He then began drawing raindrops very haphazardly all over the paper. He drew himself in a house, then drew more raindrops, using a lot of pressure with the crayon, over himself and in the house. When asked about the drawing, he stated that he was "running without a raincoat" and that he "open[ed] the door to the house and just made it, but RAIN!" He became increasingly agitated and added, "It is inside the house! Rain crashed and came inside the house!" A drawing of this nature may be indicative of anxiety and feelings of limited coping resources. Tamik is in the middle of several familial transitions and may lack the necessary coping mechanisms to deal with these changes.

- Tamik was also administered a sentence completion task as part of the CSPRI. His responses included themes of death, abandonment, and punishment. Some significant responses included the following (Tamik's sentence completions are italicized):

Most kids think I'm *a jerk.*
When my parents fight, I think that *they will kill like hit each other.*
I don't understand why my parents *let me go outside and die.*
I wish I could *spend more time with my friend. I don't have a friend.*
In 10 years, I hope I'll *be a police and kick somebody's butt.*

Additionally, when asked what makes him happy, Tamik stated, "To jump in a pool and die. I would be a little bit happy, but still bored." During this portion of the assessment, Tamik had periods of focused cooperation and periods of anger and resistance. The results of interviews and the projective assessment suggest that Tamik is demonstrating markedly different social and emotional development from that of his peers.

Behavior Assessment System for Children, Third Edition

The Behavior Assessment System for Children, Third Edition (BASC-3), is a multimethod, multidimensional system used to evaluate the behavior and self-perceptions of children and young adults ages 2–25 years. Separate rating scales are available for completion by teachers and by parents. Scores are reported as T scores with a mean of 50 and a standard deviation of 10.

The Parent Rating Scale (PRS) is a comprehensive measure of adaptive and problem behaviors observed by parents at home and in the community. The PRS assesses most of the clinical and adaptive skills domains that the Teacher Rating Scale (TRS) measures. In addition, the PRS includes a scale (Activities of Daily Living) that the TRS does not measure.

Table 18.A.4 is provided as a guide to BASC-3 T-score interpretation. Scores in the clinically significant range indicate a high level of maladaptive behavior or lack of adaptive behavior. Scores in the at-risk range indicate the presence of significant problems that, while requiring intervention, may not be severe enough to warrant a formal diagnosis. A selection of Tamik's BASC-3 scores is reported in Table 18.A.5.

TABLE 18.A.4. Interpretation of BASC-3 *T* Scores

Clinical scales	Adaptive scales
70+: Clinically significant	70+: Very high
60–69: At risk	60–69: High
41–59: Average	41–59: Average
31–40: Low	31–40: At-risk
30 or below: Very low	30 or below: Clinically significant

The BASC-3 PRS was completed by Ms. Parris, Tamik's mother. The TRS was completed by his teacher, Ms. Talapatra. Variance in environmental structure and rater expectations may account for discrepancies in score reports.

An analysis of Tamik's BASC-3 scores indicates that he is exhibiting clinically significant behaviors across areas assessed in the school setting. In contrast, his mother reported no areas of concern. This is inconsistent with the parent interview, where Ms. Parris reported that Tamik exhibits noncompliant and disruptive behavior at home and in the community. Specifically, she reported that she cannot get him to go to bed; that he becomes loud and disruptive in stores, often running from her; and that he is aggressive with her almost daily when she brings him to school.

Ms. Talapatra's scores yielded an F index ("fake bad") score in the "extreme caution" range. High scores on the F index indicate the possibility that a respondent has reported items in an overly negative manner. However, children who present in acute psychological distress may also score highly on this scale. Based on the severity of behaviors observed by Ms. Talapatra in the school and those reported by Ms. Parris in the home and community, the TRS scores may constitute an accurate report of Tamik's current behavioral functioning. Based on observations and related information presented in this report (all of which support the presence of maladaptive behaviors), it is likely that Ms. Talapatra was being forthright and candid in describing Tamik's behavior, and that the ratings represent a valid assessment of Tamik's social-emotional functioning.

TABLE 18.A.5. Selected BASC-3 Rating Scale Scores for Tamik

BASC-3 scales	Description	Teacher	Parent
Behavioral Symptoms Index	Overall rating of the individual's behavior	117	52
Externalizing Problems		104	51
Hyperactivity	Tendency to be overly active, rush through work, and act without thinking	89	51
Aggression	Tendency to act in a hostile manner (either verbal or physical) that is threatening to others	116	52
Conduct Problems	Tendency to engage in antisocial and rule-breaking behavior	95	49
Internalizing Problems		94	46
Anxiety	Tendency to be nervous, fearful, or worried about real/imagined problems	62	47
Depression	Feelings of unhappiness, sadness, or stress that may result in an inability to carry out everyday activities	120	51
Somatization	Tendency to be overly sensitive to and complain about relatively minor physical problems or discomforts	69	42
School Problems (TRS)		74	—
Attention Problems	Tendency to be easily distracted	74	59
Learning Problems (TRS)	Presence of academic difficulties, esp. in understanding or completing work	70	—
Additional clinical scales			
Atypicality	Tendency to behave in ways that are considered strange, such as being disconnected from or unaware of normal surroundings	120	52
Withdrawal	Tendency to evade others to avoid social contact	100	42
Adaptive Skills			
Adaptability	The ability to adapt readily to changes in the environment	24	48
Social Skills	Skills needed to interact successfully with peers/adults in home, school, and community settings	28	48
Leadership	Skills associated with accomplishing academic, social, or community goals, including the ability to work well with others	38	63
Study Skills (TRS)	Skills that are conducive to strong academic performance, including organizational skills and good study habits	31	—
Functional Communication	The ability to express ideas and communicate in ways others can easily understand	24	56
Activities of Daily Living (PRS)	The skills associated with performing basic, everyday tasks in an acceptable and safe manner	—	49

CONCLUSIONS AND RECOMMENDATIONS

Tamik is a 6-year-old boy who currently resides with his biological mother and older sister. His family has recently moved from another state, and his parents have recently divorced. Available history indicates that Tamik was the product of a typical pregnancy and delivery. It was reported that all developmental milestones were met within normal limits.

Current results indicate that Tamik's intellectual functioning is within the average range. Academic assessment revealed that Tamik has age-appropriate academic skills in all areas, despite his inability to demonstrate these in the classroom. Behavior checklists and interviews reveal that Tamik is exhibiting significant social-emotional distress. This could be attributed to the number of adverse childhood experiences he has endured. In addition to witnessing domestic violence in early childhood, Tamik has lost access to a parent as a result of divorce and has now relocated across the country. The information gathered in this evaluation should be reviewed by the eligibility team, in conjunction with progress on his behavior intervention plan and previous school records (when these become available), to determine Tamik's need for special education support. This evaluation has found that Tamik demonstrates an inability to build and maintain interpersonal relationships, a general pervasive mood of unhappiness, and inappropriate behaviors under normal circumstances.

At the current time, available information suggests that Tamik meets diagnostic criteria for other specified trauma- and stressor-related disorder (F43.8) in the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5). At the time, not enough information is available to determine whether he meets full DSM-5 criteria for a more specific trauma- and stressor-related disorder. However, the symptoms he is exhibiting are believed to be related to his multiple adverse childhood experiences.

The following remediation strategies may be helpful:

- Individual and conjoint therapy with Tamik's mother, focused on helping him navigate his adverse childhood experiences, is recommended. Tamik may benefit from cognitive restructuring, positive self-talk, and relaxation techniques. Tamik's treatment might also include helping him build insight into his own and others' feelings and address the trauma in his family.
- Tamik would benefit from structured, supportive classroom environments. Rules and expectations should be simply, clearly, and consistently presented and reviewed frequently. Consequences for both appropriate and inappropriate behavior should be enforced on a consistent basis through providing positive reinforcement (rewarding appropriate behavior) and ignoring inappropriate behavior. This could be managed in the form of a token economy, which should be coordinated with a home token economy. This plan should be coordinated with Tamik's mother to ensure consistent application in both home and school settings, including target behaviors and reinforcers.
- Positive feedback should be provided to Tamik when he responds in an appropriate way to new adults, situations, and places. His mother and teacher should talk with him beforehand about new situations and adults, so he knows what may happen or what to expect.
- Adults should use physical proximity to encourage Tamik to pay attention to tasks.
- The teacher can channel Tamik's overactivity by assigning him classroom duties (e.g., sharpening pencils, watering plants, delivering notes, handing out papers) that provide opportunities for seat breaks.
- Instances of appropriate behavior should be acknowledged through recognition and encouragement.
- Tamik should be assisted to develop social skills such as cooperating with others, sharing, giving/receiving compliments, entering a group, and standing up for his own rights without being aggressive. Instructional approaches such as modeling and role playing should also be helpful.
- Close supervision during unstructured periods is recommended.

REFERENCES

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). *Stan-*

dards for educational and psychological testing. Washington, DC: AERA.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Andrews, J. J. W. (2007). Review of the test Reynolds Intellectual Assessment Scales. *Journal of Psychoeducational Assessment, 25,* 402–408.

Angoff, W. H., & Robertson, G. R. (1987). A procedure for standardizing individually administered tests, normed by age or grade level. *Applied Psychological Measurement, 11,* 33–46.

Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Anneé Psychologique, 11,* 191–244.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since bias in mental testing. *School Psychology Quarterly, 14*(3), 208–238.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Deary, I. J. (2012) Intelligence. *Annual Review of Psychology, 63,* 453–482.

Dombrowski, S. C., & Mrazik, M. (2008). Test Review: Reynolds, CR, & Kamphaus, RW (2003). RIAS: Reynolds Intellectual Assessment Scales. Lutz, FL: Psychological Assessment Resources, Inc. *Canadian Journal of School Psychology, 23*(2), 223–230.

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing, 10,* 295–317.

Goldstein, S., & Reynolds, C. R. (Eds.). (2011). *Handbook of neurodevelopmental and genetic disorders in children.* New York: Guilford Press.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Gustafsson, J. E. (1999). Measuring and understanding *g:* Experimental and correlational approaches. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 275–291). Washington, DC: American Psychological Association.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270.

Jensen, A. (1998). The suppressed relationship between IQ and the reaction time slope parameter of the Hick function. *Intelligence, 26*(1), 43–52.

Kamphaus, R. W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kamphaus, R. W., & Reynolds, C. R. (2015). *Reynolds Intellectual Screening Test, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75). Mahwah, NJ: Erlbaum.

Messer, M. A. (2014). *Academic Achievement Battery.* Lutz, FL: Psychological Assessment Resources.

New shorter Oxford English dictionary. (1999). Portsmouth, NH; Oxford University Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Reynolds, C. R. (1998). Fundamentals of measurement and assessment in psychology. In A. Bellack & M. Hersen (Series Eds.) & C. R. Reynolds (Vol. Ed.), *Comprehensive clinical psychology: Vol. 4. Assessment* (pp. 33–56). New York: Pergamon Press.

Reynolds, C. R. (2000). Methods for detecting and evaluating cultural bias in neuropsychological tests. In E. Fletcher-Janzen, T. Strickland, & C. R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 249–286). New York: Plenum Press.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and Reynolds Intellectual Screening Test.* Lutz, FL: Psychological Assessment Resources.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales, Second Edition, and Reynolds Intellectual Screening Test, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Reynolds, C. R., Lowe, P. A., & Saenz, A. (1998). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed.). New York: Wiley.

Riccio, C. A., & Hynd, G. W. (2000). Measurable biological substrates to verbal–performance differences in Wechsler scores. *School Psychology Quarterly, 15*(4), 386–399.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Author.

Vernon, P. E. (1950). *The structure of human abilities.* New York: Wiley.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition.* San Antonio, TX: Pearson.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44*(2), 98–104.

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology, 41,* 86–94.

CHAPTER 19

The NEPSY-II

Robb N. Matthews
John L. Davis

STRUCTURE

The NEPSY-II (Korkman, Kirk, & Kemp, 2007) is an integrated battery of subtests designed to assess a broad range of functioning in children ages 3–16 years (Kemp & Korkman, 2010). Case conceptualization often requires a broad scope of evaluation: Assessment professionals must go beyond traditional measures of general characteristics (e.g., intellectual functioning) and include targeted assessment of specific traits or skills (e.g., divided attention, ideational fluency). In such cases, practitioners must choose assessment techniques that effectively target the various aspects of traits or skills under consideration, in order to gather relevant information and to develop ecologically valid recommendations related to the examinees' needs (Rhodes, D'Amato, & Rothlisberg, 2009). As a stand-alone measure or as a complement to other techniques typically used in assessing the neurological functioning of children, the NEPSY-II provides one of the few avenues for targeted assessment of specific cognitive skills in younger populations (Brooks, Sherman, & Strauss, 2010; Titley & D'Amato, 2008). Based on the general neuropsychological principles put forth by Luria (1980) and operationalized through a flexible multidomain design, the NEPSY-II allows examination of the subcomponents of complex cognitive processes. The structure and development of the NEPSY-II have been informed by current literature surrounding child development, child psychology, and pediatric neuropsychology (Korkman et al., 2007). As such, this battery provides a method for merging quantitative measures and qualitative observations of performance to create a comprehensive clinical perspective on an examinee's functioning.

The NEPSY-II is intended to assist clinicians of varying proficiency levels in better understand complex aspects of examinees' cognitive processing. The flexible structure allows the clinicians to address the specific issues raised in referral questions, to identify basic and interactive cognitive subcomponents of functional skills, and to cultivate informed recommendation based on the examinees' functioning across targeted domains (Kemp & Korkman, 2010; Luria, 1980).

The NEPSY-II structure is based on six theoretically derived domains of cognitive processing (Kemp & Korkman, 2010). These cognitive domains are thought to inform and predict a child's success in meeting the expectations across diverse environments (Brooks, Sherman, & Strauss, 2010; Hooper, 2010). The NEPSY-II has been intentionally constructed to allow for flexible administration of individual subtests. The particular set of subtests administered can be based either on a predetermined battery as specified in the manual, or on a clinician's conceptualization of a referral question and its related primary functional components (e.g., attention) or secondary subcomponents (e.g., memory); however, administration of the entire 36-subtest battery is not required. Suggested referral batteries listed in the scoring assistant and assessment planner are combinations of subtests the test developers believe may be useful for particular "classes" of difficulties and those components most likely to be affected (Korkman et al., 2007). In some populations, a short-battery or a fixed-battery approach can be an efficient way to answer referral questions. The flexible design of the NEPSY-II allows a clinician to use other information and the presenting problem to select which domains or combinations of subtests will be most useful for a specific referral. Although the NEPSY-II is not designed to assess for specific diagnoses (e.g., learning disorders), its flexible structure may contribute to a diagnostic conclusion based on identified patterns (i.e., diagnostic clusters) of cognitive processing consistent with a particular condition (Kemp & Korkman, 2010; Titley & D'Amato, 2008). Most importantly, the structure of the NEPSY-II allows for more in-depth examination of specific cognitive processing areas that may be helpful in selection of well-designed intervention programs (Titley & D'Amato, 2008).

THEORETICAL UNDERPINNINGS

The NEPSY-II, like the original NEPSY (Korkman, Kirk, & Kemp, 1998) and the original NEPS in Finnish, are based on Luria's (1980) model of functional systems and an integrated theory of brain function. As such, the initial 36 subtests and 30 performance areas were intended to correspond to components of complex cognitive functions (Korkman, 1999). A *functional system* is made up of multiple brain regions and their interconnections; all these working together form the basis for a complex psychological function or process (e.g., attention, language). The integration of the individual components (i.e., brain structures and connections) is what allows for complex thought and processes (Luria, 1980). From a clinical perspective, the nature of the individual's impaired performance is determined by the specific location of damage or developmental difference within the affected functional system(s), as well as by the effects of this damage or difference on the rest of the functional systems (Bauer, 2000; Pramuka & McCue, 2000).

Luria (1980) believed that the fundamental purpose of a neuropsychological assessment is to better describe the function of the neuropsychological symptoms (i.e., the presenting problems that result in referral for assessment). In effect, the assessment is designed to present various opportunities, thus identifying those contexts in which an individual's deficits become significant (Bauer, 2000). In this regard, Luria's approach is considered to be client- and problem-centered, with examination of performance across and within the various functional systems. As part of this process, there is a need to examine performance or function at simple, complex, and integrated levels, as well as with varying input and output demands. The interpretation is to some extent qualitative; it is not based on the number of problems solved or on a total score, but rather on which problems were and were not solved. The emphasis on complex patterns of functioning (i.e., strengths or weaknesses) is a key component of the Luria model. It is the examination of the pattern of strengths and weaknesses, believed to represent intact or disrupted systems, on which the interpretation is based (Bauer, 2000).

As noted above, specific combinations of the NEPSY-II subtests are intended to represent the various functional systems in an effort to better understand and quantify which brain functions are intact and which are impaired. Five of the six content domains of the NEPSY-II were also in the original NEPSY and correspond to specific functional systems identified by Luria. Specifically, these domains are Attention and Executive Functioning, Language, Memory and Learning, Visuospatial Processing, and Sensorimotor. For the NEPSY-II, the sixth domain of Social Perception, representing some aspects of social-emotional functioning, was created to address concerns specific to the autism spectrum. For each domain, the NEPSY-II offers subtest-level scores rather than aggregate domain-level scores.

It is important to note that these six domains are theoretically based, rather than statistically based (Korkman et al., 2007); this is one reason why it was determined that the domain scores were not appropriate. In effect, the test authors recognized that no single subtest only measures one of the functional domains, and that because of the differing components measured by each subtest, the correlations within and across domains vary considerably. With subtests measuring some overlapping skills, one strength of the theoretical rather than statistical design is the ability to assess deficits underlying impaired performance both within and across functional domains (Korkman et al., 2007). Consistent with developmental considerations, differing subtests or formats are deemed appropriate for differing age levels within each domain. Given that specific functional systems have been implicated for differing neurodevelopmental disorders, there is also flexibility in the combinations of subtests that can be administered, based on the presenting problem. Although domain scores are no longer calculated, the provision of contrast scores (an indication of whether differences between subtest scores are meaningful) allows for examination of patterns (i.e., the syndrome analysis).

ORGANIZATION AND FORMAT

Within the six domains, the NEPSY–II includes 32 individually administered and autonomous subtests, with an additional 4 delayed-recall subtests. These wide-ranging subtests are designed to assess neuropsychological functioning across the six domains, as presented in Table 19.1. Each subtest is theoretically derived to assess specific skill strengths and weaknesses. Input, output, and processing demands vary across subtests and within domains; the combination of subtest scores and differences in performance is what is important. The manual provides suggested batteries to assist in planning assessments for specific referral questions. These eight tailored batteries are based on information obtained in special-group studies for deficits in reading, math, attention, social/interpersonal skills, behavioral management, school readiness, and perceptual–motor delays. These suggested batteries are empirically based, and are designed to align with profiles of children known to possess deficits in the aforementioned areas. The format of the NEPSY-II is designed to offer flexible administration of subtests to explore specific facets of neurological functioning in children.

In terms of the demands placed on examinees, the subtests range from oral responding to individual paper-and-pencil work. The NEPSY-II is very well designed in terms of providing test material that promotes engagement across the range of ages it is intended to assess. Due to the wide range of potential subtests, users should carefully consider each child's profile and assessment needs prior to administration. Examiners should also pay special attention to the age range of examinees appropriate for each subtest, as the NEPSY-II varies considerably in format in this area. This variation is a strength of the NEPSY-II: Because it offers a wide range of tests that are appropriate across different ages and developmental levels, it affords examiners complete coverage of the intended construct. However, this variation has the potential to create confusion or misadministration of subtests. The NEPSY-II offers 17 subtests for ages 3–4, 22 subtests and 2 delayed tasks for ages 5–6, 23 subtests and 2 delayed tasks for ages 7–12, and 24 subtests and 3 delayed tasks for ages 13–16. For specific information on the age range for each subtest, see Table 19.1.

SCORING

Most subtests of the NEPSY-II provide multiple score options that allow for a variety of applications, including general performance, error rates, or measures of subcomponent skills necessary for task completion. Scores in the NEPSY-II include (but are not limited to) primary scores, process scores, combined scaled scores, and contrast scaled scores. Primary scores describe the overall ability assessed by each subtest. Process scores are intended to provide additional insight into the child's ability or error rates for the subtest. The combined scaled scores were developed to assist the clinician with understanding the unique contributions of specific subcomponents of a process and with determining whether they fall within expected limits. Combined scaled scores are similar to index scores on many other popular instruments, but interpretation is achieved by considering both the domain being measured and the subcomponents being combined in the score (Korkman et al., 2007). Several subtests of the NEPSY-II provide multiple primary scores; therefore, the contrast scaled scores allow comparison of scores within subtests. Contrast scaled scores were designed to hold the impact of one characteristic constant while performance on another is considered. These scores indicate the degree to which examinees' performance is within expectations, given their performance on a related subcomponent or skill (Kemp & Korkman, 2010). This aspect of scoring provides clinicians with additional information to discriminate between higher-level and lower-level skill deficits within a single subtest.

Additional scores available from the NEPSY-II include percentile ranks, cumulative percentages, and behavioral observations. Korkman and colleagues (2007) used percentile ranks in place of standard scores on some subtests, due to the level of skewness present in the distribution. In other

TABLE 19.1. NEPSY-II Subtests and Descriptions

Domain	Subtest	Descriptions
Attention and Executive Functioning	Animal Sorting	Assesses the ability to group pictures based on self-initiated sorting criteria.
	Auditory Attention	Assesses selective auditory attention and the ability to sustain attention.
	Response Set	Assesses the ability to shift and maintain attention.
	Clocks	Assesses the ability to organize, execute, and recreate the expression of time on an analog clock.
	Design Fluency	Assesses the child's ability to generate unique designs presented in structured or random arrays.
	Inhibition	Assesses the ability to inhibit automatic responses in favor of novel responses and the ability to switch between response types.
	Statue	Assesses motor persistence and inhibition.
Language	Body Part Naming and Identification	Assesses expressive and receptive language for identifying body parts with self-prompts and with picture prompts.
	Comprehension of Instructions	Assesses the ability to follow oral instruction with increasing complexity.
	Oromotor Sequences	Assesses oral–motor coordination.
	Phonological Processing	Assesses the child's understanding of phonological processing through repeating or recreating words with substitute words or phonemes.
	Repetition of Nonsense Words	Assesses phonological decoding skills.
	Speeded Naming	Assesses the child's ability to rapidly name common material when presented visually.
	Word Generation	Assesses the child's ability to rapidly generate verbal responses.
Memory and Learning	List Memory	Assesses verbal learning and memory, rate of learning, and the role of interference in recall for words.
	List Memory Delayed	Assesses long-term memory for words.
	Memory for Designs	Assesses spatial memory for novel visual material.
	Memory for Designs Delayed	Assesses long-term visual–spatial memory.
	Memory for Faces	Assesses discrimination and recognition of facial features.
	Memory for Faces Delayed	Assesses long-term memory for faces.
	Memory for Names	Assesses the child's memory for names, with three trials to rehearse.
	Memory for Names Delayed	Assesses long-term memory for names.
	Narrative Memory	Assesses memory for verbally presented material with free-recall and cued conditions.
	Sentence Repetition	Assesses rote memory for sentences that increase in complexity and length.
	Word List Interference	Assesses working memory and recall for verbally presented material.

(continued)

TABLE 19.1. *(continued)*

Domain	Subtest	Descriptions
Sensorimotor	Fingertip Tapping	Assesses the child's finger dexterity and motor speed.
	Imitating Hand Positions	Assesses the ability to imitate hand/finger positions.
	Manual Motor Sequences	Assesses the ability to imitate hand movement sequences.
	Visuomotor Precision	Assesses motor speed and accuracy.
Social Perception	Affect Recognition	Assesses the child's ability to recognize affect from a picture cue.
	Theory of Mind	Assesses ability to understand and properly attribute mental states.
Visuospatial Processing	Arrows	Assesses the ability to identify and infer line directionality.
	Block Construction	Assesses the ability to construct three-dimensional models from a two-dimensional picture prompt.
	Design Copying	Assesses the ability to reproduce images from a two-dimensional picture prompt.
	Geometric Puzzles	Assesses mental rotation and visual–spatial skills.
	Picture Puzzles	Assesses visual discrimination and visual–spatial skills.
	Route Finding	Assesses visual planning and directionality.

words, subtests yielding only percentile ranks tend to assess those skills generally seen early in the typical developmental process and were therefore generally at an advanced level in the standardization sample (Kemp & Korkman, 2010). NEPSY-II percentile ranks are presented in ranges, so that scores falling in the 26th–50th or 51st–75th percentiles should be interpreted as falling in the expected range of performance. Cumulative percentages represent the base rates or probability of occurrence of a particular performance, characteristic, or profile (Sattler, 2008), and should be interpreted with regard to the "rareness" of a particular outcome; this is discussed further in the section on interpretation.

In keeping with the Lurian tradition, the NEPSY-II also allows for structured qualitative observations of an examinee's performance. Some of the coded observations are simply descriptive of performance, while others are formalized observation measures. The set of behavioral observations coded on the NEPSY-II protocol are based on the clinical experience of the test authors and the behavior of typically developing children in the normative sample (Korkman et al., 2007). For example, cumulative percentages permit clinicians to compare instances of "out-of-seat behavior" with those for the normative group by age. Reviewing behavioral observations can give a reference point from which to develop hypotheses about an examinee's performance (Kemp & Korkman, 2010). Those behaviors tallied and summed in the protocol are converted to cumulative percentages (or base rates) and are included in an appendix in the test manual.

PSYCHOMETRIC PROPERTIES

Item Generation

With the revision of the NEPSY-II, several modifications were made to the original NEPSY, including the addition of several subtests and the omission of others. The manual appropriately describes the theoretical basis that necessitated these modifications, as well as the development of new subtests and additional items to existing subtests. The examination of items was conducted in multiple phases. This allowed the test authors to examine the psychometric impact of the changes in items and subtests. Specific item evaluation procedures or item evaluation outputs are not available in the manual.

Ceiling and Floors

One component of item generation involves assurance of appropriate floors and ceilings of each subtest. This is important in order to ensure assessment sensitivity across a broad range of abilities in children (Korkman et al., 2007). To this goal, the NEPSY-II provides additional items to ensure appropriate floors and ceilings in every subtest. For younger ages and lower levels of ability, additional subtests have been specifically included (e.g., Body Part Naming and Identification for 3- to 4-year-olds). Each six primary domains offer developmentally appropriate items for children between 3 and 16 years old.

Standardization

The NEPSY-II normative data were collected from 2005 to 2006. The normative data for the NEPSY-II were derived from a sample that closely matched the U.S. census data of children ages 3–16. An analysis of data gathered in the October 2003 census provided the basis for stratification across the variables of age, race/ethnicity, geographic region, and parent education level. Twelve hundred examinees were assessed for the normative sample, with 100 children from each of the 12 age groups ranging from 3 through 16. Ages 3–12 were divided by 6-month intervals, with 50 cases collected from children in the first 6 months of each year of age, and 50 cases collected from children in the last 6 months of that same year. For each age group in the normative sample, examinees were further separated by race/ethnicity categories. Each child in the normative sample was categorized as belonging to one of the following groups: white (i.e., European American), African American, Hispanic, or other.

The groups were also stratified by gender to include 50% males and 50% females. Children were also stratified by four major geographic regions of the United States: Northeast, Midwest, South, and West. Children were further selected for the normative group to represent the proportions of children living in each region. Finally, the sample was stratified by parental educational level; the children were grouped according to the parents' reported educational attainment. For children in two-parent homes, the average of the parents' educational level was used (Korkman et al., 2007).

In the renorming of the NEPSY-II, several new subtests were added (Animal Sorting, Clocks, Inhibition, Memory for Designs, Word List Interference, Affect Recognition, Theory of Mind, Geometric Puzzles, and Picture Puzzles). For many of the subtests in the NEPSY-II, items were carried over from the original NEPSY and renormed. For several subtests in the NEPSY-II (Design Fluency, Repetition of Nonsense Words, List Memory, List Memory Delayed, and Imitating Hand Positions), the normative data from the original NEPSY continue to be used, with no renorming conducted. The rationale provided was that these subtests were not expected to be subject to the Flynn effect (Flynn, 1984, 1987) or changes in the population demographics; however, no empirical examination is offered to support this decision.

Reliability

Data are available in the manual for interrater reliability in scoring protocols, test–retest stability, decision consistency of classification, and internal consistency. Each of these reliability coefficients are listed for both the primary and process scores for subtests that were evaluated (Korkman et al., 2007). Interrater reliability in scoring the NEPSY-II protocol was assessed through percent agreement. Agreement rates ranged from 93% to 99%, with Word Generation at the lowest level (93%) and Memory for Names at the highest level (99%).

Test–retest stability was calculated on both the primary and process scores. A diverse group of 165 children (52% male, 48% female) took the NEPSY-II on two occasions. The sample was then divided into six age groups: 3–4 years, 5–6 years, 7–8 years, 9–10 years, 11–12 years, and 13–16 years. Test–retest intervals ranged from 12 to 51 days, with a mean of 21 days between administrations. Several indices for evaluating test–retest stability are provided. Test–retest scores showed generally adequate stability across time for all age groups in subtests assessed. Despite a very thorough treatment of subtest stability, the manual does not provide stability data for two of the subtests in the Sensorimotor domain (Manual Motor Sequences and Fingertip Tapping). Stability estimates for several subtests (Design Fluency, Repetition of Nonsense Words, List Memory, List Memory Delayed, and Imitating Hand Positions) are based on data collected for the original NEPSY, as these were not renormed (Korkman et al., 2007).

A decision consistency index was also used to document reliability for several subtests. This alternative to standard test–retest reliability estimates was used because of skewed score distribution or restricted variance within particular subtests that might artificially lower reliability estimates. For these subtests, raw scores were converted to percentile ranks, combined scaled scores, or contrast scaled scores. Decision consistency shows the agreement between converted raw scores by two separate administrators. The test authors set two classification ranges as the criteria for judging decision consistency. For percentile ranks, the authors categorized scores as either above or below the 10th percentile. Combined or contrast scores were categorized as above or below a scaled score of 6. To achieve reliability with this index, a child would be assessed by two separate administrators, and the raw score would be converted to a percentile rank, combined score, or contrast score (where appropriate). If the resulting converted score was in the 6th percentile on the first administration and the 9th on the second, the subtest was said to be reliable. Decision consistency between raters was moderate to high on each subtest across all age groups. For three subtests (Oromotor Sequences, Manual Motor Sequences, and Route Finding), the analysis of decision consistency from the NEPSY is reported; notably, these subtests have the lowest decision consistency (Korkman et al., 2007).

In addition to stability coefficients and decision consistency procedures, split-half (Spearman–Brown) and alpha methods were used to calculate internal consistency when appropriate. The manual reports reliability coefficients for primary and process scores across individual age groups, and provides an average across the six age bands noted earlier (3–4, 5–6, 7–8, 9–10, 11–12, and 13–16). The reliability data indicated adequate to high internal consistency for a majority of subtests (r_{12} range = .21–.91). The highest internal-consistency scores were found on Comprehension of Instructions, Design Copying, and Fingertip Tapping. The lowest internal-consistency scores were found on the Inhibition and Memory for Designs subtests (Korkman et al., 2007).

Validity

Due to the wide array of specific skill areas examined on the NEPSY-II, evidence of concurrent validity is provided through a series of correlation studies on separate instruments designed to measure cognitive ability, academic achievement, neuropsychological functioning, and behavior. For example, concurrent validity of intellectual functioning was assessed with the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003), the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007), and the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006). Correlations between these instruments suggested that the NEPSY-II scores correlate well with cognitive performance in both verbal and nonverbal applications. Correlations with the Verbal Comprehension Index of the WISC-IV ranged from a low on the Auditory Attention subtest (–.02) to a high on the Narrative Memory subtest (.58). Concurrent validity in academic domains was assessed with the Wechsler Individual Achievement Test—Second Edition (WIAT-II; Psychological Corporation, 2001). Results yielded a low to moderate link between the NEPSY-II Attention and Executive Functioning domain and the WIAT-II tests of Mathematics Reasoning (r = .09–.43), Oral Expression (r = .03–.52), and Written Language (r = .08–.47). Within the NEPSY-II Memory and Learning domain, the Sentence Repetition subtest strongly correlated with the WIAT-II tests of Reading Comprehension (r = .87) and Pseudoword Decoding (r = .87). The Narrative Memory subtest of the NEPSY-II varied in correlation (r = .04–.61) with the Children's Memory Scale (CMS; Cohen, 1997). Moderate correlations were shown between subtest scores of the CMS and the Auditory Attention subtest (r = .03–.42). The variability reflects the range of constructs the NEPSY-II is designed to assess. It would not be expected that all NEPSY-II subtests would correlate with any specific measure; however, the presence of high to moderate correlations for each of the constructs mentioned above provides reasonable evidence of validity.

Additional concurrent and construct validity coefficients were derived from a variety of other measures. Results of the NEPSY-II ranged in correlation (r = –.45–.65) to the content of various measures, including the Delis–Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001); the Bracken Basic Concept Scale—Third Edition: Receptive (BBCS-3:R; Bracken, 2006a); the Bracken Basic Concept Scale—Expressive (BBCS:E: Bracken, 2006b); the Devereux Scales of Mental Disorders (DSMD; Naglieri, LeBuffe, & Pfeiffer, 1994); the Children's Communication Checklist—Second Edition (CCC-2; Bishop, 2006); the Brown Attention-Deficit Disorder Scales for Children and Adolescents (Brown ADD Scales; Brown, 2001); and the Adaptive Behavior Assessment System—Second Edition (ABAS-II; Harrison & Oakland, 2003). The highest levels of association were found between the Memory for Designs Delayed subtest of the NEPSY-II and the School Readiness (.61) and Receptive Total (.64) composites of the BBCS-3:R. In addition, the Inhibition–Switching subtest of the NEPSY-II correlated highly (.59) with the Color Word Interference subtest of the D-KEFS. In contrast, correlation coefficients were moderate but negative for the Affect Recognition subtest of the NEPSY-II in relation to the Conduct scale (–.45) and Externalizing composite (–.40) of the DSMD. The Inhibition–Switching combined scaled score also had a moderate negative correlation (–.41) with the Focus cluster of the ABAS-II.

The authors also conducted several studies of "special groups" to test criterion validity, or the scale's clinical utility in yielding information that supports a diagnosis or disability classification. Children with the following diagnoses/educational classifications were included: attention-deficit/hyperactivity disorder (ADHD), reading disorder, language disorder, mathematics disorder, mild intellectual disability, autistic disorder, Asperger syndrome, traumatic brain injury, deafness/hearing impairment, and emotional disturbance. Small-group studies compared each of the score indices for each of the subtests between a group of children with a specific condition and a control group. The control group was matched with the normative sample on demographic categories. By examining differences in mean scale scores between children identified with a particular condition and the matched controls, the test authors identified specific subtests where children within an identified diagnosis/classification group diverged from the norm. These studies helped to form the empirical basis for recommended subtests for each condition. The NEPSY-II does not provide a diagnostic recommendation for any particular condition; rather, the test authors recommend using the information from suggested batteries as a method to conceptualize deficits commonly associated with a particular condition. The authors emphasize that the NEPSY-II should not be used as the sole source to diagnose or classify a child for educational purposes (Korkman et al., 2007).

INTERPRETATION

General Principles

Consistent with common assessment practice, levels of test performance are typically described and categorized statistically, to allow comparison between an individual's performance and levels of performance in the general population (Sattler, 2008). At the same time, the importance of understanding that mild deficits (e.g., low average functioning) may significantly interfere with daily functioning cannot be understated in interpreting performance on measures such as the NEPSY-II (Riccio, Sullivan, & Cohen, 2010; Yeates, Ris, Taylor, & Pennington, 2010). Furthermore, the probability that any given technique or instrument samples (and that scores are therefore affected by) more than one closely related skill set or domain is quite high. Paramount to skilled interpretation of performance is an understanding of the component processes being sampled, their interrelation, and how these processes may be manifest across circumstances or domains. Therefore, familiarity with current cognitive theories as well as with ongoing research is essential to the adequate interpretation and integration of results of the NEPSY-II.

Differentiating specific narrow characteristics or comorbid conditions can be difficult, and may result in several plausible conclusions about an examinee's performance or the relationships between the examinee's skills. Consequently, hypotheses or conclusions regarding particular performances or the interaction of characteristics may also vary among clinicians. As the number of subtests administered, observations recorded, or comparisons made between scores increases, it is also important to remember that the likelihood that spurious results will arise also increases (Crocker & Algina, 2006). To offset the impact of spurious results, consistency among the findings across measures or techniques is the most stable method (i.e., the most resistant to erroneous conclusions) of conceptualizing results in a coherent and practical manner that is useful for describing an examinee's needs, drawing diagnostic conclusions, and developing effective recommendations (Korkman et al., 2007; Riccio et al., 2010; Sattler, 2008; Yeates et al., 2010).

Moreover, the NEPSY-II allows a clinician to make a developmental comparison of the subcomponents of a given characteristic by examining underlying individual processes, although it was not designed to examine all potential characteristics or interactions of characteristics within or across cognitive domains (Kemp & Korkman, 2010; Korkman et al., 2007). Considering the developmental differences between a beginning reader's skills and an established reader's skills elucidates these principles. The beginning reader is deliberate in applying sound–symbol association, directed attention, and working memory, whereas the established reader probably uses only sight recognition of overlearned material (Kemp & Korkman, 2010). Thus the chain of skills used by both these readers may have been similar in the beginning, but has shifted with maturation, so that the readers now differ in their application and level of automaticity of reading skills. Considering only the outcome of a reading measure would be insufficient evidence on which to base a conclusion or develop recommendations. Targeted examination of skills allows a clinician to clarify the nature of an examinee's primary and secondary difficulties (according to Luria's model), and subsequently allows for improved recommendations and intervention development. The ability to differentiate the root cause (primary) of a deficit, such as poor phonological processing, from its impact (secondary), such as slow processing speed, effectively becomes a function of the clinician's knowledge and expertise in the areas under consideration. Thus, although clinicians of varying levels of training and experience can administer and score the NEPSY-II with proficiency (Brooks, Sherman, & Strauss, 2010), both choosing the appropriate subtests and interpreting performance across subtests or domains require a higher degree of expertise in neuropsychological constructs, developmental theory, and the professional literature (Titley & D'Amato, 2008).

Results

With these basic principles in mind, the next consideration for interpretation is at the level of the scores generated. The successful interpretation of NEPSY-II results begins with looking for patterns of scores at the subtest level, then moving to combined scaled scores, percentile ranks, and finally behavioral observation base rates (Kemp & Korkman, 2010). Although each of these scores is standardized along the same metric (mean of 10, standard deviation of 3), and therefore could result in the same descriptions of range (e.g., low average, average, high average), interpretation of the scores varies considerably (Korkman et al., 2007). In keeping with conventional assessment results, scaled scores compare an individual's performance to that of same-age peers (Sattler, 2008). These scores can be used to compare the examinee's skills to the population, identify patterns of strengths and weaknesses, and/or identify a pattern of performance consistent that may suggest potential intervention strategies. The conventional method of score analysis may be the final level of analysis for evaluators with little training in neuropsychology, or the beginning of analysis for evaluators with a greater depth of training (Kemp & Korkman, 2010). Beyond the scale scores, techniques such as consideration of error types (i.e., error analyses) or examination of differences in performance across subtests or domains (i.e., profile analyses) may help to clarify an examinee's needs. This higher level of analysis is supported by examining the examinee's overall profile as it is related to the identified difficulties; the combined and contrast scaled scores are useful in this regard.

The combined scaled scores allow the clinician to consider not only the outcome of performance, but the chain of events leading to that outcome (Korkman et al., 2007). The contrast scaled scores then allow the clinician to understand the examinee's range of functioning and more clearly identify potential contributing factors. Caution is warranted in the interpretation of contrast scores, however, as they only represent *differences* in level of functioning. Thus a performance on one subtest in the superior range (e.g., Memory for Faces) versus a performance on another subtest in the average range (e.g., Memory for Faces Delayed) can result in a similar contrast scaled score to that produced with performance differences of average versus below average. Although the contrast scaled score values will be similar, the implications for functioning in these two scenarios are significantly different.

Another consideration involves the cumulative percentage scores or the base rates, as these provide important information in the interpretation process. Although the observed difference between scores can be statistically significant, the base rate indicates how meaningful that difference is. For a difference to be considered rare (i.e., clinically meaningful), Sattler (2008) has suggested a base rate of no more than 15%, while Kemp and Korkman (2010) have supported using a more stringent rate of 10%. Finally, in addition to interpretation of the various scores provided, the behavioral observations need to be considered to lend additional qualitative insights into an examinee's functioning.

Integration

NEPSY-II results are best used as a supplement to the results of other standardized measures, as well as to historical, observational, and functional information (Titley & D'Amato, 2008). Integrating NEPSY-II results with other psychometric and functional data into a comprehensive discussion of an examinee's strengths and weaknesses can be a difficult task, especially for less experienced clinicians (Hooper, 2010). Selecting a battery of tests/techniques designed to address the referral issue(s) efficiently and effectively is the first step in this process. Appropriate selection allows for comparison across subtests, within and across domains, and with other measures (e.g., intellectual) to identify the patterns of performance. Poor performance in a single area or on a single measure, without corroborating evidence, is insufficient for identifying a developmental deviation or a diagnostic conclusion; typical children show developmental differences over time, without significant impact on their daily life (Meltzer, 2007). Significant findings across measures of similar or related characteristics (e.g., subtests and coded observations) give informational consistency and internal support to hypotheses and conclusions (Kemp & Korkman, 2010). Selection of measures and methods should ensure coverage of all functional domains, with specific subtests of the NEPSY-II providing supplemental coverage.

One approach to integration would follow the model of cross-battery assessment (Flanagan, Ortiz, & Alfonso, 2013), which aligns various subtests across measures based on broad and narrow abilities. The NEPSY-II was co-normed with the WISC-IV, and is currently included as a supplemental assessment in cross-battery assessment paradigms (Flanagan et al., 2013). For clinicians deriving their own combinations of subtests for narrow and broad abilities, the information provided in Table 19.1 can be helpful in placing the NEPSY-II subtests into this type of integrative model. Alternatively, the cognitive hypothesis-testing model (e.g., Flanagan, Fiorello, & Ortiz, 2010; Flanagan, Ortiz, Alfonso, & Mascolo, 2006)—which identifies normative deficits, considers plausible contributing or causal factors (e.g., lack of motivation vs. a cognitive deficit), and links current results to ecological deficits identified in the referral question—is another model that can be used.

Regardless of the model employed, the general idea is that of moving from the simple to the complex, with emphasis on consistencies, to facilitate the identification of patterns of performance and the integration of performance on the NEPSY-II with other information gathered during the evaluation process. To facilitate the integrative process, the clinician will need to be familiar with research on the chosen instruments or techniques (i.e., what is being measured) and how the instruments relate to each other (Hooper, 2010).

Some examples from research on the NEPSY or NEPSY-II reinforce the idea that assessment of higher-level cognitive processes is far more complex than simply administering one or more subtests and generating a score in a particular range. In their review of bilingual children's performance on NEPSY subtests, Garratt and Kelly (2008) found significant performance differences from monolingual children. Bilingual children demonstrated lower visual attention, naming speed, and verbal comprehension; however, their performance on the Imitating Hand Positions and Design Copying subtests was superior to that of their monolingual peers. In addition, Dixon and Kelly (2001) found potential cultural influences on the NEPSY, as British children were noted to demonstrate better phonological processing, visual attention, and verbal comprehension skills than their American counterparts. These differences may be in part due to differences in educational strategies in different cultures, to the use of theoretically versus empirically derived domains, due to homogeneity in the norming samples, or to the interaction of these or other unidentified issues. Regardless of the causes of these differences, these findings underscore the need to consider results of the NEPSY-II in the broader context of all other information.

Gifford, Mahaney, and Gorman (2008) found that children with ADHD performed significantly better than children with autism spectrum disorder (ASD) on NEPSY-II tests of motor speed and dexterity, whereas children with ASD outperformed children with ADHD on tests of visual–spatial analysis and attention to detail. Gifford and colleagues (2008) commented that these findings diverged from their expectations, as no significant differences between these groups were noted on NEPSY-II tests of attention and concentration or of executive skills. In a second study, Gifford, Mahaney, and Gorman (2009) found that children with ADHD performed better on visual memory tasks than children with ASD and children with seizure disorders, while the children with ASD performed better than children with seizure disorders on tests of phonological processing. Among children with ASD, differences were also found for subgroups on the spectrum; the meaning of these differences is the subject of theoretical discussion and additional research.

Finally, Brooks, Sherman, and Iverson's (2010) research supports the notion that typically developing children's performance can fall well below expectations on a given day or subtest. When the seven subtests for 3- to 4-year-olds were considered together, 71.5% of the sample had one or more scores below the 25th percentile, and 40% of the sample had two or more scores below the 10th percentile. The number of low scorers in this group was mediated by parental education level, with lower levels of parental educational attainment being associated with increased rates of NEPSY-II scale scores falling below the expected range. Stamina also appears to play a role in typically developing children's NEPSY-II performance. When a 1-hour battery consisting of eight subtests was used, 70.3% of participants ages 5–6 had one or more scale scores below the 25th percentile, and 37.2% of participants had two or more scale scores below the 10th percentile. When a 2-hour battery consisting of 12 scale scores was used, 82.6% of these participants had one or more scores below the 25th percentile, and 49.3% had two or more scores below the 10th percentile. Once again, the base rates of children with low scores decreased as a function of their parents' educational attainment. The authors found similar relationships between parental education and the lowered score base rate in the 7- to 16-year-old group as well.

BEYOND TRADITIONAL INTELLECTUAL ASSESSMENT

Traditional intellectual assessment is concerned with quantifying general ability (*g*), and is most often used in determining intellectual disability or as one component of identifying learning disability. The NEPSY-II is not a measure of intelligence, but a supplemental measure to increase the extent to which the assessment process examines all functional systems. The majority of intellectual assessment instruments do not effectively measure the full range of cognitive abilities (Flanagan et al., 2006); thus the use of a traditional intelligence test with the NEPSY-II provides the means for obtaining a comprehensive assessment of domains of function (Pramuka & McCue, 2000). It provides additional information on the narrow abilities that may otherwise be ignored and are important for development of appropriate interventions.

The NEPSY-II, in combination with a general measure of ability, is consistent with a focus on the need to integrate cognitive and behavioral data. It thus increases clinicians' ability to deal with the multidimensionality of individuals (Riccio & Reynolds, 2013) and to translate this understanding into educationally relevant information (Goldstein & Naglieri, 2008). As noted earlier, the NEPSY-II permits systematic variation of the inputs, outputs, and levels of complexity, to maximize the likelihood of dissociating the potential underlying problem that is presenting as impaired performance. It is not solely a deficit-based approach; it is consistent with Luria's assertion that neuropsychological assessment is a valuable approach for determining not only deficits, but individual strengths or intact cognitive functions, as they relate to everyday functioning. Focusing on strengths lends itself to compensatory models that focus not on the underlying impairment, but on ways to compensate for the impairment to improve everyday living (Glisky & Glisky, 2002). Compensation approaches identify methods to bypass deficit skills through the use of intact functions or alternative methods of reaching the same goal (Anderson, 2002).

A more unified picture of the individual should increase the predictive value of the assessment as it relates to achievement as well as realistic life planning (Silver et al., 2006; Teeter & Semrud-Clikeman, 1997). By not only emphasizing how well a student does or doesn't do, but gaining an understanding of the types of problems (i.e., inputs, outputs) that are most difficult for the individual, a clinician can identify interventions that focus on circumventing the problems (i.e., accommodations, modifications). The more comprehensive picture typically associated with neuropsychological assessment examines individual performance across a range of functional domains, including linguistic, perceptual, sensory–motor, attention, memory, learning, executive control/planning, speed of processing, and emotional functioning (Riccio & Reynolds, 2013; Silver et al., 2006).

Following in the traditions of Luria, neuropsychological assessment such as that provided by the NEPSY-II looks beyond the failure to attain a specific skill to the underlying brain–behavior relationship that contributes to this difficulty. This is seen as important, in that this same problem within the functional system may indicate increased likelihood of failure in attaining other academic, functional, social, or behavioral skills (Riccio & Reynolds, 1998, 2013). At the same time, Kemp and Korkman (2010) have warned against drawing conclusions about specific brain function based on NEPSY-II results: A child's brain is still developing, and the long-term effects of a congenital or acquired injury on brain development will differ, depending on a number of factors that may vary over the lifespan.

Neuropsychological assessment (e.g., with the NEPSY-II) may be appropriate to establish initial functioning, as well as to track progress; it may serve to clarify intervention needs and result in referrals to other specialists (Berkelhammer, 2008). Intervention planning goes beyond the labeling or eligibility/placement process to include the identification of specific management or rehabilitation techniques, medical management approaches, outcome-related goals, and modifications that need to be addressed (Silver et al., 2006). Changes in functioning on the various subtests of the NEPSY-II, due to their developmental sensitivity, may be more apparent than on the traditional full-scale score of an intelligence test. The flexibility of being able to select specific subtests, rather than having to administer the full battery, also renders the NEPSY-II more suitable for follow-up evaluation.

CONCLUSIONS

The NEPSY-II is one of the few flexible, child-friendly batteries of tests available for the assessment of higher-level cognitive functions (Brooks, Sherman, & Strauss, 2010; Davis & Matthews, 2010). It is best used to supplement rather than to replace traditional standardized intelligence tests, in order to obtain a more comprehensive picture of multilayered abilities. It should be kept in mind that theoretical as well as statistical and psychometric bases need to be considered in using the NEPSY-II. Developmental and cultural contexts must also be considered. Without considering expected developmental or cultural differences, a clinician could easily draw erroneous conclusions and develop baseless recommendations. Similarly, convergent validity (i.e., consistency across subtests and measures) is important in decreasing the likelihood of potentially spurious results.

Although some may rightly argue against various statistical properties of this battery or the norming choices made by the authors (Titley & D'Amato, 2008), the NEPSY-II ultimately demonstrates excellent clinical utility through an integrated battery of subtests and a variety of scores allowing for in-depth assessment of an examinee's skill set (Hooper, 2010). The NEPSY-II represents continuing strides forward in the integration of qualitative and quantitative information in the comprehensive evaluation of children. Building on the foundation of the NEPSY, the NEPSY-II continues to enhance the breadth and depth of available assessment instruments for children and adolescents. However, several areas in need of additional consideration remain.

Defining the NEPSY-II domains via factor analysis would lend significant power to the battery's predictive validity. The problem with such a task is the lack of agreement in the literature regarding the definitions or methods of assessment for particular skills. For instance, although executive functions are frequently discussed in the literature, and are generally agreed to be of great importance, a widely accepted formal definition of these skills has not been established (Jurado & Rosselli, 2007; Kenworthy, Black, Harrison, Rosa, & Wallace, 2009). Furthermore, common differences in the definition of executive skills leads to differences in the validity of particular evaluation methods, as well as the ecological validity of resulting recommendations.

Developing suggested batteries and assessment practices based on research defining expected performance could also enhance the predictive and diagnostic power of the NEPSY-II. Although social perception is a concern in a variety of conditions, how would a child with a given condition perform on the NEPSY-II? Could the degree of impairment related to a given condition be ascertained on the basis of NEPSY-II performance? What impact would the level of chronicity have on a performance profile? Broadening the stratification to factors beyond those typically employed (e.g., race, socioeconomic status) could also lend insight into NEPSY-II performance. As discussed above, different cultural or familial experiences can significantly affect a performance profile. Along this line, the applicability of the NEPSY-II would probably be broadened if it were considered in relation to instruments developed for assessing English-language learners for instance. Moreover, how do the methods of assessment used in NEPSY-II subtests (e.g., Theory of Mind) influence expected performance? How should methods of assessment be expected to vary by language or cultural background? Finally, how do NEPSY-II results translate into ecologically valid recommendations? Would a particular pattern of performance suggest one recommendation or strategy over another? These types of questions (and surely others) require ongoing research related to underlying neuropsychological characteristics of the individuals being evaluated, as well as how the NEPSY-II measures those characteristics and how those measurements translate into improved recommendations.

REFERENCES

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8,* 71–82.

Bauer, R. M. (2000). The flexible battery approach to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 419–448). Mahwah, NJ: Erlbaum.

Berkelhammer, L. D. (2008). Pediatric neuropsychological evaluation. In M. Hersen & A. M. Gross (Eds.), *Handbook of clinical psychology* (pp. 497–519). Hoboken, NJ: Wiley.

Bishop, D. V. M. (2006). *Children's Communication Checklist—Second Edition* (U.S. ed.). San Antonio, TX: Harcourt Assessment.

Bracken, B. A. (2006a). *Bracken Basic Concept Scale: Expressive.* San Antonio, TX: Harcourt Assessment.

Bracken, B. A. (2006b). *Bracken Basic Concept Scale—Third Edition: Receptive.* San Antonio, TX: Harcourt Assessment.

Brooks, B. L., Sherman, E. M. S., & Iverson, G. L. (2010). Healthy children get low scores too: Prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology, 25,* 182–190.

Brooks, B. L., Sherman, E. M. S., & Strauss, E. (2010). NEPSY-II: A Developmental Neuropsychological Assessment, Second Edition. *Child Neuropsychology, 16,* 80–101.

Brown, T. E. (2001). *Brown Attention-Deficit Disorder Scales for Children and Adolescents.* San Antonio, TX: Psychological Corporation.

Cohen, M. J. (1997). *Children's Memory Scale.* San Antonio, TX: Psychological Corporation.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory.* Mason, OH: Wadsworth.

Davis, J. L., & Matthews, R. N. (2010). Test review: NEPSY—Second Edition (NEPSY-II). *Journal of Psychoeducational Assessment, 28*(2), 175–182.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System.* San Antonio, TX: Psychological Corporation.

Dixon, L. A., & Kelly, T. P. (2001). A comparison of the performance of preschool children from England and the USA on the NEPSY: Developmental Neuropsychological Assessment. *Clinical Neuropsychological Assessment, 2,* 43–60.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third" method approach to specific learning disabilities identification. *Psychology in the Schools, 47,* 739–760.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *The achievement test desk reference: A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Garratt, L. C., & Kelly, T. P. (2008). To what extent does bilingualism affect children's performance on the NEPSY? *Child Neuropsychology, 14,* 71–81.

Gifford, K. A., Mahaney, T. R., & Gorman, P. W. (2008, October). *Differentiating between seizure disorder, ADHD, and PDD in children using the NEPSY-II.* Poster presented at the annual conference of the National Academy of Neuropsychology, New York.

Gifford, K. A., Mahaney, T. R., & Gorman, P. W. (2009, November). *Using the NEPSY-II to assess ADHD, seizure disorder, and autism spectrum disorders in children.* Poster presented at the annual conference of the National Academy of Neuropsychology, New Orleans, LA.

Glisky, E. L., & Glisky, M. L. (2002). Learning and memory impairments. In P. J. Eslinger (Ed.), *Neuropsychological interventions: Clinical research and practice* (pp. 137–162). New York: Guilford Press.

Goldstein, S., & Naglieri, J. A. (2008). The school neuropsychology of ADHD: Theory, assessment, and intervention. *Psychology in the Schools, 45,* 859–874.

Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System—Second Edition.* San Antonio, TX: Psychological Corporation.

Hooper, S. R. (2010). Strengths and weaknesses of the NEPSY-II. In S. L. Kemp & M. Korkman, *Essentials of NEPSY-II assessment* (pp. 227–247). Hoboken, NJ: Wiley.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychological Review, 17,* 213–233.

Kemp, S. L., & Korkman, M. (2010). *Essentials of NEPSY-II assessment.* Hoboken, NJ: Wiley.

Kenworthy, L., Black, D. O., Harrison, B., Rosa, A. D., & Wallace, G. L. (2009). Are executive control functions related to autism symptoms in high-functioning children? *Child Neuropsychology, 15,* 425–440.

Korkman, M. (1999). Applying Luria's diagnostic principles in the neuropsychological assessment of children. *Neuropsychology Review, 9,* 89–105.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY.* San Antonio, TX: Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY—Second Edition (NEPSY-II).* San Antonio, TX: Harcourt Assessment.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

Meltzer, L. (2007). *Executive function in education: From theory to practice.* New York: Guilford Press.

Naglieri, J., LeBuffe, P., & Pfeiffer, S. I. (1994). *Devereux*

Scales of Mental Disorders (DSMD). San Antonio, TX: Psychological Corporation.

Pramuka, M., & McCue, M. (2000). Assessment to rehabilitation: Communicating across the gulf. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 337–356). Mahwah, NJ: Erlbaum.

Psychological Corporation. (2001). *Wechsler Individual Achievement Test—Second Edition*. San Antonio, TX: Author.

Rhodes, R. L., D'Amato, R. C., & Rothlisberg, B. A. (2009). Utilizing a neuropsychological paradigm for understanding common educational and psychological tests. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (pp. 321–348). New York: Springer.

Riccio, C. A., & Reynolds, C. R. (1998). Neuropsychological assessment of children. In A. S. Bellack & M. Hersen (Series Eds.) & C. R. Reynolds (Ed.), *Comprehensive clinical psychology* (Vol. 4, pp. 267–301). New York: Pergamon Press.

Riccio, C. A., & Reynolds, C. R. (2013). Principles of neuropsychological assessment in children and adolescents. In D. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 331–347). New York: Oxford University Press.

Riccio, C. A., Sullivan, J. R., & Cohen, M. J. (2010). *Neuropsychological assessment and intervention for childhood and adolescent disorders*. Hoboken, NJ: Wiley.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Author.

Silver, C. H., Blackburn, L. B., Arffa, S., Barth, J. T., Bush, S. S., Koffler, S. P., et al. (2006). The importance of neuropsychological assessment for the evaluation of childhood learning disorders: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 21,* 741–744.

Teeter, P. A., & Semrud-Clikeman, M. (1997). *Child neuropsychology: Assessment and interventions for neurodevelopmental disorders*. Boston: Allyn & Bacon.

Titley, J. E., & D'Amato, R. C. (2008). Understanding and using the NEPSY-II with young children, children, and adolescents. In R. C. D'Amato & L. C. Hartlage (Eds.), *Essentials of neuropsychological assessment: Treatment planning for rehabilitation* (pp. 149–172). New York: Springer.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Harcourt Assessment.

Yeates, K. O., Ris, M. D., Taylor, H. G., & Pennington, B. F. (2010). *Pediatric neuropsychology: Research, theory, and practice* (2nd ed.). New York: Guilford Press.

CHAPTER 20

The Universal Nonverbal Intelligence Test—Second Edition

A Multidimensional Nonverbal Alternative for Cognitive Assessment

R. Steve McCallum
Bruce A. Bracken

The Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) and the revised and newly standardized edition, the Universal Nonverbal Intelligence Test—Second Edition (UNIT2; Bracken & McCallum, 2016b), are language-free tests of cognitive ability that require no receptive or expressive language from the examiner or the examinee for administration. The need for nonverbal assessment in the United States has grown steadily over the past 30 years and continues to expand as the country becomes ever more diverse. According to estimates from the U.S. Census Bureau (2016), minority populations increased steadily in the United States from 2010 to 2015. Specifically, the percentages increased from 12.6 to 13.3%, from 0.9 to 1.2%, from 4.8 to 5.6%, and from 16.3 to 17.6% for blacks/African Americans, American Indians/Alaska Natives, Asians, and Hispanics/Latinos, respectively. The percentage of white non-Hispanics decreased from 63.6 to 61.6%.

Taken together, minority children constitute an ever-increasing percentage of public school children, particularly in many large cities. For example, minorities constitute an overwhelming percentage of the school populations in Miami (approximately 84%), Chicago (89%), and Houston (88%). In addition, the population with limited English proficiency (LEP) is the fastest-growing group in the nation. There are more than 200 languages spoken in the greater Chicago area (Pasko, 1994), and 140 in California (Unz, 1997). A recent report from the U.S. Department of Education (2016) showed 19 prominent languages spoken by students in English-language learner (ELL) programs, with Spanish being the most prominent (71% of ELL students nationally). Other prominent languages included Chinese, Arabic, Vietnamese, Haitian-Creole, Russian, Navajo, and 12 other languages ranked second or third in frequency, depending on the state. Moreover, within the last few years, an increasing number of refugees from the war-ravaged Middle East have come into the United States. Many students from these minority populations may be disenfranchised when assessed on English-language-loaded tests. Moreover, there are many students who cannot be assessed optimally with language-loaded tests (e.g., those who have significant language and communication deficits limiting their ability to respond effectively). In some cases, it is difficult to determine whether such students' limited communication skills are the result of language-related deficits or limited cognitive ability. Such a distinction is useful in determining appropriate diagnoses and interventions, and nonverbal cognitive tests such as the UNIT2 help practitioners make such determinations.

Another population often misdiagnosed when language-loaded tests are used consists of individuals with hearing impairments. According to figures from the National Institutes of Health (2010), 28,600,000 Americans are deaf or have other significant hearing impairments. Still other students have neurological and psychiatric conditions that inhibit effective verbal communications (e.g., autism spectrum disorder, traumatic brain injury, verbal learning disabilities, severe depression, and selective mutism). In a chapter focusing on nonverbal personality assessment, Wasserman (2017a) concluded that nonverbal personality assessment may be indicated for (1) individuals with neurologically based acquired language disorders (e.g., aphasia, language-based learning disabilities); (2) individuals with varied cultural, linguistic, or national backgrounds (e.g., non-English-speaking persons); (3) individuals who are illiterate or poorly educated; (4) individuals who are deaf or hard of hearing; (5) individuals with forms of emotional disturbance that are manifested through an inability or unwillingness to produce an adequate and unconstrained sample of verbal behavior; and (6) individuals who are prone to misrepresent themselves on verbal self-report or interview measures. Individuals from these same populations would likely benefit from nonverbal cognitive assessment as well.

Since its publication, the original UNIT has provided examiners a viable option for assessing the cognitive abilities of examinees who cannot be assessed optimally with a language-loaded instrument. Numerous assessment experts have explored the psychometric properties of the UNIT, and the scrutiny overall has been positive (see Athanasiou, 2000; Bandalos, 2001; Fives & Flanagan, 2002; Maller, 2000). In addition, the UNIT has been used to operationalize general intellectual ability among students with limited language abilities and other who are at risk due to language-related issues, as well as individuals who are not from at-risk populations (e.g., Brinton, Spackman, Fujiki, & Ricks, 2007; Fujiki, Spackman, Brinton, & Illig, 2008; Hughes & Zhang, 2007; Lienemann, Graham, Leader-Janssen, & Reid, 2006; Noland, 2009; Liew, McTigue, Barrois, & Hughes, 2009; Pendarvis & Wood, 2009).

As the authors of the original UNIT, we have retained many of the positive features of the original test in its revision, but added many valuable revision features and updates. As a result of a thoughtful and comprehensive revision, there is every reason to believe that the UNIT2 will be even more positively acclaimed than the original test. In the following sections, we describe the theory and structure of the UNIT2, as well as its administration/scoring directions, interpretive strategies, psychometric properties, and salient strengths and weaknesses.

THEORY AND STRUCTURE

The UNIT2 contains six subtests, as did the UNIT, and retains the focus on the assessment of memory and reasoning and the intent to assess cognition using *two organizational strategies* (symbolic and nonsymbolic processing). However, the UNIT2 provides a broader assessment of cognition than the original test by including two measures of quantitative thinking, one symbolic and one nonsymbolic.

Within each of the two fundamental organizational strategies, the three cognitive abilities of memory, reasoning, and quantitative thinking are assessed in a hierarchical arrangement, such that memory is considered to be a more foundational building block of cognition (see Jensen, 1980). There are two memory subtests, Symbolic Memory and Spatial Memory; two reasoning subtests, Cube Design and Analogic Reasoning; and two quantitative subtests, Nonsymbolic Quantity and Number Series. Three of the subtests require some motoric manipulation (i.e., Symbolic Memory, Spatial Memory, and Cube Design); the remaining subtests (i.e., Analogic Reasoning, Nonsymbolic Quantity, and Number Series) require only a pointing response. With one exception (Cube Design, which requires design constructions), the subtests that require motoric manipulation can be adapted to allow for a pointing response only, if necessary.

As mentioned previously, the original UNIT employed subtests that feature both symbolic and nonsymbolic task demands. *Symbolic* organization strategies require the use of concrete and abstract symbols to conceptualize the environment; these symbols are typically language-related (e.g., words), although the symbols may take on any form (e.g., numbers, statistical equations, rebus characters, flags). Symbols are eventually internalized and come to label, mediate, connote, and (over time) make meaningful our experiences. *Nonsymbolic* strategies require the ability to perceive and make meaningful judgments about the physical relationships within our environment; this ability is symbol-free (or relatively so) and is closer to fluid-like intellectual abilities. The UNIT2 continues that tradition; however, Symbolic and Nonsymbolic Quotients are not provided for the UNIT2, as both processes predict academic performance about equally well when group data are analyzed.

Within each of the two fundamental organizational categories included in the UNIT (i.e., nonsymbolic and symbolic), problem solution requires one of three types of cognitive abilities—*memory, fluid reasoning,* or *quantitative thinking.* That is, some of the items require primarily symbolic organization and rely heavily on memory (e.g., those included on the Symbolic Memory subtest). Other items require considerable symbolic organization and reasoning skills, but less short-term or working memory (e.g., those included on the Analogic Reasoning subtest, which requires long-term memory and reasoning). Some items seem to require nonsymbolic organization strategies and memory primarily (e.g., those on the Spatial Memory subtest). Others require nonsymbolic organization and reasoning, but little memory (Cube Design). Number Series requires symbolic organization and quantitative ability; whereas Nonsymbolic Quantity requires nonsymbolic ability and quantitative skills. The UNIT2 theoretical model is depicted in Figure 20.1.

The rationale for conceptualizing the use of symbolic versus nonsymbolic strategies to operationalize intelligence is supported in the literature. For example, Wechsler (1939) emphasized the importance of distinguishing between highly symbolic (verbal) and nonsymbolic (perceptual or performance) means of expressing intelligence. Jensen (1980) provided a rationale for a two-tiered hierarchical conceptualization of intelligence, consisting of the two subconstructs of memory (level I) and reasoning of a fluid and quantitative nature (level II). However, in contrast to many low-*g*-loaded level I memory tasks designed to require reproduction or recall of simple content, the UNIT2 memory tasks were developed to require more complex memory functioning (e.g., working memory).

The theoretical organization of the UNIT is consistent with portions of several instruments that adopt the Cattell–Horn–Carroll [CHC] model), as described by Cattell (1963), Horn (1968, 1974), Carroll (1993), and others (e.g., Woodcock, 1990; Woodcock, McGrew, & Mather, 2001; Schrank, McGrew, & Mather, 2014; see also Schneider & McGrew, Chapter 3, this volume). Kevin McGrew (personal communication, September 2016) has identified components of strata I and II within the UNIT2 model, including fluid reasoning, visual processing, and quantitative reasoning from stratum II, as well as visual memory, visualization, visual memory, quantitative reasoning, and induction from stratum I.

Intelligence consists primarily of a pervasive and fundamental general cognitive ability, *g,* which provides a base for the development of other unique or specialized skills. Although there are many methods for determining levels of intelligence, it makes little sense to conceptualize intelligence in a dichotomized manner as being either verbal or nonverbal. Rather, there are verbal and nonverbal means available to assess intelligence. Consequently, the UNIT2 should be considered a nonverbal measure of intelligence, not a measure of nonverbal intelligence, as it was designed to be a strong measure of *g,* underpinned by symbolic and nonsymbolic processes.

In addition to retaining the most salient theoretical features of the UNIT, we have addressed limitations of the original test in the revision. For example, the norms were updated and the age range extended to 21 years, 11 months; test floors were extended for younger examinees; the administration format was improved by allowing a one-way easel presentation; and pictured stimuli were colorized. In addition, there are now two Standard Battery administration options available: one that does not include short-term memory, and one with

	Symbolic Content	**Nonsymbolic Content**
Memory	Symbolic Memory	Spatial Memory
Fluid Reasoning	Analogic Reasoning	Cube Design
Quantitative Thinking	Numerical Scales	Nonsymbolic Quantity

FIGURE 20.1. Conceptual model for the UNIT2.

short-term memory. Also, a strong two-subtest Abbreviated Battery may be administered (Analogic Reasoning and Nonsymbolic Quantity).

SUBTESTS AND COMPOSITES

In addition to subtest scores and a Full Scale IQ (FSIQ), three composites can be calculated for the UNIT2 Full Scale Battery: Memory, Reasoning, and Quantitative composites. As mentioned previously, examiners may elect to administer the Standard Battery with the Memory composite (SBIQ-M) or without it (SBIQ). The UNIT2 subtests (Symbolic Memory, Nonsymbolic Quantity, Analogic Reasoning, Spatial Memory, Number Series, and Cube Design) and their administration are described in the section that follows.

ADMINISTRATION AND SCORING

Administration of Subtests

Administration of all six UNIT2 subtests requires approximately 45–50 minutes. Norms allow for two-, four-, and six-subtest batteries. The shorter batteries require approximately one-third to two-thirds less time than the full battery.

Although administration of the UNIT2 subtests is 100% nonverbal, the examiner may communicate with the examinee (if they have a common language) to establish rapport, discuss extra-test issues, and the like, as well as to reduce the awkwardness that may otherwise occur with a nonverbal interaction. The examiner must, however, present the UNIT2 stimuli nonverbally, using eight nonverbal gestures (which were used during standardization and are presented in the manual) to communicate task demands. To aid in teaching task demands, the examiner is instructed to make use of demonstration items, sample items, and "checkpoint" items; checkpoint items allow the examiner to provide feedback only if the item is failed, but unlike demonstration and sample items, checkpoint items are scored. Specific descriptions and subtest directions follow.

Symbolic Memory stimulus pages are located in Stimulus Book 1 and are presented to the examinee for 5 seconds each. Symbolic Memory items depict one or more universal human figures (i.e., baby, girl, boy, woman, and man) in varying combinations; each human figure is produced in both green and black. Examinees younger than 8 years must identify and select the matching image from options presented below the stimulus picture. For examinees older than 8 years, the stimulus page is presented for 5 seconds and then removed. The examinee then is instructed through modeling and gestures to replicate the sequenced order shown on the stimulus page. The examinee uses 1.5″ × 1.5″ response cards, each depicting one of the universal human figures, to reproduce the array shown on the stimulus page. The examinee's response has no time limits or bonus credit given for rapid performance. Materials needed include Stimulus Book 1, 10 Symbolic Memory Response Cards, and a stopwatch for timing stimulus exposure. The subtest is discontinued after the examinee obtains three consecutive scores of 0 (i.e., three consecutive failed items).

For the Nonsymbolic Quantity subtest, the examiner presents an easel-bound stimulus page in Stimulus Book 1, which includes an array of white and black domino-like objects with various numerical values. Each domino figure creates a numerical sequence, equation, analogy, or mathematical problem to be solved. The examiner points to the stimulus figure series, which ends with a red question mark. The examiner then waves a hand over the response options at the bottom of the page, points to the question mark, and shrugs to ask the examinee how the item should be completed. The examinee points to the response option presented below the stimulus that best completes the conceptual or numerical analogy, sequence, or problem. The subtest is discontinued after the examinee obtains three consecutive scores of 0 (i.e., three consecutive failed items).

The Analogic Reasoning subtest requires the examinee to solve analogies presented in a matrix format. The examinee is directed to indicate which one of several options best completes a two-cell or a four-cell analogy. Task solution requires the examinee to determine the relationships between objects. For example, in a four-cell matrix, the first cell may depict a fish, and the second water; the third cell may show a bird, and the fourth cell is blank. The examinee selects from several options the picture that best completes the matrix. In this case, a picture of the sky is a correct response. This reasoning subtest is not timed. Each correct response is assigned 1-point credit.

For the Spatial Memory subtest, the examiner briefly presents a series of grids on stimulus plates located on an administration easel. The grids show one or more green or black polka dots placed within cells (on the grid). The less difficult items use a 3 × 3 grid; the more difficult items require a

4 × 4 grid. The stimulus plate is shown for 5 seconds and then is removed from view. The examinee places response chips on a blank grid that is placed on the table top in front of him or her. Spatial Memory has no time limits, except for the 5-second exposure. Each correct response is assigned 1-point credit.

To administer the Numerical Series subtest, the examiner presents a stimulus page in Stimulus Book 3 with arrays of numbers or mathematical symbols that create analogies, sequences, or problems. The examiner waves a hand over the depicted numbers and motions to a red question mark. After pointing to the question mark, the examiner shrugs, asking the examinee to select the option that best completes the series or solves the problem. The examiner discontinues the subtest after the examinee obtains three consecutive scores of 0 (i.e., three consecutive failed items).

The Cube Design subtest requires the examinee to use up to nine cubes to replicate three-dimensional designs shown on a stimulus plate. Each cube has six facets: two white sides, two green sides, and two sides that contain diagonals (triangles), one green and one white. These cubes can be arranged to replicate the three-dimensional figures depicted on the stimulus plates. This task is timed, but the time limits are liberal, to emphasize assessment of ability rather than speed. Except for the very early items, which are scored either correct or incorrect and yield a maximum of 1 point per item, examinees may earn up to 3 points (per item). Each facet of the three-dimensional construction is judged to be either correct or incorrect. Each correct facet is assigned 1-point credit.

Materials and Scoring

In addition to the manipulatives (e.g., blocks, response chips, response grids,), general materials required to administer the UNIT2 include a stopwatch, a #2 lead pencil, a test booklet for recording the student's performance, and relevant demographic information. The stimulus plates for all subtests are contained in three easels.

Scoring is user-friendly and straightforward. Correct responses are printed on the test booklet to facilitate ease of scoring. Raw scores for each item are summed to provide a subtest total raw score, which are easily transformed to standard scores with either the tables provided in the UNIT2 manual or the UNIT2 Compuscore software (Bracken & McCallum, 2016a). The Abbreviated Battery is used for screening purposes; one of the Standard Batteries for most purposes, including placement decisions; and the Extended Battery for diagnostic purposes.

PSYCHOMETRIC PROPERTIES

Norming Procedures, Standardization Characteristics, and Technical Data

The UNIT2 was standardized and normed on a nationally representative sample of 1,603 students from 33 states; these individuals ranged in age from 5 to 21 years, with 50.8% male and 49.2% female representation. The stratification of the sample closely represented the U.S. population on all relevant variables, including sex, race, ethnicity, Hispanic origin, geographic region, special education status, parental education attainment, and household income (see the examiner's manual for details).

Reliability

Average UNIT2 coefficient alphas range from .89 to .98 for subtests and scales across the full age range; coefficients for composite scores are higher, ranging from .93 (Memory composite) to .98 (Standard Battery without Memory and Full Scale Battery). The average subtest reliability coefficients across age groups range from .89 (Spatial Memory) to .96 (Nonsymbolic Quantity, Analogic Reasoning, and Numerical Designs); Full Scale reliability coefficients range from .97 (at ages 5, 6, 7, 8, and 12) to .99 (ages 16 and 19). For the Standard Battery with Memory, the composite reliability coefficients range from .95 (ages 5 and 6) to .98 (age 17); for the Standard Battery without Memory, composite reliability coefficients range from .95 (age 5) to .99 (ages 16 and 21). For the Abbreviated Battery, composite reliability coefficients range from .94 (ages 5 and 6) to .98 (ages 12, 13, 14, 16, 17, 19, 20, and 21). FSIQ reliability coefficients are very high by race and ethnicity: All are .98, with the exception of the coefficient for American Indian/Eskimo examinees (.99). Full Scale composite reliability coefficients are all well above the recommended minimum (i.e., .90) for scores used in guiding selection/placement decisions (see Bracken, 1987; Bracken & McCallum, 1998; Wasserman & Bracken, 2013).

As an indicator of fairness, internal-reliability estimates are reported in the UNIT2 examiner's manual for special populations (e.g., children with learning disabilities or speech and language impairments) and for the important decision-making points (i.e., FSIQ of 70 ± 10 or 130 ± 10). These coefficients are similarly impressive and comparable to those reported for the entire standardization sample.

In order to assess stability, a sample of 199 participants was first divided into four age groups. Test–retest reliability over an average interval of 17.8 days was reported for the four groups and the combined sample. Practice effects for the combined age sample were small and averaged 1.58 points for the Abbreviated Battery, 2.66 points for the Standard Battery with Memory, 2.15 points for the Standard Battery without Memory, and 2.45 points for the Full Scale Battery. Obtained coefficients and those corrected for restriction and/or expansion in range are reported in the examiner's manual. Corrected subtest stability coefficients ranged from .75 (Spatial Memory) to .94 (Cube Design); corrected composite stability coefficients ranged from .86 (Memory) to .90 (Reasoning) for the overall sample. Stability coefficients for total test composites across the four batteries ranged from .85 (Abbreviated Battery) to .93 (Standard Battery without Memory and Full Scale Battery). All of the stability coefficients for the composites exceeded .80, which is indicative of strong test–retest reliability and stability. The typical test–retest gain score was small, consistently less than 0.30 standard deviation.

Interrater scorer consistency was also assessed by having two members of the publishing company staff (PRO-ED) independently score 50 protocols drawn at random from the normative sample. The resulting interrater coefficients for various subtests and composites ranged from .98 to .99, indicating excellent scoring consistency.

Validity

Data from the original UNIT provide a starting point for establishing UNIT2 validity, given the similarities across the two tests. Additional studies conducted as part of the UNIT2 norming process provide further evidence of content description validity, criterion prediction validity, and construct identification validity. For the initial determination of UNIT2 content validity, careful selection procedures were used to choose items; the items were then subjected to differential item functioning analyses to investigate the presence or absence of bias in the test's items. To examine criterion prediction validity, the UNIT2 was correlated with seven major intelligence tests. The correlations between the UNIT2 composites and criterion tests ranged from moderate to very high levels, suggesting that the UNIT2 correlates significantly and meaningfully with other tests of general intelligence.

Initially, UNIT and UNIT2 scores were correlated. The corrected correlation between the UNIT2 Standard Battery with Memory and the UNIT Standard Battery was .96, indicating that the two editions are highly related in content and constructs assessed. Consequently, the literature exploring the validity of the UNIT has considerable relevance to the findings for and use of the UNIT2. Correlations between the UNIT2 Full Scale Battery and the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014) Full Scale were .66 for the CAS2 Standard Battery and .69 for the CAS2 Extended Battery. While these are large (and significant) correlations, they are slightly lower than those obtained when the UNIT2 was compared to other measures, probably because the UNIT2 and CAS2 were developed from slightly different models of intelligence. The correlations between the UNIT2 Abbreviated Battery, Standard Battery with Memory, Standard Battery without Memory, and Full Scale Battery and the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003) were .70, .83, .83, and .84, respectively. Correlations between the four UNIT2 batteries and the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003) Abbreviated Battery IQ (ABIQ) were .59 (UNIT2 Abbreviated Battery), .69 (UNIT2 Standard Battery with Memory), .73 (UNIT2 Standard Battery without Memory), and .73 (UNIT2 Full Scale Battery). Correlations between the UNIT2 Abbreviated Battery, Standard Battery with Memory, Standard Battery without Memory, and Full Scale Battery and the Comprehensive Test of Nonverbal Intelligence—Second Edition (CTONI-2; Hammill, Pearson, & Wiederholt, 2009) were .84, .82, .85, and .85, respectively. The UNIT2 Abbreviated Battery, Standard Battery with Memory, Standard Battery without Memory, and Full Scale Battery correlated with the Woodcock–Johnson III Tests of Cognitive Abilities (WJ III COG; Woodcock et al., 2001) in the large to very large range (.64, .74, .82, and .79, respectively). These studies indicate that the UNIT2 is a sound measure of global intelligence, but that it may not be strongly correlated with measures of processing speed. Speed is a factor downplayed in the UNIT and UNIT2 because processing speed is not valued in many cultures, at least to the extent that it is valued in mainstream U.S. culture. These findings support our intention to put less emphasis on speed as a measure of intelligence within the UNIT2. The UNIT2 Abbreviated Battery, Standard Battery with Memory, Standard Battery without Memory, and Full Scale Battery also correlated with the Universal Multidimensional Abilities Scales (UMAS; McCallum & Bracken, 2012a) at moderate to very large levels (.52, .63, .65, and .72, respectively).

Support for UNIT2 construct validity was provided also by comparing means and standard deviations for different examinee age groups (i.e., growth curves), comparing the performance of different groups to the normative sample, correlating the UNIT2 with measures of academic achievement, and conducting factor analyses to compare subtests to the constructs inherent in the UNIT2 model. The UNIT2 shows a strong relationship with age, and expected mean differences between various groups (individuals with low IQ, high IQ, autism spectrum disorder, language disorders, etc.), suggesting that the UNIT2 is an effective and fair tool for assessing various populations of students. In addition, validity studies described in the examiner's manual show relationships between the UNIT2 and various achievement tests across several populations, ranging typically from .54 to .79, with a few exceptions below or above this range.

Because the UNIT2 is based on a specific theoretical model, confirmatory factor analysis (CFA) was used to assess the degree of fit with the proposed model. As described in the UNIT2 examiner's manual, the structural validity of the UNIT2 was empirically investigated by contrasting four CFA models across five age ranges (ages 5–7 years, 8–10 years, 11–13 years, 14–17 years, and 18–21 years) and the total sample, using maximum-likelihood CFA. The four models examined included a one-factor model, a two-factor Reasoning × Memory model, a two-factor Reasoning × Quantitative model, and a three-factor Reasoning × Memory × Quantitative model. The results for these models were assessed by using multiple indices of fit: (1) Wheaton, Muthén, Alwin, and Summers (1977) relative chi-square (chi-square divided by degrees of freedom); (2) Tucker and Lewis's (1973) index of fit; (3) Bentler's (1990) comparative fit index; and (4) Browne and Cudeck's (1993) root mean square error of approximation. All four models fit the UNIT2 reasonably well, which supports interpreting the test as a measure of general ability, as well as interpreting its various Standard and Abbreviated Battery options. This factor analysis also supports an acceptable model at each age range examined, as well as the organization of subtests to scales on the UNIT2 (Bracken & McCallum, 2016b).

Fairness

The burden of ensuring fairness in testing is particularly salient for authors of nonverbal tests, in part because of an increasingly diverse society and the need to ensure sensitive and equally valid assessment for a wide variety of populations. The UNIT2 manual includes an entire chapter entitled and dedicated to "Fairness," which describes extensive efforts to ensure that the test is appropriate for use with all children in the United States (i.e., that construct-irrelevant variance is minimized for all relevant populations).

The UNIT and UNIT2 were formulated, and the tests were developed, on the basis of five core fairness concepts: (1) A language-free test is less susceptible to bias than a language-loaded test; (2) a multidimensional measure of cognition is fairer than a unidimensional one; (3) a test that minimizes the influence of acquired knowledge (i.e., crystallized ability) is fairer than one that does not; (4) a test that minimizes speeded performance is fairer than one with greater emphasis on speed; and (5) a test that relies on a variety of response modes is more motivating and thereby fairer than one relying on a unidimensional response mode. Several other steps were taken to ensure fairness. For example, in the initial item development phase, items were submitted to a panel of "bias experts"—individuals sensitive to inclusion of items that might be offensive to or more difficult for individuals within certain populations (e.g., Native Americans, Hispanics). Items identified by these individuals and those identified via statistical item bias analyses were removed. In addition, a number of statistical procedures were undertaken to help ensure fairness, including calculation of separate reliabilities, factor structure statistics, mean-difference analyses, and so forth for subpopulations, as described within the UNIT2 manual's "Fairness" chapter (e.g., presentation of reliability and internal and external validity data for several populations of interest, such as African Americans, Hispanic Americans, Native Americans, Asian Americans, and individuals with hearing impairments; reliability coefficients calculated separately for diverse groups for the Full Scale Battery composites; etc.). Importantly, reliability coefficients were .97 or above across genders, 6 different ethnicities, and 12 different exceptionalities. Of interest to many users of nonverbal tests are mean score differences between minority samples and matched nonminority samples. For example, the median score differences between a sample of 224 black/African American students and a matched sample of white/European American examinees drawn from the standardization sample was 1.56 for the subtests and 10.40 for composite scores. The median score differences for a sample of 215 Hispanic examinees and matched controls were 0.78 for the subtests and 4.95 for the composites. Of note, there have been several reviews addressing fairness of the original UNIT, and these have been very positive in general (e.g., Bandalos, 2001; Braden & Athanasiou, 2005; Fives & Flanagan, 2002; Sattler, 2008). These studies have relevance for establishing the UNIT2's fairness, given the theoretical and statistical evidence showing overlap between the two instruments.

As is apparent, considerable effort was expended to establish fairness for the populations of interest to users of nonverbal tests. Readers interested in additional information about UNIT2 assessment fairness and equity may want to explore the many studies described in the UNIT2 examiner's manual.

INTERPRETING THE UNIT2

As we noted in an earlier version of this chapter (McCallum & Bracken, 2012b), multidimensional test interpretation is complicated, in part because it requires examiners to engage in a number of steps, consult numerous tables, consider a variety of cognitive models, consider carefully the limitations of the instruments they use, and (finally and most importantly) make the test results relevant for real-world application. There are three basic interpretative strategies: *normative, base rate,* and *ipsative*. Although normative interpretation is the most common strategy for determining examinee strengths and weaknesses, it is possible to interpret performance ipsatively as well. Normative interpretation relies on comparing an examinee's standard scores to those of peers, using age-based or grade-based normative data. By contrast, ipsative interpretation relies on intraindividual comparisons of scores (e.g., comparing each subtest and/or composite scores to the examinee's subtest and/or composite averages). Subtest/composite scores that deviate from the examinee's means to a greater extent than would be expected by chance are considered personal strengths or weaknesses. Finally, base rate interpretation relies on determining how rare difference scores are in the population, as defined by the test standardization sample. Most test authors show base rates of subtest pairwise difference scores and base rates associated with differences between particular subtest scores and subtest averages. Some experts assume that difference scores occurring within the population no more frequently than 15% of the time should be considered rare; others are of the opinion that difference scores so large that they occur in the population only 10% of the time should be considered rare. Interpretation based on these strategies has been made somewhat more user-friendly recently because of computer-based scoring and interpretation software. But we caution that unquestioning reliance on these software solutions can lead to misleading conclusions, and requires that examiners use the computer-generated data as a starting point only. In the next section, we describe these strategies in more detail within the context of UNIT2 scores.

As described previously, the UNIT2 features four battery options: the Abbreviated (twosubtest) Battery, the Standard (foursubtest) Battery with Memory, the Standard (four-subtest) Battery without Memory, and the Full Scale (sixsubtest) Battery. The various batteries were designed to assess memory, reasoning, and quantitative reasoning, using both symbolic and nonsymbolic task demands. Interpretation of the UNIT2 begins with the examiner's consideration of which battery should be administered. Making a choice among the four batteries depends on several issues, including the purpose of the assessment (e.g., screening, diagnostic testing, placement), the estimated attention span of the student, the time available to conduct the assessment, and related concerns. Once the choice of batteries has been made and the UNIT2 has been administered, actual test interpretation is conducted in multiple steps that consider data successively, from the most global and reliable sources (e.g., FSIQ, scale scores) to increasingly specific yet less reliable sources (e.g., subtests, items).

UNIT2 results are interpreted from both inter- and intrachild (normative and ipsative, respectively) perspectives. Both procedures have been employed by a variety of authors over the years and have become commonplace for the interpretation of psychoeducational tests (Bracken, 1984, 1992, 1993, 1998, 2006a, 2006b; Bracken & McCallum, 1998; Kaufman, 1979; Kaufman & Kaufman, 1983; Kaufman & Lichtenberger, 1999; McCallum, 1991; Sattler, 1988, 1992). The following discussion for interpreting the UNIT2 flows from principles described by these experts and reflects specific strategies and guidelines outlined in the UNIT2 manual (Bracken & McCallum, 2016b).

Specific Interpretation Guidelines

Traditional normative and ipsative interpretation should proceed from the most comprehensive and reliable scores to the most specific, least reliable scores. Test composites (e.g., FSIQs and scale scores) tend to be the most reliable scores because they include sources of variation from all of the subtests and scales that comprise the test. As such, these molar data are more reliable than the more molecular scores from individual subtests. Composite cognitive ability scores also are the best predictors of important "real-world" outcomes, particularly academic and vocational success (Sattler, 2008). Consequently, the most defensible interpretive strategy is to initially address the overall composite score and stop the interpretive process. However, whenever there is considerable variability among examinees' performance across individual subtests in a battery, the overall composite is not an ideal reflection of an examinee's true *overall* ability. When significant subtest and scale variation occurs, further interpretation of the test is warranted (Kaufman, 1979; Kaufman & Lichtenberger, 1999). Therefore, the UNIT2 manual presents the following three-step sequence for interpreting results: (1) Interpret the global intelligence score; (2) interpret the construct-specific scores; and (3) interpret subtest performance, including pairwise subtest comparisons and ipsative subtest comparisons. These three interpretation steps are described below.

Step 1: Interpret the Global Intelligence Score

First, an examiner should describe an examinee's performance at the composite level on the Abbreviated, Standard, or Full Scale Battery composites both quantitatively (e.g., standard scores, confidence intervals, percentile ranks) and qualitatively (e.g., descriptive classifications), and within their bands of error (i.e., confidence interval). Quantitative descriptions are based on interpretation of obtained scores relative to population parameters. Scores on the UNIT2 conform to the traditional normal "bell curve," and UNIT2 standard scores can be compared to global scores on other tests using the same metric (i.e., M = 100, *SD* = 15), such as the various Wechsler scales and the Woodcock–Johnson cognitive and achievement batteries (e.g., see Schrank et al., 2014).

Score variability comes from two sources: reliable variance (shared and specific), and error variance. Because random error is normally distributed, obtained scores should be considered within a band of confidence that frames the obtained score by one or more standard error(s) of measurement (*SEM*), as determined by the level of confidence desired (e.g., 68%, 95%, 99%). Confidence intervals built around obtained scores define the probability that a given range of scores would include the examinee's "true" score with a given level of confidence. In addition to the *SEM*, the UNIT2 also reports bands of error associated with the "estimated true score," which takes into account regression toward the mean. As such, bands determining estimated true scores become more elliptical as scores move toward the extremes. The UNIT2 band of error (standard error of the estimate) can be found in Table 6.2 in the examiner's manual. Finally, qualitative descriptions can be used to describe levels of examinee functioning, using classifications provided in the UNIT2 examiner's manual. Qualitative classifications for the UNIT2 range from *very superior* to *very delayed*.

Step 2: Interpret the Construct-Specific Scores

The next step of UNIT2 interpretation focuses on variability between the Memory, Reasoning, and Quantitative composites as they contribute to the estimate of overall cognitive functioning. If scores on these scales produce significant variability (i.e., significant differences between themselves), the global intelligence scale will serve as a limited estimate of the examinee's global ability, and performance on the UNIT2 construct-specific composites should be interpreted.

The Memory, Reasoning, and Quantitative composites should be described both quantitatively and qualitatively. These scores should be examined for statistically significant and meaningful differences between each other. If a difference between two scales is statistically significant (i.e., so large that it would not be likely to occur by chance), such a difference should be considered important, at least initially. As suggested by Kaufman and Lichtenberger (1999), a probability level of .05 is recommended to determine statistical significance; however, significant differences are not necessarily clinically meaningful or rare. If significant differences exist, their rarity within the general population should be considered (see Step 3). Tables 6.1 and 6.2 in Appendix E of the examiner's manual present scale deviation values considered significant.

Step 3: Interpret Subtest Performance

As authors of the UNIT2, we recommend interpretation at the most global level possible; however, significant, meaningful variability between scales should lead the examiner to consider individual subtest variability. Section 7 of the Examiner Record Form provides space to calculate normative and ipsative pairwise subtest comparisons. With the ipsative approach, a statistically significant difference necessitates further analysis. A subtest score that is significantly greater than the mean subtest score reflects a potential area of relative strength, while a subtest score that is significantly lower than the mean score reflects a potential area of relative weakness. The abilities associated with individual subtests should be used to generate hypotheses or possible explanations for individual subtest variations. Cautious interpretation of differences between subtests is recommended because their reliabilities, while robust, are lower than those for the composites. Item response patterns within a subtest can also be examined for clues about specific areas of ability or challenges.

The UNIT2 manual provides hypotheses describing examinees with particular strengths or weaknesses on the global scale scores. For example, examinees who have stronger memory (than reasoning) may reproduce visual stimuli better than they can solve problems based on the recall of stimulus juxtapositions and relationships. Table 4.11 in the UNIT2 manual shows hypotheses related to scale variations. Table 4.12 in the UNIT2 manual also presents the primary and secondary abilities assessed by each UNIT2 subtest, to assist with interpretation of strengths and weaknesses. Finally, the UNIT2 examiner's manual provides two case studies to help examiners learn to interpret UNIT2 scores.

Sound test interpretation can be conducted only when tests possess reasonably good psychometric properties. Several authors (e.g., Bracken, 1987; Bracken & McCallum, 1998; Wasserman & Bracken, 2013) have recommended basic rule-of-thumb criteria for acceptable psychometric characteristics. For example, global scores used for making placement decisions should evidence reliability at a level of .90 or better; scores used for screening purposes should have reliability at a level of .80 or better. Also, subtest and scale floors, ceilings, and item gradients should be sufficiently sensitive to capture small differences in actual ability, at a range of ±2 standard deviations. In addition, subtest specificity must meet commonly accepted criteria before subtests can be interpreted as measures of unique abilities or skills, in addition to their contribution to their scale or the total test score. That is, even though subtests within an instrument contribute to the measurement of general cognitive ability, each subtest may be a reasonably good measure of some specific cognitive skill or ability.

Finally, intelligence test scores should not be used in isolation. Critics of subtest interpretation (e.g., McDermott, Fantuzzo, & Glutting, 1990) typically have not considered the clinical value of subtest analysis when it is employed as one aspect of data analysis that may be confirmed or refuted through other data sources (i.e., triangulation of data). Thus UNIT2 subtest analysis should be conducted to generate hypotheses about a child's unique intellectual strengths and weaknesses, and should *never* be used without additional extratest information that will allow the examiner to further evaluate the hypotheses that are generated. UNIT2 interpretation is facilitated by using the worksheets and tables printed in the test booklet. The booklet also provides a graph for charting an examinee's subtest and composite profile, as well as a checklist for indicating test validity and related notes and observations (e.g., comments about the examinee's physical appearance, mood, activity level, attention/concentration, visual–motor skills, problem-solving attack skills, language use), and for providing an assessment of the testing situation. We present a case in an appendix to this chapter as an illustration of how the UNIT2 can be used.

ADDITIONAL CLINICAL APPLICATIONS

The UNIT2 can be administered to non-English-speaking populations easily, without the traditional language demands of conventional intelligence tests or costly translations. Although no gestures are completely universal, the gestures chosen for use in the UNIT and UNIT2 (e.g., affirmative head nods and pointing) provide ubiquitous modes of communication across most cultures. Also, an effort was made to employ universal item content (i.e., objects found in all industrialized cultures). Thus use of the UNIT2 with children who come into the United States from other countries is facilitated. In addition, the format is appropriate for children with deafness/hearing impairments and for those who have other types of language deficits (e.g., selective mutism, severe dyslexia, speech articulation difficulties).

A final clinical application of UNIT2 results is described by Wilhoit (2017). Specifically, Wilhoit builds on the cross-battery assessment (XBA) procedures described by McGrew and Flanagan (1998) and revised/updated by Flanagan, Ortiz, and Alfonso (2007, 2013) to describe application of XBA to a number of nonverbal instruments.

One important assumption of XBA is that subtests can be selected from different batteries and used to assess particular cognitive constructs, thereby increasing assessment precision and efficiency. This technique is particularly useful when there is no need to administer and interpret a particular test in its entirety (e.g., the referral question does not require that an FSIQ be obtained from a specific cognitive test). To aid in the application of XBA, McGrew and Flanagan provided a cognitive nomenclature based on the work of several researchers, particularly Cattell (1963), Horn (1968, 1994), and Carroll (1993). This nomenclature—referred to in recent years as the CHC system or model—is embedded in a three-tier hierarchical model. Stratum III represents *g*, the general cognitive energy presumed to underlie performance across all tasks individuals undertake. Stratum II represents relatively broad abilities that can be operationalized fairly well as factors from a factor analysis (e.g., short-term memory, long-term memory, fluid ability, acquired knowledge, visual processing, auditory processing, and processing speed). Stratum I represents abilities at a more specific level, and can be assessed relatively purely by many existing subtests; two or more of these subtests can be used to operationalize stratum II abilities. Using this system, McGrew and Flanagan (1998) characterized subtests from most existing batteries as measures of stratum I and stratum II abilities, and provided several caveats about the use of these operationalizations. Application of XBA is somewhat detailed; it requires the use of worksheets containing the names of tests and subtests, and the broad stratum II and stratum III abilities those subtests measure. Using the worksheets, an examiner can determine strengths and weaknesses according to operationalization of the CHC model by subtests from various nonverbal measures. Assessment of stratum II abilities is the primary focus. Typically, each subtest from nonverbal tests assesses a narrow stratum I ability, and two or more can be used to provide a good assessment of stratum II. The six stratum II abilities assessed by nonverbal tests include fluid intelligence (Gf), crystallized intelligence (Gc), visual processing (Gv), short-term memory (Gsm), long-term memory (Glr), and processing speed (Gs). The other ability typically included in XBA, auditory processing (Ga), is not assessed by nonverbal tests and is not included on the worksheets. The XBA worksheets allow an examiner to calculate the mean performance by averaging scores from all subtests. Each stratum II ability score (determined by averaging two or more stratum I measures within that stratum II ability) can be compared to the overall stratum II average in an ipsative fashion. Assuming that all subtests use a mean of 100 and a standard deviation of 15 (or have been converted accordingly), each average stratum II ability score that is more than 15 points from the overall mean may be considered a strength or a weakness, depending on the direction of the difference. Wilhoit (2017) provides the worksheets for this application of the UNIT2 (and other nonverbal tests). Importantly, the stratum II abilities have been linked to several important real-world products (e.g., processing speed and short-term memory underpin the ability to learn to decode words quickly, according to Mather & Jaffe, 2002, 2011). Consequently, using XBA can aid in diagnosing academic and other problems.

In summary, the UNIT2 offers a more comprehensive assessment of intelligence than the original UNIT. The two newly added subtests expand the UNIT's representation of the CHC model, by adding a quantitative factor. Thus the UNIT2 addresses more of the dimensions included in the CHC model, but does so in a completely nonverbal manner, which allows practitioners a sound substitute for existing verbally loaded CHC assessment procedures.

INNOVATIONS IN THE MEASUREMENT OF COGNITIVE ABILITIES

Several features set the UNIT2 apart from all or most existing nonverbal scales.

1. The UNIT2 is administered solely through the use of examiner demonstrations and gestures. The liberal use of sample, demonstration, and (unique) checkpoint items ensures that the examinee understands the nature of each task prior to attempting the subtest for credit.

2. The test comprises a variety of subtests that provide the opportunity for both motoric and motor-reduced (i.e., pointing) responses. Admin-

istration of UNIT subtests can be modified so that only a pointing response is required on five of the six subtests. The use of motoric and motor-reduced subtests facilitates administration by optimizing motivation and rapport. For example, a very shy child may be encouraged initially to point only; later, as rapport is gained, other, more motorically involved responses may be possible. Also, use of the motor-reduced subtests may be indicated for children with limited motor skills.

3. Subtests contain items that are as culturally fair as possible. We have included colorized line drawings and objects that are recognizable to most individuals from all cultures.

4. The test is model-based. That is, we have included subtests designed to assess reasoning (a higher-order mental processing activity), as well as complex memory and quantitative thinking. Also, we have included symbolically loaded subtests as well as less symbolically laden ones. Interpretation of the UNIT is facilitated because of these theoretical underpinnings.

5. The UNIT2 manual contains results of multiple studies comparing the UNIT2 to other tests and showing how important populations function on the test relative to individuals from the mainstream culture (African Americans, Native Americans, Asian Americans, Hispanic Americans, individuals who are deaf or hard of hearing, or those whose primarily language is not English).

6. Administration time can be controlled by the examiner, depending on the number of subtests administered. The UNIT2 includes three administration formats: a two-subtest version, a four-subtest (standard) version, and a six-subtest (extended) version.

7. Reliability estimates were calculated for two critical cut points (i.e., for those with FSIQs around 70 and those with FSIQs around 130).

8. An unprecedented array of support resources were created for the original UNIT, including a training video, a university training manual, and a computerized scoring and interpretation software program. Given the similarities between the UNIT and the UNIT2, these resources will help novice examiners become familiar with the UNIT2. Of course, the computerized scoring system has been revised and updated on the basis of the new standardization data.

9. No test is completely free of the influence of culture and language, and the UNIT2 is no exception. Any test completely devoid of these influences is unlikely to predict important outcomes. However, the UNIT2 was designed to minimize language and cultural influences on test scores, given the goal of assessing those with language deficits and those from culturally different environments. The UNIT2 manual provides the authors information regarding the extent of cultural loading and linguistic demands on each subtest, based on the model described by Flanagan and colleagues (2007, 2013). Most subtests reflect low cultural and linguistic demands, as expected.

10. Not all experts agree that a nonverbal test provides an optimal assessment of at-risk populations—that is, those with language deficits and those from culturally different environments (e.g., see the discussion in Flanagan et al., 2013, Chapter 5). However, most experts note the potential advantages of using such tests for individuals in certain at-risk populations (e.g., see the chapters by Braden; Frisby; Green, Bardos, & Doropoulou; Jaquett & Kirkpatrick; Johnsen; Moore, McCallum, & Bracken; Roid & Koch; and Wasserman [two chapters] in the 2017 *Handbook of Nonverbal Assessment*).

11. Examiners who are philosophically opposed to using a nonverbal test that relies on memory have an option: They can now use the four-subtest Standard Battery without Memory within the UNIT2. This objection, however, applied minimally to the original UNIT because its memory subtests assessed complex memory and each subtest had a strong *g* loading (above .60).

Appendix 20.1 includes a case study that demonstrates how the UNIT2 and four other test batteries can be integrated into a comprehensive evaluation. All personally identifiable information has been altered in the case study report to preserve the confidentiality of the examinee.

APPENDIX 20.1

Brief Case Study

Name: Ryan Daniel Ross

Age: 7 years, 6 months

Date of birth: 02/14/09 Grade: Entering 1

School: Knox Elementary

Date(s) of assessment: 07/14/16; 07/30/16

Examiner: Sherry Mee Bell, PhD, NCSP,
Licensed Psychologist

REASON FOR REFERRAL AND BACKGROUND INFORMATION

Ryan was referred to determine his current cognitive, academic, and social-emotional functioning and to obtain information to facilitate his educational planning. He lives with his parents, who provided background information, and an older sister, Lauren. Both parents completed a 4-year college degree and are employed full-time; English is the only language spoken in the home. Ryan's parents reported a healthy pregnancy and delivery. Ryan's birth weight was within normal limits, although his developmental milestones were mildly delayed. Ryan walked independently at age 1 year, 5 months (1:5), and talked at about age 3; his speech and language development was significantly delayed. Ryan was toilet-trained (at about 3 years, bladder, and between ages 4 and 5, bowels). His tonsils and adenoids were removed at age 4:6, providing relief of chronic middle-ear infections. Ryan's vision and hearing have been tested within the last year, and his auditory and visual acuity are within normal limits. Family history is generally negative for psychiatric and learning problems, although two first cousins (maternal side) have received speech and language services in school.

Due to apparent delays in speech and language skills, Ryan was evaluated at the University Hearing and Speech Center at age 4, during the summer of 2013. According to this assessment, Ryan exhibited a communication disorder characterized by delayed receptive and expressive language skills, as well as irregularities and delay in his development of articulation skills. He exhibited difficulty naming and identifying objects and following directions; in addition, he exhibited instances of echolalia. Ryan's spontaneous speech was characterized by strings of unintelligible, reduplicative utterances, interspersed with some intelligible words. Based on the speech–language evaluation, Ryan was deemed eligible for and has been receiving special education services through the local school system (because of developmental delays) since August 2013.

ASSESSMENT INSTRUMENTS

Because of Ryan's documented language delays, a combination of language-loaded and non-language-loaded tests was administered, in part to rule in or out disorders often associated with significant language delays (e.g., autism, intellectual disability). These included the Universal Nonverbal Intelligence Test—Second Edition (UNIT2), a nonverbal, multidimensional assessment of cognitive abilities; the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), a measure of receptive language skills; the Expressive Vocabulary Test, Second Edition (EVT-2), a measure of expressive language skills; selected subtests from the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH), a battery assessing academic achievement; and the Clinical Assessment of Behavior (CAB), a comprehensive assessment of social, emotional, and behavioral functioning.

RELEVANT TEST OBSERVATIONS AND BEHAVIORS

Ryan is short for his chronological age and is slender. He presented as somewhat shy, but he separated from his parents upon request. Ryan seemed to put forth good effort during both assessment sessions and responded well to praise and encouragement. However, he had difficulty following oral directions at times, and some instructions had to be repeated or rephrased. Occasionally he whispered his answers, especially when he seemed unsure of himself. As a result of his cooperative behavior, the test results are considered to represent a valid estimate of Ryan's current level of functioning.

ASSESSMENT TEST RESULTS

On the UNIT2, Ryan achieved a Memory index score of 100, a Reasoning index score of 94, a Quantitative index score of 103, and a Full Scale Battery score of 99 (47th percentile). The range of scores from 96 to 102 captures Ryan's true Full Scale IQ (FSIQ) with 90% confidence. The composite indices of the UNIT2 have a mean of 100 and a standard deviation of 15, consistent with mainstream intelligence tests. The UNIT2 subtests have a mean of 10 and a standard deviation of 3; Ryan's subtest scores were as follows: Symbolic Memory, 9; Nonsymbolic Quantity, 11; Analogic Reasoning, 9; Spatial Memory, 11; Numerical Series, 10; and Cube Design, 9.

Ryan displayed relatively little variability on these nonverbal cognitive tasks. He performed somewhat more strongly (a little more than half a standard deviation higher) on quantitative versus reasoning tasks, and his memory scores were solidly average. The relative strength in quantitative versus reasoning abilities is consistent with Ryan's deficits in language, given that language comprehension and expression require reasoning and problem solving. Ryan's results on the UNIT2 indicate overall average cognitive abilities, which are significantly stronger than previous estimates of his language skills.

Ryan was administered the PPVT-4, Form A. Results yielded a standard score of 75, which is ranked at the 5th percentile and yields an age equivalent of 4:9. Similarly, on the EVT-2 (Form A), Ryan achieved a standard score of 78, which is ranked at the 7th percentile and yields an age-equivalent of 5:6. Results indicated that Ryan's receptive and expressive language skills continue to be somewhat delayed relative to those of his peers and to his overall cognitive abilities (approximately 2 to 2½ years below his chronological age). Furthermore, his receptive and expressive skills are relatively commensurate, indicating generally even, though delayed, language development.

Ryan was administered several subtests from the WJ IV ACH. These tests have a population mean of 100 and standard deviation of 15. Two measures each of reading, math, and writing were administered. The Word Attack subtest was not administered, due to Ryan's significant speech articulation irregularities. Age equivalents (AE), standard scores (SS), and standard score confidence bands are reported in Table 20.A.1.

TABLE 20.A.1. Ryan's WJ IV ACH (Form A and Extended) Results

CLUSTER/subtest	AE	SS (68% band)
READING	5:4	58 (55–61)
MATHEMATICS	6:9	90 (87–92)
WRITTEN LANGUAGE	6:3	80 (77–84)
ACADEMIC SKILLS	5:9	69 (67–71)
ACADEMIC APPLICATIONS	6:3	79 (76–82)
BRIEF ACHIEVEMENT	5:9	69 (66–71)
Letter–Word Identification	5:1	55 (52–59)
Applied Problems	7:0	92 (87–98)
Spelling	5:9	72 (67–76)
Passage Comprehension	5:7	64 (60–68)
Calculation	6:7	87 (84–90)
Writing Samples	6:7	88 (83–93)

Note. Norms based on age 7:6 were used. Composite (cluster) scores are reported in ALL CAPS.

The WJ IV ACH scores were based on age norms to allow direct comparisons with the other assessments administered (i.e., the UNIT2, PPVT-4, and EVT-2). However, it should be noted that Ryan spent an additional year in kindergarten (following participation in special education preschool) and has not been exposed to the first-grade curricula experienced by typical 7-year-olds. Nonetheless, these scores are consistent with Ryan's identified language delays. Ryan's reading (letter–word recognition and beginning reading comprehension) and spelling are significantly weaker than age expectations. His beginning writing skills are also low average. These areas of achievement are consistent with his performance on language tests, but much weaker than his overall cognitive abilities as measured by the UNIT2. His math skills (math reasoning and calculations) are in the average to low average range and more consistent with his UNIT2 quantitative thinking performance.

Ryan's mother completed the CAB, Extended Form. Her ratings yielded typical scores on the Respondent Veracity Scales, which support the validity of her responses, and her ratings yielded scores generally in the average range. CAB scores have a population mean of 50 and standard deviation of 10. Ryan's overall CAB Behavior Index was 54 (64th percentile, in the average range), indicating no significant global emotional or behavioral difficulties. Ryan's scores on the Clinical Scales were all in the average range: Internalizing Behaviors (59, 80th percentile); Externalizing Behaviors (43, 24th percentile); and Critical Behaviors (43, 25th percentile). His scores on the Adaptive Scales were as follows: Social Skills (42, 21st percentile); Competence (39, 14th percentile); and Adaptive Behaviors (45, 27th percentile). These results indicate a mild adaptive weakness on the Competence scale, which includes questions on language competence. Ryan's cluster score for Anxiety on the CAB was 60 (84th percentile, indicating mild clinical risk); his cluster scores for Depression, Anger, Aggression, Bullying, Conduct Problems, Attention Deficit/Hyperactivity, Autistic Spectrum Behaviors, Learning Disability, and Intellectual Disability were all in the average range. Furthermore, scores on the Executive Functioning scale and the Gifted and Talented scale were in the average range. In general, these scores suggest that Ryan's mother perceives that her son exhibits no significant emotional or behavioral difficulties. He has developed a tendency to be slightly anxious and withdrawn; this is likely related to his language delays, which impede communication. Ryan's CAB results do not suggest the presence of behaviors consistent with global cognitive delay (intellectual disability) or autism.

SUMMARY AND RECOMMENDATIONS

Assessment results tentatively suggest a *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5; American Psychiatric Association, 2013) diagnosis of language disorder, with both receptive and expressive delays (F80.2). Ryan is being referred for a medical examination to assist in determining the exact nature of his developmental delay and/or to rule out any organic or medical etiology.

Results of both the current assessment and the medical exam should be shared with school personnel to aid in educational planning for Ryan. At school, participation in a regular classroom with special educational support is recommended. Ryan decidedly appears eligible for intensive direct and consultative speech and language services. In addition, he is likely to need support from the school resource teacher, in either an inclusion or pull-out format, to make adequate progress on academic assignments. Modifications and adaptations in Ryan's assignments will be needed. A multisensory approach (particularly heavy use of visual, pictorial, and graphic representations of content to be learned) should be beneficial for Ryan. Grading modifications will be needed, and teachers are encouraged to conduct error analysis (with Ryan) to determine which kinds of tasks give him more difficulty. In the classroom, Ryan may benefit from being paired with a "study buddy" who can prompt Ryan on how to complete tasks and assignments. Ryan is at risk for developing a specific learning disability in reading, related to his language delays. At school and at home, he needs frequent exposure to children's literature and other content that is motivating and that will support development of vocabulary and store of knowledge. Ryan's educational and developmental progress should be monitored routinely.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Athanasiou, M. S. (2000). Current nonverbal assessment instruments: A comparison of psychometric integrity and test fairness. *Journal of Psychoeducational Assessment, 18,* 211–299.

Bandalos, D. L. (2001). Review of the Universal Nonverbal Intelligence Test. In B. S. Plake & J. C. Impara (Eds.), *Fourteenth mental measurements yearbook* (pp. 1296–1298). Lincoln, NE: Buros Institute.

Bentler, P. M. (1990). Comparative fit index in structural models. *Psychological Bulletin, 107,* 238–246.

Bracken, B. A. (1984). *Bracken Basic Concept Scale.* New York: Psychological Corporation.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5,* 313–326.

Bracken, B. A. (1992). *Multidimensional Self Concept Scale.* Austin, TX: PRO-ED.

Bracken, B. A. (1993). *Assessment of Interpersonal Relations.* Austin, TX: PRO-ED.

Bracken, B. A. (1998). *Bracken Basic Concept Scale—Revised.* San Antonio, TX: Psychological Corporation.

Bracken, B. A. (2006a). *Bracken Basic Concept Scale: Expressive.* San Antonio, TX: Psychological Corporation.

Bracken, B. A. (2006b). *Bracken Basic Concept Scale: Receptive.* San Antonio, TX: Psychological Corporation.

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test.* Itasca, IL: Riverside.

Bracken, B. A., & McCallum, R. S. (2016a). *UNIT2 Compuscore.* Austin, TX: PRO-ED.

Bracken, B. A., & McCallum, R. S. (2016b). *Universal Nonverbal Intelligence Test—Second Edition.* Austin, TX: PRO-ED.

Braden, J. P. (2017). Best practices in assessing those who are deaf or hard-of-hearing. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 47–58). New York: Springer.

Braden, J. P., & Athanasiou, M. S. (2005). A comparative review of nonverbal measures of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 557–577). New York: Guilford Press.

Brinton, B., Spackman, M. P., Fujiki, M., & Ricks, J. (2007). What should Chris say?: The ability of children with specific language impairment to recognize the need to dissemble emotions in social situations. *Journal of Speech, Language, and Hearing Research, 50,* 798–811.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: SAGE.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Fives, C. J., & Flanagan, R. (2002). A review of the Universal Nonverbal Intelligence Test (UNIT): An advance for evaluating youngsters with diverse needs. *School Psychology International, 23,* 425–448.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Frisby, C. L. (2017). The nonverbal assessment of academic skills. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 251–268). New York: Springer.

Fujiki, M., Spackman, M. P., Brinton, B., & Illig, T. (2008). Ability of children with language impairment to understand emotion conveyed by prosody in a narrative passage. *International Journal of Language and Communication Disorders, 43*(3), 330–345.

Green, A., Bardos, A., & Doropoulou, M. (2017). The General Ability Measure for Adults. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 207–216). New York: Springer.

Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (2009). *Comprehensive Test of Nonverbal Intelligence—Second Edition.* Austin, TX: PRO-ED.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242–259.

Horn, J. L. (1974). Theory of function represented among auditory and visual test performances. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory.* New York: Academic Press.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: Macmillan.

Hughes, J. N., & Zhang, D. (2007). Effects of the structure of classmates' perceptions of peer's academic abilities on children's perceived cognitive competence, peer acceptance, and engagement. *Contemporary Educational Psychology, 32,* 400–419.

Jaquett, C. M., & Kirkpatrick, B. A. (2017). Wechsler Nonverbal Scale of Ability. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 151–166). New York: Springer.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Johnsen, S. K. (2017). Test of Nonverbal Intelligence: A language-free measure of cognitive ability. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 185–206). New York: Springer.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children: Administration and scoring manual.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: Wiley.

Lienemann, T. O., Graham, S., Leader-Janssen, B., & Reid, R. (2006). Improving the writing performance of struggling writers in second grade. *Journal of Special Education, 40,* 66–78.

Liew, J., McTigue, E. M., Barrois, L., & Hughes, J. N. (2008). Adaptive and effortful control and academic self-efficacy beliefs on achievement: A longitudinal study of 1st through 3rd graders. *Early Childhood Research Quarterly, 23,* 515–526.

Maller, S. J. (2000). Item invariance in four subtests of the Universal Nonverbal Intelligence Test (UNIT) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18,* 240–254.

Mather, N., & Jaffe, L. E. (2002). *Woodcock–Johnson III: Reports, recommendations, and strategies.* New York: Wiley.

Mather, N., & Jaffe, L. E. (2011). *Comprehensive evaluation: Case reports for psychologists, diagnosticians, and special educators.* Hoboken, NJ: Wiley.

McCallum, R. S. (1991). The assessment of preschool children with the Stanford–Binet Intelligence Scale: Fourth edition. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (2nd ed., pp. 107–132). Boston: Allyn & Bacon.

McCallum, R. S., & Bracken, B. A. (2012a). *Universal Multidimensional Abilities Scales.* Austin, TX: PRO-ED.

McCallum, R. S., & Bracken, B. A. (2012b). The Universal Nonverbal Intelligence Test: A multidimensional nonverbal alternative for cognitive assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 357–375). New York: Guilford Press.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8*(3), 290–302.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

Moore, A. F., McCallum, R. S., & Bracken, B. A. (2017). The Universal Nonverbal Intelligence Test: Second Edition. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 105–126). New York: Springer.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition.* Austin, TX: PRO-ED.

National Institutes of Health. (2010). *Hearing disorders and deafness.* Bethesda, MD: National Library of Medicine. Retrieved from *www.nlm.nih.gov/medlineplus/hearingdisordersanddeafness.html.*

Noland, R. M. (2009). When no bilingual examiner is available: Exploring the use of ancillary examiners as a viable testing solution. *Journal of Psychoeducational Assessment, 27,* 29–45.

Pasko, J. R. (1994). Chicago—don't miss it. *Communique, 23,* 2.

Pendarvis, E., & Wood, E. W. (2009). Eligibility of historically underrepresented students referred for gifted education in a rural school district: A case study. *Journal for the Education of the Gifted, 32*(4), 495–514.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Roid, G. H., & Koch, C. (2017). Leiter-3: Nonverbal cognitive and neuropsychological assessment. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 127–151). New York: Springer.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Author.

Sattler, J. M. (1992). Assessment of children (3rd ed., rev. and updated). San Diego, CA: Author.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Author.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–8.

U.S. Census Bureau. (2016, July 1). QuickFacts, United States: Population estimates, July 1, 2016. Retrieved from *www.census.gov/quickfacts/table/PST045215/00.*

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2016). Table 204.27. English language learner (ELL) students enrolled in public and elementary schools, by grade and home language: Selected years, 2008–09 through 2013–14. Retrieved from *http://nces.ed.gov/programs/digest/d15/tables/dt15_204.27.asp.*

Unz, R. (1997, October 19). Perspective on education: Bilingual is a damaging myth, a system that ensures failure is kept alive by the flow of federal dollars. A 1998 initiative would bring change. *Los Angeles Times,* Section M, p. 5.

Wasserman, J. D. (2017a). Nonverbal assessment of personality and psychopathology. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 231–250). New York: Springer.

Wasserman, J. D. (2017b). Nonverbal neuropsychological assessment. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 287–310). New York: Springer.

Wasserman, J. D., & Bracken, B. A. (2013). Fundamental psychometric considerations in assessment. In I. B. Weiner (Series Ed.) & J. R. Graham & J. A. Naglieri (Vol. Eds.), *Handbook of psychology: Vol. 10: Assessment psychology* (2nd ed., pp. 50–81). Hoboken, NJ: Wiley.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (pp. 84–136). San Francisco: Jossey-Bass.

Wilhoit, B. (2017). Best practices in cross-battery assessment of nonverbal cognitive ability. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 59–76). New York: Springer.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

PART IV

Relevance of Tests of Intelligence, Cognitive Abilities, and Neuropsycholgical Processes in Understanding Individual Differences

CHAPTER 21

Use of Intelligence Tests in the Identification of Giftedness

David E. McIntosh
Felicia A. Dixon
Eric E. Pierson

This chapter begins by providing a brief historical overview of giftedness, with an emphasis on identifying the major figures in the field. The theoretical formulation of intelligence and the conceptual links between intelligence and giftedness are also discussed from a historical perspective. The evolution of multitrait theories of intelligence, the impact of advances in psychometrics on gifted assessment, and the increase in the number of theory-based cognitive measures are examined. The refinement of theory of intelligence centering on the Cattell–Horn–Carroll (CHC) model is discussed.

A significant portion of the chapter focuses on the definition of *giftedness*. The distinction between the terms *gifted* and *talented* is made. The most recent definition of *gifted and talented* provided within the Every Student Succeeds Act of 2015 (ESSA) is presented and discussed, along with the broader and more inclusive definition offered by the National Association for Gifted Children (NAGC). Renzulli's (1978) multitrait definition, Tannenbaum's (1983) conception of giftedness as a psychosocial construct, Sternberg's implicit theory of giftedness (Sternberg & Zhang, 1995), and the differentiated model of giftedness and talent (Gagné, 2004) are reviewed. The implicit links among definitions, cognitive theory, and cognitive assessment are explored.

The issues of multiple intelligences versus many factors of intelligence are also considered. Specifically, Sternberg's triarchic theory of intelligence (Sternberg, Chapter 5, this volume; Sternberg & Williams, 2002) and Gardner's theory of multiple intelligences (Chen & Gardner, Chapter 4, this volume; Gardner, 1999) are reviewed. An extensive review of factor-analytic research is provided, exploring the multidimensional abilities of gifted children with commonly used measures of intelligence. Although factor-analytic research with gifted samples has for the most part supported Kaufman's (1975, 1979) two-factor model, there is a growing body of confirmatory-factor-analytic research studying more complex models of intelligence. The methodological and statistical issues contributing to differences across studies are also examined.

The process of gifted identification is then reviewed. Clark's (1997) and Borland's (1989) recommended approaches for identification are discussed. The importance of demonstrating links among a school's definition of giftedness, the identification process, and educational programming is considered. In addition, important issues to consider during the identification of cognitively gifted minority children are explored within the chapter. Alternative approaches are provided for identifying gifted children with learning disabilities.

A major portion of the chapter focuses on discussing special issues related to the intellectual assessment of gifted children. The importance of considering the theoretical differences among intelligence measures, and the importance of using theory-based measures, are both discussed. The need for a better understanding of the relationships between screening procedures and the final decision outcomes based on intelligence tests is emphasized. Implications of using a single composite IQ score and setting specific IQ "cutoff" scores when one is making identification decisions are examined. The chapter concludes with specific recommendations on the effective use of intelligence measures in the identification of giftedness.

HISTORICAL OVERVIEW OF INTELLIGENCE AND GIFTEDNESS

Although America's interest in giftedness has been magnified since the late 1800s, giftedness has been of general interest to virtually all societies in recorded history (Colangelo & Davis, 1997). However, until Francis Galton (1822–1911) established the conceptual link between intelligence and giftedness, there had been little research studying intellectual differences among humans (Clark, 1997). Using the work of his cousin Charles Darwin (1859) as a basis, Galton developed the theory of *fixed intelligence,* which essentially ignored the effects of the environment and emphasized the hereditary basis of intelligence. This theory of fixed intelligence dominated the literature for nearly half a century. Not until the mid-1950s, when research conducted by Jean Piaget, Maria Montessori, Beth Wellman, G. Stanley Hall, Arnold Gesell, and others was published, did researchers begin to question the fixed-intelligence model and begin to consider an interactive view of intelligence.

Along with the theoretical formulation of intelligence came the need to develop ways to assess intelligence. Interestingly, although Francis Galton was credited with developing the first intelligence test, Alfred Binet is more widely known for developing the first intelligence test in 1905, with the specific goal of differentially placing children in special education or regular classrooms. Binet has also been credited with establishing the concepts of *mental age* and *intelligence quotient* (IQ). As Colangelo and Davis (1997) deftly point out, it was Binet's concept of mental age that had implications for the identification of giftedness. Essentially, the concept of mental age implied that children demonstrate growth in intelligence; therefore, children may be behind, consistent with, or ahead of their peers intellectually (Colangelo & Davis, 1997). Consequently, some children identified will demonstrate advanced levels of intelligence.

In 1916, Lewis Terman published the famous Stanford–Binet Intelligence Scale, which has seen four revisions to date. Not only was Terman recognized for developing one of the most popular measures of intelligence, but he was also instrumental in one of the most significant longitudinal studies on giftedness of the 20th century, earning him distinction as the "father of gifted education" (Clark, 1997; Colangelo & Davis, 1997; Davis, Rimm, & Siegle, 2011). Soon after the development of the Stanford–Binet Intelligence Scale, the popularity of intelligence testing soared. In fact, it soared so greatly that many schools based educational placement decisions solely on IQ scores. Unfortunately, many schools continue to identify gifted students by using a single measure of cognitive ability. The benefits and limitations of such an approach are discussed later.

Another landmark event that was instrumental in the public's interest in giftedness occurred in 1957, when the Soviet Union launched the world's first human-made satellite, *Sputnik.* The fact that the Soviets had both the scientific and technological power to accomplish this feat was viewed by some Americans as a shocking defeat to U.S. education. The results were an increase in the focus on educating gifted children in more advanced classes, especially in science and mathematics, and a call to action for a "total talent mobilization" (Davis & Rimm, 1998). In the United States, *Sputnik* was a wake-up call; new programs and schools were designed for high-ability students, with the purpose of keeping ahead globally. Modern societal concerns about annual academic progress and "failing schools" echo these historic concerns.

Over the last 30 years, new theories of intelligence, advances in psychometrics, and advances in technology have resulted in a renewed focus on the identification of gifted individuals. The focus on neurobiological data and mental processes has resulted in the development of new theories of intelligence. Gardner (1983, 1993, 1999; see also Chen & Gardner, Chapter 4, this volume) has proposed a theory of *multiple intelligences,* which focuses on eight areas of intellect (with a ninth one, *existential,* proposed): *linguistic, musical, logical–mathematical, spatial, bodily–kinesthetic, interpersonal, intrapersonal,* and *naturalistic.* The significance of his theory lies not so much in the eight identified areas of intellect as in the underlying assumptions that form the basis of his theory. To be specific, Gardner has emphasized the neurobiological influences on intelligence and the importance of better understanding the interaction between genetics and environment in the development of intelligence. Likewise, the *triarchic theory* of intelligence developed by Sternberg (1985; see also Sternberg, Chapter 5, this volume) has focused on better understanding three kinds of mental processes related to giftedness: *analytic, synthetic,* and *practical.* The difficulty with these theories of intelligence has been in determining how to assess and apply the various constructs presented by the authors. However, there have been several attempts to apply Gardner's multiple-intelligences theory and Sternberg's triarchic theory within school settings, with varying results (Coleman & Cross, 2005).

There has also been a renewed interest in the use of Luria's theory of neuropsychology in the development of cognitive tests (Kaufman & Kaufman, 2004; Naglieri & Das, 1997). Although many psychologists and neuropsychologists find this theory clinically helpful in understanding patterns of deficits in individuals, it is important to recognize that Luria's theory was drawn from a tradition of analyzing individuals with head injury. Most often, people reflect upon Luria's work with individuals with head trauma. However, Luria (1987) also wrote and studied individuals with amazing talents and abilities in memory. Those working with gifted and talented students may want to consider possible applications of this theory in contrast to the dominant CHC theory.

Long before Gardner and Sternberg developed their theories, L. L. Thurstone (1938) proposed his theory of *primary mental abilities.* According to Thurstone's theory, seven primary intelligence factors or abilities were measured in tests of intelligence: (1) *word fluency,* the ability to think of a lot of words, given a specific stimulus; (2) *verbal comprehension,* the ability to derive meaning from words; (3) *number* or *numerical ability,* the ability involved in all arithmetic tasks; (4) *memory,* the ability to use simple or rote memory of new material, both in verbal and in pictorial form; (5) *induction,* the ability to examine verbal, numerical, or pictorial material and derive from it a generalization, rule, concept, or principle; (6) *spatial perception,* the ability to see objects in space and to visualize varying arrangements of those objects; and (7) *perceptual speed,* the ability to discern minute aspects of elements of pictures, letters, and words as rapidly as possible. Feldhusen (1998) has suggested that close examination of these seven abilities reveals some parallels to the popular multiple intelligences of Gardner in the types of abilities mentioned. Thurstone's theory of primary mental abilities proposed that students' achievement in school could best be understood by their relative amount of ability in these seven areas. The end of the 1990s and the beginning of the new century saw a consolidation of theory that resembles the model originally proposed by Thurstone. Specifically, this period saw the consolidation of Gf-Gc theory and the work of Carroll into the CHC model of intelligence (Schneider & McGrew, Chapter 3, this volume; Strauss, Sherman, & Spreen, 2006). Similar to Thurstone's model, CHC theory suggests that intelligence can be thought of as having seven primary ability areas: Gf (*fluid intelligence*), Gc (*crystallized intelligence*), Gsm (*short-term memory*), Gv (*visual processing*), Ga (*auditory processing*), Glr (*long-term storage and retrieval*), and Gs (*processing speed*).

Advances in psychometrics and technology have led to the development of cognitive measures that are theory-driven and reflect the multitrait nature of intelligence. In the past, cognitive measures were developed with little or no consideration of intelligence theory. Consequently, there was an overreliance on the unitary construct of intelligence, *g* (Flanagan & Ortiz, 2002). Following the convergence of theories surrounding the development of the CHC theory, many of the most commonly administered tests related to gifted assessments in the past (the Stanford–Binet and the Wechsler series) have been aligned with CHC theory by the test publishers (O'Donnell, 2009; Strauss et al., 2006; Warne, 2016). These are discussed in more detail later.

While current trends in test publishing and research have moved toward an increased emphasis on instruments built around theories of intelligence, it is important to keep in mind that earlier measures used with the assessment of giftedness had as their primary goal the classification of individuals along a unitary dimension of intelligence. As a result of changes in how estimates of IQ are obtained and in the nature of the tests, a child who was classified as gifted in the 1930s–1970s may look very different intellectually from what a psychologist looking at a gifted child in the current decade would see.

There is a renewed interest in providing gifted students with specific programming to meet their educational needs. At the heart of this renewed interest is an often hotly debated issue: How should children be identified for these special programs? What should guide the process of identification? Central to this issue is the importance of establishing a clear definition of *giftedness* to guide the development of services for the population defined.

THE ISSUE OF DEFINITION

The terms *gifted* and *talented* have been used in a variety of ways to describe individuals who perform at a superior intellectual level. *High-ability* or *high-functioning* have been terms frequently used in order to avoid using either *gifted* or *talented* because these terms have been problematic over the years. Borland (1989) has stated that there is a rupture between the word *gifted* in its various usages and a clearly and consensually defined group of children in schools. Hence the dichotomy between what the term actually means and how it is frequently used has caused some confusion about how to regard the dimensions of giftedness. Similarly, people have often disagreed over what *gifted* and *talented* mean in terms of individual characteristics, behaviors, and need for services.

Tannenbaum (1997) has defined *giftedness* as potential for becoming acclaimed producers. Gagné (1999) has differentiated between *gifts,* which he calls *aptitudes,* and *talents,* which he calls *expressions of systematically developed abilities or skills* in at least one field of human activity. According to Gagné, catalysts—environmental, intrapersonal, and motivational—transform intellectual, creative, socioaffective, and sensory–motor gifts (abilities, aptitudes) into talents (performances) in the academic, technical, artistic, interpersonal, and athletic areas (see also Davis & Rimm, 1998). Cox, Daniel, and Boston (1985) have avoided the term *gifted,* preferring to call these students *able learners.* Renzulli and Reis (1997) prefer *gifted behaviors,* which can be developed in certain students at certain times and in certain circumstances. Treffinger (1995) likes the term *talent development,* calling the shift a fundamentally new orientation to the nature of the field. Best-selling authors (e.g., Coyle, 2009; Gladwell, 2008) have embraced talents and gifts in their books, defining these constructs outside the cognitive dimensions present in most of the previous definitions: "[Talent is] the possession of repeatable skills that don't depend on physical size" (Coyle, 2009, p. 11), and "Outliers are those who have been given opportunities—and who have had the strength and presence of mind to seize them" (Gladwell, 2008, p. 267). For Gladwell (2008), at least, the gift is the opportunity as much as (or even more than) the innate traits. Many others have used the terms *gifted* and *talented* synonymously. If one standard definition were always used, the confusion would be nonexistent. This confusion in definitions has indeed been a major issue to both those concerned with studying gifted individuals and those concerned with educating them.

The definition of *gifted and talented* provided within the Every Student Succeeds Act of 2015 (ESSA; Pub. L. 114-95, 2015) is as follows:

> Students, children, or youth who give evidence of high achievement capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who need services and activities not ordinarily provided by the school in order to fully develop those capabilities. (p. 398)

Although this definition has served the purpose of informing schools of what areas should be considered in serving gifted students, it really does not inform anyone of specific ways to find these people and is perhaps too broad for most school districts to operationalize effectively. In contrast, it does demonstrate the current trend in widening the perspective in order to allow more people to be served. Indeed, multitrait definitions tend to be the norm today.

In contrast, the definition from the National Association for Gifted Children (NAGC; 2010) is broader and more inclusive than the one provided in the ESSA, and it includes a wider range of skills and abilities than are usually addressed in schools:

> Gifted individuals are those who demonstrate outstanding levels of aptitude (defined as an exceptional ability to reason and learn) or competence (documented performance or achievement in top 10% or rarer) in one or more domains. Domains include any structured area of activity with its own symbol system (e.g., mathematics, music, language) and/or set of sensorimotor skills (e.g., painting, dance, sports). (p. 1)

This definition may also be too broad for schools in terms of directions for identifying the population to be served, but it does focus on the areas of exceptionality that are considered important to serve.

Renzulli (1978) has proposed a multitrait definition of giftedness that focuses on three interlocking clusters of traits: above-average, but not necessarily superior, ability; motivational traits that Renzulli calls *task commitment*; and creativity. According to Renzulli, "it is the interaction among the clusters that . . . [is] the necessary ingredient for creative/productive accomplishment" (p. 182). The form of giftedness characterized by high scores on standardized tests and model classroom behavior has been termed *schoolhouse giftedness* by Renzulli and Reis (1997).

Tannenbaum (1983) has defined *giftedness* as a psychosocial construct. He states that gifted individuals are those "with the potential for becoming critically acclaimed performers or exemplary producers of ideas in spheres of activity that enhance the moral, physical, emotional, social, intellectual, or aesthetic life of humanity" (p. 86). The key to this definition is the focus on the gifted individual as a producer of ideas.

Sternberg and Zhang (1995) have developed an implicit theory of giftedness that embodies five criteria: *excellence, rarity, productivity, demonstrability,* and *value.* Stating that implicit theories are relativistic because what is perceived as giftedness is based on the values of the particular time period or place in existence, Sternberg and Zhang have argued for the need for implicit theories to fill in the gaps left by explicit theories (i.e., those that specify the content of what it means to be gifted). "The problem is that in the science of understanding human gifts, we do not have certainties. There are no explicit theories known to be totally and absolutely correct, nor are there likely to be any in the foreseeable future" (p. 91).

In contrast to multitrait or all-encompassing definitions of giftedness are the definitions centering around the cognitive aspects of reasoning and judgment that can be found in a test score. To this end, the Binet–Simon test (Binet & Simon, 1905) was developed as an early paper-and-pencil test that attempted to measure intelligence. A revision of this test later became known as the Stanford–Binet Intelligence Scale (Terman, 1916), and this IQ-based definition is still widely accepted today (Coleman & Cross, 2005). Terman (1925a), the father of the gifted movement in this country, was perfectly content with defining giftedness as the possession of a very high IQ.

Using the Stanford–Binet to identify a population of gifted children, Lewis Terman and his research team were interested in investigating intelligence and achievement in a group of high-functioning children in the 1920s. Terman (1925b) wrote the following about gifted children:

> When the sources of our intellectual talent have been determined, it is conceivable that means may be found which would increase our supply. When the physical, mental and character traits of gifted children are better understood, it will be possible to set about their education with better hope of success. . . . In the gifted child, Nature has moved far back the usual limits of educability, but the realms thus thrown open to the educator are still *terra incognita.* It is time to move forward, explore, and consolidate. (pp. 16–17)

Specifically, Terman's team asked the following questions: Do precocious children become exceptional adults? Do high-IQ adults exhibit a disproportionate degree of mental health problems? Are brilliant children also physically superior? Does having a high IQ correlate with excellent school performance? Can gifted children be expected to display exceptional adult career achievements as eminent scientists, scholars, artists, and leaders? If high-ability children become extraordinary adults, what can be learned from the personal and educational antecedents that seem to nurture their development? (Sabotnik & Arnold, 1994). Terman's group's research focused on the lives of those high-functioning individuals who had scored in the top 1% on the Stanford–Binet Intelligence Scale, and on what could be learned about their lives to create educational opportunities that would serve similar people. They concluded from their research that superior children apparently became superior adults (Oden, 1968).

Leta Hollingworth's (1942) work with high-IQ students (i.e., children with IQs over 180) at the Speyer School in New York City was also important in informing researchers about the impact of an enrichment program for gifted students on their adult achievement and values. In 1981, several adults who had formerly attended the school were interviewed concerning the school's impact on their lives (White & Renzulli, 1987). Among those interviewed, three were from the group in Hollingworth's early study. They stated that the school provided lifelong love for learning, pleasure in independent work, and joy in interacting with similar high-ability students (Sabotnik & Arnold, 1994). Hence, both Terman's and Hollingworth's noteworthy research depended in large part on the IQ score measured for each individual and on what these high-ability individuals became later in life.

On the contrary, Simonton (2008) used Winner's (1996) definition of giftedness—"A gifted child or adolescent is someone who masters a particular domain at a faster rate than the average youth" (p. 253)—in his historiometric study of 291 eminent African Americans. Multiple-regression analyses indicated that adulthood eminence and creative achievement were positively correlated with early giftedness. Simonton offered two main implications, one theoretical and one practical, for this study. Theoretically, his inquiry established an impressive developmental continuity across the lifespan: Precocious development in childhood and adolescence predicts the magnitude of eminence and achievement in adulthood. Practically, his study indicated that giftedness must not be evaluated according to a "one-size-fits-all" procedure, but rather according to the occurrence of precocious behaviors that are specific to a given culture and achievement domain. The variety of gifts manifested in these precocious individuals would not have been identified by a score on a standard intelligence test.

A new paradigm has recently been developed that addresses the issue of identification in a different way. This paradigm is called *advanced academics* (Peters, Matthews, McBee, & McCoach, 2014); the idea is that instead of identifying giftedness as a stable trait, the authors advocate assessing students' need for a particular program as well as their probability of success in the program. The overarching goal is to better match instruction and teaching with each student's current level of mastery and need. Thus the advanced academics paradigm ignores the controversial question of who is or is not gifted, or even if such a thing as giftedness exists. Instead, academic needs form the foundation on which programs and services are built. The authors state:

> Academic need is not a property of individuals, but rather emerges from an interaction between a student and a particular academic environment—a particular teacher, subject, curriculum, and peer group. As such, academic need is not expected to be a stable trait across contexts, schools, or teachers, but rather a (hopefully) temporary condition that arises when the instructional pacing, depth, and/or content are not sufficiently rigorous as to require full engagement and effort from the student. (p. 40)

Dai and Chen (2013) examine three paradigms in gifted education that currently exist: (1) the *gifted child* paradigm, (2) the *talent development* paradigm, and (3) the needs-based *differentiation* paradigm. Identification of the three paradigms is based on historical and theoretical grounds. Each paradigm is carefully defined and describes the current trends in serving gifted students according to how gifted educators best define giftedness.

Gridley, Norman, Rizza, and Decker (2003) have proposed a definition of giftedness based on the CHC theory of intelligence. This theory combines Cattell and Horn's model of Gf (fluid) and Gc (crystallized) intelligence with Carroll's standard multifactorial model. Carroll's model (see Schneider & McGrew, Chapter 3, this volume) suggests that cognitive abilities exist at three levels or *strata*: (1) a lowest or first stratum composed of numerous narrow abilities; (2) a second stratum consisting of about 8–10 broad abilities; and (3) a third stratum comprising a single general intellectual ability, commonly called *g*. Gridley and colleagues' definition is as follows:

> Intellectually gifted students are those who have demonstrated 1) Superior potential or performance in general intellectual ability (Stratum III) and/or 2) Exceptional potential or performance in specific intellectual abilities (Stratum II) and/or 3) Exceptional general or specific academic aptitudes (Strata I and II). (p. 291)

The practicality of Gridley and colleagues' definition is that it suggests that giftedness can be measured by a test. The authors state that they do not "focus on the genetic causes of gifts, but rather . . . on gifts as intellectual abilities and talents as special academic aptitudes being of equal value in their need for nurturing and development" (pp. 290–291).

Most professionals regard giftedness in school as an academic need to be served (e.g., Coleman & Cross, 2005; Rizza, McIntosh, & McCunn, 2001). In order to be served appropriately, students must be identified, and standardized tests are the major methods used for identification purposes. Although a standardized test is available to measure each dimension in the federal definition, most programs for gifted individuals are particularly interested in intelligence tests because most gifted programs focus on serving students of high cognitive ability.

Gallagher (1995) has stated that an IQ test is merely one measure of the development of intellectual abilities at a given time. It gives an indication of a child's current development, so that children can be compared to one another on such characteristics as their store of knowledge, reasoning ability, and ability to associate concepts—all of which are important predictors of academic success. IQ tests still remain the single most effective predictors of academic success that we have today. There is evidence (Rindermann, 2007) to suggest that one of the reasons why measures of intelligence are such good predictors of academic success is that they measure a single innate construct overlapping with academic achievement. It is also likely that this single innate construct has shared measurement error with common measures of academic achievement. Pyryt (1996) agrees with this focus on the best measures currently available, stating that IQ tests are very useful for making legal decisions regarding the eligibility for participation in gifted programs. IQ tests still serve as important tools for recognizing the special education needs of intellectually gifted students.

The modern IQ test, with its age-based normative comparison, allows for students' level of giftedness to be classified by a method such as that suggested by Gagné (2004) and encouraged by Baer and Kaufman (2004). It is important to recognize that this classification system may be difficult to implement, as the item ceilings for different measures of intelligence may limit the ability of individuals to be classified at the highest level or to maintain a high level of classification at different ages.

THE ISSUE OF ONE VERSUS MANY FACTORS IN INTELLIGENCE

Defining Intelligence(s)

The IQ score, a unidimensional construct used for many purposes, has been historically very important in identifying and understanding giftedness. In fact, as noted earlier, the idea that a child is intellectually precocious has often been synonymous with a high IQ score. Those arguing against the idea of an IQ score have stated that this measure leads to a narrow view of intelligence that is tied to the skills most valued in schools—linguistic and logical–mathematical skills (Ramos-Ford & Gardner, 1997). In addition, Ramos-Ford and Gardner note that a majority of children are still admitted to specialized educational programs for gifted students on the basis of an IQ score of 130, or two standard deviations above the mean on an intelligence test. A score of 129, virtually the same score, will keep another student out of such a program. This cutoff score process is problematic and all too prevalent in school programming for gifted students.

Arguing for a theory of multiple intelligences, Gardner (1999) has defined *intelligence* as an ability or set of abilities that permits an individual to solve problems or fashion products that are of consequence in a particular cultural setting. Ramos-Ford and Gardner (1997) conclude, "A multiple intelligences approach to assessment and instruction strives toward identifying and supporting the 'gifts' in every individual" (p. 65).

Sternberg's triarchic theory of intelligence suggests that intelligence includes "applying component processes to novel tasks for the purposes of adaptation to, shaping of, and selection of environments" (Sternberg & Williams, 2002, p. 148). Sternberg has described both his triarchic theory of intelligence and Gardner's theory of multiple intelligences by using a systems metaphor. Sternberg's metaphor suggests that to understand the various aspects of intelligence working together as a system, one needs to understand the integration within the system itself (Sternberg & Williams, 2002). Although these theories have gained popularity in recent years, they lack empirical data to support their effectiveness (Sternberg & Williams, 2002).

Sternberg and Williams (2002) state, "Perhaps the most difficult challenge in the study of intelligence is figuring out the criteria for labeling a thought process or a behavior as intelligent" (p. 1). One must establish criteria to use in trying to decide what constitutes intelligence. Early experts suggested that intelligence is based on adaptation to the environment (e.g., Colvin, 1921, and Pintner, 1921; both cited in Sternberg & Kaufman, 2001). Later, Boring (1923; cited in Sternberg & Kaufman, 2001) suggested that intelligence could and should be defined operationally as that which intelligence tests test. Current definitions by both experts and laypersons suggest that adaptation to the environment, whether with practical problem-solving ability or academic skills, is still the essential theme in defining intelligence. Sternberg and Williams have further suggested three criteria to understand the mental processes and behaviors that can be labeled intelligent: correlation of a target thought or behavior with cultural success, or *cultural adaptation*; mental skills development, or *cultural and biological adaptation*; and evolutionary origins and development, or *biological adaptation*.

With these emphases in mind, then, individuals who are called gifted will be those who can best adapt to their environments, and the purpose of finding these individuals through identification processes in schools will be to maximize their abilities in doing so.

Factor-Analytic Research

Extensive factor-analytic research has been conducted with the goal of exploring the multidimensional nature of intellectual abilities among gifted children. The majority of this research has used the Wechsler Intelligence Scale for Children—Revised (WISC-R; Wechsler, 1974), the WISC-III (Wechsler, 1991), and the WISC-IV (Wechsler, 2003) (Brown & Yakimowski, 1987; Macmann, Plasket, Barnett, & Siler, 1991; Mishra, Lord, & Sabers, 1989; Reams, Chamrad, & Robinson, 1990; Rowe, Dandridge, Pawlush, Thompson, & Ferrier, 2014; Watkins, Greenwalt, & Marcell, 2002). In general, factor-analytic studies have consistently found support for the Verbal Comprehension and Perceptual Organization two-factor model (Karnes & Brown, 1980; Sapp, Chissom, & Graham, 1985; Watkins et al., 2002) proposed by Kaufman (1975, 1979). Among the two-factor models, the Verbal Comprehension factor was typically composed of the Similarities, Vocabulary, Comprehension, and Information subtests. More variability was displayed across studies in the composition of the Perceptual Organization factor. The majority of the studies found that the Block Design and Object Assembly subtests loaded on the Perceptual Organization factor, while the Picture Completion and Picture Arrangement subtests were found to load inconsistently across studies on this factor.

Although factor-analytic studies generally have supported a two-factor model, there has been varying support for a three-factor model (Brown, Hwang, Baron, & Yakimowski, 1991; Brown & Yakimowski, 1987; Karnes & Brown, 1980; Macmann et al., 1991). The specific composition of the third factor has varied significantly across studies. For example, several studies found that the Information, Arithmetic, and Coding subtests primarily composed the third factor (Brown & Yakimowski, 1987; Brown et al., 1991), whereas Sapp and colleagues (1985) found that the Information, Arithmetic, Vocabulary, and Block Design subtests primarily composed the third factor. In addition, Karnes and Brown (1980) noted that the Arithmetic and Picture Completion subtests composed the third factor (Freedom from Distractibility).

Only limited factor-analytic research has been conducted among gifted children with cognitive measures other than versions of the WISC (Cameron et al., 1997). Cameron and colleagues (1997) conducted a confirmatory factor analysis using the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983). Although they compared four models of intelligence, they determined that Horn and Cattell's theory of fluid–crystallized intelligence provided the broadest understanding of the cognitive functioning of children referred for gifted services (Cameron et al., 1997).

The factor structures of cognitive measures have also been studied among gifted members of ethnic minority groups (Greenberg, Stewart, & Hansche, 1986; Masten, Morse, & Wenglar, 1995; Mishra et al., 1989). Factor-analytic research conducted by Greenberg and colleagues (1986) supported the WISC-R Verbal Comprehension and Perceptual Organization two-factor model with a sample of gifted black children. Another study, which examined the factor structure of the WISC-R with Mexican American children referred for intellectually gifted assessment (Masten et al., 1995), was unable to adequately replicate the factor structure proposed by Kaufman (1975). In contrast, the cognitive constructs of gifted Navajo children were similar to the Freedom from Distractibility and Perceptual Organization factors identified in research based on the standardization sample of the WISC-R (Mishra et al., 1989).

The variability in results among factor-analytic studies appears to stem primarily from methodological and statistical differences. Macmann and colleagues (1991) noted that restriction in variance due to sample selection might have contributed to differences in the composition of factors. There also appears to be great disparity across studies related to sample sizes. Although the majority of studies used large samples (e.g., Macmann et al., 1991; Watkins et al., 2002), several studies used samples with fewer than 150 participants. Factor-analytic research using gifted ethnic minority children utilized the smallest samples, with some using fewer than 100 participants (e.g., Masten et al., 1995; Mishra et al., 1989).

A lack of consistency across studies in the criteria used for determining giftedness has also made it difficult to compare factors across studies. The criteria for inclusion ranged from a WISC Full Scale IQ of 120 and higher to 130 and higher. In addition, it was not uncommon for the criteria for inclusion to include participants with a WISC Full Scale IQ, Verbal IQ, and/or Performance IQ of 130 or higher. Gifted eligibility for some studies included children who did not meet the stated IQ criteria but did demonstrate advanced academic performance. The study conducted by Brown and Yakimowski (1987) demonstrates how selection criteria can influence the composition and the number of factors identified. They studied the WISC-R scores for three different groups of children: children who scored in the average range (IQ score between 85 and 115), children who scored 120 or higher (high-IQ group), and children in gifted programs (gifted group). The average group displayed the two-factor model commonly associated with the WISC-R; however, a four-factor solution was identified for the gifted group, and a five-factor solution was identified for the high-IQ group. The additional factors suggested that children in the high-IQ and gifted groups processed information differently from the children with average cognitive abilities (Brown & Yakimowski, 1987). Thus the composition of the sample appears to have had an influence on the number and types of factors generated.

The use of different combinations of WISC subtests in factor analyses also contributed to the different composition of the factors and to whether two- or three-factor models were generated. The 10 regularly administered WISC-R subtests were consistently utilized in the factor analyses, while the Digit Span and Mazes subtests were often excluded. Studies that included the Digit Span subtest found it to load consistently on the same factor with the Arithmetic subtest (Brown & Yakimowski, 1987; Watkins et al., 2002).

The type of extraction method used, the criteria used to determine the number of factors to interpret, and the criteria used to determine the composition of factors also varied greatly across studies. The type of extraction method used (e.g., maximum-likelihood, principal-components, principal-axis) could have influenced the number of factors generated and the composition of those factors. In addition, a vast array of criteria was used across studies for determining the number of factors to interpret. Specifically, the scree test, the chi-square statistic, eigenvalues greater than 1.00, and various combinations of these techniques were used by researchers for making the determination of how many factors were identified. Differences were also found across studies in the type of rotation methods (e.g., varimax, oblique), resulting in differences among factors. There were considerable differences in the criteria used to determine whether a subtest loaded on a specific factor. Although some studies failed to indicate the criteria used for identifying significant factor loadings, the studies that did provide criteria for significant loadings tended to range from .30 to .50. Given these differences in methodology and statistical techniques across the factor-analytic studies, it is not surprising to find some differences in the cognitive constructs of gifted children on intelligence measures.

Some of the consistency of the findings related to the structure of intelligence in gifted samples may also reflect the structure of the instruments that were used to assess it. The WISC-R and WISC-III were not closely tied to any theory of intelligence (O'Donnell, 2009). As a result, much of the structure of the WISC at that time focused narrowly on measures of Gf and Gc, while underrepresenting other abilities. Consequently, even if the individuals had possessed differing factor structures in these areas, they were not measured and would not emerge in the analysis.

These earlier factor-analytic structures were important in helping to improve our understanding of the continuity and similarity for the structure of intelligence in gifted samples with the rest of the population. It is important to recognize that studies using an exploratory approach are apt to dismiss more complex models in favor of simpler parsimonious models (two- or four-factor models vs. a seven- or eight-factor model). In part, the exploratory approach may sometimes hide existing factors. These occasions are more likely to occur when sample size is relatively limited and when the variance of the sample is limited, both of which have occurred in the past in the literature on giftedness. From a statistical viewpoint, then, the inclusion of individuals who may not have met all previous gifted criteria may actually have improved the ability of the models to fit because it added needed variance to the samples.

In summary, factor-analytic research using gifted samples has for the most part confirmed the presence of Kaufman's two-factor model. In addition, the WISC-R/WISC-III subtests that compose the Verbal Comprehension factor have been replicated across numerous studies, suggesting significant stability of this factor with gifted children. Less stability has been shown in the composition of the Perceptual Organization factor, and even less stability has been shown in relation to a third factor. Although the majority of the research has been exploratory, a few confirmatory-factor-analytic studies have been published (Brown et al., 1991; Cameron et al., 1997). However, there is a continued need to study hierarchical models of intelligence with gifted samples. There is also a need to demonstrate the utility of considering multiple cognitive constructs in identifying giftedness.

THE ISSUE OF IDENTIFICATION

The Process of Identification as Related to the Definition

Several authors previously mentioned have discussed the importance of cognitive ability measures for the identification process. A very controversial part of serving gifted students is the process of locating the population to be served—a process known as *identification*. Clark (1997) has suggested the following considerations in a comprehensive identification program:

- Evidence that students demonstrate extraordinary ability in relationship to their age-level peers.
- Evidence of the range of capabilities and needs.
- Processes that measure potential as well as achievement.
- Methods that seek out and identify students from varying linguistic, economic, and cultural backgrounds, and special populations.
- Implications for educational planning.

This comprehensive list of services has opened the door to much controversy as to what a school should do for this special population. Borland (1989) has cautioned that defining the target population is the first and most important step in programming for gifted students. In other words, if a school selects a narrow definition or one based exclusively on cognitive ability, then the school's program should reflect this definition. On the other hand, if the school chooses to adopt the ESSA's definition, then a very comprehensive array of services should be available. A major issue in identification of gifted and talented youth has been the validity of the identification process with respect to program goals and services (Feldhusen & Jarwan, 1993). Because placement decisions are the goal of identification, all measures used are very important.

Identification generally begins with a screening procedure, in which students are selected first on the basis of their performance on a group achievement test. Students who score the highest on this general group test, according to the school's criteria, form a talent pool. Next, the talent pool's members take a more selective test (perhaps a more precise instrument with a lower standard error of measurement). High-ability students often score very highly on group tests, and a *ceiling effect* may occur, in which their scores cluster near the very top of possible scores on the test. It is sometimes incorrectly assumed that all of these children are equally talented and need a similar program. However, this is not always the case. More precise tests that address this ceiling effect are preferred, so that those identified for the program are truly those students most capable (if that is the definition of the target population for the specific school's program). After this second screening step, all measures to be considered in identification are added, and the top students are identified for the program. One approach to addressing the ceiling effect common with group measures may be to use individually administered intelligence tests, such as the WISC-V (Wechsler, 2014a); the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003); or the Woodcock–Johnson III (WJ III; Woodcock et al., 2001) or WJ IV (Schrank, McGrew, & Mather, 2014). However, whatever test is administered, the ceiling effect must be considered in selecting tests that truly measure the abilities to be served. Another strategy that has received some popularity over the years is the use of out-of-age tests with students believed to be gifted. The higher ceilings of these instruments allow for students to be challenged and fit with a historical definition of advanced mental age. One problem with this approach is that because the tests were given to children not in the normative sample, age-based norms cannot be used, and comparisons cannot be made. As a result, application of a proportion estimate such as that suggested in Gagné's (2004) classification system is inappropriate.

Zhu, Cayton, Weiss, and Gabel (2008) published an extended set of normative tables for use in assessing gifted children. These tables are designed for use with children who reach the ceiling of the traditional normative set for their age on one or more subtests of the WISC-IV (Wechsler, 2003). These extended tables are intended to help differentiate between gifted and highly gifted individuals. Given the frequency of individuals with IQ scores above 150, the degree to which clinicians will need to resort to these tables is limited. In addition, the tables are believed to be more beneficial in tracking progress or growth of gifted children, as they should be more sensitive to performance above the original ceiling of the test.

Recently, Raiford, Drozdick, Zhang, and Zhou (2015) developed expanded index tables for the

WISC-V (Wechsler, 2014), which may be desired by some who are assessing gifted students. The two expanded index tables are the Verbal (Expanded Crystallized) Index (VECI) and the Expanded Fluid Index (EFI). These indexes allow for standard scores up to 155. The advantage of these tables is that they provide a broader assessment of two historically important areas of gifted assessments. Also, others have attempted to reduce ceiling effects for children and adolescents with other instruments, including the Stanford–Binet, by creating a Gifted Composite (McGowan, Holtzman, Coyne, & Miles, 2016).

Identification of Gifted Minority Students

A major issue in identification of gifted students relates to the representation of minority students in gifted programs. According to Ford and Harris (1999), projections are that minority students will account for almost half (46%) of all public school students by the year 2020. This increase in minority students at the national level has not been reflected in gifted education. In fact, according to Gallagher (2002), national surveys indicate that only 10% of students performing at the highest levels are culturally, linguistically, and ethnically diverse students, even though these diverse students represent 33% of the school population. A major focus of attention in gifted education is the goal of parity in gifted programs among all members of society, but this parity has not been easy to achieve. Although the Jacob Javits Act of 1988 helped initiate programs for gifted racial minority students from economically disadvantaged areas, the problem of finding and/or developing useful tools to identify these students still exists. Efforts to find more valid, reliable, and useful instruments to assess giftedness and potential among minority students, and to increase teacher training in identification and assessment so as ultimately to increase the referral of minority students to gifted services, are paramount in helping to find and serve these underrepresented students (Ford & Harris, 1999).

In July 1997, the NAGC adopted a policy statement on testing and assessment of gifted students, in which it called for more equitable identification and assessment instruments and procedures. The notions of fairness and accountability underlie the proposal. In this position paper on the use of tests in the identification and assessment of gifted students, the following issues were addressed:

> Given the limitations of all tests, no single measure should be used to make identification and placement decisions. That is, no single test or instrument should be used to include a child in or exclude a child from gifted education services. The most effective and equitable means of serving gifted students is to assess them—to identify their strengths and weaknesses, and to prescribe services based on these needs. Testing situations should not hinder students' performance. Students must feel comfortable, relaxed, and have a good rapport with the examiner. Best practices indicate that multiple measures and different types of indicators from multiple sources must be used to assess and serve gifted students. Information must be gathered from multiple sources (caregivers/families, teachers, students, and others with significant knowledge of the students), in different ways (e.g., observations, performances, products, portfolios, interviews) and in different contexts (e.g., in-school and out-of-school settings). (NAGC, 1997, p. 52)

This call for a different, more inclusive way to find and serve gifted minority students is a call for diversity in programs that have typically been labeled "elitist" by many. To widen the representation, school personnel must be educated to use multiple measures to find these underrepresented students. A change in identification practices must encourage the examination of gifted individuals in cultural and environmental contexts, and must provide a basis for recognizing talents without penalizing students for certain learning styles and expressions (Frasier et al., 1995).

Assessments used to identify gifted and talented students may represent a clash between cultures, in which the mainstream culture is unable to recognize or underestimates the abilities of underrepresented minority students (Briggs & Reis, 2004). One issue that has prevented the identification of gifted minority students is that of test bias. Reynolds and Kaiser (1990), in discussing the issue of content validity and its relation to test bias, have stated:

> An item or subscale of a test is considered to be biased in content when it is demonstrated to be relatively more difficult for members of one group than for members of another in a situation where the general ability level of the groups being compared is held constant and no reasonable theoretical rationale exists to explain groups' differences on the item (or subscale) in question. (p. 625)

Items on tests often tap experiences that are relevant to middle-class students. Those from impoverished families may simply not comprehend such items, and therefore may miss questions because of environmental deficiencies rather than actual lack of intelligence. Such problems point to the bias that underlies *content validity* in both achievement and intelligence tests.

In addition, bias in *construct validity* is of concern. The fact that different groups define giftedness and intelligent behaviors in a variety of ways makes the measurement of these constructs difficult and often invites bias. Again, if the construct in question is always defined in middle-class terms, impoverished students may not be found and served.

Finally, bias is also seen in terms of *predictive validity*—the extent to which an instrument predicts the future success of a person in various situations. If teachers read the results of an intelligence test and therefore judge a student's future worth on the basis of these results, the student may not fare well. In fact, teachers' expectations may diminish because of perceived deficiencies that may not be accurate indicators of the student's ability. For these reasons, assessment issues with minority students are of major concern in the identification of students for programs.

Ford and Harris (1999) have suggested the following options when evaluators are considering how best to assess ability and potential in linguistically, racially, and culturally diverse students:

- Adapt instruments (e.g., modify the instruments in terms of their language demands).
- Renorm the selected instruments on the basis of local norms and needs.
- Modify predetermined cutoff scores for minority students.
- Use an alternative nonverbal cognitive measure thought to assess the same construct.

The issue of diversity in membership in gifted programs is currently a major issue. Educators must understand its importance and must respond to the need for alternative identification tools if they desire to provide high-quality education for all.

Identification of Gifted Students with Learning Disabilities

Another group of students who are often missed in the identification process is the group of those who are gifted but have learning disabilities. Although this description seems to be an oxymoron, Davis and Rimm (1985; Davis et al., 2011) have noted that estimates of the number of such students in U.S. schools range from 120,000 to 180,000. Identification of students with both talents and disabilities is problematic and challenges educators (Olenshak & Reis, 2002; Sternberg & Grigorenko, 1999). Historically, most school personnel have relied on discrepancy formulas between intelligence and ability test scores; analyses of intelligence test results for differences across subtests (*scatter*); and multidimensional approaches that incorporate qualitative data, such as structured interviews and observations (Lyon, Gray, Kavanagh, & Krasnegor, 1993). With the ESSA revisions and the filtered changes likely to be enacted through different states to comply with this law, school districts will depend more heavily on dynamic and locally normed assessments. As a result, the provision of special education services to children who have learning disabilities and are also gifted is likely to increase the delay in identification, with the exception of those individuals with severe disabilities or multiple areas of disability.

Furthermore, the identification of these students is complicated because their gifted abilities often mask their disabilities, or, conversely, their disabilities may disguise their giftedness (Volker, Lopata, & Cook-Cottone, 2006). These problems may exclude students from inclusion in either programs for gifted individuals or programs for those with learning disabilities (Baum, Owen, & Dixon, 1991; Olenshak & Reis, 2002). This is also true for students with other exceptionalities, such as attention-deficit/hyperactivity disorder and high-functioning autism spectrum disorder. Astute educators are aware of these major issues in identification and search for ways to include rather than exclude all gifted students. As a result of the tendency for gifted students to exhibit higher levels of functioning globally when compared with other students, even in areas of suspected disability, it may be more beneficial for psychologists and committees to take an ipsative approach to analyzing test data. This method is consistent with that recommended by Volker and colleagues (2006); it is also a pattern approach consistent with the more recent ESSA.

Identification of Gifted English-Language Learners

English-language learners (ELLs) are underrepresented in programs that serve gifted students (Lohman, Korb, & Lakin, 2008). Indeed, according to Harris, Plucker, Rapp, and Martinez (2009), the lack of attention to giftedness in underrepresented populations such as ELLs is a critical weakness in the identification literature because of the concomitant rapid increase in the number of ELLs in the United States. For example, in 1979, approximately 1 in 10 school-age children spoke a language other than English at home; by 2003, the proportion had risen to nearly 1 in 5 (9.9 million) children (National Center for Education Statistics, 2005). Harris and colleagues state that between the 1989–1990 and 2004–2005 school years, ELL enrollment in public schools more than doubled, from 2,030,451 to 5,119,561 students (National Center for Education Statistics, 2008).

This problem of identifying a rapidly growing population provides challenges for administrators seeking unbiased identification methods. Coleman and Cross (2005) argue that the identification of culturally different gifted students is a perplexing problem because their average performance on aptitude and achievement tests tends to be one standard deviation below the mean when general norms are used. Ortiz and Dynda (2005) state that if there is to be any validity to conclusions drawn by practitioners, four issues concerning test bias must be understood: (1) acknowledging the cultural content embedded in any given test; (2) understanding the linguistic demands inherent in any given test; (3) appreciating meaningful differences in norm sample representation for diverse individuals; and (4) recognizing the limitations of "nonverbal" assessment. (See also Ortiz, Piazza, Ochoa, & Dynda, Chapter 25, this volume.)

Administrators of gifted programs in schools have looked for a suitable means for identifying ELL students that would increase representation of these learners without relying solely on language as the major requisite for testing. Nonverbal tasks have long been present in intelligence tests, providing one indicator of ability for students who are native speakers of the language, while perhaps serving as the only indicator of ability for examinees who are not fluent speakers of the language (Lohman et al., 2008). Nevertheless, finding the right test and using the appropriate norm group for comparisons are essential aspects of identifying gifted ELL students.

Multiple Means of Assessment

For all these reasons, multiple means of assessing students are often used. Test scores are one group of determinants. Others include nominations by teachers, students, parents, and peers. In addition, checklists often tap areas of strength in students. Performance assessments, such as portfolios and auditions, are often valuable in the identification of gifted and talented students for programs. Of course, the parameters of the program must be considered in the design of any identification criteria. In addition, determining a reasonable formula that takes into account all the criteria and determines the students who then emerge as those who qualify for the program is a difficult and challenging problem for educators who manage gifted programs in schools.

SPECIAL ISSUES

Competing Theoretical Approaches

Although a lack of consideration of theory when evaluators are selecting cognitive measures is a concern in intellectual assessment with all children (Flanagan & Ortiz, 2002), an atheoretical approach appears to predominate within the schools when cognitive measures to identify gifted children are being chosen. This pattern is not unique to school professionals and reflects the greater historical need for these instruments to be able to differentiate between those who will succeed and those who will not despite educative efforts (Strauss et al., 2006). The selection of cognitive measures is often based on ease of administration, cost, and familiarity with the test. In addition, many current psychologists choose to employ theory-based instruments. Examples include the WJ IV (Schrank et al., 2014), the SB5 (Roid, 2003), the WISC-V (Wechsler, 2014a), and the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014). All but the last is based on or mapped onto the CHC theory of intelligence, and the CAS2 is based on the *planning, attention, simultaneous, sequential* (PASS) model. The Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004) draws upon the theoretical model of Luria's neuropsychological theory and CHC theory.

Although the current advances in instrument development have led to a plethora of choices, this plethora can create additional confusion when evaluators are trying to select a measure for identifying cognitively gifted children. The problem is that different measures of intelligence, regardless of theory, assess different skills; the result is that some children are not offered opportunities to participate in gifted programs (Simpson, Carone, Burns, Seidman, & Sellers, 2002; Tyler-Wood & Carri, 1991). The implications of using a specific measure of intelligence should be considered prior to its selection. For example, subtests on the CAS2 were selected to focus primarily on intelligence as a problem-solving and reasoning ability that generalizes across areas, whereas the WJ IV may assess this problem-solving ability in specific ability areas (e.g., auditory processing). Characteristics of the gifted children identified, available programming that addresses the specific characteristics of the children identified, and the extent to which nonmodal gifted children (e.g., children from impoverished backgrounds, gifted children with disabilities) are excluded should all be considered when a specific measure of intelligence is chosen. Therefore, it is essential that the theoretical differences in intelligence measures be considered—and, more important, that theory-based measures be utilized during the identification process. This view is consistent with that of Flanagan and Ortiz (2002), who advocate theory as the center of all intellectual assessment activities. The process of identifying cognitively gifted children is complex enough without starting the identification process with an atheoretical or outdated measure of intelligence.

Linking Screening Procedures with Intelligence Measures

The majority of schools have developed some type of system for identifying gifted children, albeit some systems are better than others. Usually included within the system is a procedure for screening children and thus reducing the number who are eventually referred for more comprehensive testing. Unfortunately, the comprehensive testing typically includes the administration of a single standardized measure of intelligence. And in any case, many evaluators fail to study the accuracy of screening procedures related to the final decision on whether children receive gifted programming, which is often based on individualized measures of intelligence.

Although many would consider screening to be the crucial point in the identification process, predictive validity must be established between the screening procedure and the intellectual measure(s) used to ensure the accuracy and utility of the identification process. The difficulty with demonstrating predictive validity is that screening procedures can vary from teacher nominations to the use of group intelligence tests (Coleman & Cross, 2005). Considering the wide variety of screening methods used, relationships between screening procedures and intellectual measures can range from very low to very high. Screening processes and tools to identify gifted and talented abilities unlikely to be identified through the use of a standardized intelligence test also need to be selected and utilized. To be clear, a screening procedure that is more highly related to a selected intelligence test will result in a higher level of agreement at different points in the identification process. However, those who are gifted and talented in such domains as leadership or music may not be identified as such on an IQ measure. Coleman and Cross (2005) recommend using a fairly liberal screening threshold, to avoid missing children who may qualify for gifted programming. This approach seems advisable, given the lack of research exploring the relationships between many of the screening procedures used and the final decision outcomes based on intelligence tests. This process is consistent with Gagné's (2004) suggestion that those in the top 10% be considered gifted.

Use of a Single Test Composite Score

The use of a single cognitive test composite score as the primary criterion for determining giftedness is highly common within schools. In the past, the WISC-R (Wechsler, 1974) and the fourth edition of the Stanford–Binet (SB-IV; Thorndike, Hagen, & Sattler, 1986) were the most commonly used cognitive measures in the schools (Coleman & Cross, 2005). Coleman and Cross (2005) also note that one of these measures was commonly used as the final decision criterion for determining giftedness. In fact, school districts and states have defined giftedness solely on the basis of WISC IQ cutoff scores (Fox, 1981; Karnes, Edwards, & McCallum, 1986). Others have suggested that the use of strict IQ cutoff scores is too restrictive and does not consider other characteristics of giftedness (Renzulli & Delcourt, 1986; Renzulli, Reis, & Smith, 1981). However, because of the overwhelming use of intelligence tests, there is a need to discuss the implications of using a single IQ score for making decisions on giftedness.

Implications of Using Cutoff Scores

One of the crucial decisions made by any school system is where to set the IQ "cutoff" score. It is clear from reviewing the literature that there is little, if any, consensus on where the cutoff score should be set. This makes it extremely difficult not only to evaluate decision outcomes across studies, but also to interpret research on giftedness in general. In the literature, the inconsistency is quite evident. For example, Karnes and Brown (1980) used a cutoff score of 119; Hollinger (1986) used a cutoff score of 130; and Fishkin, Kampsnider, and Pack (1996) used a cutoff score of 127. However, most studies use an overall IQ score that is at least two standard deviations above the mean. Currently, this would translate into a WISC-V and an SB5 Full Scale IQ of 130. It should be noted that the SB5 has a standard deviation of 15. It is also important to understand that the historic method of calculating IQ used in early work on giftedness was built around a mental age formula; as a result, gifted individuals might obtain an IQ of 200, but today that is not possible. The point of this lengthy review of cutoff scores is that, regardless of where a school system sets a cutoff score, it is still too rigid an approach.

Many school systems fail to understand the basic psychometric process of how scores are derived, or the pitfalls of placing so much weight on a single score. To be specific, when placement decisions are being made, it is important to consider the standard error of measurement (*SEM*), which allows an evaluator to estimate a range of scores based on the obtained score. For example, suppose a child obtains a Full Scale IQ of 129 on a test with an *SEM* of ±4 points. The examiner can be 68% confident that the next time the child is administered the same test, the child's score will fall somewhere within the range from 125 to 133. Therefore, if the criterion for placement into a gifted program is rigidly set at 130, this child will be denied services. However, if the child is later administered the same intelligence test and obtains a Full Scale IQ score of 131, the child will then be recommended for gifted services. A sounder approach would be to make a decision regarding eligibility and the test scores within the context of the child's history, academic progress, and behavior. Repeated assessments over time will truly differentiate the groups.

It also is important to note that many test manuals now report confidence intervals using the method of regression to the mean, instead of rigidly applying the ±1 *SEM*. Although many might suggest that in considering the *SEM*, all that is being proposed is to lower the cutoff score, in fact this is not what is being proposed. What is being suggested is to allow for flexibility in making identification decisions, even when cutoff scores are being used. Therefore, identification decisions should be made with a proper understanding of the underlying psychometric characteristics of standardized intelligence tests, and also with consideration of other performance variables (e.g., academic achievement).

Furthermore, it is important to recognize that children with an estimated IQ in the range of 131–134 may quite possibly obtain scores in the range of 128–131 if reassessed. The children would be less likely to obtain scores of 133–135, as these would be further from the mean.

Recent updates and versions of testing manuals also indicate that contemporary assessment tools may produce lower IQ estimates for gifted samples than previous test versions have produced. As part of the development phase for tests, many test publishers use gifted samples to help validate the cognitive measures. While measures such as the WISC-V, the SB5, the WJ IV, and the KABC-II all show significant differences between matched samples and gifted samples, the range of scores for previously identified gifted students may fall in the high average range (WJ-IV, WISC-V, KABC-II). The range of scores for the WISC-V also includes the traditional cut-off of 130 in the confidence interval for previously identified youth at the upper end (Wechsler, 2014b). It is important to recognize that the developers of these tests used different methods to identify students, including previous test results, history of services, or independent determination by a school. This pattern of results is consistent with the Flynn effect (Flynn, 1987) and suggests that high-ability students are not immune to this phenomenon. For example, the technical manual for the SB5 discusses the Flynn effect as one possible reason why previously identified gifted students displayed a mean SB5 Full Scale IQ of 123.7 (Roid, 2003). Essentially, it is more difficult for previously identified gifted individuals to retain scores of 130 or higher, due to the change in IQ values over time. Examiners identifying gifted students through the use of an IQ test are advised to use a new instrument in tandem with estimates from an instrument that has been available for 5 years or more and with which the examiners are well familiar.

The work of Ziegler and Ziegler (2009) adds a further wrinkle to the use of cutoff scores from a single test or even from multiple tests. An analysis of the fundamental basis of measurement error and estimation of the construct of intelligence led these authors to note that high levels of cognitive abilities (Full Scale IQs greater than 130) are likely to be underestimated by measures of intelligence. This prediction appears to be supported by the data from the previously discussed studies. What is advocated as a result is a low-threshold test that will catch all potentially gifted students, followed by an intelligence test that will underestimate but have a lower error rate as a result of a previous screen (Ziegler & Ziegler, 2009). One strategy that may help to reduce this possible problem would be the application of a Bayesian approach to gifted identification. Essentially, the essence of the idea would be to use the existing datasets regarding the predicted level of cognitive ability expected on a new version of a test, and to estimate the probability that an individual would fall into a gifted sample similar to the one used in the validation studies by the test publishers. Similar strategies have been advocated in other areas recently for helping to improve classification with high-stakes testing (Wagenmakers, Morey, & Lee, 2016).

Beyond the Composite IQ Score

Interpretation of intelligence test results beyond the full test's composite score is also a recommended practice in identifying gifted children. Depending on an intelligence test's theoretical model, it can provide a multiple-cluster (e.g., verbal ability, thinking ability) index (e.g., Verbal Comprehension Index, Visual Spatial Index) or composite scores beyond the global composite score. Ignoring a child's performance on these other indexes may preclude him or her from receiving gifted services. As an example, if a child attains a WISC-V Full Scale IQ of 127, a Verbal Comprehension Index of 142, a Visual Spatial Index of 117, a Fluid Reasoning Index of 117, a Working Memory Index of 121, and a Processing Speed Index of 100, the child may be overlooked for gifted services if the sole criterion is the Full Scale IQ. Here is a child who obviously displays verbal abilities within the gifted range, and the Full Scale IQ fails to account for these specific cognitive skills.

Moreover, it is important to consider the demands and skills required by specific subtests that contribute to index, cluster, or composite scores. In the aforementioned example, the child's lowest WISC-V index score is in Processing Speed. The subtests that contribute to the Processing Speed Index rely on speed, short-term visual memory, cognitive flexibility, and concentration. Although this child demonstrates an average level of processing speed (which is considerably lower than his or her Verbal Comprehension Index score), the score may be more a function of how the child has approached each specific subtest contributing to the Processing Speed Index.

Many gifted children have learned to sacrifice speed for accuracy and perform less well on speed-related tasks than on others. Kaufman (1992) has noted that the highly reflective and perfectionistic nature of many gifted children affects their performance on measures of intelligence where speed is rewarded. Also, variances between composite scores are fairly common among gifted individuals (Malone, Brock, Brounstein, & Shaywitz, 1991). Another theoretical interpretation of this discrepancy would argue that the Processing Speed Index is a much less *g*-loaded index of ability, and therefore less indicative of intelligence and more indicative of motoric speed (Reynolds & Kamphaus, 2003). The key issue here is not to rely solely on the overall composite score in understanding the cognitive skills of gifted children, but to encourage professionals to evaluate the clinical, emotional, and behavioral data leading to a performance estimate.

McGrew, Schrank, and Woodcock (2007) indicate that samples of gifted students tend to perform better on tasks of Gc and Gf than on tasks of long-term retrieval (Glr), working memory (Gsm), or visual–spatial thinking. Results from the KABC-II manual suggest that gifted students perform best on tasks designed to measure Gc (Kaufman & Kaufman, 2004).

The growing awareness that individuals who are gifted often show lower levels of performance on less *g*-loaded activities prompted the development of a Gifted Composite for the SB5 (McGowan et al., 2016). The composite removes working memory from the Full Scale IQ estimate. When independently examined with a sample of individuals referred for gifted evaluations, the Gifted Composite increased for 80% of the sample and by an average of 5 points (McGowan et al., 2016). This strategy to a large degree mirrors the strategy advocated by others who prefer to use the Wechsler tests' General Ability Index (GAI) or the EFI or VECI as opposed to the Full Scale IQ. The use of the GAI removes both Processing Speed and Working Memory subtests (Wechsler, 2014a).

Lack of Ceilings

With many measures of intelligence, there are not enough items at the upper end of subtests to fully discriminate the cognitive abilities of gifted individuals. Although individual items used within specific subtests are selected for their ability to discriminate between children at different levels of cognitive ability, for many subtests it is difficult to establish a ceiling, which makes it difficult to obtain an accurate estimate of cognitive ability with gifted children. When gifted children are assessed, it is common for them not to obtain a ceiling on several subtests of a specific cognitive measure. When this occurs, the overall test composite score is likely to be an underestimate of their true level of cognitive functioning. Many would suggest that this does not matter because their level of cognitive functioning is so high that they would qualify for gifted services anyway. Here the issue is not so much one of identification as of matching children with appropriate gifted programming. It is important to accurately assess the unique cognitive skills of even highly gifted students, to meet their educational needs and interests.

Recommendations on the Effective Use of Intelligence Tests

A few general recommendations on the use of intelligence tests in the identification of giftedness are warranted. First, it is recommended that school systems develop an operational definition of giftedness that incorporates the term *cognitively gifted.* The use of this term is suggested if a school system's primary criterion for placement in its gifted program is based on an individualized measure of intelligence. It is important to recognize that most current measures of cognitive abilities today are aligned with the CHC model of intelligence, and that this model can have important implications for the education of children in gifted programs beyond just entry (Warne, 2016). If other measures (e.g., rating scales, nominations, achievement tests) or characteristics (e.g., leadership skills, motivation) are considered along with results on an intelligence test, the definition should address the role of the intelligence test in relation to the other measures and characteristics tapped during the identification process. Second, it is recommended that the school system develop specific procedures related to the referral, screening, testing, and placement of gifted children. Again, if the school's primary identification criterion is based on an individualized measure of intelligence, and the goal is to identify cognitively gifted children, then the screening measure should be highly correlated with the selected intelligence test.

Third, it is recommended that school systems become thoroughly aware of the specific advantages and limitations of using standardized intelligence tests to identify giftedness. Coleman and Cross (2005) note that intelligence tests are highly reliable, are individually administered, allow an examiner to observe a child's behavior directly, and allow for a broader sampling of behavior than screening or group intelligence tests do. Other advantages of intelligence tests are their abilities to identify exceptionally gifted individuals with special educational needs (Gross, 1993); children who do not display the stereotypical high verbal ability, high achievement, and high motivation often associated with giftedness (Pyryt, 1996); and children who are "twice exceptional" (e.g., children identified as gifted and with learning disabilities). Children who are identified for gifted programs as a result of high *g* have improved life and health outcomes across the lifespan (Warne, 2016). As for limitations, this chapter has primarily focused on assisting schools in making informed decisions regarding the use and limitations of intelligence tests with gifted students.

Fourth, it is recommended that school systems use theory as a primary guide in selecting intelligence tests. As Flanagan and Ortiz (2002) astutely note, the use of a modern and valid theory of intelligence at the beginning of the assessment process is critical in facilitating accurate measurement and interpretation. Given the recent theoretical advances in intelligence testing, there is no excuse for ignoring theory and blindly selecting intelligence tests to identify giftedness. To be specific, understanding that one measure of intelligence assesses the constructs of verbal comprehension and perceptual organization, while another assesses verbal ability, working memory, and processing speed, is important to consider when evaluators are choosing intelligence measures for making identification decisions. Also, it is important to understand how these psychological constructs are measured. Do the specific constructs use speeded tasks, visual tasks, or verbal tasks? Are students penalized for less *g*-loaded constructs such as working memory (McGowan et al., 2016)? Without an understanding of the underlying theory behind the aforementioned constructs, it will be difficult to interpret an individual's results on an intellectual measure. Also, if a cognitive measure is found to assess primarily verbal ability, it is likely that only gifted children with strong verbal skills will be identified, while children with excellent perceptual organization skills, strong speed-of-information-processing skills, or excellent working memory skills will be missed during the identification process.

Last, it is recommended that school systems consider using a cross-battery approach when assessing for giftedness. Specifically, the CHC cross-battery approach (Flanagan et al., Chapter 27, this volume; Flanagan, Ortiz, Alfonso, & Mascolo, 2002; McGrew & Flanagan, 1998) should be considered. This approach is a method for utilizing separate batteries of tests and ensuring a broader range of theoretical constructs in the assessment of children. Although research is needed to support using the CHC cross-battery approach in the identification of gifted children, it does demonstrate how multiple batteries can be used to decrease the reliance on a single measure of cognitive ability.

SUMMARY

This chapter has explored some of the central issues related to using intelligence tests in the identification of giftedness. Although identification of giftedness is complex and has been richly debated, our goal has been to advocate for school systems' making informed decisions about the role of intelligence tests in the identification process. Although many strategies exist for identifying gifted and talented children for services, it appears that the use of a specific measure of cognitive ability dominates the identification process within the schools. It is important to design programs that begin with a definition of giftedness and build from there. The definition precedes and guides the identification process, and provides the rationale for what instruments to use in order to find and then to serve the appropriate students. This chapter has focused on the need for precision in choosing effective theory-driven intelligence tests for the identification process. Many programs for gifted students are focused on cognitive ability, and yet misuse test data in making critical decisions. The current availability of cognitive assessments for identifying superior cognitive ability is better now than ever before. It remains to be seen whether those who administer the programs will choose instruments wisely, in order to obtain the information needed to identify those students who are in need of a different type of program because of their demonstrated cognitive ability.

REFERENCES

Baer, J., & Kaufman, J. (2004). Considering the DMGT: Something old, something new. *High Ability Studies, 15*(2), 149–150.

Baum, S., Owen, S. V., & Dixon, J. (1991). *To be gifted and learning disabled: From definitions to practical intervention strategies*. Mansfield Center, CT: Creative Learning Press.

Binet, A., & Simon, T. (1905). New methods for the diagnosis of intelligence in the child. *L'Année Psychologique, 11,* 191–244.

Borland, J. H. (1989). *Planning and implementing programs for the gifted*. New York: Teachers College Press.

Briggs, C. J., & Reis, S. M. (2004). An introduction to the topic of cultural diversity and giftedness. In C. Tomlinson, D. Ford, S. Reis, C. Briggs, & C. Strickland (Eds.), *In search of the dream: Designing schools and classrooms that work for high potential students from diverse cultural backgrounds* (pp. 5–32). Washington, DC: National Association for Gifted Children.

Brown, S. W., Hwang, M. T., Baron, M., & Yakimowski, M. E. (1991). Factor analysis of responses to the WISC-R for gifted children. *Psychological Reports, 69,* 99–107.

Brown, S. W., & Yakimowski, M. E. (1987). Intelligence scores of gifted students on the WISC-R. *Gifted Child Quarterly, 31,* 130–134.

Cameron, L. C., Ittenbach, R. F., McGrew, K. S., Harrison, P. L., Taylor, L. R., & Hwang, Y. R. (1997). Confirmatory factor analysis of the K-ABC with gifted referrals. *Educational and Psychological Measurement, 57,* 823–840.

Clark, B. (1997). *Growing up gifted*. Upper Saddle River, NJ: Prentice Hall.

Colangelo, N., & Davis, G. A. (1997). Introduction and overview. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 3–9). Boston: Allyn & Bacon.

Coleman, L. J., & Cross, T. L. (2005). *Being gifted in school* (2nd ed.). Waco, TX: Prufrock Press.

Cox, J., Daniel, N., & Boston, B. O. (1985). *Educating able learners*. Austin: University of Texas Press.

Coyle, D. (2009). *The talent code: Greatness isn't born. It's grown. Here's how.* New York: Bantam.

Dai, D. Y., & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly, 57,* 151–168.

Darwin, C. (1859). *On the origin of species*. London: Murray.

Davis, G. A., & Rimm, S. (1985). *Education of the gifted and talented*. Englewood Cliffs, NJ: Prentice-Hall.

Davis, G. A., & Rimm, S. (1998). *Education of the gifted and talented* (4th ed.). Boston: Allyn & Bacon.

Davis, G. A., Rimm, S., & Siegle, D. (2011). *Education of the gifted and talented* (6th ed.). Upper Saddle River, NJ: Pearson.

Every Student Succeeds Act of 2015 (ESSA), Pub. L. No. 114-95, 20 U.S.C. 6301 et seq. (2015).

Feldhusen, J. F. (1998). A conception of talent and talent development. In R. C. Friedman & K. B. Rogers (Eds.), *Talent in context: Historical and social perspectives on giftedness* (pp. 193–209). Washington, DC: American Psychological Association.

Feldhusen, J. F., & Jarwan, F. A. (1993). Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 233–251). New York: Pergamon Press.

Fishkin, A. S., Kampsnider, J. J., & Pack, L. (1996). Exploring the WISC-III as a measure of giftedness. *Roeper Review, 18,* 226–231.

Flanagan, D. P., & Ortiz, S. O. (2002). Best practices in intellectual assessment: Future directions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1351–1372). Bethesda, MD: National Association of School Psychologists.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities.* Boston: Allyn & Bacon.

Flynn, J. R. (1987). Massive gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Ford, D. Y., & Harris, J. J. (1999). *Multicultural gifted education.* New York: Teachers College Press.

Fox, L. H. (1981). Identification of the academically gifted. *American Psychologist, 36,* 1103–1111.

Frasier, M. M., Hunsaker, S. L., Lee, J., Finely, V. S., Frank, E., Garcia, J. H., et al. (1995). *Educators' perceptions of barriers to the identification of gifted children from economically disadvantaged and limited English proficient backgrounds* (Report No. RM95216). Storrs: University of Connecticut, National Research Center on the Gifted and Talented.

Gagné, F. (1999). My convictions about the nature of abilities, gifts, and talents. *Journal for the Education of the Gifted, 22,* 109–136.

Gagné, F. (2004). Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies, 15*(2), 119–147.

Gallagher, J. J. (1995). Education of gifted students: A civil rights issue? *Phi Delta Kappan, 76*(5), 408–410.

Gallagher, J. J. (2002). *Society's role in educating gifted students: The role of public policy* (Report No. RM02162). Storrs: University of Connecticut, National Research Center on the Gifted and Talented.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H. (1999). Are there additional intelligences?: The case for naturalist, spiritual, and existential intelligences. In J. Kane (Ed.), *Education, information, and transformation* (pp. 111–131). Upper Saddle River, NJ: Prentice Hall.

Gladwell, M. (2008). *Outliers: The story of success.* New York: Little, Brown.

Greenberg, R. D., Stewart, K. J., & Hansche, W. J. (1986). Factor analysis of the WISC-R for white and black children evaluated for gifted placement. *Journal of Psychoeducational Assessment, 4,* 123–130.

Gridley, B. E., Norman, K. A., Rizza, M. G., & Decker, S. L. (2003). Assessment of gifted children with the Woodcock–Johnson III. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 285–317). San Diego, CA: Academic Press.

Gross, M. U. M. (1993). Nurturing the talents of exceptionally gifted individuals. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 473–490). New York: Pergamon Press.

Harris, B., Plucker, J. A., Rapp, K. E., & Martinez, R. S. (2009). Identifying gifted and talented English-language learners: A case study. *Journal for the Education of the Gifted, 32,* 368–393.

Hollinger, C. L. (1986). Beyond the use of full scale IQ scores. *Gifted Child Quarterly, 30,* 74–77.

Hollingworth, L. S. (1942). *Children above 180 IQ.* Yonkers, NY: World Book.

Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. *Journal of Educational Psychology, 72,* 197–199.

Karnes, F. A., Edwards, R. P., & McCallum, R. S. (1986). Normative achievement assessment of gifted children: Comparing the K-ABC, WRAT, and CAT. *Psychology in the Schools, 23,* 346–352.

Kaufman, A. S. (1975). Factor structure of the WISC-R at eleven age levels between 6½ to 16½ years. *Journal of Consulting and Clinical Psychology, 43,* 135–147.

Kaufman, A. S. (1979). WISC-R research: Implications for interpretation. *School Psychology Digest, 18,* 5–27.

Kaufman, A. S. (1992). Evaluation of the WISC-III and WPPSI-R for gifted children. *Roeper Review, 14,* 154–158.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition manual.* Circle Pines, MN: American Guidance Service.

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly, 52,* 275–296.

Luria, A. R. (1987). *The mind of a mnemonist: A little book about a vast memory.* Cambridge, MA: Harvard University Press.

Lyon, G. R., Gray, D. B., Kavanagh, J. F., & Krasnegor, N. A. (Eds.). (1993). *Better understanding learning disabilities: New views from research and their implications for education and public policies.* Baltimore: Brookes.

Macmann, G. M., Plasket, C. M., Barnett, D. W., & Siler, R. F. (1991). Factor structure of the WISC-R for children of superior intelligence. *Journal of School Psychology, 29,* 19–36.

Malone, P. S., Brock, V., Brounstein, P. J., & Shaywitz, S. E. (1991). Components of IQ scores across levels of measured ability. *Journal of Applied Social Psychology, 21,* 15–28.

Masten, W. G., Morse, D. T., & Wenglar, K. E. (1995). Factor structure of the WISC-R for Mexican-American students referred for intellectually gifted assessment. *Roeper Review, 18*(2), 130–131.

McGowan, M. R., Holtzman, D. R., Coyne, T. B., & Miles, K. L. (2016). Predictive ability of the SB5 Gifted Composite versus the Full-Scale IQ among children referred for gifted evaluations. *Roeper Review, 38*(1), 40–49.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross battery assessment.* Boston: Allyn & Bacon.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual: Woodcock–Johnson III Normative Update.* Rolling Meadows, IL: Riverside.

Mishra, S. P., Lord, J., & Sabers, D. L. (1989). Cognitive processes underlying WISC-R performance of gifted and learning disabled Navajos. *Psychology in the Schools, 26,* 31–36.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System.* Itasca, IL: Riverside.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition.* Austin, TX: PRO-ED.

National Association for Gifted Children (NAGC). (1997). *The use of tests in the identification and assessment of gifted students.* Washington, DC: Author.

National Association for Gifted Children (NAGC). (2010). *Redefining giftedness for a new century: Shifting the paradigm.* Washington, DC: Author. Retrieved from *www.nagc.org/sites/default/files/Position%20Statement/Redefining%20Giftedness%20for%20a%20New%20Century.pdf.*

National Center for Education Statistics. (2005). *The condition of education 2005* (NCES Publication No. 2005-094). Washington, DC: U.S. Government Printing Office.

Oden, M. H. (1968). The fulfillment of promise: 40 year follow-up of the Terman gifted group. *Genetic Psychology Monographs, 77,* 3–93.

O'Donnell, L. (2009). The Wechsler Intelligence Scale for Children—Fourth Edition. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 153–190). Hoboken, NJ: Wiley.

Olenshak, F. R., & Reis, S. M. (2002). Gifted students with learning disabilities. In M. Neihart, S. M. Reis, N. Robinson, & S. Moon (Eds.), *The social and emotional development of gifted children* (pp. 177–191). Waco, TX: Prufrock Press.

Ortiz, S. O., & Dynda, A. M. (2005). Use of intelligence tests with culturally and linguistically diverse populations. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 545–556). New York: Guilford Press.

Peters, S. J., Matthews, M. S., McBee, M. T., & McCoach, D. B. (2014). *Beyond gifted education.* Waco, TX: Prufrock Press.

Pyryt, M. C. (1996). IQ: Easy to bash, hard to replace. *Roeper Review, 18,* 255–258.

Raiford, S. E., Drozdick, L., Zhang, O., & Zhou, X. (2015). *WISC-V: Expanded index scores* (Technical Report No. 1). Bloomington, MN: Pearson.

Ramos-Ford, V., & Gardner, H. (1997). Giftedness from a multiple intelligences perspective. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 54–66). Boston: Allyn & Bacon.

Reams, R., Chamrad, D., & Robinson, N. (1990). The race is not necessarily to the swift: Validity of the WISC-R bonus points for speed. *Gifted Child Quarterly, 34,* 108–110.

Renzulli, J. S. (1978). What makes giftedness: Reexamining a definition. *Phi Delta Kappan, 60,* 18–24.

Renzulli, J. S., & Delcourt, M. A. (1986). The legacy and logic of research on the identification of gifted persons. *Gifted Child Quarterly, 30,* 20–23.

Renzulli, J. S., & Reis, S. M. (1997). *The schoolwide enrichment model: A how-to guide for educational excellence.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., Reis, S. M., & Smith, L. H. (1981). *The revolving door identification model.* Mansfield Center, CT: Creative Learning Press.

Reynolds, C. R., & Kaiser, S. M. (1990). Bias in assessment of attitude. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 611–653). New York: Guilford Press.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales.* Lutz, FL: Psychological Assessment Resources.

Rindermann, H. (2007). The *g*-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667–706.

Rizza, M. G., McIntosh, D. E., & McCunn, A. (2001). Profile analysis of the Woodcock–Johnson III Tests of Cognitive Abilities with gifted students. *Psychology in the Schools, 38,* 447–455.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Rowe, E. W., Dandridge, J., Pawlush, A., Thompson, D. F., & Ferrier, D. E. (2014). Exploratory and confirmatory factor analyses of the WISC-IV with gifted students. *School Psychology Quarterly, 29,* 536–552.

Sabotnik, R. F., & Arnold, K. D. (1994). *Beyond Terman: Contemporary longitudinal studies of giftedness and talent.* Norwood, NJ: Ablex.

Sapp, G. L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. *Psychological Reports, 57,* 947–951.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014).

Woodcock–Johnson IV. Rolling Meadows, IL: Riverside.

Simonton, D. K. (2008). Childhood giftedness and adult genius: A historimetric analysis of 291 eminent African Americans. *Gifted Child Quarterly, 52,* 243–255.

Simpson, M., Carone, D. A., Jr., Burns, W. J., Seidman, D. M., & Sellers, A. (2002). Assessing giftedness with the WISC-III and the SB-IV. *Psychology in the Schools, 39,* 515–524.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge, UK: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (1999). *Our labeled children.* Reading, MA: Perseus Books.

Sternberg, R. J., & Kaufman, J. C. (2001). *The evolution of intelligence.* Mahwah, NJ: Erlbaum.

Sternberg, R. J., & Williams, W. (2002). *Educational psychology.* Boston: Allyn & Bacon.

Sternberg, R. J., & Zhang, L. (1995). What do we mean by giftedness?: A pentagonal implicit theory. *Gifted Child Quarterly, 39,* 88–94.

Strauss, E., Sherman, E. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Tannenbaum, A. J. (1983). *Gifted children: Psychological and educational perspectives.* New York: Macmillan.

Tannenbaum, A. J. (1997). The meaning and making of giftedness. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 27–42). Boston: Allyn & Bacon.

Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.

Terman, L. M. (1925a). *Genetic studies of genius: Mental and physical traits of a thousand gifted children* (Vol. 1). Stanford, CA: Stanford University Press.

Terman, L. M. (1925b). *Genetic studies of genius: Mental and physical traits of a thousand gifted children* (Vol. 4). Stanford, CA: Stanford University Press.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Treffinger, D. J. (1995). Talent development: An emerging view. *Understanding Our Gifted, 7*(4), 3.

Tyler-Wood, R., & Carri, L. (1991). Identification of gifted children: The effectiveness of various measures of cognitive ability. *Roeper Review, 14,* 63–64.

Volker, M. A., Lopata, C., & Cook-Cottone, C. (2006). Assessment of children with intellectual giftedness and reading disabilities. *Psychology in the Schools, 43*(8), 855–869.

Wagenmakers, E., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions In Psychological Science, 25*(3), 169–176.

Warne, R. T. (2016). Five reasons to put the *g* back into giftedness. *Gifted Child Quarterly, 60*(1), 3–15.

Watkins, M. W., Greenwalt, C. G., & Marcell, C. M. (2002). Factor structure of the Wechsler Intelligence Scale for Children—Third Edition among gifted students. *Educational and Psychological Measurement, 62,* 164–172.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children—Fifth Edition. Technical and Interpretive Manual.* Bloomington MN: Pearson.

White, W. L., & Renzulli, J. S. (1987). A forty year follow-up of students who attended Leta Hollingworth's School for Gifted Children. *Roeper Review, 10*(2), 89–94.

Winner, E. (1996). *Gifted children: Myths and realities.* New York: Basic Books.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside.

Zhu, J., Cayton, T., Weiss, L., & Gabel, A. (2008). *Wechsler Intelligence Scale for Children—Fourth Edition: WISC-IV extended norms* (Technical Report No. 7). San Antonio, TX: Pearson.

Ziegler, A., & Ziegler, A. (2009). The paradoxical attenuation effect in tests based on classic test theory: Mathematical background and practical implications for the measurement of high abilities. *High Ability Studies, 20*(1), 5–14.

CHAPTER 22

Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition

Dawn P. Flanagan
Vincent C. Alfonso
Michael Costa
Katherine Palma
Meghan A. Leahy

> There are no rules for converting concepts to operational definitions. Therefore, operational definitions are judged by *significance* (i.e., is it an authoritative marker of the concept?) and *meaningfulness* (i.e., is it a rational and logical marker of the concept?).
>
> —KAVALE, SPAULDING, AND BEAM (2009, p. 41)

Historically, identification of specific learning disabilities (SLD) has almost always included a consideration of an individual's overall cognitive ability, as well as his or her unique pattern of strengths and weaknesses (Sotelo-Dynega, Flanagan, & Alfonso, 2018). However, in recent years intelligence tests and tests of specific cognitive abilities and neuropsychological processes have come under harsh attack as useful tools in the identification of SLD (for a review, see Schneider & Kaufman, 2017). Although "IQ" tests have always had their critics, it was not until the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA, 2004) and the publication of its attendant regulations (U.S. Department of Education, 2006) that such criticism became more widespread (Fletcher-Janzen & Reynolds, 2008).

It is beyond the scope of this chapter to review all the issues surrounding the debate about the utility (or lack thereof) of cognitive and neuropsychological tests in the identification of SLD. The interested reader is referred to Alfonso and Flanagan (2018) for a comprehensive treatment of these issues. Based on our review of the literature and our clinical experience, we find inherent utility in cognitive and neuropsychological assessment for SLD identification and treatment. Therefore, the purposes of this chapter are to describe Flanagan, Ortiz, and Alfonso's (2013, 2017) operational definition of SLD, and to highlight the importance and utility of gathering data from cognitive and neuropsychological tests (among other quantitative and qualitative data sources) within this framework (see also Flanagan, Alfonso, Sy, et al., 2018).

This chapter was adapted from Flanagan, Alfonso, Sy, Mascolo, McDonough, and Ortiz (2018). Copyright © John Wiley & Sons, Inc. Adapted by permission All rights reserved.

A BRIEF PERSPECTIVE ON THE DEFINITION OF SLD

IDEA 2004 defines *SLD* as "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. Such terms include such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia," and SLD "does not include learning problems that are primarily the result of visual, hearing, or motor disabilities; of intellectual disability; of emotional disturbance; or of environmental, cultural, or economic disadvantage." Many researchers in the field (e.g., Kavale, Spaulding, & Beam, 2009) have argued that the federal definition of SLD in IDEA 2004 and its regulations do not reflect the best thinking about the SLD construct because it has not changed in well over 30 years. This fact is astonishing, as several decades of inquiry into the nature of SLD resulted in numerous (but unsuccessful) proposals over the years to modify the definition. If the field of SLD is to recapture its status as a reliable entity in special education and psychology, then more attention must be paid to the federal definition (Kavale & Forness, 2000). To bring clarity to the definition, Kavale and colleagues (2009) specified the boundaries of the term and the class of things to which it belongs. In addition, their definition delineates what SLD is and what it is not. Although their description is not a radical departure from the federal definition, it provides a more comprehensive description of the nature of SLD. The Kavale and colleagues definition is as follows:

> Specific learning disability refers to heterogeneous clusters of disorders that significantly impede the normal progress of academic achievement . . . The lack of progress is exhibited in school performance that remains below expectation for chronological and mental ages, even when [the student is] provided with high-quality instruction. The primary manifestation of the failure to progress is significant underachievement in a basic skill area (i.e., reading, math, writing) that is not associated with insufficient educational, cultural/familial, and/or sociolinguistic experiences. The primary severe ability–achievement discrepancy is coincident with deficits in linguistic competence (receptive and/or expressive), cognitive functioning (e.g., problem solving, thinking abilities, maturation), neuropsychological processes (e.g., perception, attention, memory), or any combination of such contributing deficits that are presumed to originate from central nervous system dysfunction. The specific learning disability is a discrete condition differentiated from generalized learning failure by average or above (>90) cognitive ability and a learning skill profile exhibiting significant scatter indicating areas of strength and weakness. The major specific learning disability may be accompanied by secondary learning difficulties that also may be considered when [educators are] planning the more intensive, individualized special education instruction directed at the primary problem. (p. 46)

Kavale and colleagues state that their richer description of SLD "can be readily translated into an operational definition providing more confidence in the validity of a diagnosis of SLD" (p. 46). In the following section, we describe an operational definition of SLD that captures the nature of SLD as reflected in the federal definition and in Kavale and colleagues' definition. In addition, the reasons why operational definitions are important and necessary for SLD identification are highlighted.

THE NEED FOR AN OPERATIONAL DEFINITION OF SLD

An operational definition of SLD is needed to provide more confidence in the validity of the SLD diagnosis (Flanagan, Fiorello, & Ortiz, 2010; Flanagan & Schneider, 2016; Kavale et al., 2009). An *operational definition* provides a process for the identification and classification of concepts that have been defined formally (see Sotelo-Dynega, Flanagan, & Alfonso, 2018, for a summary). With no change in the federal definition of SLD, the field has turned to articulating ways to operationalize SLD, with the intent of improving the clinical identification of this condition (Alfonso & Flanagan, 2018; Flanagan et al., 2013; Flanagan, Ortiz, Alfonso, & Mascolo, 2002, 2006; Kavale & Flanagan, 2007; Kavale & Forness, 2000; Kavale et al., 2009; Schneider & Kaufman, 2017; Swanson, 1991).

For more than three decades, the main operational definition of SLD was the so-called "discrepancy criterion." Discrepancy was first introduced in Bateman's (1965) definition of learning disabilities (LD) and later was formalized in federal regulations as follows:

> (1) The child does not achieve commensurate with his or her age and ability when provided with appropriate educational experiences, and (2) the child has

> a *severe discrepancy between achievement and intellectual ability* in one or more areas relating to communication skills and mathematics abilities. (U.S. Office of Education, 1977, p. 65083; emphasis added)

Several problems with the traditional ability–achievement discrepancy approach to SLD identification have been discussed extensively in the literature and are highlighted elsewhere (e.g., Hale, Wycoff, & Fiorello, 2011; see also Fiorello & Wycoff, Chapter 26, this volume); therefore, they are not repeated here.

With the reauthorization of IDEA in 2004, and the corresponding deemphasis on the traditional ability–achievement discrepancy criterion for SLD identification, there have been several attempts to operationalize the federal definition, many of which can be found in Alfonso and Flanagan (2018). Table 22.1 provides examples of how the 2004 federal definition of SLD has been operationalized.

One of the most comprehensive operational definitions of SLD was described nearly 20 years ago by Kavale and Forness (2000). These researchers critically reviewed the available definitions of learning disability and methods for their operationalization, and found them to be largely inadequate. Therefore, they proposed a modest, hierarchical operational definition that reflected current research (at the time) on the nature of SLD. Their operational definition is illustrated in Figure 22.1.

In their definition, Kavale and Forness (2000) attempted to incorporate the complex and multivariate nature of SLD. Figure 22.1 shows that SLD is determined through evaluation of performance at several "levels," each of which specifies particular diagnostic conditions. Furthermore, each level of the evaluation hierarchy depicted in Figure 22.1 represents a necessary, but not sufficient, condition for SLD determination. Kavale and Forness contended that it is only when the specified criteria are met at all five levels of their operational definition that SLD can be established as a "discrete and independent condition" (p. 251). Through their operational definition, Kavale and Forness provided a much more rational and defensible approach to the practice of SLD identification than that which had been offered previously. In short, their operationalization of SLD used "foundation principles in guiding the selection of elements that explicate the nature of LD" (p. 251); this represented both a departure from and an important new direction for current practice.

Flanagan and colleagues (2002) identified some aspects of Kavale and Forness's (2000) operational definition that they believed needed to be modified. For example, although Kavale and Forness's operational definition captured the complex and multivariate nature of SLD, it was not predicated on any particular theoretical model, and it did not specify what methods might be used to satisfy criteria at each level. In addition, the hierarchical structure depicted in Figure 22.1 seems to imply somewhat of a linear approach to SLD identification, whereas the process is typically more recursive and iterative. Consequently, Flanagan and colleagues proposed a similar operational definition of SLD, but based their definition primarily on the Cattell–Horn–Carroll (CHC) theory and its research base. In addition, these researchers provided greater specification of methods and criteria that may be used to identify SLD (e.g., Flanagan et al., 2013).

Because operational definitions represent only temporary assumptions about a concept, they are subject to change (Kavale et al., 2009). Flanagan and colleagues modify their operational definition routinely to ensure that it reflects the most current theory, research, and thinking with regard to (1) the nature of SLD; (2) the methods of evaluating various elements and concepts inherent in SLD definitions (viz., the federal definition); and

TABLE 22.1. Examples of How the IDEA 2004 Federal Definition of SLD Has Been Operationally Defined

- Absolute low achievement (for a discussion, see Burns, Maki, Warmbold-Brann, & Preast, 2018; Fletcher & Miciak, 2018; Lichtenstein & Klotz, 2007)
- Ability–achievement discrepancy (see Zirkel & Thomas, 2010, for a discussion)
- Failure to respond to scientifically based intervention (see Balu et al., 2015; Fletcher, Barth, & Stuebing, 2011; Fletcher, Lyon, Fuchs, & Barnes, 2007; Fletcher & Miciak, 2018; Fuchs & Fuchs, 1998; Hosp, Hosp, & Howell, 2007)
- Pattern of strengths and weaknesses (PSW; also called *alternative research-based approach* or *third-method* approach) (see Flanagan, Ortiz, & Alfonso, 2013, 2017; Hale, Flanagan, & Naglieri, 2008; Hale, Wycoff, & Fiorello, 2011; see also Alfonso & Flanagan, 2018, for a review of prominent PSW methods)

Note. All examples in this table include a consideration of exclusionary factors as specified in the federal definition of SLD.

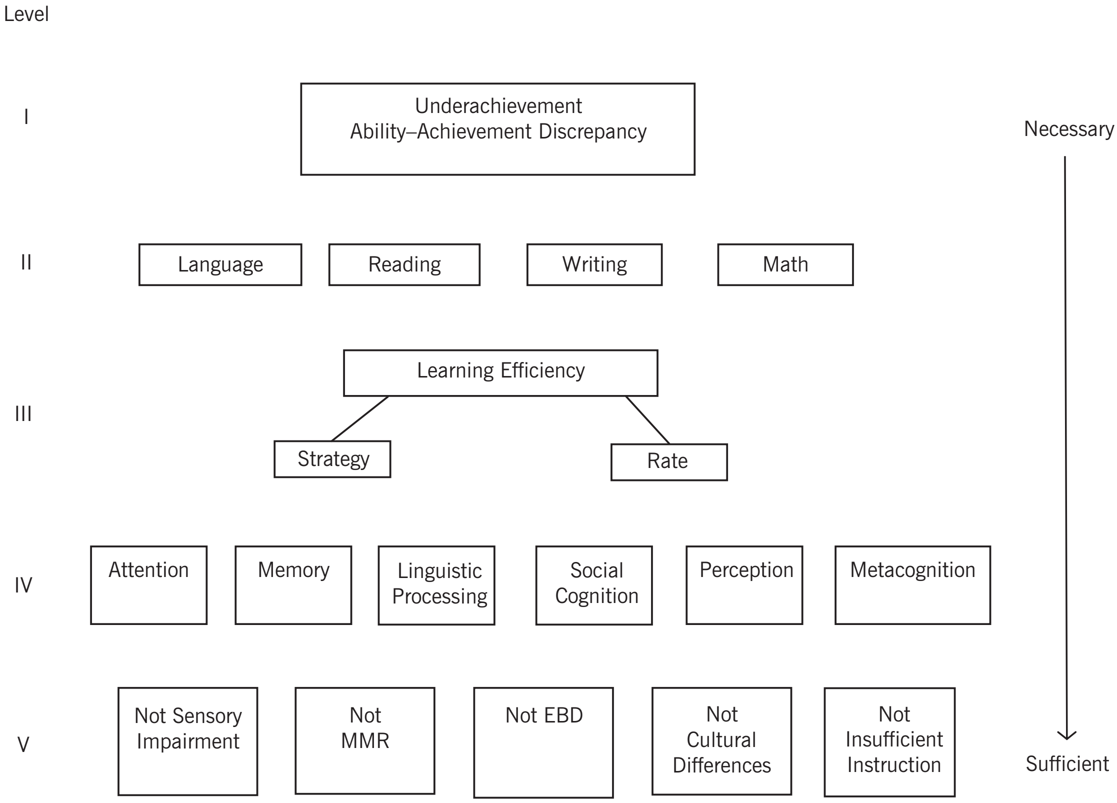


FIGURE 22.1. Kavale and Forness's operational definition of SLD. MMR, mild mental retardation (now called mild intellectual disability); EBD, emotional or behavioral disorder. From Kavale and Forness (2000). Copyright © by SAGE Publications, Inc. Reprinted by permission of SAGE Publications, Inc.

(3) criteria for establishing SLD as a discrete condition separate from undifferentiated low achievement and overall below-average ability to think and reason, particularly for the purpose of acquiring, developing, and applying academic skills (Flanagan et al., 2017; Flanagan, Alfonso, Sy, et al., 2018). Their operational definition of SLD is now referred to as the *dual-discrepancy/consistency* (DD/C) method (Flanagan et al., 2013, 2017; Flanagan, Alfonso, Sy, et al., 2018) and is presented in Figure 22.2. Flanagan and colleagues' approach to SLD identification encourages a continuum of data-gathering methods, beginning with curriculum-based measurement (CBM) and progress monitoring, and culminating in norm-referenced tests of cognitive abilities and neuropsychological processes for students who demonstrate an inadequate response to intervention.

Figure 22.2 shows that the DD/C operational definition of SLD is arranged according to levels, as is Kavale and Forness's (2000) definition. At each level, the definition includes (1) defining characteristics regarding the nature of SLD (e.g., an individual has difficulties in one or more areas of academic achievement); (2) the focus of evaluation for each characteristic (e.g., academic achievement, cognitive abilities/neuropsychological processes, exclusionary factors); (3) examples of evaluation methods and relevant data sources (e.g., standardized, norm-referenced tests and educational records, respectively); and (4) the criteria that need to be met to establish that an individual possesses a particular characteristic of SLD (e.g., below-average performance in an academic area, such as basic reading skill; cognitive processing weaknesses that are related to the academic skill weaknesses). As may be seen in Figure 22.2, the "Nature of SLD" column includes a description of what SLD is and what it is not. Overall, the levels represent adaptations and extensions of the recommendations offered by Kavale and colleagues (e.g., Kavale & For-

Level	Nature of SLD[a]	Focus of Evaluation	Examples of Evaluation Methods and Data Sources	Criteria for SLD	SLD Classification and Eligibility
I	Difficulties in one or more areas of academic achievement, including (but not limited to)[b] basic reading skill, reading comprehension, reading fluency, oral expression, listening comprehension, written expression, math calculation, and math problem solving.	**Academic Achievement:** Performance in specific academic skills (e.g., Grw [reading decoding, reading fluency, reading comprehension, spelling, written expression], Gq [math calculation, math problem solving], and Gc [communication ability, listening ability]).	Response to high-quality instruction and intervention via progress monitoring; performance on norm-referenced; standardized achievement tests; evaluation of work samples; observations of academic performance; teacher/parent/student interview; history of academic performance; and data from other members of the multidisciplinary team (MDT) (e.g., speech–language pathologist, interventionist, reading specialist).	Performance in one or more academic areas is *weak or deficient*[c] (despite attempts at delivering higher-quality instruction) as evidenced by converging data sources. Note that low scores are not sufficient to meet this condition. These scores must also represent *unexpected underachievement* (a condition determined by X-BASS, based on an individual's unique pattern of scores).	**Necessary**
II	SLD does not include a learning problem that is the result of visual, hearing, or motor disabilities; of intellectual disability; of social or emotional difficulty or disorder; or of environmental, educational, cultural, or economic disadvantage.	**Exclusionary Factors:** Identification of potential primary causes of academic skill weaknesses or deficits, including intellectual disability, cultural or linguistic difference, sensory impairment, insufficient instruction or opportunity to learn, organic or physical health factors, social-emotional or psychological difficulty or disorder.	Data from the methods and sources listed at levels I and III; behavior rating scales; medical records; prior evaluations; interviews with current or past professionals (counselors, psychiatrists, etc.)	Performance is not *primarily* attributed to these exclusionary factors, although one or more of them may contribute to learning difficulties. (Consider using the Exclusionary Factors form, which is included in this chapter as Figure 22.3.)	
III	A disorder in one or more of the basic psychological/neuropsychological processes involved in understanding or in using language, spoken or written; such disorders are presumed to originate from central nervous system dysfunction.	**Cognitive Abilities and Processes:** Performance in cognitive abilities and processes (e.g., Gv, Ga, Glr, Gsm, Gs), specific neuropsychological processes (e.g., attention, executive functioning, orthographic processing, rapid automatic naming), and learning efficiency (e.g., associative memory, free-recall memory, meaningful memory).	Performance on norm-referenced tests; evaluation of work samples; observations of cognitive performance; task analysis; testing limits; teacher/parent/student interview; history of academic performance; and records review.	Performance in one or more cognitive abilities and/or neuropsychological processes (related to academic skill deficiency) is *weak or deficient*[c] as evidenced by converging data sources. Note that low scores are not sufficient to meet this condition. The cognitive ability or process in question must also be *domain-specific* (a condition determined	

				by X-BASS, based on an individual's unique pattern of scores).	↓
IV	The SLD is a discrete condition differentiated from generalized learning deficiency by generally average or better ability to think and reason and a learning skill profile exhibiting significant variability, indicating a pattern of cognitive and academic strengths and weaknesses.	**Pattern of Strengths and Weaknesses (PSW) Marked by a Dual-Discrepancy/Consistency (DD/C) Situation:** Determination of whether academic skill weaknesses or deficits are *unexpected* and related to *domain-specific* cognitive weaknesses or deficits; pattern of data reflects a below-average aptitude–achievement *consistency* with at least *average ability* to think and reason.	Data gathered at all previous levels, as well as any additional data following a review of initial evaluation results (e.g., data gathered for hypothesis testing; data gathered via demand analysis and limits testing).	Circumscribed below-average aptitude–achievement *consistency*; circumscribed ability–achievement and ability–cognitive aptitude *discrepancies*, with at least average ability to think and reason; clinical judgment supports the impression that the student's overall ability to think and reason will enable him or her to benefit from tailored or specialized instruction/intervention, compensatory strategies, and accommodations, such that his or her performance rate and level will likely approximate those of more typically achieving, nondisabled peers. The DD/C PSW analysis is conducted by X-BASS, based on an individual's unique pattern of strengths and weaknesses.	**Sufficient for SLD Identification**
V	The SLD has an adverse impact on educational performance.	**Special Education Eligibility:**[d] Determination of least restrictive environment (LRE) for delivery of instruction and educational resources.	Data from all previous levels and MDT meetings.	Student demonstrates significant difficulties in daily academic activities that cannot be remediated, accommodated, or otherwise compensated for *without* the assistance of individualized special education services.	**Necessary for Special Education Eligibility**

[1] This column includes concepts inherent in the federal definition (IDEA, 2004), Kavale, Spaulding, and Beam's (2009) definition, Harrison and Holmes's (2012) consensus definition, and other prominent definitions of SLD (see Sotelo-Dynega, 2018). Thus, he most salient prominent SLD markers are included in this column.

[2] Poor spelling with adequate ability to express ideas in writing is often typical of dyslexia and/or dysgraphia. Even though IDEA 2004 includes only the broad category of written expression, poor spelling and handwriting are often symptomatic of a specific writing disability and should not be ignored (Wendling & Mather, 2009).

[3] Weak performance is typically associated with standard scores in the 85–89 range, whereas deficient performance is often associated with standard scores that are greater than 1 *SD* below the mean. Interpretations of weak or deficient performance based on standard scores that fall in the weak and deficient ranges are bolstered when they have ecological validity (e.g., when there is evidence that the abilities or processes identified as weak or deficient manifest in everyday classroom activities that require these abilities and processes).

[4] The major SLD may be accompanied by secondary learning difficulties that should be considered in planning the more intensive, individualized special education instruction directed at the primary problem. For information on linking assessment data to intervention, see Mascolo, Alfonso, and Flanagan (2014).

FIGURE 22.2. A CHC-based operational definition of SLD; this approach is now known as the DD/C method. X-BASS, Cross-Battery Assessment Software System (Flanagan, Ortiz, & Alfonso, 2017). Based on Flanagan and Alfonso (2017) and Flanagan, Ortiz, and Alfonso (2013).

ness, 2000; Kavale et al., 2009), but they also include concepts from various other researchers (e.g., Berninger, 2011; Decker, Bridges, & Vetter, 2018; Fletcher-Janzen & Reynolds, 2008; Geary, Hoard, & Bailey, 2011; Geary et al., 2017; Hale & Fiorello, 2004; Hale et al., 2010; Mazzocco & Vukovic, 2018; Nelson & Wiig, 2018; Reynolds & Shaywitz, 2009a, 2009b; Siegel, 1999; Stanovich, 1999; Vellutino, Scanlon, & Lyon, 2000).

The DD/C operational definition of SLD presented in Figure 22.2 differs from the one presented by Kavale and Forness (2000; see Figure 22.1) in four important ways. First, it is grounded in a well-validated contemporary theory on the structure of abilities (i.e., CHC theory; see Schneider & McGrew, Chapter 3, this volume, for a description). Second, in lieu of the traditional ability–achievement discrepancy method, a specific pattern of cognitive and academic ability and neuropsychological processing strengths and weaknesses is used as a defining characteristic or marker for SLD.[1] (It is important to understand that any pattern used for SLD determination should be supported by research on the relations among CHC abilities, processes, and academic outcomes—and, where possible, evidence on the neurobiological correlates of learning disorders in reading, math, and writing; see McDonough, Flanagan, Sy, & Alfonso, 2017.) Third, the evaluation of exclusionary factors occurs earlier in the SLD identification process in our operational definition, to prevent individuals from having to undergo additional testing. Fourth, we emphasize that SLD assessment is a recursive process rather than a linear one, and that information generated and evaluated at one level may inform decisions made at other levels. The recursive nature of the SLD identification process is reflected by the circular arrows in Figure 22.2. Each level of the CHC-based operational definition of SLD is described in more detail in the next section.

THE DD/C OPERATIONAL DEFINITION OF SLD

> A diagnosis identifies the nature of a specific learning disability and has implications for its probable etiology, instructional requirements, and prognosis. Ironically, in an era when educational practitioners are encouraged to use evidence-based instructional practices, they are not encouraged to use evidence-based differential diagnoses of specific learning disabilities.
>
> —BERNINGER (2011, p. 204)

According to the U.S. Department of Education (2006) regulations, there are three permissible methods for the identification of SLD: (1) traditional ability–achievement discrepancy, (2) response to intervention, and (3) alternative research-based approaches. The DD/C operational definition of SLD is consistent with the alternative research-based "third option" for SLD identification. The DD/C definition is grounded primarily in CHC theory, but has been extended to include important neuropsychological functions that are not explicit in CHC theory (e.g., executive functions, orthographic processing, cognitive efficiency). The essential elements in evaluation of SLD in the DD/C definition include (1) academic ability analysis, (2) evaluation of mitigating and exclusionary factors, (3) cognitive ability and processing analysis, (4) pattern of strengths and weaknesses (PSW) analysis, and (5) evaluation of interference with learning for purposes of special education eligibility.

It is assumed that the levels of evaluation depicted in Figure 22.2 are undertaken after it has been determined that the student is demonstrating an inadequate response to high-quality instruction and pre-referral interventions (consistent with tiers 1 and 2 of a response-to-intervention [RTI] approach or a multi-tiered system of support [MTSS]) have been conducted with little or no success, and therefore a focused evaluation of specific abilities and processes through standardized testing of cognitive and academic abilities is deemed necessary (Flanagan, Fiorello, et al., 2010; see also McCloskey, Slonim, & Rumohr, Chapter 39, this volume). Evaluation for the presence of an SLD assumes that an individual has been referred for testing specifically because of observed learning difficulties. Moreover, before a formal SLD assessment is begun, other data from multiple sources may have (and probably should have) already been uncovered within the context of intervention implementation. These data may include results from CBM, progress monitoring, informal testing, direct observation of behaviors, work samples, reports from people familiar with the child's difficulties (e.g., teachers, parents), and information provided by the child him- or herself. This type of systematic approach to understanding learning difficulties can emanate from any well-researched theory (e.g., Decker et al., 2018; Hale, Wycoff, & Fiorello, 2011; McCloskey, Whitaker, Murphy, & Rogers, 2012; McDonough & Flanagan, 2016). A summary of each level of the DD/C operational definition of SLD follows.

Level I: Difficulties in One or More Areas of Academic Achievement

SLD is marked by dysfunction in learning. That is, learning is somehow disrupted from its normal course by an underlying ability and processing deficits. Although the specific mechanism that inhibits learning is not directly observable, one can proceed on the assumption that it manifests in observable phenomena, particularly academic achievement. Thus level I of the operational definition involves documenting that some type of *learning* deficit exists. In the DD/C definition, the presence of a *weakness* or *normative weakness/deficit* (Table 22.2) —established through standardized testing of the major areas of academic achievement (e.g., reading, writing, math, oral language), and supported through other means, such as clinical observations of academic performance, work samples, and parent and teacher reports—is a necessary but insufficient condition for SLD determination. A finding of a weakness in academic achievement is not sufficient for SLD identification because this condition alone may be present for a variety of reasons, only one of which is SLD. Furthermore, the academic area of weakness must also meet criteria for *unexpected underachievement,* as discussed later in this chapter.

The academic areas that are generally assessed at level I in the DD/C operational definition include the eight areas of achievement specified in the federal definition of SLD (IDEA, 2004). These eight areas are basic reading skills, reading fluency, reading comprehension, math calculation, math problem solving, written expression, listening comprehension, and oral expression. Most of the skills and abilities measured at level I represent an individual's stores of acquired knowledge. These specific knowledge bases (e.g., quantitative knowledge [Gq], reading and writing ability [Grw], vocabulary knowledge [Gc]) develop largely as a function of formal instruction, schooling, and educationally related experiences (Carroll, 1993). Typically, the eight areas of academic achievement are measured via standardized, norm-referenced tests. In fact, many comprehensive achievement batteries, such as the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), the Woodcock–Johnson IV Tests of Achievement (Schrank, Mather, & McGrew, 2014), and the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014), measure all eight areas (see Table 22.3). It is important to realize that data on academic performance should come from multiple sources (see Figure 22.2, level I, column 4).

If weaknesses or deficits in the child's academic achievement profile are not identified, then the issue of SLD may be moot because such weakness-

TABLE 22.2. Definition of Weakness and Normative Weakness or Deficit

Term or concept	Meaning within the context of DD/C	Comments
Weakness	Performance on standardized, norm-referenced tests that falls *below average* (where average is defined as standard scores between 90 and 110 [inclusive], based on a scale having a mean of 100 and standard deviation of 15). Thus a weakness is associated with standard scores of 85 to 89 (inclusive).	Interpreting scores in the very narrow range of 85–89 requires clinical judgment, as abilities associated with these scores may or may not pose significant difficulties for an individual. Interpretation of any cognitive construct as a weakness for the individual should include ecological validity (i.e., evidence of how the weakness manifests itself in real-world performances, such as classroom activities).
Normative weakness or deficit	Performance on standardized, norm-referenced tests that falls greater than one standard deviation below the mean (i.e., standard scores <85). This type of weakness is often referred to as *population-relative* or *interindividual.* The terms *normative weakness* and *deficit* are used interchangeably.	The range of 85–115, inclusive, is often referred to as the range of *normal limits* because it is the range in which nearly 70% of the population falls on standardized, norm-referenced tests. Therefore, scores within this range are sometimes classified as *within normal limits* (WNL). As such, any score that falls outside and below this range is a normative weakness *as compared to most people*. Notwithstanding, the meaning of any cognitive construct that emerges as a normative weakness is enhanced by ecological validity.

TABLE 22.3. Correspondence between Eight Areas of SLD and WIAT-III, WJ IV Tests of Achievement, and KTEA-3 Subtests

Areas in which SLD may be manifested (listed in IDEA 2004)	WIAT-III subtests	WJ IV subtests	KTEA-3 subtests
Oral expression	Oral Expression		Association Fluency Object Naming Facility Oral Expression
Listening comprehension	Listening Comprehension		Listening Comprehension
Written expression	Alphabet Writing Fluency Sentence Composition Essay Composition Spelling	Spelling Writing Samples Sentence Writing Fluency Editing	Spelling Written Expression
Basic reading skills	Early Reading Skills Word Reading Pseudoword Decoding	Letter–Word Identification Word Attack	Decoding Fluency Letter and Word Recognition Nonsense Word Decoding Phonological Processing
Reading fluency skills	Oral Reading Fluency	Sentence Reading Fluency Oral Reading Word Reading Fluency	Silent Reading Fluency Word Recognition Fluency
Reading comprehension	Reading Comprehension	Passage Comprehension Reading Recall Reading Vocabulary	Reading Comprehension Reading Vocabulary
Mathematics calculation	Numerical Operations Math Fluency—Addition Math Fluency—Subtraction Math Fluency—Multiplication	Calculation Math Facts Fluency	Math Computation Math Fluency
Mathematics problem solving	Math Problem Solving	Applied Problems Number Matrices	Math Concepts and Applications

es are a necessary component of the definition. Nevertheless, some children who struggle academically may not demonstrate academic weaknesses or deficits on standardized, norm-referenced tests of achievement; this is particularly true of very bright students, for a variety of reasons. For example, some children may have figured out how to compensate for their processing deficits. Therefore, it is important not to assume that a child with a standard score in the upper 80s or low 90s on a "broad reading" composite is "OK," particularly when a parent, a teacher, or the student him- or herself expresses concern. Under these circumstances, a more focused assessment of the CHC and neuropsychological processes related to reading should be conducted. Conversely, the finding of low scores on norm-referenced achievement tests does not guarantee that there will be corresponding low scores on norm-referenced cognitive tests in areas that are related to the achievement area—*an important fact that was ignored in a relatively recent investigation of the DD/C method* (i.e., Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016b). Below-average achievement may be the result of a host of factors, only one of which is weaknesses or deficits in related cognitive processes and abilities. Most practitioners know this to be true. See Flanagan and Schneider (2016) for a discussion. When weaknesses or deficits in academic performance are found, and are corroborated by other data sources, the process advances to level II.

Level II: Exclusionary Factors as Potential Primary or Contributory Reasons

Level II involves evaluating whether any documented weaknesses or deficits found through level I evaluation are or are not *primarily* the result of factors that may be, for example, largely external to the child, noncognitive in nature, or the result of a disorder other than SLD. Because there can be many reasons for weak or deficient academic performance, causal links to SLD should not be ascribed prematurely. Instead, reasonable hypotheses related to other potential causes for academic weaknesses should be developed. For example, cultural and linguistic differences are two common factors that can affect both test performance and academic skill acquisition adversely and can result in achievement data that appear to suggest SLD (see Ortiz, Melo, & Terzulli, 2018; Ortiz, Piazza, Ochoa, & Dynda, Chapter 25, this volume). In addition, lack of motivation, social-emotional disturbance, performance anxiety, psychiatric disorders, sensory impairments, and medical conditions (e.g., hearing or vision problems) also need to be ruled out as potential explanatory correlates to (or *primary* reasons for) any weaknesses or deficits identified at level I. Figure 22.3 provides an example of a form that may be used to document systematically and thoroughly that the exclusionary factors listed in the federal definition of SLD were evaluated.

Note that because the process of SLD determination does not necessarily occur in a strict linear fashion, evaluations at levels I and II often take place concurrently, as data from level II are often necessary to understand performance at level I. The circular arrows between levels I and II in Figure 22.2 are meant to illustrate the fact that interpretations and decisions that are based on data gathered at level I may need to be informed by data gathered at level II. Ultimately, at level II, the practitioner must judge the extent to which any factors other than cognitive impairment can be considered the *primary* reason for the academic performance difficulties. The form in Figure 22.3 provides space for documenting this judgment. If performance cannot be attributed primarily to other factors, then the second criterion necessary for establishing SLD according to the operational definition is met, and assessment may continue to the next level.

It is important to recognize that although factors such as having English as a second language may be present and may affect performance adversely, SLD can also be present. Certainly, children who have vision problems, chronic illnesses, limited English proficiency, and so forth, may also have SLD. Therefore, when these or other factors at level II are present, or even when they are determined to be *contributing* to poor performance, SLD should not be ruled out. Rather, only when such factors are determined to be *primarily* responsible for weaknesses in learning and academic performance—not merely contributing to them—should SLD be discounted as an explanation for dysfunction in academic performance. Examination of exclusionary factors is necessary to ensure fair and equitable interpretation of the data collected for SLD determination and, as such, is not intended to *rule in* SLD. Rather, careful examination of exclusionary factors is intended to rule out other possible explanations for deficient academic performance.

One of the major reasons for placing evaluation of exclusionary factors at this (early) point in the SLD assessment process is to provide a mechanism that is efficient in both time and effort, and that may prevent the unnecessary administration of additional tests. However, it may not be possible to rule out all the numerous potential exclusionary factors completely and convincingly at this stage in the assessment process. For example, the data gathered at levels I and II may be insufficient to draw conclusions about such conditions as developmental disabilities and intellectual disability (ID; see Farmer & Floyd, Chapter 23, this volume), which often requires more thorough and direct assessment (e.g., administration of an intelligence test and adaptive behavior scale). When exclusionary factors—at least those that can be evaluated at this level—have been examined carefully and eliminated as possible *primary* explanations for poor academic performance, the process may advance to the next level.

Level III: Performance in Cognitive Abilities and Neuropsychological Processes

The criterion at level III is like the one specified in level I, except that it is evaluated with data from an assessment of cognitive abilities and neuropsychological processes. Analysis of data generated from the administration of standardized tests represents the most common method available by which cognitive abilities and neuropsychological processes in children are evaluated. However, other types of information and data are relevant to

Evaluation and Consideration of Exclusionary Factors for SLD Identification

An evaluation for specific learning disabilities (SLD) requires an evaluation and consideration of factors *other* than a disorder in one or more basic psychological processes that may be the primary cause of a student's academic skill weaknesses and learning difficulties. These factors include (but are not limited to) vision/hearing[a] or motor disabilities, intellectual disability (ID), social-emotional or psychological disturbance, environmental or economic disadvantage, cultural and linguistic factors (e.g., limited English proficiency), insufficient instruction or opportunity to learn, and physical/health factors. These factors may be evaluated via behavior rating scales, parent and teacher interviews, classroom observations, attendance records, social/developmental history, family history, vision/hearing exams, medical records, prior evaluations, and interviews with current or past counselors, psychiatrists, and paraprofessionals who have worked with the student. Noteworthy is the fact that students with (and without) SLD often have one or more factors (listed below) that *contribute* to academic and learning difficulties. However, the practitioner must rule out any of these factors as being the *primary* reason for a student's academic and learning difficulties to maintain SLD as a viable classification/diagnosis.

Vision (check all that apply):

☐ Vision test recent (within 1 year)
☐ Vision test outdated (>1 year)
☐ Passed
☐ Failed
☐ Wears glasses

☐ History of visual disorder/disturbance
☐ Diagnosed visual disorder/disturbance
☐ Name of disorder: ______________________________
☐ Vision difficulties suspected or observed (e.g., difficulty with far- or near-point copying; misaligned numbers in written math work; squinting or rubbing eyes during visual tasks such as reading, computer use)

Notes: ____

____

____

____

____

Hearing (check all that apply):[b]

☐ Hearing test recent (within 1 year)
☐ Hearing test outdated (>1 year)
☐ Passed
☐ Failed
☐ Uses hearing aids

☐ History of auditory disorder/disturbance
☐ Diagnosed auditory disorder/disturbance
☐ Name of disorder: ____________________
☐ Hearing difficulties suggested in the referral (e.g., frequent requests for repetition of auditory information; misarticulated words; attempts to self-accommodate by moving closer to sound source; obvious attempts to read speech)

Notes: ____

____

____

____

____

(continued)

FIGURE 22.3. Form for documenting evaluation of exclusionary factors in the SLD identification process. Developed by Jennifer T. Mascolo and Dawn P. Flanagan. This form may be reproduced and disseminated.

Motor Functioning (check all that apply):

☐ Fine motor delay/difficulty

☐ Gross motor delay/difficulty

☐ Improper pencil grip (specify type: ________________)

☐ Assistive devices/aids used (e.g., weighted pens, pencil grip, slant board)

☐ History of motor disorder

☐ Diagnosed motor disorder

☐ Name of disorder: ________________________________

☐ Motor difficulties suggested in the referral (e.g., illegible writing; issues with letter or number formation, size, spacing; difficulty with fine motor tasks such as using scissors, folding paper)

Notes: ____

____

Cognitive and Adaptive Functioning (Check All That Apply):

☐ Significantly “subaverage intellectual functioning” (e.g., IQ score of 75 or below)

☐ Pervasive cognitive deficits (e.g., weaknesses or deficits in many cognitive areas, including Gf *and* Gc)

☐ Deficits in adaptive functioning (e.g., social skills, communication, self-care)

☐ Areas of significant adaptive skill weaknesses (check all that apply):

☐ Motor skills

☐ Daily living skills

☐ Communication

☐ Behavioral/emotional skills

☐ Socialization

☐ Other

Notes: ____

____

Social-Emotional/Psychological Factors (check all that apply):

☐ Diagnosed psychological disorder (specify: ________________________________)

☐ Date of diagnosis: ____

☐ Family history significant for psychological difficulties

☐ Disorder presently treated (specify treatment modality—e.g., counseling, medication): ____________

☐ Reported difficulties with social-emotional functioning (e.g., social phobia, anxiety, depression)

☐ Social-emotional/psychological issues suspected or suggested by referral

☐ Home–school adjustment difficulties

☐ Lack of motivation/effort

☐ Emotional stress

☐ Autism

☐ Present medications (type, dosage, frequency, duration):

☐ Prior medication use (type, dosage, frequency, duration):

☐ Hospitalization for psychological difficulties (date[s]: ________________________________)

☐ Deficits in social, emotional, or behavioral [SEB] functioning (e.g., as assessed by standardized rating scales) —significant scores from SEB measures:

Notes: ____

____

(continued)

FIGURE 22.3. *(continued)*

Environmental/Economic Factors (check all that apply):

☐ Limited access to educational materials in the home
☐ Caregivers unable to provide instructional support
☐ Economic considerations precluded treatment of identified issues (e.g., filling a prescription, replacing broken glasses, tutoring)
☐ Temporary crisis situation
☐ History of educational neglect
☐ Frequent transitions (e.g., shared custody)
☐ Environmental space issues (e.g., no space for studying, sleep disruptions due to shared sleeping space)

Notes: ____

____

Cultural/Linguistic Factors (check all that apply):[c]

☐ Limited number of years in U.S. (________)
☐ No history of early or developmental problems in primary language
☐ Current primary-language proficiency: (Date: ________ Scores: ________)
☐ Acculturative knowledge development (Circle one: High Moderate Low)
☐ Language(s) other than English spoken in home
☐ Lack of or limited instruction in primary language (# of years: ________)
☐ Current English-language proficiency: (Date: ________ Scores: ________)
☐ Parental educational and socioeconomic level (Circle one: High Moderate Low)

Notes: ____

____

Physical/Health Factors (check all that apply):

☐ Limited access to health care
☐ Minimal documentation of health history/status
☐ Chronic health condition (specify: ________________________________)
☐ Temporary health condition (date/duration: ________________________________)
☐ Hospitalization (dates: ________________________________)
☐ History of Medical Condition (date diagnosed: ________________________________)
☐ Medical treatments (specify:________________________________)
☐ Repeated visits to doctor/school nurse
☐ Medication (type, dosage, frequency, duration:________________________________)

Notes: ____

____

Instructional Factors (check all that apply):

☐ Interrupted schooling (e.g., midyear school move) Specify why: ________________
☐ New teacher (past 6 months)
☐ Nontraditional curriculum (e.g., home-schooled)
☐ Days absent/tardy: ____________
☐ Retained or advanced a grade or more
☐ Accelerated curriculum (e.g., AP classes)

Notes: ____

(continued)

FIGURE 22.3. *(continued)*

Determination of Primary and Contributory Causes of Academic Weaknesses and Learning Difficulties (check one):

☐ Based on the available data, it is reasonable to conclude that one or more factors are *primarily* responsible for the student's observed learning difficulties. Specify: ____________________

☐ Based on the available data, it is reasonable to conclude that one or more factors *contribute* to the student's observed learning difficulties. Specify: ____________________

☐ *No* factors listed here appear to be the primary cause of the student's academic weaknesses and learning difficulties.

[a] For a vision or hearing disorder, it is important to understand the nature of the disorder, its expected impact on achievement, and the time of diagnosis. It is also important to understand what was happening instructionally at the time the disorder was suspected and/or diagnosed.

With regard to hearing, even mild loss can impact initial receptive and expressive skills as well as academic skill acquisition. When loss is suspected, the practitioner should consult professional literature to further understand the potential impact of a documented hearing issue (see the American Speech–Language–Hearing Association guidelines, *www.asha.org*).

With regard to vision, refractive error (i.e., hyperopia and anisometropia), accommodative and vergence dysfunctions, and eye movement disorders are associated with learning difficulties, whereas other vision problems (e.g., constant strabismus and amblyopia) are not. As such, when a vision disorder is documented or suspected, the practitioner should consult professional literature to further understand the impact of the visual disorder (e.g., see American Optometric Association, *www.aoa.org*).

[b] When there is a history of hearing difficulties and an SLD diagnosis is being considered, hearing testing should be recent (i.e., conducted within the past 6 months).

[c] When evaluating the impact of language and cultural factors on a student's functioning, the practitioner should consider whether and to what extent other individuals with similar linguistic and cultural backgrounds as the referred student are progressing and responding to instruction in the present curriculum. For example, if an LEP student with limited English proficiency is not demonstrating academic progress or is not performing as expected on a class- or districtwide assessment when compared to his or her peers who possess a similar level of English proficiency and acculturative knowledge, it is unlikely that cultural and linguistic differences are the sole or primary factors for the referred student's low performance. In addition, it is important to note that as the number of cultural and linguistic differences in a student's background increase, the likelihood becomes greater that poor academic performance is attributable primarily to such differences rather than to a disability.

Note. All 50 U.S. states specify eight exclusionary criteria. Namely, learning difficulties cannot be primarily attributed to (1) visual impairment; (2) hearing impairment; (3) motor impairment; (4) intellectual disability; (5) emotional disturbance; (6) environmental disadvantage; (7) economic disadvantage; and (8) cultural difference. Certain states have adopted additional exclusionary factors including autism (CA, MI, VT, and WI), emotional stress (LA and VT), home or school adjustment difficulties (LA and VT), lack of motivation (LA and TN), and temporary crisis situation (LA, TN, and VT). We have integrated these additional criteria under "Social-Emotional/Psychological Factors" and "Environmental/Economic Factors," and have added two further categories (namely, "Instructional Factors" and "Physical/Health Factors") to this form. This form may be reproduced and disseminated.

Note. Developed by Jennifer T. Mascolo and Dawn P. Flanagan. This form may be reproduced and disseminated.

FIGURE 22.3. *(continued)*

cognitive performance (see Figure 22.2, level III, column 4). Practitioners should seek out and gather data from other sources as a means of providing corroborating evidence for standardized test findings. For example, when test findings are found to be consistent with a child's performance in the classroom, more confidence may be placed on test performance because interpretations of cognitive deficiency have ecological validity—an important condition for any diagnostic process (Flanagan et al., 2013; Flanagan, Alfonso, Sy, et al., 2018; Hale & Fiorello, 2004). Table 22.4 provides an example of the cognitive abilities and neuropsychological processes measured by the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014), the Woodcock–Johnson IV Tests of Cognitive Abilities (Schrank, McGrew, & Mather, 2014), and the Kaufman Assessment Battery for Children, Second Edition Normative Update (Kaufman & Kaufman, 2018). For similar information on all major intelligence tests and neuropsychological instruments, see Flanagan and colleagues (2017).

A particularly salient aspect of the DD/C operational definition of SLD is the concept that a weakness or deficit in a cognitive ability or process underlies difficulties in academic performance or skill development. Because research demonstrates

TABLE 22.4. Cognitive Abilities and Neuropsychological Processes Measured by the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V), Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG), and Kaufman Assessment Battery for Children II Normative Update (KABC-II NU) Subtests

	CHC broad and narrow abilities							Neuropsychological domains							
Subtest	Gf	Gc	Gsm	Gv	Gs	Ga	Glr	Sensory–motor	Speed and efficiency	Attention	Visual-spatial (RH) and detail (LH)	Auditory–verbal	Memory and/or learning	Executive	Language^{a}
					WISC-V										
Arithmetic^{b}	✓ (RQ)		✓ (MW)							✓		✓	✓	✓	✓^{R}
Block Design				✓ (Vz)				✓			✓			✓	
Cancellation					✓ (P)			✓	✓	✓	✓			✓	
Coding					✓ (R9)			✓	✓	✓	✓		✓	✓	
Comprehension		✓ (K0)										✓	✓		✓$^{E/R}$
Delayed Symbol Translation							✓ (MA)			✓			✓	✓	
Digit Span			✓ (MS, MW)							✓		✓	✓	✓	
Figure Weights	✓ (RG)									✓	✓			✓	
Information		✓ (K0)										✓	✓		✓^{E}
Immediate Symbol Translation							✓ (MA)			✓			✓	✓	
Letter–Number Sequencing			✓ (MW)							✓		✓	✓	✓	
Matrix Reasoning	✓ (I)										✓			✓	
Naming Speed Literacy							✓ (NA)		✓	✓	✓		✓		
Naming Speed Quantity					✓ (N)				✓	✓	✓		✓		
Picture Concepts	✓ (I)									✓	✓		✓		

(continued)

TABLE 22.4. *(continued)*

	CHC broad and narrow abilities							Neuropsychological domains							
Subtest	Gf	Gc	Gsm	Gv	Gs	Ga	Glr	Sensory–motor	Speed and efficiency	Attention	Visual-spatial (RH) and detail (LH)	Auditory–verbal	Memory and/or learning	Executive	Language[a]
Picture Span			✓ (MS)								✓		✓	✓	
Recognition Symbol Translation							✓ (MA)			✓			✓	✓	
Similarities		✓ (VL)										✓	✓	✓	✓^{E}
Symbol Search					✓ (P)			✓	✓	✓	✓			✓	
Visual Puzzles				✓ (Vz)						✓	✓			✓	
Vocabulary		✓ (VL)										✓	✓		✓^{E}
					WJ IV COG										
Analysis–Synthesis	✓ (RG)										✓	✓	✓	✓	✓^{R}
Concept Formation	✓ (I)										✓	✓	✓	✓	✓^{R}
General Information		✓ (K0)										✓	✓		✓$^{R/E}$
Letter–Pattern Matching					✓ (P)				✓	✓	✓				
Memory for Words			✓ (MS)							✓		✓	✓		
Nonword Repetition			✓ (MS)			✓ (UM)				✓		✓			
Number–Pattern Matching					✓ (P)				✓	✓	✓				
Numbers Reversed			✓ (MW)							✓		✓	✓	✓	
Number Series	✓ (RQ)								✓		✓				
Object–Number Sequencing			✓ (MW)							✓		✓	✓		

(continued)

TABLE 22.4. *(continued)*

	CHC broad and narrow abilities							Neuropsychological domains							
Subtest	Gf	Gc	Gsm	Gv	Gs	Ga	Glr	Sensory–motor	Speed and efficiency	Attention	Visual-spatial (RH) and detail (LH)	Auditory–verbal	Memory and/or learning	Executive	Language^a
Oral Vocabulary		✓ (VL)										✓	✓		✓^{E}
Pair Cancellation					✓ (P)				✓	✓	✓				
Phonological Processing						✓ (PC)				✓		✓			
Picture Recognition				✓ (MV)						✓	✓		✓		
Story Recall							✓ (MM)			✓		✓			✓^{E}
Verbal Attention			✓ (MW)							✓		✓			
Visual–Auditory Learning							✓ (MA)			✓			✓	✓	
Visualization					✓ (Vz)					✓	✓			✓	
					KABC-II NU										
Atlantis							✓ (MA)				✓		✓	✓	
Atlantis Delayed							✓ (MA)				✓		✓	✓	
Block Counting					✓ (Vz)						✓				
Conceptual Thinking	✓ (I)				✓ (Vz)						✓			✓	
Expressive Vocabulary		✓ (VL)										✓	✓		✓^{E}
Face Recognition					✓ (MV)					✓	✓		✓		
Gestalt Closure					✓ (CS)						✓		✓		
Hand Movements			✓ (MS, MV)					✓		✓	✓		✓		

(continued)

TABLE 22.4. *(continued)*

	CHC broad and narrow abilities							Neuropsychological domains							
Subtest	Gf	Gc	Gsm	Gv	Gs	Ga	Glr	Sensory–motor	Speed and efficiency	Attention	Visual-spatial (RH) and detail (LH)	Auditory–verbal	Memory and/or learning	Executive	Language[a]
Number Recall			✓ (MS)							✓		✓	✓		
Pattern Reasoning	✓ (I)				✓ (Vz)						✓			✓	
Rebus							✓ (MA)					✓	✓	✓	
Rebus Delayed							✓ (MA)					✓	✓	✓	
Riddles	✓ (RG)	✓ (VL)										✓	✓	✓	✓$^{R/E}$
Rover	✓ (RG)			✓ (SS)							✓			✓	
Story Completion	✓ (RG)	✓ (K0)									✓		✓	✓	
Triangles				✓ (Vz)				✓			✓			✓	
Verbal Knowledge		✓ (VL, K0)										✓	✓		✓^{R}
Word Order			✓ (MS, MW)							✓		✓	✓	✓	

Note. Gf, fluid intelligence; Gc, crystallized intelligence; Gsm, short-term memory; Gv, visual processing; Gs, processing speed. RQ, quantitative reasoning; MW, working memory; SR, spatial relations; Vz, visualization; P, perceptual speed; R9, rate of test taking; K0, general (verbal) knowledge; LD, language development; MS, memory span; I, induction; RG, general sequential reasoning; CF, flexibility of closure; VL, lexical knowledge. The following Cattell–Horn–Carroll (CHC) broad abilities are omitted from this table because none is a primary ability measured by the WISC-V: Glr (long-term storage and retrieval), Ga (auditory processing), Gt (decision/reaction time or speed), and Grw (reading and writing ability). Most CHC test classifications are from Flanagan, Ortiz, and Alfonso (2017). Classifications according to neuropsychological domains were based on our readings of neuropsychological texts (e.g., Fletcher-Janzen & Reynolds, 2008; Hale & Fiorello, 2004; Lezak, 1995; Miller, 2007, 2010) and are also found in Flanagan, Alfonso, and Mascolo (2011).

[a] E, expressive; R, receptive.

[b] Cognitive ability classifications for the Arithmetic subtest are based on analyses conducted by Keith, Fine, Taub, Reynolds, and Kranzler (2006; viz., Gf:RQ). It is important to note that the Keith et al. analyses did not include any other measures of math achievement; therefore, Gq was not represented adequately in their study. Arithmetic has been identified in many other studies as a measure of Gq, particularly math achievement (A3) (see, for discussions, Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009).

that the relationship between the cognitive dysfunction and the manifest learning problems is causal in nature[2] (e.g., Flanagan & Schneider, 2016; Fletcher, Lyon, Fuchs, & Barnes, 2007; Fletcher, Taylor, Levin, & Satz, 1995; Hale & Fiorello, 2004; Hale et al., 2010; Wagner & Torgesen, 1987), data analysis at this level should seek to ensure that identified weaknesses or deficits on cognitive and neuropsychological tests bear an empirical relationship to those weaknesses or deficits in achievement identified previously. It is this very notion that makes it necessary to draw upon cognitive and neuropsychological theory and research to inform operational definitions of SLD and increase the reliability and validity of the SLD identification process. Theory and its related research base not only specify the relevant constructs that ought to be measured at levels I and III, but predict the way they are related. Furthermore, application of current theory and research provides a substantive empirical foundation from which interpretations and conclusions may be drawn. Tables 22.5 through 22.7 provide summaries of the relations between CHC cognitive abilities and processes and reading, math, and writing achievement, respectively, based on findings from multiple literature reviews (Berninger, 2011; Flanagan, Ortiz, Alfonso, & Mascolo, 2006; Flanagan et al., 2013; McDonough et al., 2017; McGrew & Wendling, 2010; Niileksela, Reynolds, Keith, & McGrew, 2016). These tables also provide summaries of the literature on the etiology of academic difficulties (see McDonough et al., 2017, for a discussion).

The information contained in Tables 22.5 through 22.7 may be used to guide how practitioners organize their assessments at this level. That is, prior to selecting cognitive and neuropsychological tests, a practitioner should have knowledge of those cognitive abilities and processes that are most important for understanding a child's academic performance in the area(s) in question (i.e., the area[s] identified as weak or deficient at level I). Evaluation of cognitive performance should be comprehensive in the areas of suspected dysfunction. Because evidence of a cognitive weakness or deficit is a necessary condition for SLD determination, if no weaknesses or deficits in cognitive abilities or processes are found, then an essential criterion for SLD determination is not met. When the criterion at level III is not met, an evaluation of whether the obtained cognitive data represent an evaluation that was sufficient in breadth and depth vis-à-vis what is known about the relations between abilities, processes, and academic skill acquisition and development is warranted. Furthermore, a more in-depth exploration of exclusionary factors evaluated at level II may be warranted.

Also, because new data are gathered at level III, it is now possible to evaluate the exclusionary factors that could not be evaluated earlier (e.g., ID). The circular arrows between levels II and III in Figure 22.2 are meant to illustrate that interpretations and decisions based on data gathered at level III may need to be informed by data gathered at level II. Likewise, data gathered at level III are often necessary to rule out (or in) one or more factors listed at level II in Figure 22.2. Reliable and valid identification of SLD depends in part on being able to understand academic performance (level I), cognitive performance (level III), and the many factors that may facilitate or inhibit such performances (level II).

Level IV: Data Integration—The DD/C Pattern of Strengths and Weaknesses

The fourth level of evaluation involves an analysis of the individual's PSW. It revolves around a theory- and research-guided examination of performance across academic skills, cognitive abilities, and neuropsychological processes to determine whether the child's PSW is consistent with the SLD construct.

Figure 22.4 provides an illustration of three common components of the PSW method for identification of SLD. First, individuals with SLD have cognitive processing weaknesses or deficits. These weaknesses are depicted by the bottom left oval in the figure. Second, individuals with SLD have academic skill weaknesses or deficits. These weaknesses are depicted by the bottom right oval in the figure. Third, individuals with SLD have areas of cognitive strength. These strengths are depicted in the top center oval in the figure. In addition to these three components, the relationships between these ovals are important. The double-headed arrows between the top oval and the two bottom ovals in the figure indicate the presence of statistically significant discrepancies in measured performance between cognitive strengths and the areas of academic and cognitive weakness. These discrepancies denote that the differences are reliable and not due to chance. The double-headed arrow between the two bottom ovals reflects a consistency between the cognitive and academic areas of weakness. The consistency

TABLE 22.5. Summary of Relations among CHC Domains, Reading Achievement, and the Etiology of Reading Functions

CHC broad ability	Reading achievement	Etiology of reading functions
Gf	Inductive (I) and general sequential reasoning (RG) abilities play a moderate role in reading comprehension. Executive functions (EF), such as planning, organization, and self-monitoring, are also important.	Several cortical and subcortical structures are frequently implicated in basic reading skills and word-reading accuracy. Recent work appears to identify dysfunction in a left-hemispheric network that includes the occipito-temporal region, inferior frontal gyrus, and inferior parietal region of the brain (Fletcher, Simos, Papanicolaou, & Denton, 2004; Richlan, 2012; Richlan et al., 2009; Shaywitz et al., 2000; Silani et al., 2005). Numerous imaging studies have also found that dysfunctional responses in the left inferior frontal and temporo-parietal cortices play a significant role with regard to phonological deficits (Skeide et al., 2015). Similar brain regions are activated on tasks involving reading fluency, but additional activation is observed in areas involved in eye movement and attention (Jones, Ashby, & Branigan, 2013). Furthermore, there is also evidence for increased activation in the left occipito-temporal region, in particular the occipito-temporal sulcus, which is important for rapid processing of letter patterns (Dehaene & Cohen, 2011; Shaywitz et al., 2004).
Gc	Language development (LD), lexical knowledge (VL), and listening ability (LS) are important at all ages for reading acquisition and development. These abilities become increasingly important with age. Oral language, listening comprehension, and EF (planning, organization, self-monitoring) are also important for reading comprehension.	
Gwm	Memory span (MS) and working memory capacity (WM) or attentional control are important for overall reading success. Phonological memory or WM for verbal and sound-based information may also be important. WM is important for reading comprehension, which involves holding words and sentences in awareness, while integrating prior knowledge with incoming information.	
Gv	Orthographic processing (often measured by tests of perceptual speed that use orthographic units as stimuli) is related to reading rate and fluency. Orthographic processing involves the ability to process units of words based on visual long-term memory representations, which is critical for automatic word recognition.	Brain regions often associated with reading comprehension include the anterior temporal lobe, inferior temporal gyrus, inferior frontal gyrus, inferior frontal sulcus, and middle and superior frontal and temporal regions (Ferstl et al., 2008; Gernsbacher & Kaschak, 2003). More recent research has revealed a relationship between listening and reading comprehension and activation along the left superior temporal sulcus, which has referred to by some as the *comprehension cortex* (Berl et al., 2010). However, broader pathways are also activated in reading comprehension, reflecting increased cognitive demand compared to listening.
Ga	Phonetic coding (PC) or phonological awareness/processing is very important during the elementary school years for the development of basic reading skills and word-reading accuracy. Phonological memory or WM for verbal and sound-based information may also be important.	
Glr	Naming facility (NA) or rapid automatic naming (RAN; also called speed of lexical access) is very important during the elementary school years for reading rate and fluency or word recognition skills. Associative memory (MA) is also important.	Family and genetic factors have long been identified as crucial in reading achievement, with some researchers suggesting that a child with a parent with a reading disability is eight times more likely to have dyslexia than a child in the general population (Pennington & Olson, 2005).
Gs	Perceptual speed (P) abilities are important throughout school, but particularly during the elementary school years.	Shared environmental factors include language and literacy environment during childhood (Wadsworth et al., 2000) and quality of reading instruction.

Note. Information in this table was culled from the following sources: Flanagan, Ortiz, and Alfonso (2013); Flanagan, Ortiz, Alfonso, and Mascolo (2006); McDonough, Flanagan, Sy, and Alfonso (2017); McGrew and Wendling (2010); Niileksela, Reynolds, Keith, and McGrew, 2016. All references cited in the "Etiology . . ." column may be found in these sources.

TABLE 22.6. Summary of Relations among CHC Domains, Math Achievement, and the Etiology of Math Functions

CHC broad ability	Math achievement	Etiology of math functions
Gf	Reasoning inductively (I) and deductively with numbers (RQ) is very important for math problem solving. Executive functions such as set shifting and cognitive inhibition are also important.	The intraparietal sulcus in both hemispheres is widely viewed as crucial in processing and representing numerical quantity (number sense), although there may be differences in activation as a function of age (Ansari & Dhital, 2006; Ansari, Garcia, Lucas, Hamon, & Dhital, 2005; Dehaene et al., 2004; Kaufmann et al., 2006; Kucian, von Aster, Loenneker, Dietrich, & Martin, 2008; Mussolin et al., 2010; Price & Ansari, 2013).
Gc	Language development (LD), lexical knowledge (VL), and listening ability (LS) are important at all ages for math problem solving. These abilities become increasingly important with age. Number representation (e.g., quantifying sets without counting, estimating relative magnitude of sets) and number comparisons are related to overall number sense.	Regions of the left fronto-parietal cortex, including the intraparietal sulcus, angular gyrus, and supramarginal gyrus, have been consistently associated with math calculation (Ansari, 2008; De Smedt, Holloway, & Ansari, 2011; Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene et al., 2004). The dorsolateral prefrontal cortex has also been found to show increased activation during calculation, implying that executive functioning and working memory may be playing a role in the process (Davis et al., 2009).
Gwm	Memory span (MS) and working memory capacity (WM) or attentional control are important for math problem solving and overall success in math.	
Gv	Visualization (VZ), including mental rotation, is important primarily for higher-level math (e.g., geometry, calculus) and math problem solving.	A left-hemisphere network that includes the precentral gryus, inferior parietal cortex, and intraparietal sulcus is often implicated in math fact retrieval (Dehaene & Cohen, 1992, 1997; Dehaene et al., 1999). Furthermore, some researchers believe that rote math facts are retrieved from verbal memory, thereby requiring activation of the angular gyrus and other regions associated with linguistic processes (Dehaene, 1992; Dehaene & Cohen, 1995; Dehaene et al., 1999).
Ga		
Glr	Naming facility (NA; also called speed of lexical access) and associative memory (MA) are important for memorization and rapid retrieval of basic math facts and for accurate and fluent calculation.	
Gs	Perceptual speed (P) is important during all years, especially the elementary school years for math calculation fluency.	Prevalence of math disabilities is about 10 times higher in those with family members who had math disabilities than in the general population (Shalev et al., 2001).
		Environmental factors, including motivation, emotional functioning (e.g., math anxiety), and suboptimal or inadequate teaching, may also contribute to math difficulties (Szucs & Goswami, 2013; Vukovic et al., 2013). Furthermore, math achievement may be associated with cultural or gender-based attitudes that may be transmitted in the family environment (e.g., Chiu & Klassen, 2010; Gunderson et al., 2011).

Note. Information in this table was culled from the following sources: Flanagan, Ortiz, and Alfonso (2013); Flanagan, Ortiz, Alfonso, and Mascolo (2006); McDonough, Flanagan, Sy, and Alfonso (2017); McGrew and Wendling (2010); and McGrew et al. (2014). All references cited in the "Etiology . . ." column may be found in these sources.

TABLE 22.7. Summary of Relations between CHC Domains and Writing Achievement and the Etiology of Writing Functions

CHC broad ability	Writing achievement	Etiology of writing functions
Gf	Inductive (I) and general sequential reasoning (RG) are consistently related to written expression at all ages. Executive functions such as attention, planning, and self-monitoring are also important.	Neural correlates of writing are less understood, but some studies have suggested that the cerebellum and parietal cortex, particularly the left superior parietal lobe, may be involved (Katanoda et al., 2001; Magrassi et al., 2010). In addition, the frontal lobes have also been implicated and are considered crucial in planning, brainstorming, organizing, and goal setting, which are important for written expression (Shah et al., 2013). Functional neuroimaging studies have provided substantial evidence for the role of the ventral-temporal inferior frontal gyrus and the posterior inferior frontal gyrus in spelling (Rapp et al., 2015; van Hoorn et al., 2013). Other areas that have been identified include the left ventral cortex, bilateral lingual gyrus, and bilateral fusiform gyrus (Planton et al., 2013; Purcell et al., 2014; Richards et al., 2005, 2006). However, many of these regions have also been associated with reading and are not distinct to spelling/writing disorders. Although there is a significant genetic component involved in the development of writing skills, this etiology is often shared with a broad variety of reading and language skills (Olson et al., 2013).
Gc	Language development (LD), lexical knowledge (VL), and general information (K0) are important primarily after second grade and become increasingly important with age. Level of knowledge of syntax, morphology, semantics, and VL has a significant impact on clarity of written expression and text generation ability.	
Gwm	Memory span (MS) is important to writing, especially spelling skills, whereas working memory (WM) has shown relations with advanced writing skills (e.g., written expression, synthesizing multiple ideas, ongoing self-monitoring).	
Gv	Orthographic processing (often measured by tests of perceptual speed that use orthographic units as stimuli) is particularly important for spelling.	
Ga	Phonetic coding (PC) or phonological awareness/processing is very important during the elementary school years (primarily before fifth grade) for both basic writing skills and written expression.	
Glr	Naming facility (NA; also called speed of lexical access) has demonstrated relations with writing fluency. Storing and retrieving commonly occurring letter patterns in visual and motor memory are needed for spelling.	
Gs	Perceptual speed (P) is important during all school years for basic writing skills and is related to written expression at all ages.	

Note. Information in this table was culled from the following sources: Flanagan, Ortiz, and Alfonso (2013); Flanagan, Ortiz, Alfonso, and Mascolo (2006); McDonough, Flanagan, Sy, and Alfonso (2017); McGrew and Wendling (2010); and McGrew et al. (2014). All references cited in the "Etiology . . ." column may be found in these sources.

means that underlying cognitive processing weaknesses or deficits impede the typical development of basic academic skills in individuals with SLD. The cognitive and academic PSW represented in Figure 22.4 retains the component of unexpected underachievement that has historically been synonymous with the SLD construct (Kaufman, 2008; Kavale & Forness, 2000; Sotelo-Dynega et al., 2018), and adds an underlying cognitive processing component that was missing from traditional discrepancy approaches. The manner in which all components of the pattern are defined varies, sometimes quite substantially, between models. Figure 22.4 includes wording that is most consistent with the DD/C model, and each component of this PSW model is described next.

When the process of SLD identification has reached level IV, three *preliminary* criteria for SLD identification have been met: (1) one or more weaknesses or deficits in academic performance; (2) one or more weaknesses or deficits in cognitive abilities and/or neuropsychological processes; and (3) exclusionary factors determined not to be the primary causes of the academic and cognitive weaknesses or deficits. What has not been determined, however, is whether the pattern of results is marked by an empirical or ecologically valid relationship between the identified cognitive and academic weaknesses; whether the individual's cognitive weakness is *domain-specific*; whether the individual's academic weakness (underachievement) is *unexpected*; and whether the individual displays at least average ability to think and reason. These four conditions form a specific PSW that is marked by two discrepancies and a consistency (DD/C). Within the context of the DD/C operational definition, the nature of unexpected underachievement suggests that not only does a child have specific, circumscribed, and related academic and cognitive weaknesses or deficits—referred to as a *below-average cognitive aptitude–achievement consistency*—but that these weaknesses exist along with generally average or better ability to think and reason (i.e., overall average cognitive ability). These four conditions form a specific PSW that is marked by two discrepancies and a consistency (DD/C). The Cross-Battery Assessment Software System (X-BASS; Flanagan et al., 2017) is needed to determine whether the data demonstrate the DD/C pattern because specific formulae, regression equations, correction for false negatives, and so forth are necessary to make the determination (for explanations of how X-BASS conducts a PSW analysis, see Flanagan et al., Chapter 27, this volume; Flanagan, Alfonso, Sy, et al., 2018). Each of these four conditions is described below.

Consistency between Cognitive and Academic Weaknesses

A student with an SLD has specific cognitive and academic weaknesses or deficits. When these weaknesses are related empirically, or when there is an ecologically valid relationship between them, the relationship is referred to as a *below-average cognitive aptitude–achievement consistency* in the DD/C operational definition. This consistency is a necessary marker for SLD because SLD is caused in part by cognitive processing weaknesses or deficits. Thus there is a need to understand and identify the underlying cognitive ability or processing problems that contribute significantly to the individual's academic difficulties.

The term *cognitive aptitude* within the context of the DD/C operational definition represents the specific cognitive ability or neuropsychological processing weaknesses or deficits that are related empirically to the academic skill weaknesses or deficits. For example, if a child's basic reading skill deficit is related to cognitive deficits in phonological processing (a narrow Ga ability) and rapid automatic naming (a narrow Gr ability), then the combination of below-average Ga and Gr performances represents his or her *below-average cognitive aptitude for basic reading,* meaning that these below-average performances *raise the risk* of a weakness in basic reading skills (Flanagan & Schneider, 2016). Moreover, the finding of below-average performance on measures of phonological processing, rapid automatic naming, and basic reading skill represents a *below-average cognitive aptitude–achievement consistency* (or, more specifically, a below-average *reading* cognitive aptitude–*reading* achievement consistency). The concept of below-average cognitive aptitude–achievement consistency reflects the notion that there are documented relationships between specific cognitive abilities and processes and specific academic skills (see Tables 22.5–22.7). Therefore, the finding of below-average performance in related cognitive and academic areas is an important marker for SLD in the DD/C operational definition, and in other alternative research-based approaches (e.g., McCloskey et al., 2012; see Alfonso & Flanagan, 2018, for support of this SLD marker in other PSW models).

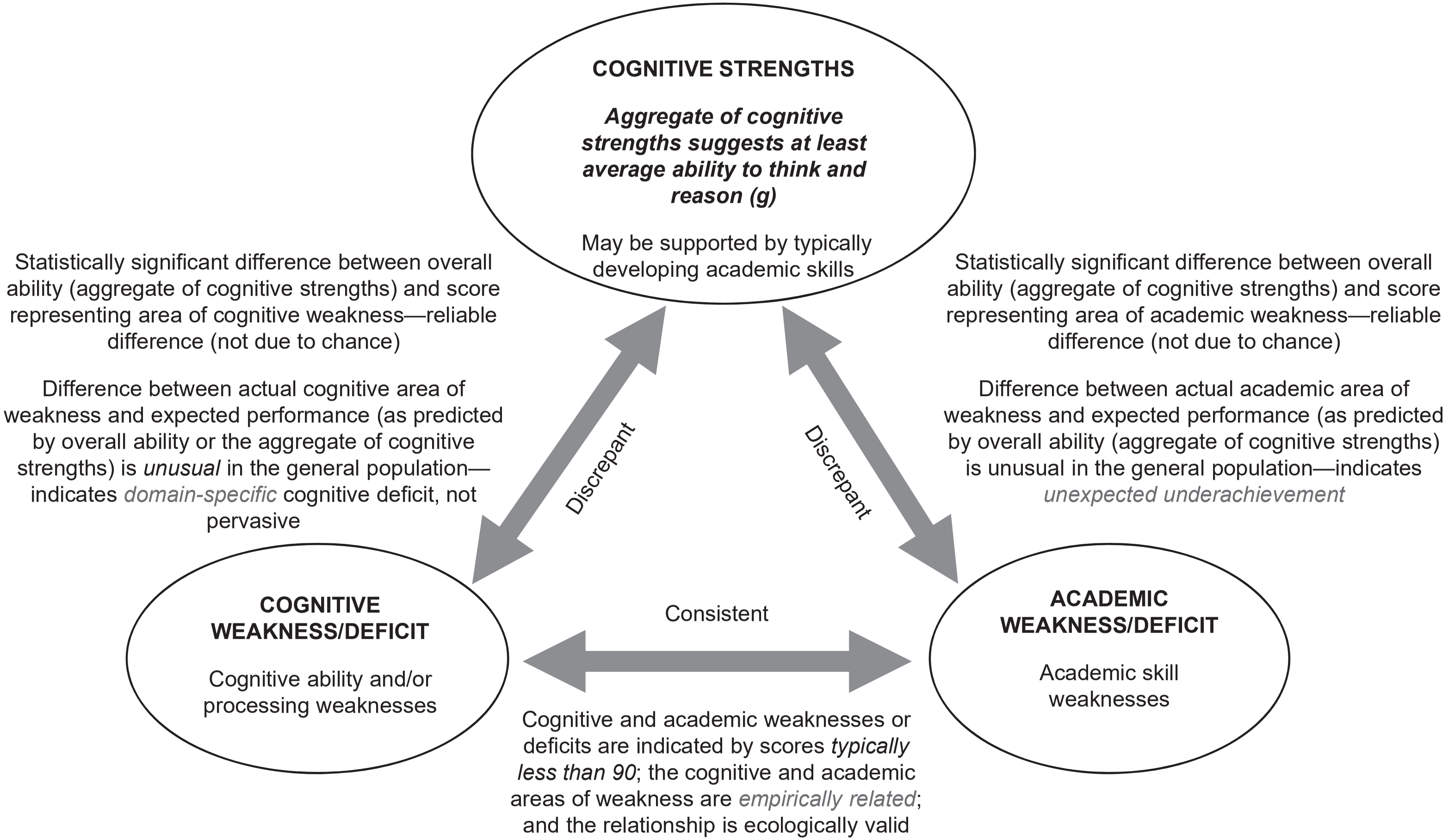


FIGURE 22.4. Conceptual understanding of the dual-discrepancy/consistency (DD/C) method. This figure is based on the work of Flanagan, Fiorello, and Ortiz (2010) and Hale, Flanagan, and Naglieri (2008). However, the wording in this figure is consistent wiht Flanagan, Ortiz, and Alfonso's (2013) DD/C model.

In the DD/C definition, the criteria for establishing a below-average cognitive aptitude–achievement *consistency* are as follows:

1. "Below-average" performance (i.e., scores of less than 90, and more typically at least a standard deviation or more below the mean) in the specific cognitive *and* academic areas that are considered weaknesses or deficits; and
2. Evidence of an empirical relationship between the specific cognitive and academic areas of weakness and/or an ecologically valid relationship between these areas. To validate the relationship between the cognitive and academic areas of weakness, practitioners can document the way each cognitive weakness or deficit manifests itself in real-world performances (see Mascolo, Alfonso, & Flanagan, 2014, for guidance).

It is important to understand that these criteria are operationalized further in X-BASS. Table 22.8 provides a more detailed explanation of how a consistency between cognitive and academic weaknesses is determined by X-BASS.

When the criteria for a below-average cognitive aptitude–achievement consistency are met, there may or may not be a nonsignificant difference between the scores that represent the cognitive and academic areas of weakness. That is, in the DD/C definition, *consistency* refers to the fact that an empirical or ecologically valid relationship exists between the areas of identified cognitive and academic weakness, but not necessarily a nonsignificant difference between these areas. While a nonsignificant difference between the areas of cognitive and academic weakness would be expected, it need not be an inclusionary criterion for SLD. Because many factors facilitate and inhibit performance, a student may perform better or worse academically than his or her cognitive weaknesses may suggest (for a discussion, see Flanagan & Schneider, 2016; Flanagan et al., 2013; Flanagan, Alfonso, Sy, et al., 2018).

Discovery of consistencies among cognitive abilities and/or processes and academic skills in the below-average (or lower) range could result from ID, developmental disabilities, or generally below-average cognitive ability (this would negate two important markers of SLD: that cognitive weaknesses are domain-specific and that underachievement is unexpected) (see Lovett & Kilpatrick, 2018). Therefore, identification of SLD cannot rest on below-average cognitive aptitude–achievement consistency alone. A child with SLD typically has many cognitive capabilities. Therefore, in the DD/C definition, the child must also demonstrate at least average ability to think and reason (i.e., standard scores generally ≥ 85), despite cognitive weaknesses or deficits. For example, in the case of a young child with reading decoding difficulties, it would be necessary to determine that performance in areas less related to this skill (e.g., Gf, math ability) are about average or better. Such a finding would suggest that the related weaknesses in cognitive and academic domains are not due to a more pervasive form of cognitive dysfunction, thus supporting the notion of *unexpected underachievement*—that the child would be likely to perform within normal limits (e.g., at or close to grade level) in whatever achievement skill he or she was found to be deficient in, if not for *specific* cognitive ability or processing weaknesses or deficits. Moreover, because the child has generally average or better ability to think and reason, the academic skill deficiency is indeed unexpected. In sum, the finding of a pattern of circumscribed and related weaknesses (i.e., below-average cognitive aptitude–achievement consistency), despite at least average ability to think and reason (or a pattern of strengths) is convincing evidence of SLD—particularly when the student who demonstrates this pattern did not respond well to high-quality instruction, and when exclusionary factors were ruled out as the primary causes of the deficits.

At Least Average Ability to Think and Reason

An SLD is just what its full name indicates: *specific*. It is not general. As such, the below-average cognitive aptitude–achievement consistency ought to be circumscribed and represent a level of functioning significantly different from the student's cognitive capabilities or strengths in other areas. Indeed, the notion that students with SLD are of generally average or better overall cognitive ability is well known and has been written about for decades (e.g., Hinshelwood, 1917; Orton, 1937). In fact, the earliest recorded definitions of learning disability were developed by clinicians, based on their observations of individuals who experienced considerable difficulties with the acquisition of basic academic skills, despite their average or above-average general intelligence. According to Monroe (1932), "The children of superior mental capacity who fail to learn to read are, of course, spectacular examples of specific reading difficulty since they have such

TABLE 22.8. Description of the Below-Average Aptitude–Achievement Consistency of the DD/C Model and How It Is Determined by Using X-BASS

DD/C	X-BASS	Comments
Areas of cognitive and academic weakness are below average, and there is an empirical and/or ecologically valid relationship between them.	For this component of the PSW analysis, X-BASS answers two specific questions and based on the answers to those questions, provides a statement about the presence of below-average aptitude–achievement consistency. The first question is “Are the scores that represent the cognitive and academic areas of weakness actually weaknesses as compared to these areas in most people (i.e., below average or lower compared to same-age peers from the general population)?” The program parses the cognitive and academic weakness scores into three levels: < 85, 85–89 inclusive, and ≥ 90. Scores less than 85 are considered *normative* weaknesses; scores between 85 and 89 (inclusive) are considered weaknesses because they are below average; and scores of 90 or higher are not considered to be weaknesses. Next, the two scores (academic and cognitive) are examined relative to each other. When both scores are less than 85, the program will report a “Yes,” meaning that both scores are normative weaknesses. If one score is less than 85 and the other is between 85 and 89, the program will report “Likely.” If both scores are between 85 and 89 (inclusive), the program reports “Possibly” (because the scores are within normal limits, despite being classified as below average). The program will also report “Possibly” when one score is less than 85 and one is 90 or higher. If one score is between 85 and 89 (inclusive) and the other is 90 or higher, the program reports “Unlikely” and when both scores are 90 or higher, the program reports “No,” indicating that the scores cannot be considered weaknesses as compared to most people. The second question is “Are the areas of cognitive and academic weakness related empirically?” The strength of the relationship between the cognitive and academic areas of weakness is reported automatically by X-BASS as either LOW (median intercorrelation $< .30$), MODERATE or MOD) (median intercorrelation between .30 and .50), or HIGH (median intercorrelation $> .50$), based on a review of the literature (see Flanagan, Ortiz, & Alfonso, 2013; McGrew & Wendling, 2010) and the technical manuals of cognitive and intelligence batteries (e.g., WJ IV, WISC-V). Information regarding where the cognitive and academic weakness scores fall as compared to those for most people, and the strength of the relationship between the two areas, is used to answer the question “Is there a below-average aptitude–achievement consistency?” The answer automatically generated by X-BASS will be either “Yes, Consistent,” “No, Not Consistent,” or “Possibly, Use Clinical Judgment.” For example, if the cognitive and academic areas selected by the evaluator as weaknesses are associated with scores that fall below 85 and if the strength of the relationship between the areas of cognitive and academic weakness is moderate or high, then the program will report “Yes, Consistent.”	In some cases, the answer to the question of whether or not an individual’s PSW is marked by a below-average aptitude–achievement consistency may not be clear from the quantitative data alone. As such, it is always important to interpret an individual’s PSW within the context of all available data sources (e.g., exclusionary factors, behavioral observations, work samples) and render a judgment about SLD based on the totality of the data.

obvious abilities in other fields" (p. 23; cf. Mather, 2011). Indeed, "all historical approaches to SLD *emphasize the spared or intact abilities* that stand in stark contrast to the deficient abilities" (Kaufman, 2008, pp. 7–8; emphasis added).

Current definitions of SLD also recognize the importance of generally average or better overall ability as a characteristic of individuals with SLD. For example, the official definition of learning disability of the Learning Disabilities Association of Canada states in part, "Learning Disabilities refer to a number of disorders which may affect the acquisition, organization, retention, understanding or use of verbal or nonverbal information. These disorders affect learning in individuals who *otherwise demonstrate at least average abilities essential for thinking and/or reasoning*" (*www.ldac-acta.ca/learn-more/ld-defined*, emphasis added; see also Harrison & Holmes, 2012).

Unlike some definitions of SLD, such as Canada's, the IDEA 2004 definition does not refer to overall cognitive ability level. However, the 2006 federal regulations contain the following phrasing: "(ii) The child exhibits a pattern of strengths and weaknesses in performance, achievement, or both, relative to age, State-approved grade-level standards, or intellectual development, that is determined by the group to be relevant to the identification of a specific learning disability" (U.S. Department of Education, 2006). Given the vagueness of the wording in the federal regulations, one could certainly infer that this phrase means that the cognitive and academic areas of concern are significantly lower than what is expected, relative to those of same-age peers or relative to otherwise average intellectual development. Indeed, there continues to be considerable agreement that a student who meets criteria for SLD has *some* cognitive capabilities that are at least average in relation to those of most people (e.g., Alfonso & Flanagan, 2018; Berninger, 2011; Feifer, 2012; Fiorello, Flanagan, & Hale, 2014; Flanagan & Alfonso, 2017; Geary et al., 2011; Hale & Fiorello, 2004; Hale et al., 2011; Harrison & Holmes, 2012; Kaufman, 2008; Kavale & Flanagan, 2007; Kavale & Forness, 2000; Kavale et al., 2009; Mather & Wendling, 2011 and Chapter 28, this volume; McCloskey et al., 2012; Naglieri & Feifer, 2018; Shaywitz, 2003). Moreover, the criterion of overall average or better cognitive ability (despite specific cognitive processing weaknesses) is necessary for differential diagnosis (see Lovett & Kilpatrick, 2018).

When a student does not meet criteria specified in the DD/C operational definition of SLD, it is possible that the student exhibits *slow learning* (SL; i.e., below-average ability to learn and achieve). However, by failing to differentially diagnose SLD from SL or other conditions that impede learning (such as ID or pervasive developmental disorders), the SLD construct loses its meaning, and there is a tendency (albeit well intentioned) to accept anyone under the SLD category who has learning difficulties for reasons other than specific cognitive dysfunction (e.g., Kavale & Flanagan, 2007; Kavale, Kauffman, Bachmeier, & LeFever, 2008; Lovett & Kilpatrick, 2018; Mather & Kaufman, 2006; Reynolds & Shaywitz, 2009a, 2009b). According to Kavale and colleagues (2008, p. 145), "About 14% of the school population may be deemed SL, but this group does not demonstrate unexpected learning failure, but rather an achievement level consonant with IQ level . . . slow learn[ing] has never been a special education category, and 'What should not happen is that a designation of SLD be given to a [child with] slow learning' (Kavale, 2005, p. 555)." Although the underlying and varied causes of the learning difficulties of all students who struggle academically *should be investigated and addressed,* an accurate SLD diagnosis is necessary because it informs instruction (e.g., Hale et al., 2010). As such, it seems prudent for practitioners to adhere closely to the DD/C operational definition of SLD (or other alternative research-based models), so that SLD can be differentiated from other disorders that also manifest themselves in academic difficulty (Berninger, 2011; Della Toffalo, 2010; Lovett & Kilpatrick, 2018).

Although it may be some time before consensus is reached on what constitutes "at least average overall cognitive ability" or "at least average ability to think and reason" for SLD identification, a child who has SLD, *generally speaking,* ought to be able to perform academically at a level approximating that of his or her more typically achieving peers *when provided with individualized instruction as well as appropriate accommodations, and instructional and curricular modifications alongside remedial interventions.* In addition, for a child with SLD to reach performances (in terms of both rate of learning and level of achievement) approximating those of his or her nondisabled peers, the child must possess the ability to learn compensatory strategies and apply them independently; this often requires higher-level thinking and reasoning, including intact executive processes (see Maricle & Avirett, Chapter 36, this volume; McCloskey, Perkins, & Van Divner, 2009). Individuals with SLD can minimize the effects of cognitive processing weaknesses on their ability to access instruction and the curriculum under certain circumstances (Mascolo et al., 2014). Special education provides the mechanism to assist a child with SLD in minimizing or bypassing his or her processing deficits through individualized instruction and intervention and through the provision of appropriate adaptations, accommodations, remediation, and compensatory strategies. However, to succeed in minimizing the effects of an individual's cognitive processing weaknesses in the educational setting to the point of achieving at or close to grade level, at least average overall ability to think and reason is very likely to be requisite, especially in upper elementary school and beyond (see Fuchs & Young, 2006, for a discussion of the mediating effects of IQ on response to intervention). Of course, it is important to understand that while at least average overall ability to think and reason is probably necessary for a child with SLD to be successful at minimizing his or her cognitive processing deficits, many other factors may facilitate or inhibit academic performance, including motivation, determination, perseverance, familial support, quality of individualized instruction, student–teacher relationships, and existence of comorbid conditions (see Flanagan et al., 2012, for a discussion; see also Flanagan & Schneider, 2016).

Determining at least average ability to think and reason for a child who has a below-average cognitive aptitude–achievement consistency is not a straightforward task, and there is no agreed-upon method for determining this condition or even requiring this condition as part of a state's or district's SLD identification guidelines (i.e., it is part of some methods of SLD identification, but not all methods). The main difficulty in determining whether an individual with *specific* cognitive weaknesses has at least average ability to think and reason (as determined by an estimate of *g*)[3] is that the global ability score or scores available on a cognitive or intelligence battery may be attenuated by the cognitive processing weakness(es). Most batteries have a total test score that is an aggregate of *all* (or nearly all) abilities and processes measured by the instrument. As such, in many instances, an individual's specific cognitive weaknesses or deficits attenuate the total test score on these instruments. This problem with ability tests was noted as far back as the 1920s, when Orton stated, "It seems probably that psychometric tests as ordinarily employed give an entirely erroneous and unfair estimate of the intellectual capacity of these [learning disabled] children" (1925, p. 582; cf. Mather, 2011). Perhaps for this reason, intelligence and cognitive ability batteries have become more differentiated, offering a variety of specific cognitive ability composites and options for global ability estimates. Nevertheless, there is increasing agreement that a child who meets criteria for SLD has at least some cognitive capabilities that are indeed average or better (e.g., Berninger, 2011; Flanagan et al., 2008, 2017; Geary et al., 2011; Hale & Fiorello, 2004; Hale et al., 2010; Kaufman, 2008; Kavale & Flanagan, 2007; Kavale & Forness, 2000; Kavale et al., 2009; Naglieri & Feifer, 2018).

To determine whether a child who demonstrates a below-average cognitive aptitude–achievement consistency also has at least average ability to think and reason, consistent with the DD/C model, X-BASS is used (see Flanagan et al., Chapter 27, this volume). The X-BASS provides a means of parceling out cognitive deficits from global functioning and judging the robustness of the spared abilities or cognitive strengths. This program is not meant to replace clinical judgment, but rather to inform it. Others have also developed methods and suggested formulas for determining whether individuals have cognitive strengths that are in stark contrast to their cognitive weaknesses (Naglieri, 2011; see also Fiorello & Wycoff, Chapter 26, this volume). Ultimately, the determination regarding whether a child with a below-average cognitive aptitude–achievement consistency has an SLD (and not SL or ID, for example), or exhibits unexpected (not expected) underachievement, must rely to some extent on clinical judgment.[4] Such judgment, however, is bolstered by converging data from multiple sources that were gathered via multiple methods and clinical tools (Alfonso & Flanagan, 2018; Flanagan & Alfonso, 2011).

Even when it is determined that a student has overall average ability to think and reason, along with a below-average cognitive aptitude–achievement consistency, these findings alone do not satisfy the criteria for a PSW consistent with the SLD construct in the DD/C operational definition. This is because it is not yet clear whether the differences between the score representing overall ability and those representing specific cognitive and academic weaknesses or deficits are statistically significant, meaning that such differences are reliable differences (i.e., not due to chance). Moreover, it is not yet clear whether the cognitive area or areas of weakness are domain-specific, and whether the academic area or areas of weakness (or underachievement) are unexpected.

Domain-Specific Cognitive Weaknesses or Deficits: The First Discrepancy

SLD has been described as a condition that is *domain-specific*. In other words, areas of cognitive weakness or deficit are circumscribed, meaning that while they interfere with learning and achievement, they are not pervasive and do not affect all or nearly all areas of cognition. According to Stanovich (1993), "The key deficit must be a vertical faculty rather than a horizontal faculty—a domain-specific process rather than a process that operates across a variety of domains" (p. 279). It is rare to find an operational definition that specifies a criterion for determining that the condition is "domain-specific." Some suggest that this condition is supported by a statistically significant difference between a student's overall cognitive ability and a score representing the individual's cognitive area of weakness (e.g., Hale & Fiorello, 2004; Naglieri, 2011).

However, a statistically significant difference between two scores means only that the difference is not due to chance; it does not provide information about the *rarity* or infrequency of the difference in the general population. Some statistically significant differences are common in the general population; others are not. Therefore, to determine whether the cognitive area that was identified as a weakness by the evaluator is domain-specific, the difference between the individual's actual and expected performance in this area should be uncommon in the general population.

X-BASS is needed to conduct the calculations necessary (1) to determine if a proxy for *g* can be derived, based on the cognitive areas designated as strengths; and (2) to arrive at an overall ability (*g*) estimate. X-BASS then uses the individual's unique pattern of strengths (proxy for *g*) to predict where the individual was expected to perform in the cognitive domain that is weak, and reports whether the difference between predicted and actual cognitive performance is rare relative to same-age peers (i.e., occurs in about 10% or less of the general population). A rare difference is considered a *domain-specific weakness* (see Flanagan et al., Chapter 27, this volume, for more detail).

Unexpected Underachievement: The Second Discrepancy

Traditionally, ability–achievement discrepancy analysis was used to determine whether an individual's underachievement (e.g., reading difficulty) was unexpected (i.e., the individual's achievement was not at a level commensurate with his or her overall cognitive ability). A particularly salient problem with the ability–achievement discrepancy approach is that a total test score from a cognitive or intelligence test (e.g., Full Scale IQ or FSIQ) is often used as the estimate of overall ability. However, for an individual with SLD, the total test score is often attenuated by one or more specific cognitive weaknesses or deficits, and therefore may provide an unfair or biased estimate of the individual's overall intellectual capacity. Furthermore, when the total test score is attenuated by specific cognitive weaknesses or deficits, the ability–achievement discrepancy is often not statistically significant, which frequently results in denying the student much-needed academic interventions and special education services (e.g., Aaron, 1995; Hale et al., 2011). For this reason, the WISC-V includes the General Ability Index (GAI) as an alternative to the FSIQ and the WJ IV includes the Gf-Gc composite for use in comparison (discrepancy) procedures—an alternative that Flanagan and her colleagues have advocated for many years (e.g., see Appendix H in Flanagan, McGrew, & Ortiz, 2000; Appendix H in Flanagan et al., 2013).

The DD/C operational definition circumvents the problem that plagued traditional ability–achievement discrepancy methods by determining whether the individual has at least average ability to think and reason, *despite one or more cognitive areas of weakness*. As stated above, X-BASS calculates a proxy for *g* when an individual's designated areas of strength are sufficient for this purpose. X-BASS then uses this value to predict where the individual was expected to perform in the academic domain that is weak, and reports whether the difference between predicted and actual academic performance is rare relative to same-age peers (i.e., occurs in about 10% or less of the general population). A rare difference is considered *unexpected underachievement* (see Flanagan et al., Chapter 27, this volume, for more detail).

Level V: SLD's Adverse Impact on Educational Performance

When a child meets criteria for an SLD diagnosis (i.e., when criteria for levels I through IV are met), it is typically obvious that the child has difficulties in daily academic activities that need to be addressed. The purpose of the fifth and final level of evaluation is to determine whether the identified condition (i.e., SLD) impairs academic functioning to such an extent that special education services are warranted.

The legal and diagnostic specifications of SLD necessitate that practitioners review the whole of the collected data and make a professional judgment about the extent of the adverse impact that any measured deficit has on an individual's performance in one or more areas of learning or academic achievement. Essentially, level V analysis serves as a kind of quality control feature designed to prevent the application of an SLD diagnosis in cases in which "real-world" functioning is not in fact impaired or substantially limited, compared to that of same-age peers in the general population—regardless of the patterns seen in the data.

This final criterion requires practitioners to take a very broad survey not only of the entire array of data collected during the assessment, but also of the real-world manifestations and practical implications of any presumed disability. In general, if the criteria at levels I through IV have been met, it is likely that in the majority of cases, level V analysis serves only to support conclusions that have already been drawn. However, in cases where data may be equivocal, level V analysis is an important safety valve, ensuring that any representations of SLD suggested by the data are indeed manifested in observable impairments in one or more areas of functioning in real-life settings.

Children with SLD require individualized instruction, accommodations, and curricular modifications to varying degrees, based on such factors as the nature of the academic setting, the severity of the SLD, the developmental level of each child, the extent to which each child can compensate for specific weaknesses, the way instruction is delivered, the content being taught, and so forth. As such, some children with SLD may not require special education services, such as when their academic needs can be met through classroom-based accommodations (e.g., use of a word bank during writing tasks, extended time on tests) and/or differentiated instruction (e.g., allowing a student with a writing deficit to record reflections on a reading passage and transcribe them outside the classroom prior to submitting a written product). Other children with SLD may require both classroom-based accommodations *and* special education services. And in a case where a child with SLD is substantially impaired in the general education or inclusive setting, a self-contained special education classroom may be required to meet his or her academic needs adequately.

There are two possible questions at Level V that must be answered by the multidisciplinary team (MDT). First, can the child's academic difficulties be remediated, accommodated, or otherwise compensated for without the assistance of individualized special education services? If the answer is yes, then services (e.g., accommodations, curricular modifications) may be provided, and their effectiveness monitored, in the general education setting. If the answer is no, then the MDT must answer the second question: What are the nature and extent of special education services that will be provided to the child? In answering this question, the MDT must ensure that individualized instruction and educational resources are tailored to the child in the least restrictive environment. Furthermore, such interventions should be linked to assessment (i.e., the identified cognitive and academic weaknesses) and should be evidence based.

Summary of the DD/C Operational Definition of SLD

In the preceding paragraphs, we have provided a summary of the DD/C operational definition of SLD. This definition provides a research-based framework for the practice of SLD identification and is likely to be most effective when it is informed by advances in cognitive and neuropsychological theory and research that support (1) the identification and measurement of constructs associated with SLD; (2) the relationships between cognitive abilities and processes and academic skills; and (3) a defensible method of interpreting results. Among the many important components of the definition, we have focused primarily on specifying criteria at the various levels of evaluation to establish the presence of SLD in a manner consistent with IDEA 2004 and its attendant regulations. These criteria include identification of empirically related academic and cognitive abilities and processes in the below-average range, compared to those of same-age peers from the general population; determination that exclusionary factors are not the primary cause of the identified academic and cognitive deficits; and identification of a pattern of performance reflecting domain-specific cognitive weaknesses, unexpected underachievement, and at least average ability to think and reason.

When the quantitative criteria specified at each level of the operational definition are met, as determined by X-BASS, and exclusionary factors have been ruled out as the primary cause of learning difficulties, it may be concluded that the data gathered are sufficient to support a diagnosis or classification of SLD. Because the conditions outlined in Figure 22.2 are based on current SLD research, and the calculations carried out by X-BASS are psychometrically sound, the DD/C operational definition represents progress toward a more complete and defensible approach to the process of evaluating SLD than previous (and many competing) methods (see Flanagan et al., Chapter 27, this volume; Miller, Maricle, & Jones, 2016).

The PSW Approach in Perspective

Given its increasing popularity, research on the PSW approach is emerging. One emerging body of research indicates that there is a lack of agreement among PSW models. This research also suggests that PSW models are effective at determining *who does not have SLD,* but they are not as effective at determining *who does have SLD.* Valid points are made about potential weaknesses of PSW models in this literature (e.g., Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016a, 2016b; Miciak, Fletcher, Stuebing, Vaughn, & Tolar, 2014; Stuebing, Fletcher, Branum-Martin, & Francis, 2012). However, it is important to understand that among the studies that have been conducted thus far, there are misrepresentations of PSW models, faulty assumptions about PSW models, and questions about the appropriateness of the methodology used to evaluate the assumptions underlying these models (see Flanagan & Schneider, 2016). Nevertheless, those engaged in PSW research should be commended for their work and for getting the conversation going. The current research has already sparked new ideas on how to evaluate the accuracy of the PSW approach more effectively (see Schneider's contribution in Flanagan, Alfonso, & Schneider, 2018; see also Miller et al., 2016).

Another emerging body of research provides support for a neuropsychological PSW approach (Hale et al., 2010, 2016). Specifically, this research shows the relevance of PSW methods for differential diagnosis of SLD in reading (e.g., Feifer, Nader, Flanagan, Fitzer, & Hicks, 2014), math (e.g., Kubas et al., 2014), and written expression (e.g., Fenwick et al., 2016). Valid points are made about the potential strengths of PSW models in this literature. Although valid points are made both for and against the use of PSW models, the results of the studies that have been published to date are affected by methodological preferences used to analyze the data, as well as the accuracy (and inaccuracy) of the assumptions made about each PSW model (for brief discussions, see Alfonso & Flanagan, 2018; Fiorello & Wycoff, Chapter 26, this volume; Flanagan & Schneider, 2016).

NOTES

1. Most individuals have statistically significant strengths and weaknesses in their cognitive ability and processing profiles. Intraindividual differences in cognitive abilities and processes are commonplace in the general population (McGrew & Knopik, 1996; Oakley, 2006). Therefore, statistically significant variation in cognitive and neuropsychological functioning in and of itself must not be used as de facto evidence of SLD. Instead, the pattern must reflect what is known about the nature of SLD (see Figure 22.2).

2. The term *causal* as used within the context of the DD/C model has been misconstrued to mean *deterministic*. That is, if we know the causal inputs, we can predict the outcome perfectly (Kranzler et al., 2016b). However, just because the causal inputs may be known, the outcomes clearly and obviously cannot be predicted perfectly. Cognitive abilities are indeed causally related to academic abilities, but the relationship is *probabilistic,* not deterministic, and is of moderate size (Flanagan & Schneider, 2016). The finding of cognitive weaknesses raises the risk of academic weaknesses; it does not guarantee academic weaknesses (Flanagan & Schneider), as assumed by Kranzler and colleagues. Likewise, it should not be assumed that the finding of academic weaknesses means that there are related cognitive weaknesses (again, a faulty assumption made by Kranzler et al.). As most practitioners know, in many cases there are no cognitive correlates to academic underachievement. This is because academic weaknesses may be related to numerous factors, only one of which is a cognitive weakness.

3. Many scholars use the term *overall cognitive/intellectual ability* interchangeably with the first factor that emerges in a factor analysis of cognitive tests—that is, Spearman's *g.* The estimates of overall cognitive or intellectual ability (or ability to think and reason) referred to in this chapter are consistent with this conceptualization.

4. Overall average (or better) cognitive ability or at least average ability to think and reason is difficult to

determine in students with SLD because their specific cognitive deficits often attenuate their total test scores (e.g., IQ). Therefore, such decisions should be based on multiple data sources and data-gathering methods. For example, a student with an SLD in mathematics may have a below-average WISC-V Full Scale IQ, due to deficits in processing speed and working memory (Flanagan & Alfonso, 2017; Geary et al., 2011). However, if the student has an average or better WISC-V GAI and average or better reading and writing ability, for example, then it is reasonable to assume that this student has at least average ability to think and reason. Of course, the more converging data sources that are available to support this conclusion, the more confidence one can place in such a judgment. The X-BASS calculates a value called the *facilitating cognitive composite (FCC)* that summarizes the individual's cognitive integrities or strengths when such a value is considered a good proxy of *g* given its constituents. The FCC is used in the PSW analysis conducted by X-BASS.

REFERENCES

Aaron, P. G. (1995). Differential diagnosis of reading disabilities. *School Psychology Review, 24,* 345–360.

Alfonso, V. C., & Flanagan, D. P. (2018). *Essentials of specific learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bateman, B. (1965). An educational view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders* (Vol. 1, pp. 219–239). Seattle, WA: Special Child.

Berninger, V. W. (2011). Evidence-based differential diagnosis and treatment of reading disabilities with and without comorbidities in oral language, writing, and math: Prevention, problem-solving consultation, and specialized instruction. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 203–232). Hoboken, NJ: Wiley.

Burns, M. K., Maki, K. E., Warmbold-Brann, K., & Preast, J. L. (2018). Using student response to intervention to identify SLD: Requirements, recommendations, and future research. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 257–284). Hoboken, NJ: Wiley.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.

Decker, S. L., Bridges, R. M., & Vetter, T. (2018). Cognitive neuroscientific contributions to theoretical understanding of specific learning disabilities. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 285–306). Hoboken, NJ: Wiley.

Della Toffalo, D. A. (2010). Linking school neuropsychology with response-to-intervention models. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 159–183). Hoboken, NJ: Wiley.

Feifer, S. G. (2012, May 11). *Integrating RTI with cognitive neuropsychology: A scientific approach to reading.* Paper presented at the Fordham University 4th Annual Assessment Conference, New York.

Feifer, S. G., Nader, R. G., Flanagan, D. P., Fitzer, K. R., & Hicks, K. (2014). Identifying specific reading subtypes for effective educational remediation. *Learning Disabilities: A Multidisciplinary Journal, 20,* 18–30.

Fenwick, M. E., Kubas, H. A., Witzke, J. W., Fitzer, K. R., Miller, D. C., Maricle, D. E., et al. (2016). Neuropsychological profiles of written expression learning disabilities determined by concordance-discordance model criteria. *Applied Neuropsychology: Child, 5*(2), 83–96.

Fiorello, C. A., Flanagan, D. P., & Hale, J. B. (Eds.). (2014). The utility of the pattern of strengths and weaknesses approach [Special issue]. *Learning Disabilities: A Multidisciplinary Journal, 20*(1).

Flanagan, D. P., & Alfonso, V. C. (2011). *Essentials of specific learning disability identification.* Hoboken, NJ: Wiley.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment.* Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2011). A CHC-based operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 233–298). Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 459–483). New York: Guilford Press.

Flanagan, D. P., Alfonso, V. C., & Schneider, W. J. (2018). *Using WISC-V and PSW for SLD identification amidst the controversy.* Mini-skills workshop presented at the annual meeting of the National Association of School Psychologists, Chicago, IL.

Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., McDonough, E. M., & Ortiz, S. O. (2018). Dual discrepancy/consistency operational definition of SLD: Integrating multiple data sources and multiple data gathering methods. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 329–430) Hoboken, NJ: Wiley.

Flanagan, D. P., Fiorello, C., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools, 47,* 739–760.

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2017). *The cross-battery assessment software system, version 2.0* (X-BASS v2.0). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities.* Boston: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2006). *The achievement test desk reference (ATDR): A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., & Schneider, W. J. (2016). Cross-battery assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues' "Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth." *International Journal of School and Educational Psychology, 4*(3), 137–145.

Fletcher, J. M., Barth, A. E., & Stuebing, K. K. (2011). A response to intervention (RTI) approach to SLD identification. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 115–144). Hoboken, NJ: Wiley.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guilford Press.

Fletcher, J. M., & Miciak, J. (2018). A response to intervention (RTI) approach to SLD identification. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 221–256). Hoboken, NJ: Wiley.

Fletcher, J. M., Taylor, H. G., Levin, H. S., & Satz, P. (1995). Neuropsychological and intellectual assessment of children. In H. Kaplan & B. Sadock (Eds.), *Comprehensive textbook of psychiatry* (6th ed., pp. 581–601). Baltimore: Williams & Wilkins.

Fletcher-Janzen, E., & Reynolds, C. R. (Eds.). (2008). *Neuropsychological perspectives on learning disabilities in the era of RTI: Recommendations for diagnosis and intervention.* Hoboken, NJ: Wiley.

Fuchs, D., & Young, C. L. (2006). On the irrelevance of intelligence in predicting responsiveness to reading instruction. *Exceptional Children, 73,* 8–30.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13,* 204–219.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2011). How SLD manifests in mathematics. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 43–64). Hoboken, NJ: Wiley.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2017). Early conceptual understanding of cardinality predicts superior school-entry number system knowledge. *Psychological Science, 29*(2), 191–205.

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223–236.

Hale, J. B., Chen, S. A., Tan, S. C., Poon, K., Fitzer, K. R., & Boyd, L. A. (2016). Reconciling individual differences with collective needs: The juxtaposition of sociopolitical and neuroscience perspectives on remediation and compensation of student skill deficits. *Trends in Neuroscience and Education, 5*(2), 41–51.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hale, J. B., Flanagan, D. P., & Naglieri, J. A. (2008). Alternative research-based methods for IDEA (2004) identification of children with specific learning disabilities. *Communique, 36*(8), 1, 14–15.

Hale, J. B., Wycoff, K. L., & Fiorello, C. A. (2011). RTI and cognitive hypothesis testing for identification and intervention of specific learning disabilities: The best of both worlds. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 173–201). Hoboken, NJ: Wiley.

Harrison, A. G., & Holmes, A. (2012). Easier said than done: Operationalizing the diagnosis of learning disability for use at the postsecondary level in Canada. *Canadian Journal of School Psychology, 27,* 12–34.

Hinshelwood, J. (1917). Congenital word-blindness. *Lancet, 190,* 980.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement.* New York: Guilford Press.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 118 Stat. 2147 (2004).

Kaufman, A. S. (2008). Neuropsychology and specific learning disabilities: Lessons from the past as a guide to present controversies and future clinical practice. In E. Fletcher-Janzen & C. Reynolds (Eds.), *Neuropsychological perspectives on learning disabilities in an era of RTI: Recommendations for diagnosis and intervention* (pp. 1–13). Hoboken, NJ: Wiley.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement, Third Edition.* Bloomington, MN: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2018). *Kaufman Assessment Battery for Children—Second Edition Normative Update.* Bloomington, MN: Pearson.

Kavale, K. A. (2005). Identifying specific learning disability: Is responsiveness to intervention the answer? *Journal of Learning Disabilities, 38,* 553–562.

Kavale, K. A., & Flanagan, D. P. (2007). Utility of RTI and assessment of cognitive abilities/processes in evaluation of specific learning disabilities. In S. Jimerson, M. Berns, & A. Van Der Heyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 130–147). New York: Springer.

Kavale, K. A., & Forness, S. R. (2000). What definitions of learning disability say and don't say: A critical analysis. *Journal of Learning Disabilities, 33,* 239–256.

Kavale, K. A., Kauffman, J. M., Bachmeier, R. J., & LeFever, G. B. (2008). Response-to-intervention: Separating the rhetoric of self-congratulation from the reality of specific learning disability identification. *Learning Disability Quarterly, 31,* 135–150.

Kavale, K. A., Spaulding, L. S., & Beam, A. P. (2009). A time to define: Making the specific learning disability definition prescribe specific learning disability. *Learning Disability Quarterly, 32,* 39–48.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher-order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 30,* 89–119.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016a). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School and Educational Psychology, 4*(3), 124–136.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016b). Cross-battery assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School and Educational Psychology, 4*(3), 146–157.

Kubas, H. A., Drefs, M. A., Poole, J. M., Schmid, A. D., Holland, S. C., & Fiorello, C. A. (2014). Cognitive and academic profiles associated with math disability subtypes. *Learning Disabilities: A Multidisciplinary Journal, 20*(1), 31–44.

Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lichtenstein, R., & Klotz, M. B. (2007). Deciphering the Federal Regulations on identifying children with specific learning disabilities. *Communique, 36*(3), 1, 13–16.

Lovett, B. J., & Kilpatrick, D. A. (2018). Differential diagnosis of specific learning disability versus other difficulties. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 549–572). Hoboken, NJ: Wiley.

Mascolo, J. T., Alfonso, V. C., & Flanagan, D. P. (2014). *Essentials of planning, selecting, and tailoring interventions for unique learners.* Hoboken, NJ: Wiley.

Mather, N. (2011). *Let's stop monkeying around: What we know about reading disabilities.* Verona, NY: New York Association of School Psychologists.

Mather, N., & Kaufman, N. (2006). Introduction to the special issue, part one: It's about the what, the how well, and the why. *Psychology in the Schools, 43,* 747–752.

Mather, N., & Wendling, B. J. (2011). How SLD manifests in writing. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 65–88). Hoboken, NJ: Wiley.

Mazzocco, M. M. M., & Vukovic, R. (2018). How SLD manifests in mathematics. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 59–102). Hoboken, NJ: Wiley.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties.* New York: Routledge.

McCloskey, G., Whitaker, J., Murphy, R., & Rogers, J. (2012). Intellectual, cognitive, and neuropsychological assessment in three-tier service delivery systems in schools. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 852–881). New York: Guilford Press.

McDonough, E. M., & Flanagan, D. P. (2016). Use of the Woodcock–Johnson IV in the diagnosis of specific learning disabilities in school-age children. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation* (pp. 212–253). San Diego, CA: Academic Press.

McDonough, E. M., Flanagan, D. P., Sy, M., & Alfonso, V. C. (2017). Specific learning disorder. In S. Goldstein & M. DeVries (Eds.), *Handbook of DSM-5 disorders in children and adolescents* (pp. 77–104). New York: Springer.

McGrew, K. S., & Knopik, S. N. (1996). The relationship between intra-cognitive scatter on the Woodcock–Johnson Psycho-Educational Battery—Revised and school achievement. *Journal of School Psychology, 34,* 351–364.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

McGrew, K. S., & Wendling, B. (2010). Cattell–Horn–Carroll cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651–675.

Miciak, J., Fletcher, J. M., Stuebing, K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29*(1), 21–37.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment.* Hoboken, NJ: Wiley.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Miller, D. C., Maricle, D. E., & Jones, A. M. (2016). Comparing three patterns of strengths and weaknesses models for the identification of specific learning disabilities. *Learning Disabilities: A Multidisciplinary Journal, 21*(2), 31–45.

Monroe, M. (1932). *Children who cannot read: The analysis of reading disabilities and the use of diagnostic tests in the instruction of retarded readers.* Chicago: University of Chicago Press.

Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 145–172). Hoboken, NJ: Wiley.

Naglieri, J. A., & Feifer, S. G. (2018). Pattern of strengths and weaknesses made easy: The discrepancy/consistency model. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 431–474). Hoboken, NJ: Wiley.

Nelson, N. W., & Wiig, E. H. (2018). How SLD manifests in oral expression and listening comprehension. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 145–194). Hoboken, NJ: Wiley.

Niileksela, C. R., Reynolds, M. R., Keith, T. Z., & McGrew, K. S. (2016). A special validity study of the WJ IV: Acting on evidence for specific abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 65–106). Amsterdam: Elsevier Academic Press.

Oakley, D. (2006). Intra-cognitive scatter on the Woodcock–Johnson Tests of Cognitive Abilities—Third Edition and its relation to academic achievement. *Dissertation Abstracts International, 67,* 1199B.

Ortiz, S. O., Melo, K. E., & Terzulli, M. A. (2018). Difference or disorder: Assessment of SLD with an English learner. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 505–548). Hoboken, NJ: Wiley.

Orton, S. T. (1937). *Reading, writing, and speech problems in children.* New York: Norton.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition (WIAT-III).* San Antonio, TX: Author.

Reynolds, C. R., & Shaywitz, S. E. (2009a). Response to intervention: Prevention and remediation, perhaps. Diagnosis, no. *Child Development Perspectives, 3,* 44–47.

Reynolds, C. R., & Shaywitz, S. E. (2009b). Response to intervention: Ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly, 34,* 130–145.

Schneider, W. J., & Kaufman, A. (2017). Let's not do away with comprehensive cognitive assessments just yet. *Archives of Clinical Neuropsychology, 32,* 8–20.

Schrank, F. A., Mather, N. S., & McGrew, K. S. (2014). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Shaywitz, S. (2003). *Overcoming dyslexia.* New York: Knopf.

Siegel, L. S. (1999). Issues in the definition and diagnosis of learning disabilities: A perspective on *Guckenberger v. Boston University. Journal of Learning Disabilities, 32,* 304–320.

Sotelo-Dynega, M., Flanagan, D. P., & Alfonso, V. C. (2018). Overview of specific learning disabilities. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed., pp. 3–28). Hoboken, NJ: Wiley.

Stanovich, K. E. (1993). The construct validity of discrepancy definitions of reading disability. In G. R. Lyon, D. B. Gray, J. F. Kavanagh, & N. A. Krasnegor (Eds.), *Better understanding learning disabilities: New views from research and their implications for education and public policy* (pp. 273–307). Baltimore: Brookes.

Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of Learning Disabilities, 32,* 350–361.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41*(1), 3–22.

Swanson, H. L. (1991). Operational definitions and learning disabilities: An overview. *Learning Disability Quarterly, 14*(4), 242–254.

U.S. Department of Education. (2006, August 14). 34 C.F.R. Parts 300 and 301: Assistance to states for the education of children with disabilities and preschool grants for children with disabilities: Final rule. *Federal Register, 71*(156), 46539–46845.

U.S. Office of Education. (1977). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42*(250), 65082–65085.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ–achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*(2), 192–212.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wendling, B., & Mather, N. (2009). *Essentials of evidence-based academic interventions.* Hoboken, NJ: Wiley.

Zirkel, P. A., & Thomas, L. B. (2010). State laws for RTI: An updated snapshot. *Teaching Exceptional Children, 42*(3), 56–63.

CHAPTER 23

Use of Intelligence Tests in the Identification of Children and Adolescents with Intellectual Disability

Ryan L. Farmer
Randy G. Floyd

Intellectual disability (ID)—previously called *mental retardation* (see Rosa's Law, 2010)—is characterized by significant deficits in intellectual and adaptive functioning, with onset during the developmental period (American Association on Intellectual and Developmental Disabilities [AAIDD], 2010; American Psychiatric Association [APA], 2013). *Intellectual functioning* is broadly defined by hierarchical composites (most typically IQs) that assess abstract thinking, judgment, learning, planning, problem solving, and reasoning. *Adaptive functioning*, or adaptive behavior, is defined by conceptual, practical, and social domains of behavior. Deficits in both intellectual and adaptive functioning need to be present and have an impact on the individual prior to age 18 in order for these deficits to be considered ID. While broad, this definition is commensurate with all clinical diagnostic (AAIDD, 2010; APA, 2013; World Health Organization [WHO], 2010, 2016) and governmental classification (Individuals with Disabilities Education Improvement Act of 2004 [IDEA, 2004]; McNicholas et al., 2018; Social Security Administration [SSA], n.d.) procedures for ID. The impact of ID on an individual can be significant. To better understand ID, it is beneficial to consider the presentation of individuals with ID, causal factors, prevalence and incidence rates, and comprehensive assessment of ID for the purposes of diagnosis and treatment.

PRESENTATION, EPIDEMIOLOGY, AND ETIOLOGY

Individuals with ID will present with a variety of intellectual and adaptive limitations, ranging from impairments in developmental repertoires (e.g., gross motor skills) to academic and social difficulties (APA, 2013; National Academies of Sciences, Engineering, and Medicine, 2015). Behavioral excess may also manifest partially as a result of reduced functional communication (McClintock, Hall, & Oliver, 2003). For instance, Cooper and colleagues (Cooper, Smiley, Allan, et al., 2009; Cooper, Smiley, Jackson, et al, 2009) determined prevalence rates of 6.3% for physical aggression, 4.9% for self-injurious behavior, and 3.0% for destructive behavior for children with ID. In addition to behavioral excess, behavioral deficits may warrant clinical or educational supports.

The international prevalence rate of ID—that is, the proportion of the population affected by ID—is highly variable across studies, but it is estimated to be 10.37 per 1000 in the population, or 1%, with a higher prevalence rate in males (McKenzie, Milton, Smith, & Oullette-Kuntz, 2016). Historically, IQ-only approaches to the diagnosis of ID would result in a prevalence rate of approximately 2–3%, based on a Gaussian curve. Within the clinical population, approximately 85%, 10%, 3.5%, and 1.5% of cases are classified at mild, moderate, severe, and profound levels of severity, respectively (National Academies, 2015). Furthermore, in the United States between 2008 and 2013, students with ID accounted for approximately 2% of children served via special educational programs (U.S. Department of Education, 2016). In terms of identification, individuals with dysmorphic features, documented microcephaly, or significant impairment, such as deficits in gross and fine motor skills or marked language delays, are more likely to be identified with ID within the first 2 years of life (APA, 2013).

The etiology of ID varies as significantly as its presentation. Prenatal influences include all known genetic and metabolic anomalies; malformations during brain development; *in utero* exposure to toxins, teratogens, drugs, and alcohol; and exposure to maternal or placental infection. Included within this broad category of prenatal influences, genetic disorders account for a significant proportion of ID cases, with estimates between 15 and 20% (see Kaufman, Ayub, & Vincent, 2010). For this reason, genetic counseling is frequently recommended if the etiology of ID is unclear (American Academy of Pediatrics & American College of Medical Genetics and Genomics, 2013). Perinatal influences are those that occur immediately before, during, or after birth, and include a variety of childbirth-related traumas. Finally, postnatal influences are those that occur after childbirth and include infection; neurocognitive disorders, including both open and closed head injuries, as well as congenital and degenerative conditions; exposure to toxins; and sustained social deprivation. A comprehensive review of the epidemiology and etiology of ID is outside the scope of this chapter (see Walker & Johnson, 2006).

DEFINITION OF ID

ID has been defined by various diagnostic and classification criteria (see Table 23.1). Fortunately, the diagnostic criteria for ID are largely similar across organizations and have been relatively stable since 1958 (AAIDD, 2010). However, updates and minor inconsistencies may affect the assessment process and identification rate. For example, Papazoglou, Jacobson, McCabe, Kaufmann, and Zabel (2014) found that changes to the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5; APA, 2013) criteria may result in as much as a 9% reduction in diagnosis rates when compared to diagnosis rates based on the previous version of the DSM (APA, 2000).

A number of sources have outlined the differences among the various sets of criteria for ID (e.g., Floyd, Woods, Singh, & Hawkins, 2016), but few have included comparisons with (1) the *International Statistical Classification of Diseases and Related Health Problems*, 10th revision, second edition (ICD-10; WHO, 2004); (2) its probable successor, the beta draft of ICD-11 (WHO, 2016); (3) IDEA (2004) eligibility criteria; and (4) SSA (n.d.) criteria. For the most part, all agencies utilize three classification elements: deficits in (1) intellectual functioning, (2) adaptive functioning, and (3) onset during the developmental period. As with any operational definition, diagnostic criteria should be judged on the basis of their objectivity, clarity, and completeness (Hawkins & Dobes, 1977). In this case, the classification elements of the definition can be deemed present or absent in a systematic, measurable manner; are clear, without need for further interpretation; and are accompanied by boundary conditions that are delineated so that one can be included or excluded on the basis of objective data. Accordingly, the definition provided by the AAIDD (2010) is the most precise one available.

Intellectual Deficits

The AAIDD (2010), APA (2013), SSA (n.d.), and WHO (2004, 2016) define deficient intellectual functioning as an IQ at approximately two standard deviations or below (see the second column in Table 23.1). The language of these guidelines specifying that an IQ “equal to or below” two standard deviations below the mean is fairly consistent between organizations, and this criterion translates into a standard score equal to or below 70 on most intelligence tests. Only ICD-10 (WHO, 2004) requires a standard score below 70, with no provision for a diagnosis when an IQ is equal to 70. The outlier, IDEA (2004), specifies that intellectual functioning must be “significantly subaverage” with no quantifiable definition of subaverage. In response to the lack of specificity offered by the federal definition, many states have specified cutoff scores for IQ (McNicholas et al., 2018).

Adaptive Behavior Deficits

Historically, the focus in evaluating ID has been on intellectual functioning, but the most recent versions of the AAIDD (2010) and APA (2013)

TABLE 23.1. Intellectual Disability (ID) Criteria by Agency or Group

Agency or group	Intellectual deficit	Adaptive deficit	Onset	Level of impairment
AAIDD (2010)	IQ: ~ ≤2 *SD*	Overall or domain score; ≤2 *SD*	<18 years of age	Adaptive functioning and level of ongoing supports
DSM-5 (APA, 2013)	IQ: ~ ≤2 *SD*	At least one domain; sufficient impairment requiring ongoing support	Developmental period, often with early manifestation	Adaptive functioning and level of ongoing supports
ICD-10 (WHO, 2004)	IQ: <2 *SD*	Supplemental only	Developmental period implied	Intellectual impairment
ICD-11 beta draft (WHO, 2016)	IQ: ≤2 *SD*	Adaptive functioning; ≤2 *SD*	Developmental period	Adaptive and intellectual impairment
IDEA (2004)	Significantly subaverage general intellectual functioning	Adaptive behavior deficits	Developmental period	Level of impairment not delineated; impairment must impede educational performance
SSA (n.d.)	Significantly subaverage general intellectual functioning: A full-scale IQ of 70 or below, *or* a full-scale IQ of 71–75 and part score of 70 or below	Significant deficits in adaptive functioning in two broad areas	<18 years of age	Not delineated in definition

Note. IQ is used in this table to indicate composites representing psychometric *g*; such composites may not utilize the term *IQ*. AAIDD, American Association on Intellectual and Developmental Disabilities; APA, American Psychiatric Association; DSM-5, *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition; ICD-10, *International Statistical Classification of Diseases and Related Health Problems,* 10th revision, second edition; ICD-11 beta draft, *International Statistical Classification of Diseases and Related Health Problems,* 11th revision, beta draft; IDEA, Individuals with Disabilities Education Improvement Act of 2004; *SD,* standard deviation; SSA, Social Security Administration; WHO, World Health Organization.

definitions, and upcoming iterations of the WHO (2016) definition, are emphasizing adaptive functioning and the intensity of necessary supports in the home, school, and community. The SSA (n.d.) requires evidence of adaptive deficits across domains, but does not specify the type of evidence required. Overall, more variability is noted regarding the adaptive deficits category than the intellectual deficits category. These criteria are presented in the third column of Table 23.1. The AAIDD (2010), DSM-5 (APA, 2013), ICD-11 beta draft (WHO, 2016), IDEA (2004), and SSA (n.d.) require that adaptive deficits be documented. Only ICD-10 (WHO, 2004) considers adaptive behavior deficits to be supplementary information. However, even within those that require the documentation of adaptive deficits, the AAIDD and APA criteria permit clinicians to diagnose ID when a deficit is noted in only one domain of adaptive behavior (practical, social, or conceptual). The AAIDD also permits consideration of a composite adaptive score, whereas the ICD-11 beta draft (WHO, 2016) seemingly requires a composite adaptive score. Again, IDEA (2004) utilizes the least specific language, in that it only requires that deficits in adaptive behavior be present. Only about half of states specify the type of adaptive scores that should be considered; most do not specify cutoff criteria (McNicholas et al., 2018).

Age of Onset

Age of onset serves as the third functional component of the diagnostic criteria. Although there is minor variability in the terminology used across criteria, it is largely agreed that age of onset should be prior to age 18 (AAIDD, 2010). In the fourth column of Table 23.1, the onset criteria are specified. Only ICD-10 (WHO, 2004) does not specify the developmental period, which is only heavily implied in its description of ID. DSM-5 (APA, 2013) specifies that identification typically occurs during the early developmental period, and defines this period of time as prior to entering first grade.

CONSIDERATIONS IN MAKING A DIAGNOSIS

When a clinician is considering ID as a diagnosis, a number of factors may confound the assessment process. These factors fit into two broad categories: (1) competing hypotheses regarding the appropriate diagnosis, and (2) construct-irrelevant influences that reduce the validity of the assessment results. In addition to considering competing hypotheses and construct-irrelevant influences, clinicians should consider the influence of culture and language.

Differential Diagnosis

When a clinician is presented with a referral concern that warrants a comprehensive assessment for ID, common competing hypotheses that should be ruled out or considered for children include autism spectrum disorder (ASD), assorted communication disorders (CD), global developmental delay (GDD), neurodevelopmental disorders (NDD), and neurocognitive disorders (NCD). The process of addressing, considering, and accepting or ruling out competing diagnostic hypotheses is *differential diagnosis* (APA, 2013). Although these issues must be considered, it is of note that ASD, CD, and NCD can all be comorbid with ID if a comprehensive assessment reveals that the child or adolescent meets all relevant criteria. DSM-5 (APA, 2013) specifies that a diagnosis “should be made” (p. 39) when all three elements of the operational definition are met; however, clinicians should not diagnose ID solely on the presence of genetic or other primary medical conditions (e.g., Down syndrome). In general, clinicians should be knowledgeable about related diagnoses as well as appropriate procedures for ruling out competing hypotheses. Brief descriptions of major confounding diagnoses and their relation to ID are provided below; for the sake of brevity, DSM-5 nomenclature is employed throughout.

ASD is a pervasive developmental disorder that affects social-emotional functioning and is perhaps the most challenging co-occurring diagnosis to disentangle from ID, due to significant behavioral similarities between the two clinical populations (Matson & Shoemaker, 2009). The most salient factors in discriminating between ASD and ID are language development and social skills. A comprehensive assessment should be conducted by an interdisciplinary team of clinicians with sufficient expertise to determine whether the child or adolescent warrants a single diagnosis (i.e., ASD or ID) or dual diagnoses (i.e., both ASD and ID). Clinicians are strongly encouraged to engage in continuing education, due to the rapid progress of research in this area and advancement of best practices (see Volkmar et al., 2014). Of note, children with ASD may present with challenging behavioral topographies that further reduce the validity of standardized testing. Test examiners should have expertise with the types of behavior that may be present or should consult an expert in the area.

CD is a broad term encompassing language delays, disorders, and related language impairment, and has an estimated prevalence of 6–18% (American Speech–Language–Hearing Association, n.d.). As many as 4% of children with ID also present with comorbid CD (Pinborough-Zimmerman al., 2007). Children with CD may perform poorly on language-laden components of intelligence tests and may receive lower adaptive ratings in areas such as communication, preacademic skills, and social skills (American Speech–Language–Hearing Association, n.d.; Fujiki, Brinton, & Todd, 1996; Harrison & Oakland, 2015). When CD is suspected, consultation with a speech–language pathologist is strongly encouraged, and a comprehensive, interdisciplinary evaluation is necessary to make a differential diagnosis.

GDD is defined as deficits in a variety of developmental domains, such as intellectual functioning, adaptive functioning, or gross motor skills (Shevell, 2008). Pragmatically, DSM-5 (APA, 2013) has maintained this definition, which has also included children who cannot undergo diagnostic assessment for any number of reasons or who are too young to undergo such testing. That is, children ages 4 years, 11 months and younger may be diagnosed with GDD when a pattern of developmental deficits is present, but data regarding intellectual functioning or “clinical severity”

(APA, 2013, p. 41) is absent or unreliable. Similarly, for individuals older than 5 years of age, a diagnosis of unspecified intellectual disability (intellectual developmental disorder) may be provided in extenuating circumstances, such as when physical impairments or severe problem behaviors interfere with assessment (APA, 2013). GDD and unspecified ID are unique among the other disorders listed in this section, as they are exclusionary of a diagnosis of ID. These diagnostic classifications are only utilized in the event that relevant information is unavailable, but the preponderance of evidence supports a diagnosis.

Additionally, some children do not meet criteria to satisfy all classification elements of ID or other specific developmental disorders (e.g., GDD), but still have significant impairment in social, occupational, or other domains that warrants clinical services. Such children may be diagnosed with NDD. Clinicians may choose to specify the etiology of that disorder (e.g., NDD associated with prenatal alcohol exposure) or not (i.e., unspecified NDD). Differentiating GDD and unspecified ID from NDD may be challenging, but it should be based on a thorough review of all relevant criteria. NDD is, by default, a broader category than GDD or unspecified ID, but it does not include those situations described previously where valid results are unavailable. Similar to GDD and unspecified ID, NDD is exclusionary of a diagnosis of ID (APA, 2013). Uniquely, ICD-10 (WHO, 2004) also includes borderline intellectual functioning (BIF), which has the sole classification element that an individual's IQ is between 71 and 84. It is clear upon review that this version of BIF overlaps with NDD, but it does not require functional impairment in any domains. DSM-5 (APA, 2013) defines BIF similarly, but does not recognize it as a mental disorder; rather, it is considered a condition or problem that may justify assessment (e.g., cognitive testing) or treatment (e.g., discrete-trial training) procedures when the primary diagnosis does not solely warrant such procedures. It is recommended that clinicians use BIF solely as a modifier or to justify initial assessment (e.g., rule-out of ID) and not as a stand-alone diagnosis.

The group of disorders known as NCD, including traumatic brain injury, is characterized by regression in both cognitive and adaptive domains. The rate of acquired NCD (i.e., traumatic brain injury) requiring hospitalization is of significant concern among child and adolescent populations, with incidence rates as high as 74 in 100,000. Of those, the estimated prevalence of disability is as high as 20% (Thurman, 2014). The criteria for major NCD echo the first two elements of the criteria for ID: deficits in (1) intellectual functioning and (2) adaptive functioning (APA, 2013). One primary distinction between ID and NCD is that ID requires that onset of the deficits occur during the developmental period. Furthermore, the criteria for NCD only specify that the impacted domain demonstrates a noticeable decline from previous levels of functioning. When NCD results in stable cognitive and adaptive deficits that would otherwise meet the criteria for ID and the individual is within the developmental period, the child or adolescent is said to have met all of the classification elements of ID and could receive both diagnoses (APA, 2013).

Construct-Irrelevant Influences

The accurate assessment of ID is of utmost importance, and thus clinicians should act purposefully to ensure that their test results are valid (Bracken, 2000). A number of individual variables may result in *construct-irrelevant* influences, which introduce error and reduce the validity of assessment results. Stable construct-irrelevant influences include sensory (e.g., visual or auditory) impairments, physical (e.g., fine or gross motor) impairments, and speech and language impairments. Transient construct-irrelevant influences include fatigue, interfering behavioral topographies (e.g., impulsivity and noncompliance), illness, motivation, and mood lability (Bracken, 2000; Kranzler & Floyd, 2013). Identifying construct-irrelevant influences prior to assessment provides an opportunity for the clinician to consider modifying the assessment battery and to consider the use of testing accommodations. Any number of strategies and accommodations may be warranted to minimize construct-irrelevant influences or to elucidate results in light of a competing diagnostic hypothesis (see Herschell, Greco, Filcheck, & McNeil, 2002; Wechsler, 2014).

Culture and Language

Culture is a constellation of learned behaviors that are transferred between people within a given group. Culture shapes one's experiences through complex, culture-specific contingencies (Glenn, 2004). According to this logic, it is clear that all behavior is directly or indirectly shaped by cultural experiences. Based on the understanding that psychometric instruments measure samples of behavior, it is also logical that the assessment results obtained through the use of such measures are culturally shaped (Valencia & Lopez, 1992). In order to understand assessment results fully, clinicians must be familiar with the cultural backgrounds of their clients, as well as the cultures of the local community and of the local student population (Rhodes, Ochoa, & Ortiz, 2005).

As with culture, language proficiency and linguistic background can have a significant impact on the assessment results of children and adolescents from diverse backgrounds. Mather and Wendling (2014) posited that "the most important accommodation for students who are English language learners (ELL) is having an examiner who is knowledgeable about important issues relevant to second language acquisition, the assessment process, and the interpretation of test results for students" (p. 43). Clinical judgment in relation to how cultural and linguistic factors may influence assessment is paramount when a clinician is interpreting the many pieces of data obtained during the assessment process. Luckasson and Schalock (2015) provide an excellent review of the standards of clinical judgment within the field of ID, and a commentary about the necessity for considering cultural and linguistic factors.

BEST PRACTICES IN THE ASSESSMENT OF ID

A comprehensive, multidimensional assessment of ID goes beyond classification and must consider the individual from an ecological, reciprocal perspective (see Figure 23.1). The AAIDD (2010) provides a multidimensional framework of human functioning encompassing five domains: (1) intellectual abilities, (2) adaptive behavior, (3) health, (4) participation, and (5) context of the individual. Information pertaining to each domain is necessary for understanding assessment results and for

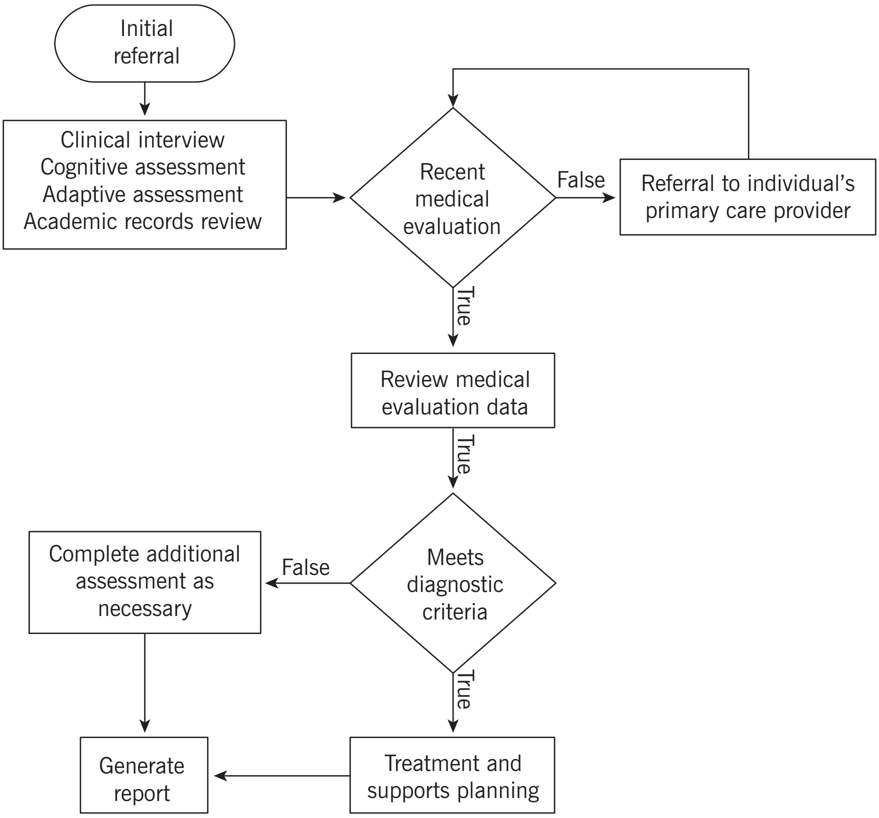


FIGURE 23.1. Process map for the assessment of ID.

successful treatment planning. In order to develop best practices in regard to ID assessment, it is necessary to consider how a clinician might go about evaluating each of the five domains. In general, a comprehensive and focused clinical interview is the foundation of the assessment process, and it should provide guidance for continued assessment across the five dimensions of human functioning.

Dimension 1: Intellectual Functioning

Intellectual functioning is typically assessed based on results from an individual, norm-referenced cognitive ability test battery like those described extensively elsewhere in this book, but we do not review such batteries in this chapter. Instead, we describe the importance of the scores these batteries yield during the assessment process. Generally speaking, best practices in the assessment of intellectual functioning include interpreting global composite scores, considering confidence intervals surrounding those scores, ensuring that the assessment instruments used are appropriate for the individual and are psychometrically sound, and completing accurate and standardized test administrations. The final section of this chapter also addresses controversial practices associated with the assessment of intellectual functioning.

Dimension 2: Adaptive Behavior

Adaptive functioning is an assortment of conceptual, practical, and social skills that an individual uses on a day-to-day basis. Adaptive behaviors make up an important dimension of human functioning as defined by the *International Classification of Functioning, Disability and Health* (ICF; WHO, 2001), and understanding an individual's adaptive skills not only constitutes the second classification element for a diagnosis of ID, but also informs treatment planning and supports. To determine whether an individual meets the criteria for a diagnosis of ID, norm-referenced instruments should be used with multiple, informed stakeholders (e.g., parent and teacher) to obtain cross-setting ratings of the individual's everyday performance. For more information on available instruments and recommendations for practice, see Floyd and colleagues (2015).

In addition to norm-referenced instruments, we encourage clinicians to consider criterion-referenced instruments (Partington, 2010), direct observations of adaptive behavior (Mannix, 2009), semistructured interviews with the client and stakeholders related to specific adaptive behavior strengths and weaknesses (Thompson et al., 2004), and combinations thereof, in order to facilitate treatment and supports planning. For young children and adolescents, we encourage clinicians to include instruments such as the Assessment of Basic Language and Learning Skills—Revised (Partington, 2010); the Verbal Behavior Milestones Assessment and Placement Program (Sundberg, 2008); and the Supports Intensity Scale (Thompson et al., 2004). These instruments, more so than others, can help to identify key areas for intervention development.

Dimension 3: Health

The AAIDD (2010), in alignment with the WHO (2001), defines *health* broadly as a client's physical health, mental health, and social well-being. Reviewing medical history, developmental history, and current medical status during a clinical interview is a key component, while consultation with the client's primary care provider provides continuity of care and ensures that an updated medical examination has been completed for the individual. Consultation with a pediatrician may yield significant information that directly affects primary and secondary diagnoses and the development of treatment goals and supports. Furthermore, engaging in multidisciplinary or interdisciplinary practice with the child or adolescent's primary care provider facilitates referrals and client care. Whenever possible, continuing communication between providers is recommended.

Dimension 4: Participation

A significant component of human functioning is the ability to engage with others socially; to take part in activities; and to assume roles in the home, school, workplace, and community. *Participation* is broadly defined as an individual's engagement in these activities in a manner typical of the person's culture and age group (AAIDD, 2010; WHO, 2001). Clinicians should consider participation during their broad screening of adaptive and intellectual functioning, and should consider how the child or adolescent's functioning affects his or her level of participation across domains. Additionally, clinicians should weigh the benefits of engaging in direct observation of behavior across settings, in an effort to better understand barriers to participation and to determine how supports may facilitate engagement.

Dimension 5: Context

An individual's context should be considered from an ecological perspective (Bronfenbrenner, 1979), wherein interactions with the individual's immediate environment, his or her community, and the encompassing culture and society are considered. These systems may affect an individual's various opportunities, expectations, resources, and social interactions. Broadly, clinicians should consider (1) personal factors, such as characteristics of the individual and family, specific strengths and weaknesses, and familial support; and (2) environmental factors, such as societal attitudes, accessibility, and community supports. Clinicians should also consider advantages and barriers across settings and within the community, the possible impact of the individual's health status on his or her functioning, and the person's level of social supports.

It is beyond the scope of this chapter to review each of these dimensions fully, but we encourage the reader to consider how the distinct components converge to provide a comprehensive perspective of the individual, his or her personal strengths and limitations, and the strengths and limitations of the environment. Not only is consideration of these five dimensions useful for the purpose of diagnosis, but it is also critical to developing appropriate and sufficient treatment and support plans.

ASSESSMENT OF INTELLECTUAL FUNCTIONING

What IQs Measure

As evident in Table 23.1, there is a historical and general consensus that intellectual deficits are best measured by global ability composites from cognitive ability test batteries that are broad in scope and multidimensional in their measurement. Although these global ability composites are no longer based on a ratio of mental age and chronological age that produces a true quotient, the term *intelligence quotient* (IQ) is often used to encapsulate a diversity of specific score labels. Throughout the remainder of this chapter, we use this term.

IQs from individually administered cognitive ability test batteries are typically derived from summing norm-referenced scores corresponding to component parts (i.e., subtests), which are often tasks containing similar items scaled in increasing difficulty. When these scores are aggregated, the effects of shared cognitive abilities are magnified, and the effects of cognitive abilities unique to each task are minimized. The product is a more general measure of intellectual abilities than any task could represent alone. In the same vein, this process of aggregation also increases the reliability of IQs; reliability is a much-valued property, especially when clinicians are making high-stakes diagnostic decisions based on these scores. IQs very frequently demonstrate reliability estimates of .95 and higher (Kranzler & Floyd, 2013), which is a very high standard not typically met by other measures in psychology. Reliability is discussed later in this chapter.

The AAIDD (2010) makes explicit references to *general mental ability* (a.k.a. psychometric *g*) as the focal point in assessment of intellectual functioning. Some of our research with various colleagues in recent years has investigated how well IQs from cognitive ability test batteries measure this psychological construct. This research used two methods to examine the "*g* saturation" of IQs derived from some of the most prominent test batteries available for children and adolescents. First, Reynolds, Floyd, and Niileksela (2013) employed confirmatory factor analysis methods applied to norming sample data from subtests from three batteries. These analyses produced a statistic called *hierarchical omega*, which represented the percentage of variance in the IQs that could be attributed to psychometric *g*. Results showed that 82–83% of this variance could be attributed to this construct. Clearly, these IQs are highly *g*-saturated.

In addition, we (Farmer, Floyd, Reynolds, & Kranzler, 2014) drew data from three validity studies, in each of which two test batteries were administered to the same sample of children and adolescents. Confirmatory factor analysis was used to create a latent variable representing the construct of psychometric *g* from one test battery. Then this latent variable was correlated with the IQ from the other test battery (administered concurrently). Results revealed very strong correlations and substantial *g* saturation across five IQs. Similar to the Reynolds and colleagues (2013) results, 77– 90% (M = 85%) of variance in the IQs could be attributed to psychometric *g*. Based on both measurement theory and psychometric analyses, the IQs yielded by these cognitive ability test batteries appear to be good representations of the construct they target. These values do not by any means indicate perfect measurement accuracy, but they do reflect remarkably high relations between a score used in the practice of psychology and the psychological entity it targets.

Reynolds and colleagues (2013) employed complete norming sample datasets, and we (Farmer et al., 2014) employed much smaller validity study samples including typically developing students. These samples were not "clinical samples," selected on the basis of some condition or diagnosis. As a result, it is unclear whether these results would generalize to those who are not typically developing—especially students with ID. Research by Detterman and Daniel (1989), Reynolds, Keith, and Beretvas (2010), and Reynolds (2013) focused on *Spearman's law of diminishing returns* (SLODR), providing keen insights into the *g* saturation of IQs with children and adolescents with (or suspected of having) ID. Sometimes called *ability differentiation,* SLODR reflects the finding that individual tasks included in cognitive ability test batteries correlate more strongly as general mental ability levels decrease across samples (and vice versa). Following SLODR, one can reason that IQs (as reflections of common variance across correlated tasks) are better measures of psychometric *g* for those who have ID than those at higher general ability levels—those with IQs in the average range and those with intellectual giftedness.

Using data from the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007) norming sample, and calculating hierarchical omega coefficients as mentioned previously, Reynolds (2013) demonstrated that the *g* saturation of a global composite formed from five DAS-II composite scores ranged from 89 to 92% across age groups for those with IQs of 70 and below. The *g* saturation was much weaker for other ability groups, as it ranged from 83 to 86% for those with IQs of approximately 100, and 54 to 66% for those with IQs of 130 or higher. (Recall that similar values from Reynolds et al., 2013, using the full ability range, spanned 82–83%.) Although these findings bring into question the validity of using IQs to represent psychometric *g* when those with high IQs are assessed, they strengthen the support for using IQs as the key measure reflecting intellectual functioning during the assessment of ID. Thus, consistent with SLODR, IQs are likely to be better measures of their targeted construct with those most likely to complete assessments targeting ID symptoms (all things being equal) than with those at higher general ability levels.

Preconditions for Testing

As noted previously, in order to be confident that IQs from the most well-developed cognitive ability test batteries measure psychometric *g* with integrity for each individual being tested, it is necessary that those engaged in assessment identify potential confounds that may surface during testing and make efforts to eliminate or accommodate them. In contrast to an assessment model that highlights understanding of the construct-relevant and construct-irrelevant reasons for score variation after testing has been completed (e.g., National Research Council, 2002; Sattler, 2008), a model that reduces or eliminates the effects of construct-irrelevant influences prior to testing would be ideal (Bracken, 2000). As many children and adolescents with ID display concomitant sensory impairments, physical disabilities, or behavioral excesses, this population may be at particular risk for underperformance during testing. In other words, because of these construct-irrelevant influences, the truest estimates of intellectual functioning may be underestimates. For example, children experiencing fine motor problems will likely perform more slowly than expected when asked to construct patterns with blocks and complete puzzles when given puzzle pieces, regardless of their "visual processing" abilities. As a result, these children may not be awarded time bonuses that are sometimes available, and they may fail items they otherwise would have completed if they had been provided more time.

To address these construct-irrelevant influences, we (along with our colleagues) developed parent and teacher ratings scales, an interview, and direct screening measures to better identify such problems before testing begins. Published in part in Kranzler and Floyd (2013), the Screening Tool for Assessment (STA) targets vision problems, color blindness, hearing problems, articulation problems, fine motor problems, and behavioral excesses that might interfere with the accurate measurement of intellectual abilities during testing. The STA parent and teacher rating scales each include 32 items that might indicate problems in these areas. The STA Direct Screening Form includes a nine-item semistructured interview that addresses readiness for testing; the prior night's sleep; and sensory, speech, and motor problems. It is followed by informal direct screening measures in which examinees are asked to identify letters that decrease in size (vision screening), label color (color blindness screening), complete simple commands to point to body parts (hearing screening), repeat words comprising letter patterns associated with common articulation problems (articulation screening), and trace lines and write a short sentence with a pencil (fine motor screening). Since the publication of the STA in 2013, a vision screening form that includes line drawings versus letters has been developed for those who have not yet learned the names of letters, but can name what is depicted in simple drawings. (The vision screener can be obtained by contacting either of us chapter authors.) Although, to our knowledge, they have not been extensively used and evaluated by experts in the field and with large numbers of students with ID, these screening forms may assist examiners in identifying construct-irrelevant influences before testing. Some modifications may be needed, however, with lower-functioning and younger students with ID.

Psychometric Considerations in Test Selection and Interpretation

Those engaged in assessment of ID should be aware of psychometric considerations when they are selecting and completing cognitive ability test batteries. Kranzler and Floyd (2013) have offered a checklist designed to guide assessors through consideration of these key features, which include the quality of the test battery's norming; subtest, test, and IQ scaling; reliability; and validity evidence. In the sections that follow, we highlight these key features and standards for evaluating scores from cognitive ability test batteries and adaptive behavior instruments.

Norming

It is important that those engaged in assessment of ID draw upon cognitive ability test batteries that are supported by the largest, most expansive, and most recent norming sample data (National Research Council, 2002). Of course, larger norming samples are likely to produce more accurate indications of the relative discrepancy between one individual's measured intellectual ability and that of their age-based peers. Like survey poll results, smaller samples are likely to be less accurate in their results than larger samples. In the case of cognitive ability test batteries, the size of the total norming sample matters less than the size of the norm group to which the individual is compared because those comparisons are uniformly age-based. Some standards suggest that 1-year norm intervals should include at least 100 individuals to be considered adequate (e.g., Hammill, Brown, & Bryant, 1992). More recently, attention has been paid to evaluating the size of norm intervals associated with more narrow segments of the norming sample (e.g., 1-month to 6-month intervals) associated with norm tables provided by the test publishers. According to Norfolk and colleagues (2015), the following evaluation scheme can be employed in evaluating norming sample size: *good* when at least 100 children and adolescents are included per norm interval, *adequate* when 30–99 children and adolescents are included, and *inadequate* when fewer than 30 children and adolescents are included. In addition to being large, the norming sample should be representative of the targeted population. Thus it is desirable that it represent the demographic characteristics of that population and be considered nationally representative. Floyd and colleagues (2015) have suggested that similar criteria can be applied in examining sampling across states: *good* with sampling across at least 35 states, *adequate* with sampling across 25–35 states, and *inadequate* with sampling across less than 25 states.

A more systematic influence on test scores that is highly relevant to ID assessment is the recency of the norming data. In accordance with the Flynn effect (Flynn, 1984; Trahan, Stuebing, Hiscock, & Fletcher, 2014), norming data collected more recently will produce lower scores than those collected decades before on contemporaneous assessments. As such, older norming sample data will inflate the estimate of an individual's intellectual functioning and potentially lead to failure to identify ID in cases in which it exists (McGrew, 2015b). Following extensive research focused on the Flynn effect, one might generally expect that IQs will be inflated 3 points for every decade between the collection of the norming data and testing, so that a test normed in 2006 could be expected to produce an IQ 3 points higher than a test that was normed in 2016 if both were administered to the same individual in 2017. Across the span of 15 years, this inflation would approximate 5 IQ points. This predicable influence due to the Flynn effect might tip the scales in undermining an otherwise accurate assessment of intellectual functioning; thus, it should be acknowledged and controlled. Following Alfonso and Flanagan (2009) and Floyd and colleagues (2015), Floyd and colleagues (2016) have offered these standards for evaluating the recency of norming samples: "*good* when at least some of the norming data (when a range of data collection was reported) were collected within the past 10 years, *adequate* when data were collected within the past 15 years, and *inadequate* when all data were collected more than 15 years ago" (p. 275).

Scaling

The range of scores produced by cognitive ability test batteries should also be considered. During the assessment of ID, it is important that the IQs be able to measure intellectual functioning at its lowest levels. Across recently published test batteries, IQs most frequently range from about 40 to about 160, and relatively few produce IQs that fall below 40 (Kranzler & Floyd, 2013). This range is generally satisfactory for identifying most cases of ID (based on the criteria of IQs of appropriately 70 or lower evident in Table 23.1); an IQ of 40 is more than two standard deviations below 70. Limitations in test batteries producing problems in measuring intellectual functioning at the lowest levels of ability are called *floor violations*, and they are typically evident at the task (e.g., subtest or test) level. Floor violations reflect the fact that there are not enough easy items associated with each subtest or test, so that some children or adolescents cannot answer at least one item correctly. According to a traditional standard (Bracken, 1987), a floor violation occurs when a subtest or test raw score of 1 does not produce a norm-referenced score at least two standard deviations below the mean. This means that a raw score of 1 should be equal to a scaled score of 4 or lower, a *T* score of 30 or lower, and a deviation IQ of 70 or lower. It is wise to examine this information (often apparent in norm tables) before testing begins, while considering the age of the examinee. Floor violations are most evident at the young age levels of test batteries, and test batteries with multiple floor violations are likely to overestimate intellectual functioning and fail to correctly identify true cases of ID.

Reliability and Confidence Intervals

Reliability refers to consistency in measurement across replications, and it is a very important characteristic to consider in the assessment of ID. The reliability of IQs is typically reported in the form of internal-consistency reliability estimates and test–retest reliability coefficients. Internal-consistency reliability estimates for composites scores like IQs are typically calculated by considering the internal consistency of the tasks (subtests or tests) contributing to the IQs (e.g., coefficient alpha values), as well as the correlations between those task scores. The resulting internal-consistency reliability estimates for IQs are, as noted previously, often .95 or higher (Kranzler & Floyd, 2013). Test–retest reliability coefficients represent (on a relative scale) consistency across time. Test–retest reliability is typically examined by administering a test battery to the same group twice across a month or two; resultant scores from each administration are correlated with one another. Due to the nature of the correlation typically employed (a Pearson product–moment correlation), differences in the absolute level of scores across administrations (which may be due to practice effects and the like) will not contribute to higher or lower correlations. Thus the resulting coefficient represents relative relations (and not absolute differences) across time.

Kranzler and Floyd (2013) have suggested that the standard for internal-consistency reliability coefficients be .95 and the standard for test–retest reliability coefficients be .90 (especially across a month or less and employing correlations corrected for restriction of range). These standards, however, may be too high when scores other than IQs are considered. Reynolds and Livingston (2014) and Floyd and colleagues (2015) have offered a more forgiving scale for both types of reliability, which holds them to a standard of .90. Regardless of the standard that is selected, a more reliable measure is usually better.

As both the AAIDD (2010) and APA (2013) reference consideration of the *standard error of measurement* (*SEM*), a general term used to describe the distribution of hypothetical true scores an examinee is likely to obtain with repeated measurement (all things being equal), the reliability of measures has real-world effects. An IQ with high reliability will produce *SEM* values (±2 points) that are far smaller than those IQs with inadequate reliability (±5 points). As a result, the confidence interval, which stems from extension of the *SEM* value, will be relatively narrow. A score with inadequate reliability can produce a band of error that easily spans more than 15 points—more than a standard deviation—which would make it incredibly difficult to determine whether (1) intellectual functioning is actually below a stated cutoff or (2) a low score is due to inaccuracy in measurement for the student being assessed.

Reliability (typically, internal-consistency reliability) is also important, as its values are used to calculate *estimated true scores* and *standard error of estimated true score* values, which form the foundation for confidence intervals offered by most cognitive ability test batteries. To our surprise, most batteries of this type do not equip users to employ *SEM* values per se (Kranzler & Floyd, 2013), as the standard error of estimated true score is usually employed instead. With individuals with low IQs relative to their peers, the estimated true score for IQs will be slightly closer to the mean than their associated obtained IQs, and the standard error of estimated true score for IQs will be slightly smaller than the traditional *SEM* for the same variable. These appropriate but slight adjustments to the confidence intervals for IQs may be substantial enough to alter the decisions made by professionals considering these scores (as discussed in the section on cutoff scores and clinical judgment that follows). Reliability clearly matters.

Validity

It behooves clinicians to consider all the evidence addressing a test's integrity in measuring its targeted constructs as well as its intended interpretations. As such, *validity* refers to "the degree to which evidence and theory support the interpretations of test scores by proposed uses of tests" (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014, p. 11). In the most recent conceptions of this measurement property, validity evidence is categorized into five types. *Evidence based on content* refers to developing and employing test items that appear to measure the targeted construct, and then examining item-level information to evaluate whether the items achieve these goals. For example, items from intelligence tests should be developed on the basis of a theoretical understanding of cognitive abilities, and item scores should be substantially correlated and be able to be scaled in terms of their difficulty. *Evidence based on response processes* refers to evaluating the behavioral and cognitive steps that those taking tests would complete to solve its items. Both intelligence tests that require advanced fine motor skills and memory tests that unduly involve mathematics knowledge could be said to require construct-irrelevant processes, indicating invalidity in assessment. In our experience, publishers of cognitive ability test batteries do not provide evidence based on test content or responses processes derived from only ID samples. This information typically applies to all test takers, and examiners need to determine whether the test's content and response processes would undermine the conclusions they draw for the students they assess, considering the students' unique characteristics.

Evidence based on internal structure reflects the correlations between parts of a test. Evidence may appear as relations between and among individual tasks (subtests and tests) within a battery, and analyses often include correlations and more sophisticated exploratory and confirmatory factor analyses. In contrast, *evidence based on relations with other variables* reflects the relations between scores from a test battery and variables not derived from that test battery. Researchers examining this source of validity evidence often refer to *criterion-related validity, concurrent validity,* and *predictive validity*. Other variables correlated with test scores in question may include demographic characteristics, such as age, gender, race, or ethnicity, as well as scores from other assessment instruments. For example, a study examining the correlations between IQs and adaptive behavior domain scores obtained from a parent interview, and a clinical group comparison study employing students with ID, provide such evidence. Although there appears to be an increase in clinical group comparison studies involving children with ID reported in test battery manuals, it is extremely uncommon to see validity studies addressing evidence based on internal relations and evidence based on relations with other variables that employ only samples with ID. These studies are not typically conducted, in part because correlational studies with samples associated with a restricted range of scores (vs. scores varying like those across the population) tend to produce weaker correlations (due to an artifact) than those found when a broad sample is targeted. Thus, rather than strengthening the body of validity evidence, studies like these (e.g., with students with ID, who display such range restriction) are associated with artifacts that appear to weaken the validity argument. As a result, few test publishers, authors, and researchers endeavor to complete them.

Evidence based on consequences is the least understood and evaluated type of validity evidence; this term refers to evidence associated with intended positive effects of testing and unintended detrimental effects of testing. For example, evidence indicating invalidity in assessment might be indicated if a new cognitive ability test battery classified only students who were from a minority ethnic background as having ID, and classified students with similar characteristics who were from the majority ethnic background as not having ID. In sum, test users should carefully review validity evidence summarized in test manuals, available online, and published in chapters, books, and peer-reviewed journals, to ensure that it supports trustworthy data representative of targeted constructs.

Controversial Practices in the Assessment of Intellectual Functioning

In this section, we briefly review five controversies associated with ID assessment and offer our informed opinions to guide the application of research to practice. Readers should refer to chapters by McGrew for discussion of score adjustments addressing the Flynn effect (McGrew, 2015b) and use of multidimensional cognitive ability test batteries (McGrew, 2015a).

Using Fixed Cutoffs versus Flexible Cutoffs

As noted previously, diagnostic systems designed to identify ID have increasingly focused on clinical judgment (AAIDD, 2010; Luckasson & Schalock, 2015) and recognition of measurement error during assessment (evident in use of confidence intervals). Thus, there is less emphasis on *fixed cutoffs* (sometimes called *bright-line cutoffs*) associated with absolute score markers than in the past. For example, with fixed cutoffs, a child or adolescent who is one or two points above a score threshold (e.g., an IQ of 70) would not meet the symptom criterion for the disorder. This practice is clearly problematic. Instead, *flexible cutoffs* associated with interpretation of standard error terms and associated confidence intervals and application of clinical judgment when determining if symptom criteria are met offer much greater sensitivity in identifying cases of ID. Although Table 23.1 indicates this trend toward flexible cutoffs, only about half of states in the United States refer to these considerations (McNicholas et al., 2018). We view flexible cutoffs as necessary and discourage use of fixed cutoffs in the assessment of ID.

Obtaining and Averaging Multiple IQs

For a variety of reasons, cognitive ability test batteries produce different IQs for many individuals (AAIDD, 2010; McGrew, 2015a). In fact, research has demonstrated that a quarter of those completing different test batteries concurrently may obtain IQs that are discrepant by 10 or more points (Floyd, Clark, & Shadish, 2008). Throughout our careers, we have heard others recommend administering two or more test batteries in sequence when it is suspected that a child or adolescent has ID, and we have engaged in this practice ourselves. It seems wise to employ at least two indicators of a deficit in intellectual functioning when ID is being considered. This strategy is particularly wise when a student's performance during initial testing indicates that a confounding variable is in play, lowering the student's score. But what should be done if the IQs from valid administrations are discrepant, and one indicates a deficit in intellectual functioning and the other does not? On a descriptive level, the statistical significance of IQ differences can be tested (see McGrew, 2015a); sizeable score differences might not be unexpected when measurement error is considered. It is likely that the average of two or more valid IQs would be a more accurate estimate of intellectual functioning than any one alone. Based on Schneider and McGrew's (2011) work, Schneider (2013) has provided the most defensible means of averaging these scores. We view this method as promising and psychometrically defensible, yet recognize that it is understudied.

Interpreting Part Scores

The National Research Council (2002) offered a number of important insights and new research findings to the discussion of ID assessment and identification. One was consideration of *part scores,* which are variables targeting more specific abilities than the IQ. As a contrast to the IQ (a total test score), part scores refer to subtest and subscale scores more generally and to factor index, clusters, and broad ability composite scores more specifically. The National Research Council (2002) first asserted that variation across part scores contributing to an IQ might be problematic by stating:

> There are occasions when a total test score may not be the best indicator of an individual's overall intellectual functioning, and the examiner must resort to interpreting one of the instrument's part scores as the best indicator of overall intellectual functioning. . . . In such cases, the instrument's total test score may offer little more than an awkward and artefactual "average" of a number of relatively disparate subtest or subscales (i.e., part scores). (p. 109)

After addressing methods to evaluate the integrity of the IQ, such as testing the statistical significance, determining the base rate of differences between part scores, and selecting subtests contributing to the IQ (e.g., based on their *g* saturation), the National Research Council (2002) explained how IQs should be considered in relation to part scores.

First, the council outlined the upper limit of the IQ range that would indicate deficits in intellectual functioning by stating that

> when a scale score discrepancy meets the previously mentioned criteria of significance and meaningfulness, the total test score may simply be too high to support a diagnosis of mental retardation. . . . The final criterion for deciding whether or not to use part scores in place of the total test score . . . is that, no matter how great the discrepancy between relevant subscales, individuals with total test scores greater than 75 should not be diagnosed. (pp. 113–114)

Next, after advocating for consideration of IQs except in instances when their validity is in question, the council stated, "In such cases, appropriate part scores may better represent the individual's true overall level of cognitive functioning" (p. 114), but added two caveats. It stated that "only part scores derived from scales that demonstrate high *g*-loadings . . . should be used in place of the composite IQ score when its validity is in doubt" (p. 115), and that "if a part score is used in place of the composite IQ score . . . the part score should not exceed 70" (p. 115). As reported in Table 23.1, these recommendations still form the foundation for diagnosis according to the SSA (n.d.).

Following the National Research Council (2002) recommendations, Bergeron and Floyd (2006, 2013) examined part score profiles of students with ID. Bergeron and Floyd (2006) targeted part score profiles of 30 students with ID receiving special education services. Seven part scores were obtained from the Woodcock–Johnson III (WJ III) Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001). Although the average IQ (the WJ III General Intellectual Ability—Extended score) was in the very low range (*M* = 54.8, *SD* = 12.5), and almost every student demonstrated an IQ of 70 or below, part scores were much more varied. In fact, 37% of the students scored in the average range (90–110) or higher on at least one part score, and 77% of students scored in the low average range (80–89) or higher on at least one part score. In a follow-up study, Bergeron and Floyd (2013) employed datasets offered as clinical validity studies of students with ID supporting the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003), the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004), and the DAS-II (Elliott, 2007). They also found that part scores were likely to indicate higher levels of functioning than the tests' IQs did. Despite the fact that the group-level WISC-IV, KABC-II, and DAS-II profiles were generally "flat," with all mean values less than 70 and the associated IQs routinely less than 70 for individuals, they found that a third or more of cases produced a part score in the low average range or higher for the WISC-IV (45%), KABC-II (52%), and DAS-II (33%).

These findings indicate two phenomena at play: *regression toward the mean* and *combinatorial probabilities*. (We highly recommend the excellent instructional video focusing on these phenomena by Schneider, 2011.) According to Kranzler and Floyd (2013), regression toward the mean refers to

> the general statistical phenomenon by which repeated measurement produces scores that are less extreme and closer to average due to chance occurrence alone. . . . Because of regression toward the mean, we would expect (1) that children and adolescents with very low IQs will exhibit some subtest or composite scores that are closer to the population mean than their IQs and (2) that the subtests and composite scores demonstrating the lowest correlations with the others will be most affected. (p. 188)

In contrast, combinatorial probabilities explain how the IQs (stemming in part from the same tasks producing the part scores) are likely to be lower than most part scores—and consistently so. Combinatorial probabilities can be explained the following way:

> Just as it is increasingly improbable that a person will roll a 5 on each consecutive roll of a die, it is increasingly improbable that an examinee will score consistently lower than his or her peers across subtests and composite scores. As a result of combinatorial probabilities, IQs will be more deviant (i.e., further away from the mean) than the average of the scores that contribute to them, representing this increasing improbability. . . . Accordingly, for individuals with very low IQs, subtest scores and composite scores will be higher than expected from their IQs. (Kranzler & Floyd, 2013, p. 188)

Consideration of part scores (especially under the assumption of a valid administration, not confounded by construct-irrelevant influences) clearly complicates consideration of deficits in intellectual functioning. It appears improbable that professionals are likely to enact the recommendation that an IQ of 75 or lower, accompanied by a part score of 70 or lower, be indicative of a deficit in intellectual functioning—especially when flexible cutoffs are increasingly applied. Consideration of part scores, however, also introduces other complications, discussed next.

Invalidating IQs When Their Components Are Discrepant

It is sometimes stated that large discrepancies between the part scores contributing to an IQ undermine its meaningfulness in representing psychometric *g*. Recommendations from test authors appearing in guidelines for interpretation included in test kits (Wechsler, 2014), and the widespread proliferation of recommendations for clinical interpretation (Sattler, 2008), reinforce this practice. It is assumed that this same practice should be applied to students with ID; in fact, the National Research Council (2002) addressed the statistical significance and base rate of part score discrepancies when it discussed identification of ID.

In a series of studies designed to evaluate the validity of this practice, Watkins and colleagues (Freberg, Vandiver, Watkins, & Canivez, 2008; Kotz, Watkins, & McDermott, 2008; Watkins, Glutting, & Lei, 2007) employed intelligence test norming sample data to examine differential prediction of IQs on achievement outcomes when children and adolescents were distinguished by whether they displayed substantial variation in part scores (variable profiles) or not (flat profiles). Watkins and colleagues (2007) demonstrated, using data from two large-scale validity studies and data obtained from practitioners, no differences between groups formed based on the extent of variation in part score profiles (yet matched on IQs) when the relations between IQs and reading and mathematics scores were compared. Freberg and colleagues (2008) replicated these findings, using data obtained from practitioners that included IQs obtained at one point in time and achievement test scores obtained later in time. All data were from students with disabilities (including ID). Prior findings were extended through the analysis of subsamples of students with larger discrepancies in their part score profiles. Finally, Kotz and colleagues (2008) completed similar analyses, using data from the original Differential Ability Scales (Elliott, 1990) and two co-normed achievement subtests. Again, as in Freberg and colleagues' study, prior findings were extended through the analysis of subsamples with larger discrepancies in their part score profiles and through examining subtypes of part score discrepant profiles (e.g., those with relative strengths evident in one part score vs. another).

More recently, McGill (2016) supported the results of these studies by Watkins and colleagues through an analysis of the KABC-II (Kaufman & Kaufman, 2004), which includes a greater number of part scores (five) than the standard batteries employed by Watkins and colleagues do. McGill formed groups based on part score profile variation (as Watkins and colleagues did), and selected data from only the group with at least one discrepant part score in the profile for further analysis. McGill demonstrated through exploratory factor analysis that the KABC-II factor structure was not undermined by such discrepancies, and showed that indices of integrity in measurement (omega coefficients) remained sound when data from groups with discrepant part score profiles were considered. In addition, predictions employing the IQ from the KABC-II (the Fluid–Crystallized Index) revealed that it remained a strong predictor of achievement outcomes when data from individuals with discrepant part score profiles were used. Most constituent part scores offered small and negligible incremental validity in prediction of achievement outcomes. Findings were replicated across two age groups (ages 7–12 and 13–18).

When we consider the results from Bergeron and Floyd (2006, 2013), it seems likely that students with ID would be just as likely—and perhaps even more likely, due to the effects of regression toward the mean on part scores—to demonstrate sizeable discrepancies between part scores, which have been claimed to undermine the meaningfulness of IQs. In addition to this evidence from Watkins and colleagues (Freberg et al., 2008; Kotz et al., 2008; Watkins et al., 2007) and McGill (2016) supporting the integrity of the measurement of the construct of psychometric *g* amidst cases with significant variation in part scores, as well as the maintenance of the predictive properties of IQs in explaining variation in achievement, we see the practice of invalidating IQs when their components are discrepant as particularly problematic when applied to cases of ID.

More generally, if global composites like IQs are cast aside due to variability in their constituents, clinicians are left with scores that possess inferior psychometric properties—especially reliability. Following this successive-levels approach (Sattler, 2008), if the part scores contributing to an IQ are discrepant, one would disregard the IQ and interpret the part scores, which typically possess lower reliability and weaker bodies of validity evidence. Then, if subtests contributing to part scores are discrepant, one would disregard the part scores and interpret subtest scores, which almost always possess lower reliability (and sometimes inadequate reliability) and usually are supported by even weaker bodies of validity evidence than part scores are. Following this logic, if there is variation in performance across items scaled in terms of difficulty, one might even disregard the subtest scores and interpret item-level scores, which almost certainly possess terribly low reliability and inadequate validity evidence.

Identifying Infants and Preschool-Age Children with ID

All diagnostic models included in Table 23.1 require onset of ID during the developmental period, which usually implies "during the first 18 years of life." Professionals tasked with identifying the reasons for developmental delays in young children face many challenges, and some of these professionals may be called upon to determine if children with severe impairments have ID. They are faced with several problems associated with measurement accuracy, as assessment of infants, toddlers, and other preschool-age children are not always responsive to individual testing (Bracken, 2000); their behaviors may range from unresponsive to aggressive. As suggested earlier in this chapter, these behaviors may be more extreme or more likely to be multitudinous in cases of children referred for ID. In addition, growth in language skills, cognitive abilities, motor skills, and attention is rapid during this period; thus changes across short periods of time should be expected. As a result, research has demonstrated that many estimates of intellectual functioning obtained during this period are not sufficiently reliable across time or maximally predictive of long-term outcomes (Jenni et al., 2015). In the same vein, many developmental screeners commonly employed during this period (especially with children under the age of 3) do not yield IQs or other scores designed to represent intellectual functioning per se. As a result, scores that only roughly reflect intellectual functioning may need to be employed.

Due to these complications in identifying ID (and other conditions) during early childhood and to the stigma associated with the diagnosis, classification systems have been extended to address developmental delays similar to ID without employing the ID label. As noted previously, DSM-5 (APA, 2013) refers to GDD when referencing a condition associated with deficits in intellectual functioning, adaptive functioning, and gross motor skills in early childhood. In addition, IDEA (2004) allows for students ages 3–9 to be identified with a developmental delay if they display delays in physical, communication, social-emotional, cognitive, or adaptive behavior development. On the basis of references to cognitive and adaptive behavior development, a child within this range with the characteristics of ID could be identified as having a developmental delay. Considering psychometric standards guiding testing practices, the effects of social labeling, and the effectiveness of early interventions to address a variety of areas of developmental delays, delaying identification of ID for young children with mild cases of the condition seems appropriate in many cases.

CONCLUSION

ID is a disabling condition associated with deficits in intellectual functioning and adaptive behavior. Those involved in the assessment of children and adolescents with ID should consider recent changes in diagnostic criteria for the condition, as well as best practices in assessment. They should carefully review the resources provided by the AAIDD (2010) and should stay abreast of research informing score interpretation and test battery selection. This chapter has been written to assist them in achieving these goals.

REFERENCES

Alfonso, V. C., & Flanagan, D. P. (2009). Assessment of preschool children: A framework for evaluating the adequacy of the technical characteristics of norm-referenced instruments. In B. Mowder, F. Rubinson, & A. Yasik (Eds.), *Evidence based practice in infant and early childhood psychology* (pp. 129–166). Hoboken, NJ: Wiley.

American Academy of Pediatrics, Committee on Bioethics, Committee on Genetics, & American College of Medical Genetics and Genomics, Social, Ethical, and Legal Issues Committee. (2013). Ethical and policy issues in genetic testing and screening of children. *Pediatrics, 131,* 620–622.

American Association on Intellectual and Developmental Disabilities (AAIDD). (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: Author.

American Educational Research Association (AERA), American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychiatric Association (APA). (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Speech–Language–Hearing Association. (n.d.). Spoken language disorders (Practice portal). Retrieved January, 17, 2017, from *www.asha.org/practice-portal/clinical-topics/spoken-language-disorders.*

Bergeron, R., & Floyd, R. G. (2006). Broad cognitive abilities of children with mental retardation: An analysis of group and individual profiles. *American Journal on Mental Retardation, 111,* 417–432.

Bergeron, R., & Floyd, R. G. (2013). Individual part score profiles of children with intellectual disability: A descriptive analysis across three intelligence tests. *School Psychology Review, 42,* 22–38.

Bracken, B. A. (1987). Limitations of preschool scales and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5,* 313–326.

Bracken, B. A. (2000). Maximizing construct relevant assessment: The optimal preschool testing situation. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp. 33–44). Boston: Allyn & Bacon.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist, 34*(10), 844.

Cooper, S.-A., Smiley, E., Allan, L. M., Jackson, A., Finlayson, J., Mantry, D., et al. (2009). Adults with intellectual disabilities: Prevalence, incidence and remission of self-injurious behavior, and related factors. *Journal of Intellectual Disability Research, 53,* 200–216.

Cooper, S.-A., Smiley, E., Jackson, A., Finlayson, J., Allan, L., Mantry, D., et al. (2009). Adults with intellectual disabilities: prevalence, incidence and remission of aggressive behaviour and related factors. *Journal of Intellectual Disability Research, 53,* 217–232.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low-IQ groups. *Intelligence, 13,* 349–359.

Elliott, C. D. (1990). *Differential Ability Scales.* San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Farmer, R. L., Floyd, R. G., Reynolds, M. R., & Kranzler, J. (2014). IQs are very strong but imperfect indicators of psychometric *g:* Results from conjoint confirmatory factor analysis. *Psychology in the Schools, 51,* 801–813.

Floyd, R. G., Clark, M. H., & Shadish, W. R. (2008). The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice, 39,* 414–423.

Floyd, R. G., Shands, E. I., Alfonso, V. C., Phillips, J., Autry, B. K., Mosteller, J. A., et al. (2015). A systematic review and psychometric evaluation of adaptive behavior scales and recommendations for practice. *Journal of Applied School Psychology. 31,* 83–113.

Floyd, R. G., Woods, I. L., Singh, L. J., & Hawkins, H. K. (2016). Use of the Woodcock–Johnson IV in the diagnosis of intellectual disability. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 272–290). San Diego, CA: Academic Press.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology, 15,* 131–139.

Fujiki, M., Brinton, B., & Todd, C. M. (1996). Social skills of children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 27,* 195–202.

Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst, 27,* 133–151.

Hammill, D. D., Brown, L., & Bryant, B. R. (1992). *A consumer's guide to tests in print* (2nd ed.). Austin, TX: PRO-ED.

Harrison, P. L., & Oakland, T. (2015). *Adaptive Behavior Assessment System, Third Edition (ABAS-3): Manual.* Bloomington, MN: Pearson.

Hawkins, R. P., & Dobes, R. W. (1977). Behavioral definitions in applied behavior analysis: Explicit or implicit? In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, method, and application* (pp. 167–188). Hillsdale, NJ: Erlbaum.

Herschell, A. D., Greco, L. A., Filcheck, H. A., & McNeil, C. B. (2002). Who is testing whom?: Ten suggestions for managing the disruptive behavior of young children during testing. *Intervention in School and Clinic, 47,* 140–148.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1401 (2004).

Jenni, O. G., Fintelmann, S., Caflisch, J., Latal, B., Rousson, V., & Chaouch, A. (2015). Stability of cognitive performance in children with mild intellectual disability. *Developmental Medicine and Child Neurology, 57,* 463–469.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: A review. *Journal of Neurodevelopmental Disorders, 2,* 182–209.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the general conceptual ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review, 37,* 261–278.

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide.* New York: Guilford Press.

Luckasson, R., & Schalock, R. L. (2015). Standards to guide the use of clinical judgment in the field of intellectual disability. *Intellectual and Developmental Disabilities, 53,* 240–251.

Mannix, D. (2009). *Life skill activities for special children* (2nd ed.). San Francisco: Jossey-Bass.

Mather, N., & Wendling, B. J. (2014). *Examiner's manual: Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30,* 107–114.

McClintock, K., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviors in people with disabilities: A meta-analytic study. *Journal of Intellectual Disability Research, 47,* 405–416.

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology,* 6(1), 49–79.

McGrew, K. S. (2015a). Intellectual functioning. In E. Polloway (Ed.), *The death penalty and intellectual disability* (pp. 85–112). Washington, DC: American Association on Intellectual and Developmental Disabilities.

McGrew, K. S. (2015b). Norm obsolescence: The Flynn effect. In E. Polloway (Ed.), *The death penalty and intellectual disability* (pp. 155–172). Washington, DC: American Association on Intellectual and Developmental Disabilities.

McKenzie, K., Milton, M., Smith, G., & Oullette-Kuntz, H. (2016). Systematic review of the prevalence and incidence of intellectual disabilities: Current trends and issues. *Current Developmental Disorders Reports, 3,* 104–115.

McNicholas, P., Floyd, R. G., Woods, I. L., Singh, L. J., Manguno, M., & Maki, K. (2018). State special education criteria for identifying intellectual disability: A review following revised diagnostic criteria and Rosa's Law. *School Psychology Quarterly, 33,* 75–82.

National Academies of Sciences, Engineering, and Medicine. (2015). Prevalence of intellectual disabilities. In T. F. Boat & J. T. Wu (Eds.), *Mental disorders and disabilities among low-incident children* (pp. 169–178). Washington, DC: National Academies Press.

National Research Council. (2002). *Mental retardation: Determining eligibility for Social Security benefits.* Washington, DC: National Academy Press.

Norfolk, P. A., Farmer, R. L., Floyd, R. G., Woods, I. L., Hawkins, H. K., & Irby, S. M. (2015). Norm block sample sizes: A review of 17 individually administered intelligence tests. *Journal of Psychoeducational Assessment, 33,* 544–555.

Papazoglou, A., Jacobson, L. A., McCabe, M., Kaufmann, W., & Zabel, T. A. (2014). To ID or not to ID?: Changes in classification rates of intellectual disability using DSM-5. *Intellectual and Developmental Disabilities, 52,* 165–174.

Partington J. W. (2010). *Assessment of Basic Language and Learning Skills—Revised.* Walnut Creek, CA: Behavior Analysts.

Pinborough-Zimmerman, J., Satterfield, R., Miller, J., Bilder, D., Hossain, S., & McMahon, W. (2007). Communication disorders: Prevalence and comorbid intellectual disability, autism, and emotional/behavioral disorders. *American Journal of Speech–Language Pathology, 16,* 359–367.

Reynolds, C. R., & Livingston, R. B. (2014). A psychometric primer for school psychologists. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology* (6th ed., pp. 281–300). Bethesda, MD: National Association of School Psychologists.

Reynolds, M. R. (2013). Interpreting intelligence test composite scores in light of Spearman's law of diminishing returns. *School Psychology Quarterly, 28,* 63–76.

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric *g* indexed by global composites?: Evidence from three popular intelligence tests. *Psychological Assessment, 25,* 1314–1321.

Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence, 38,* 231–241.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide.* New York: Guilford Press.

Rosa's Law, Pub. L. No. 111-256 § 4, 124 Stat. 2643 (2010).

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Author.

Schneider, W. J. (2011, March 29). A geometric representation of composite scores [Video file]. Retrieved from *http://assessingpsyche.wordpress.com/2011/03/29/a-geometric-representation-of-composite-scores.*

Schneider, W. J. (2013, December 29). Can't decide which IQ is best?: Make a composite score. Retrieved from *https://assessingpsyche.wordpress.com/2013/12/29/cant-decide-which-iq-is-best-make-a-composite-score.*

Schneider, W. J., & McGrew, K. S. (2011, March 24). "Just say no" to averaging IQ subtest scores. Retrieved from *www.iapsych.com/iapap101/iap10110.pdf.*

Shevell, M. (2008). Global developmental delay and mental retardation or intellectual disability: Conceptualization, evaluation, and etiology. *Pediatric Clinics of North America, 55,* 1071–1084.

Social Security Administration (SSA). (n.d.). Disability evaluation under Social Security: Part B, 112.00 mental disorders-children (Social Security Blue Book). Retrieved January 17, 2017, from *www.ssa.gov/disability/professionals/bluebook/112.00-MentalDisorders-Childhood.htm.*

Sundberg, M. L. (2008). *Verbal Behavior Milestones Assessment and Placement Program: VB-MAPP.* Concord, CA: AVB Press.

Thompson, J. R., Bryant, B. R., Campbell, E. M., Craig, E. M., Hughes, C. M., Rotholz, D., et al. (2004). *Supports Intensity Scale user's manual.* Washington, DC: American Association on Mental Retardation.

Thurman, D. J. (2014). The epidemiology of traumatic brain injury in children and youths: A review of the research since 1990. *Journal of Child Neurology, 31,* 20–27.

Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140,* 1332–1360.

U.S. Department of Education, EdFacts Data Warehouse. (2016). Number of children ages 6 through 21 served under IDEA, part B, by disability and state, 2015–2016 [Data file]. Retrieved from *www2.ed.gov/programs/osepidea/618-data/static-tables/index.html.*

Valencia, R. R., & Lopez, R. (1992). Assessment of racial and ethnic minority students: Problems and prospects. In M. Zeidner & R. Most (Eds.), *Psychological testing: An inside view* (pp. 399–439). Palo Alto, CA: Consulting Psychologists Press.

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., & State, M. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 53,* 237–257.

Walker, W. O., Jr., & Johnson, C. P. (2006). Mental retardation: Overview and diagnosis. *Pediatrics in Review, 27*(6), 204.

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the full scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14,* 13–20.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *Administration and scoring manual: Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

World Health Organization (WHO). (2001). *International classification of functioning, disability and health: ICF.* Geneva, Switzerland: Author.

World Health Organization (WHO). (2004). *International statistical classification of diseases and related health problems* (10th rev., 2nd ed.). Geneva, Switzerland: Author.

World Health Organization (WHO). (2010). *mhGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings: Mental health Gap Action Programme (mhGAP).* Geneva, Switzerland: Author.

World Health Organization (WHO). (2016). International statistical classification of diseases and related health problems (11th rev., beta draft). Retrieved from *https://icd.who.int/dev11/l-m/en.*

CHAPTER 24

Intellectual and Neuropsychological Assessment of Individuals with Sensory and Physical Disabilities and Traumatic Brain Injury

Scott L. Decker
Julia Englund Strait
Alycia M. Roberts
Joseph Ferraracci

This chapter outlines the neuropsychological basis and assessment of sensory and physical deficits, as well as traumatic brain injury (TBI), in children. These disabilities are discussed together and addressed within a neuropsychological framework because behavioral deficits resulting from these disabilities clearly originate in biological functioning. The goal of the chapter is to provide a better understanding of the clinical manifestations of these disabilities for enhancing assessment procedures, interventions, and services for children with such conditions.

Assessments of sensory and physical functioning have long been part of standard neuropsychological evaluations (S. Finger, 1994). Although basic sensory abilities are often viewed as orthogonal to higher-level abilities, sensory measures have been correlated not only with academic measures (Decker, 2004), but also with measures of intellectual functioning (Decker, 2002; Roberts, Pallier, & Goff, 1999; Stankov, Seizova-Cajic, & Roberts, 2001), and this relationship between sensory processing and cognition increases as individuals age (Humes, Busey, Craig, & Kewley-Port, 2013). Regarding physical functioning, there is a longstanding history of literature linking motor functioning (including physical activity) and cognition (Haapala, 2013; Sibley & Etnier, 2003; Tomporowski, Davis, Miller, & Naglieri, 2008). More recently, research has begun to investigate the importance of early fine motor development as a predictor of cognitive functioning (Kim, Carlson, Curby, & Winsler, 2016; Stöckel & Hughes, 2016). As such, both sensory and physical abilities are prerequisites for the input and output mechanisms of cognitive faculties and provide indicators of the cognitive system's integrity.

There are shared areas of the brain that participate in sensory–motor behaviors as well as higher cognitive abilities. For example, lesions in the left frontal area of the brain may result in contralateral motor impairment (i.e., right-sided hemiparesis) as well as language deficits (i.e., expressive aphasia), due to localized proximity of both functions in a spatially similar area of the brain (Kolb & Whishaw, 2003). The benefit of neuropsychological theory is that it provides a framework for explaining the connections between sensory deficits and intellectual deficits, as well as between sensory functions and higher cognitive faculties like language or executive functions (e.g., working memory, inhibition, motor planning). The benefits of a neuropsychological framework extend beyond assessment to intervention planning and establishing rehabilitation expectations.

In the sections that follow, we focus specifically on sensory, physical, and TBI-related disabilities. Each section provides both practical and legal definitions for each type of disability. In addition, assessment and intervention issues are addressed. Both individual child factors and test-related factors important to consider when clinicians are selecting, administering, and interpreting neuropsychological and cognitive tests are included.

SENSORY DISABILITIES

According to the U.S. Census Bureau (2010), approximately 8.1 million people in the United States, or about 3.3% of the general population, had difficulty seeing, and approximately 7.6 million people, or about 3.1% of the general population, had difficulty hearing. The *Standards for Educational and Psychological Testing* volume (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014) emphasizes fairness in testing for individuals with such disabilities and suggests common accommodations for these individuals, including modifying presentation or response format, timing, or setting; using only portions of a test; or substituting assessment instruments/using alternative tests (see also Montgomery, Torres, & Eiseman, Chapter 30, this volume). Individual factors may affect which tests are given and what modifications are made, as well as how accurately the resultant test scores reflect a child's cognitive and neuropsychological functioning. It is therefore the clinician's responsibility to ensure selection of appropriate instruments, nonbiased test administration and interpretation of results, and application and consideration of necessary modifications for each child.

In a school setting, children with visual impairment or blindness (VI/B) are served mainly under the Individuals with Disabilities Education Improvement Act of 2004 (IDEA, 2004) category of *visual impairment,* defined in federal regulations (U.S. Department of Education, 2006) as follows:

> *Visual impairment including blindness* means an impairment in vision that, even with correction, adversely affects a child's educational performance. The term includes both partial sight and blindness. (p. 46757)

According to the U.S. Department of Education (n.d.), in the 2014–2015 school year, U.S. public schools in the 50 states, outlying areas, and freely associated states served 2,959 children ages 3–5 (of 753,697 total preschool students served under all categories) and 25,567 children and youth ages 6–21 (of 5,944,241 total) under the IDEA 2004 category of visual impairment. Slightly more children and youth (9,042 ages 3–5 and 67,884 ages 6–21 in 2014–2015) were served under the combined *deafness* and *hearing impairment* IDEA categories, defined, respectively, as follows:

> *Deafness* means a hearing impairment that is so severe that the child is impaired in processing linguistic information through hearing, with or without amplification that adversely affects a child's educational performance. . . . *Hearing impairment* means an impairment in hearing, whether permanent or fluctuating, that adversely affects a child's educational performance but that is not included under the definition of deafness in this section. (U.S. Department of Education, 2006, p. 46756)

And, lastly, in 2014–2015, 165 children ages 3–5 and 1,243 children and youth ages 6–21 were served under the IDEA *deaf-blindness* definition:

> *Deaf-blindness* means concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational needs that [the children] cannot be accommodated in special education programs solely for children with deafness or children with blindness. (U.S. Department of Education, 2006, p. 46756)

Few studies focus specifically on the cognitive and neuropsychological assessment of children with deaf-blindness, which involves multiple disabilities that may or may not be interrelated, and so the populations with VI/B and deafness/hearing impairment (D/HI) are our main focus in this portion of the chapter. Briefly, however, in a recent review of individual variables for examiners to consider when evaluating children and individuals with deaf-blindness, Dalby and colleagues (2009) specified etiology as a major factor. According to Dalby and colleagues, individuals with congenital deaf-blindness are more likely to experience cognitive, adaptive, and social impairments than those whose deaf-blindness is acquired. Those with acquired deaf-blindness are more likely to use speech as their primary mode of communication. The American Association of the Deaf-Blind's website (*www.aadb.org*) offers educational resources to address the needs of children with multiple disabilities, such as listings of state organizations, service providers, and support groups; frequently asked questions and fact sheets; newsletters and magazines; and information about assistive technology. *Best Practices in School Psychology V* (Thomas & Grimes, 2008) offers broad-based information helpful for designing instruction and intervention for individuals with multiple disabilities like deafblindness. In particular, Powell-Smith, Stoner, Bilter, and Sansosti (2008) emphasize the importance of systemic collaboration, considering each child's optimal learning environment and mode of communication, and encouraging family involvement.

Etiologies: A Brief Overview

In a typically developing child, vision begins with the eye detecting light in the environment and transducing the stimulus energy into neurological impulses. These impulses are sent from the eye (retina) to the occipital lobes of the brain via the optic nerve, which travels through the lateral geniculate nucleus in the midbrain before transferring information to the primary visual cortex for basic visual processing; then information such as orientation, contrast, color, location in space, and identity is saved for higher-level processing in either the occipito-parietal dorsal visual stream or the temporo-parietal ventral visual stream, depending on information type. In some children with VI/B, abnormal pigment production in the eye (albinism, a congenital condition; Bradley-Johnson & Morgan, 2008) disrupts this process early in the stream. Other congenital causes include retinitis pigmentosa, which involves degeneration of light-sensitive retinal cells, or various forms of prenatal damage to the visual system. If the damage occurs during or after birth, the vision loss is considered acquired. Bradley-Johnson and Morgan (2008) list the following possible diseases and accidents that may result in acquired vision loss: head injury, anoxia at birth, central nervous system infections (e.g., meningitis), and medication reactions. Damage to any part of the visual pathway can lead to observable difficulties with visual tasks, but most children's vision loss can be traced to the eye or optic nerve early in the pathway. One exception is cortical visual impairment, sometimes called cortical blindness, which occurs with damage to the brain's visual system from head injuries, infections, and other accidents or diseases. Children with this condition experience problems not only with basic visual processing, but also with attention and other cognitive functions (Bradley-Johnson & Morgan, 2008). Understanding the etiology of a child's vision loss can aid in individualized and sensitive psychoeducational assessment and planning.

Children without hearing impairments first receive sound information as sound waves, or vibrations, entering the outer, middle, and then inner ear. The vibrations travel from the eardrum to three tiny bones in the middle ear, and then one of those bones, called the stirrup, sends the vibrations along the coiled cochlea in the inner ear. When the cochlea vibrates, tiny hairs called cilia move, and the sound information passes through the auditory nerve to the auditory cortex in the temporal lobes of the brain for processing. The auditory cortex organizes information from different sound frequencies in higher cortical areas, which individuals "hear" as different sounds.

According to the American Speech–Language–Hearing Association (ASHA; 2010), sensorineural hearing loss—which occurs when the cochlea or auditory nerve is damaged, is permanent, and often cannot be corrected with surgery or other medical procedures—can result from disease (e.g., viruses or tumors); exposure to toxins (e.g., drugs) or high noise levels; head injuries; inherited genetic syndromes; or perinatal injury. These children not only have trouble hearing low-level sound, but also experience difficulty understanding speech and hearing sounds clearly (ASHA, 2010). Conductive hearing loss, on the other hand, results primarily in an inability to hear faint sounds or a reduction in sound level that is often medically correctable. Conductive hearing loss can be caused by middle ear pathology (e.g., otitis media or ear infection), impacted earwax, ear canal infection, irritation from a foreign body in the ear, or malformation or absence of any physical part of the outer or middle ear (ASHA, 2010). If there is damage to or malfunction of both the outer or middle ear and the inner ear or auditory nerve, mixed (sensorineural and conductive) hearing loss can occur. In addition, hearing loss can be unilateral (occurring only in one ear) or bilateral. According to ASHA, about 1 in 1,000 children is born with unilateral hearing loss, and approximately 3% of all school-age children have it. Children with both unilateral and bilateral hearing loss are at risk for academic, speech–language, and social-emotional difficulties (ASHA, 2010).

Special Considerations for Assessment: Child Factors

A variety of individual factors can influence the test performance of children with sensory disabilities and can affect the interpretation of results. Selected chapters in the Thomas and Grimes (2008) volume offer helpful educational programming information for students with VI/B (Bradley-Johnson & Morgan, 2008) and students with D/HI (Lukomski, 2008), highlighting the diversity of these populations and the lack of a "one-size-fits-all" solution for their psychoeducational planning. For both populations, early intervention and family collaboration, as well as a multidisciplinary approach to assessment, placement, and intervention, are imperative for appropriate data-based decision making in the schools. The school psychologist must cooperate with a team of specialists and other professionals—such as speech–hearing therapists and educators, medical personnel, and VI/B educators—to develop a comprehensive, interdisciplinary educational plan for each child that addresses child–environment fit, considers preferred modes of communication, and includes objectives and strategies for improving social interaction and communication skills. Figure 24.1 summarizes additional broad guidelines for selecting, administering, and interpreting intellectual assessment measures for students with sensory disabilities.

For a child with D/HI, Lukomski (2008) stresses including the following specific elements in the child's plan: strategies for handling communication breakdowns, handling the child's fatigue, school staff training in communication strategies, and American Sign Language (ASL) parent training. In addition, decision-making and planning teams must keep in mind that most children with hearing problems have average IQs, but exhibit lags in literacy and academic achievement because of developmental differences in the acquisition and internalization of language. For children with VI/B, having a basic understanding of the variety and availability of adaptive technologies and materials—such as raised-line paper, computer programs with speech or Braille output, and closed-circuit television that electronically magnifies text for children with low vision—is crucial (Bradley-Johnson & Morgan, 2008). Since many visual cues in the environment are unavailable to these children, the school psychologist must also include strategies targeted at improving organizational skills and social communication. For children with either D/HI or VI/B, evaluation should include systematic observation of the children in a variety of settings, since social interactions, communication modes, and behavior may vary among contexts.

Next, we consider specific areas of heterogeneity in the populations with VI/B and D/HI that may affect test selection, administration, and interpretation, including the following: age at onset; quantitative and qualitative nature of the vision or hearing loss; etiology; comorbidity; and (for children with D/HI) reading and language ability, preferred mode of communication, and parental hearing status.

Individual Factors to Consider in Assessment of Children with VI/B

One major individual factor to consider in the assessment of children with VI/B is age at onset of the sensory impairment, which contributes to the heterogeneity of this population. For example, in the standardization sample for the Comprehensive Vocational Evaluation System (CVES; Dial, Chan, Mezger, et al., 1991)—an empirically developed, VI/B-specific neuropsychological battery—56% of cases were considered "congenital," with onset of vision impairment occurring from birth to 1 year of age; 8% were "early blind" (2 years to 5 years, 11 months); 9% were "school-age" (6 years to 17 years, 11 months); and the rest were adult-onset cases (Hill-Briggs, Dial, Morere, & Joyce, 2007). Joyce, Isom, Dial, and Sandel (2004) found that adults with early-onset VI outperformed those with adult-onset VI in shape and texture discrimination on tests from the Haptic Sensory Discrimination Test (HSDT) and McCarron Assessment of Neuromuscular Development, VI/B Version (MAND-VI), providing evidence that age at onset of VI substantially impacts neuropsychological test results. Conversely, adult-onset participants excelled at persistent motor control, and the school-onset group performed significantly better than early-onset individuals in bimanual dexterity (Joyce et al., 2004). On the other hand, MacCluskie, Tunick, Dial, and Paul (1998) compared Wechsler Adult Intelligence Scale—Revised (WAIS-R) and Cognitive Test for the Blind (CTB) performance for groups of adults ($N = 60$) with early-onset (before age 2) versus late-onset (after age 5) VI and found no significant differences. However, years of education significantly contributed to variance in cognitive ability scores. The authors tested a small sample of adults and did not control for degree of residual vision, so these results should be interpreted with caution. Notwithstanding, one implication for professionals assessing persons with VI/B is that while age of onset itself may not contribute to variability in cognitive test scores, number of

Selection	Administration	Interpretation
Consider results of appropriate screening and sensory assessment measures already administered (e.g., functional vision assessment [FVA] and learning media assessment [LMA] for students with VI/B).	Plan administration in collaboration with other professionals as appropriate (e.g., VI professionals, speech–language pathologists).	Include multiple sources of information, from different informants and in a variety of formats, in any interpretations and conclusions (e.g., direct observation, teacher and parent report, record review).
Be clear about, and clearly state in the report, reasons for evaluation and how results will be used (e.g., diagnosis, disability determination, intervention planning, etc.).	Ensure that examiner has appropriate training in test theory, standardized assessment, child development, and unique needs and characteristics of children with sensory impairments.	Provide clear rationale for using the selected tests, and clearly identify which subtests were included and excluded (if any) and why.
Evaluate appropriateness of educational programming, including placement and any interventions, up to the current point in time (i.e., has the student received evidence-based, appropriate core instruction?).	Plan adaptations and accommodations in collaboration with other professionals knowledgeable about sensory disabilities before administering the test.	Report IQ and factor scores as ranges or confidence intervals, rather than point estimates.
Consider medical history and current conditions, possible comorbid disability conditions, and the child's strengths.	Provide standardized administration whenever possible. Consider which tasks will require adaptations (which do not change the content or difficult level of test materials), which tasks will need significant modification, and which tasks are wholly inappropriate.	Include qualitative information about behavior and performance when available, particularly when test administration deviates from standardized conditions (e.g., testing-of-limits procedures, alternative response modalities).
Consider both individual (e.g., level of impairment or communication mode) and test factors (e.g., norm sample, language requirements).	Follow appropriate best practice and ethical guidelines, and provide appropriate documentation in the report, for all adaptations and accommodations to test materials and administration or response methods (e.g., use of interpreters, Braille or orally administered test versions).	Clearly note limitations of the currently available data, and of standardized test scores, in drawing firm conclusions about performance and level of functioning.
Consider most appropriate subtests within each test battery, given task demands (input, processing, and output) and test-author-provided suggestions.	Follow test manual recommended accommodations, if available.	Involve professionals knowledgeable about sensory disabilities (e.g., rehabilitation professionals) in generating appropriate recommendations.

FIGURE 24.1. Guidelines for assessing children with sensory disabilities. These guidelines are based on information and conclusions presented in Goodman, Evans, and Loftin (2011); Lund, Miller, and Ganz (2014); Reesman et al. (2014); and the Texas School for the Blind and Visually Impaired (2016).

years of education seems to matter. Thus it is important to consider how each examinee's current age and grade level may influence scores in individuals with both congenital and late-onset VI/B.

Another variable that can affect test performance and interpretation of results is level of blindness or amount of vision loss. For instance, an individual with congenital (onset prior to 18 months of age) blindness and no vision—a Braille or tactile learner—is at the most severe end of the spectrum, and will require the most modifications both to testing and to educational content and delivery methods; a print learner with low vision will require fewer testing and educational modifications. Consider the illustrative case of the CVES standardization sample, in which 71% of cases were considered "legally blind," 18% "visually impaired," and 11% "totally blind." It is also worth noting that even two individuals with the same optical refraction score (e.g., 20/100) may have different functional levels of vision (Hill-Briggs et al., 2007). In the Joyce and colleagues (2004) study, participants who had some residual functional vision performed significantly better on various subtests of the HSDT and MAND-VI (included in the CVES battery) than those who were "totally blind"; again, this demonstrates the importance of considering level of vision loss in assessing individuals with VI/B.

Clinicians should keep in mind that VI/B has different etiologies in different people, and that multiple etiologies may even be present within the same individual. Studies have found differences in cognitive and neuropsychological performance based on these different etiologies. For example, participants whose blindness was caused by early birth (retinopathy of prematurity, or ROP) performed significantly worse on spatial and auditory analysis (verbal–spatial cognitive abilities) and hand strength (perceptual–motor functions) tests from the CVES battery than participants with either retinitis pigmentosa or congenital cataracts as etiologies (Nelson, O'Brien, Dial, & Joyce, 2001). It was concluded that ROP caused more cognitive impairment than other common etiologies of VI/B. Similarly, McGee (1994) found that those with diabetes-related blindness performed worse on tasks measuring left-lateralized perceptual–motor functions, as well as on nonverbal cognitive tasks.

Many individuals with VI/B have comorbid disabilities or conditions that may or may not be related to their vision loss, including neuropsychological conditions (such as head injuries, cerebral palsy, and tumors) and other disabilities commonly seen in school settings (e.g., learning, hearing, or physical disabilities). In fact, throughout the CVES standardization process, only 25% of the individuals who presented with vision loss did *not* have profiles consistent with another such disorder (Hill-Briggs et al., 2007), and standardization was based only on this 25%. However, since the majority of individuals with VI/B seem to have comorbid disorders, it is debatable whether or not a group without comorbidity is truly representative of the population with VI/B in the United States. In fact, according to Miller (2007), almost all children with sensory impairments have additional impairments in academic, social, cognitive, adaptive, and/or behavioral domains. It is vital for the evaluator to keep in mind that the cognitive and neuropsychological profile observed during evaluation of an individual with VI/B and a comorbid condition may be the effect of the VI/B, of the other neuropsychological disorder, of the additive combination of both independent conditions, or of the multiplicative effects of both conditions acting together.

Individual Factors to Consider in Assessment of Children with D/HI

Age at onset should also be considered in assessing children with D/HI. Because hearing, verbal communication, and language development are linked, children with early-onset (prior to 18 months old) hearing loss and those whose hearing loss occurred after significant language development exhibit significantly different profiles of linguistic and communicative functioning (Braden, 1994; Marschark & Clark, 1993; Meadow, 1980). Although severe and profound hearing loss will have the most significant impact on test administration, any level of hearing loss can have some effect on standardized test performance (Braden, 1994). In addition, examiners should consider whether the hearing loss is progressive, as an examinee's current level of hearing loss may not be the same as it was earlier in development or will be later in life (Hill-Briggs et al., 2007). Finally, as in the evaluation of children with VI/B, clinicians must also consider the additive and interactive influences of any comorbid disability or medical condition. A learning disability, for example, may be overlooked if diagnostic overshadowing causes the examiner to erroneously attribute all verbal reasoning deficits to hearing loss. According to several recent surveys of students with D/HI in special education, about half have additional disabilities (Gallaudet Research Institute, 2007; Mitchell, 2006; Szymanski, Brice, Lam, & Hotto, 2012).

One relatively unique factor in assessing individuals with D/HI is their level of reading and language ability. According to the Gallaudet Research Institute (Holt, 2005), 18-year-olds with severe hearing loss usually perform on tests of reading comprehension at a grade 4.5 level, and those with profound hearing loss at a grade 3.8 level. Competency with reading and writing tasks may also be negatively affected by the different grammatical structures of traditional spoken English and many signing systems, such as ASL. Therefore, asking children with D/HI to read the directions for a test is not an appropriate modification. Variability in reading and language levels should also be considered when examiners are selecting tests that include high verbal loadings or rely substantially on reading instructions and/or stimuli.

A further consideration for evaluators in testing an individual with D/HI is the examinee's primary or preferred mode of communication—both current and during development. Current communication mode has implications for communication methods during testing; for example, this affects the decision of whether to use an interpreter, to seek an examiner fluent in ASL or another visual communication system, or to modify the test administration in some other way so as to accommodate the individual child's hearing loss level and preferred communication style. Several studies have also shown that mode of communication during development—for example, cued speech, oral communication, combined methods, or sign language (ASL or another signing system)—affects cognitive functioning and performance, and even lateralization of neuropsychological tasks (e.g., Bosworth & Dobkins, 1999; Cattani & Clibbens, 2005; Charlier & Leybaert, 2000; LaSasso, Crain, & Leybaert, 2003).

Parental hearing status (i.e., whether the parents also have D/HI) can also affect a child's development of language and communication (Anderson, 2006), as well as achievement and psychosocial adjustment (Polat, 2003). It can also affect cultural and identity development, as the child decides whether to identify primarily with deaf culture or to assimilate him- or herself into the hearing world (Leigh, Marcus, Dobosh, & Allen, 1998). Clinicians should not only be sensitive to these cultural issues, but remember that the hearing status of a child's parents will affect the child's cognitive functioning (Braden, 1994; Vernon, 2005) and his or her preferred or primary communication mode in adolescence and adulthood. Although debates continue regarding this issue, it has been shown that deaf children born to hearing parents function at a lower cognitive level than those born to deaf parents (Braden, 1994; Vernon, 2005).

Special Considerations for Assessment: Test Factors

In addition to addressing individual child factors, clinicians should consider the inadequacies of the assessment instruments themselves—from instructions, to stimuli, to norms and norming samples—for use with special populations, as such inadequacies have historically plagued both research and application of evaluation methods for individuals with VI/B and D/HI.

For individuals with VI/B, some tests can be administered without vision requirements, such as subtests from the Halstead–Reitan Battery (HRB), including the Tactual Performance Test, Grip Strength, and Finger Oscillation, as well as the haptic version of Raven's Progressive Matrices (Rich & Anderson, 1965). However, tests for general cognition and memory that lack vision requirements are more difficult to find. Furthermore, the appropriateness of these tests for the population with VI/B is debatable, considering inadequate norms for special populations and the need to control for examinees' varying levels of residual vision (Hill-Briggs et al., 2007). Although traditional IQ tests (e.g., the Stanford–Binet and Wechsler scales) have been adapted for individuals with VI/B, the norming samples for this group have historically been very small and included children with comorbid disabilities (Gutterman, Ward, & Genshaft, 1985). For example, in the original norming sample for the CVES battery created specifically for examinees with VI/B, 44% of those tested displayed profiles indicative of some kind of neuropsychological disorder (Dial, Chan, Tunick, Gray, & Marme, 1991). As noted earlier, the CVES norms were ultimately based on the final 25% of tested cases that had no other disability or medical condition besides VI/B (Hill-Briggs et al., 2007); therefore, practitioners should be wary of the inherent problems faced by test creators in finding and selecting appropriate norming samples. If much of the normative sample already exhibits the very class of disorders the test is designed to detect, sensitivity for doing so is severely limited.

Examiners testing children with D/HI must consider that these individuals' need for basic communication accommodations during the test automatically breaks standard administration. Oral administration without a sign language interpreter, therefore, is often impossible and is always inappropriate with individuals who do not primarily use oral communication outside the testing situation (Hill-Briggs et al., 2007). Reesman and colleagues (2014) specifically note that "when and if a hearing psychologist with limited proficiency in ASL is assessing the cognitive skills of an ASL-using child who is deaf or hard of hearing, the inclusion of measures requiring interpretation and evaluation of a child's language may be viewed as wholly inappropriate" (p. 104). Additionally, individuals who typically use sign language or other visual communication strategies may seem to have adequate oral expression abilities, but may have difficulty processing incoming auditory information. Informal communication abilities can be similarly deceiving, but the testing situation is a formal communication context with different demands and a higher imperative to ensure that each individual sound and word is being expressed and received precisely as it is intended. These individuals need to spend extra time processing auditory information, and when interpreters are used, time delays and inconsistencies in wording and meaning can have a negative impact on test performance.

Use of interpreters also makes it difficult to adhere closely to standardized administration. This is because ASL and spoken English have different syntactical rules and grammatical structure, which render verbatim or direct translation impossible. Because of these difficulties, test instructions should be given in the examinee's preferred communication mode—such as ASL, if possible—and the examiner should allow extra practice trials and modeling when necessary (Hill-Briggs et al., 2007). If the examiner is not fluent in the examinee's primary communication mode, an experienced interpreter who is familiar with the examinee's dialect or signing approaches, as well as with mental health interpreting idiosyncrasies (Miller & Vernon, 2001), should be sought. Moreover, sensory distractions—both visual and auditory—should be carefully minimized. Hill-Briggs and colleagues (2007) also caution against asking examinees with D/HI to cover their eyes or wear a blindfold during perceptual and motor tasks (since this eliminates their sole mode of communication), and recommend using visual barriers instead.

The question of which tests to use in assessing individuals with D/HI is as important as deciding how to modify them, and is further discussed below. Although the Wechsler scales have been used with such individuals in the past (Braden, 1994), "adapting" a Wechsler test by only considering the nonverbal half of the test (Performance IQ, in older editions) is problematic, as excluding the other half of the test may lead to under- or overestimation of overall cognitive ability. Furthermore, as a general rule, tests with high verbal loading should be avoided. Testing considerations for individuals with D/HI may seem complex, but with appropriate evaluation of materials and methods employed, accurate measures of functioning are possible.

Instruments for Assessing Individuals with VI/B

Clinicians might consider using a cognitive measure developed specifically for individuals with VI/B, such as the Cognitive Test for the Blind (CTB), which was standardized on and designed for such individuals. The CTB is one of several tests included in the CVES for assessing individuals with VI/B (Hill-Briggs et al., 2007), and it covers three primary neuropsychological factors: verbal–spatial cognitive abilities, perceptual–motor functions, and emotional–coping concerns. Although the CTB has acceptable reliability and validity (Nelson, Dial, & Joyce, 2002), it can only be used with individuals age 14 and older. Despite the positive results from test development of the CVES showing its utility for populations with VI/B (Chan, Lynch, Dial, Wong, & Kates, 1993; Dial, Chan, Mezger, et al., 1991; Kaskel, Dial, Chan, & Roldan, 1991), the test was developed for adults; although some measures suggested in the battery (e.g., the Wechsler Intelligence Scale for Children [WISC]) are well suited for children, the battery as a whole is not.

Although there have been attempts to create and disseminate child-friendly cognitive tests specifically for individuals with VI/B, most of these tests are no longer commercially available due to problems with test reliability and validity, as well as inadequately made comparisons between samples of children with VI/B and typically developing children. Creators of the Bielefeld Developmental Test for Blind Infants and Preschoolers, for example, attempted to create "blind-neutral" items for young children but were not successful, and the test was removed from circulation (Brambring &

Tröster, 1994). The best option remaining for practitioners working with children who have VI/B is to try to select, from the available children's standardized cognitive and neuropsychological tests, those that most easily lend themselves to modification. Select subtests from widely available batteries such as the WISC-V, for example, are appropriate for a pediatric population and can be used with these children if interpreted with caution.

The Texas School for the Blind and Visually Impaired (2010) previously suggested using only the verbal subtests from the various Wechsler scales or the Woodcock–Johnson Tests of Cognitive Abilities (WJ COG), but they stressed that clinical judgment is needed for interpreting results from only half of the entire test. Because the WJ COG, however, includes tests that are designed to be used selectively, interpretation of only verbal subtests on this measure may be preferable to using those from the Wechsler scales, in which subtests are designed to contribute to an interpretable composite. Also, minimizing the inclusion of subtests that are heavily nonverbally loaded or that comprise several tasks dependent on visual–spatial information is essential. Examiners' manuals for several popular tests like the WJ IV COG provide tables listing appropriate subtests for populations with VI/B and suggested accommodations specifically for these examinees (Mather & Wendling, 2014). According to a test manual search we conducted, however, neither the WISC-V nor the WJ IV COG manual offers VI/B norms. Table 24.1 displays commonly used cognitive tests that provide suggested accommodations and/or lists of appropriate subtests for individuals with VI/B (and other disabilities).

Flanagan, Ortiz, and Alfonso (2013) have revised their recommendations for the cross-battery assessment (XBA) approach to cognitive evaluation of children. This approach is based on the Cattell–Horn–Carroll (CHC) model of intelligence (see Schneider & McGrew, Chapter 3, this volume), which purports that general intelligence, or *g*, comprises a number of broad-stratum abilities, such as fluid reasoning (Gf), crystallized knowledge (Gc), quantitative reasoning (Gq), visual–spatial processing (Gv), and short-term memory (Gsm). Instead of administering only one cognitive test battery to children, examiners using this approach systematically select tests from more than one battery, to ensure that the abilities and processes most germane to the referral are assessed comprehensively. The XBA approach often results in a more psychometrically stable and complete picture of cognitive abilities than that which can be obtained by a single battery. This approach may aid examiners of children with VI/B in selecting test instruments, as subtests from different batteries can be administered according to both their contributions to the CHC broad abilities and their level of appropriateness for these children. If one Gf subtest, as an illustration, involves visual stimulus material—as does the WJ IV COG Analysis–Synthesis test—then Gf tests that are less dependent

TABLE 24.1. Special Testing Consideration Information Available in Cognitive Test Manuals

	Suggested accommodations provided				Norms/clinical sample scores provided				Chart of appropriate subtests provided			
Test	VI/B	D/HI	PI	TBI	VI/B	D/HI	PI	TBI	VI/B	D/HI	PI	TBI
WISC-V		✓[a]						✓[b]		✓[a]		
WJ IV COG	✓	✓	✓					✓[c]	✓	✓		
SB5	✓	✓	✓				✓			✓		
KABC-II						✓						
NEPSY-II						✓		✓				
DAS-II	✓	✓	✓			✓				✓		

Note. VI/B, visual impairment/blindness; D/HI, deafness/hearing impairment; PI, physical impairment; WISC-V, Wechsler Intelligence Scale for Children—Fifth Edition; WJ IV COG, Woodcock–Johnson IV Tests of Cognitive Abilities; SB5, Stanford–Binet Intelligence Scales, Fifth Edition; KABC-II, Kaufman Assessment Battery for Children—Second Edition; NEPSY-II, NEPSY—Second Edition; DAS-II, Differential Ability Scales—Second Edition.

[a] Available in a separately released Technical Report (Day, Adams Costa, & Raiford, 2015).

[b] n = 20, ages 7–16 years.

[c] n = 12, ages 7–17 years.

on visual information can be selected. Evaluators applying the XBA approach, however, must note that some broad-ability cluster scores (such as Gv) may be difficult or impossible to obtain or interpret, depending on the individual child and his or her level of vision loss.

Various tests from the HRB or Luria–Nebraska Battery (LNB) can be used for individuals with VI/B if careful attention is paid to the examinee's need for modifications and if these modifications are considered in interpreting results. The NEPSY-II, which may be more appropriate for school-age children, does not offer VI/B norms or suggested accommodations, but includes subtests similar to those found in the HRB and LNB and may be administered with similar accommodations. Ultimately, test selection should be a collaborative decision between a clinician and a specialist or vision teacher knowledgeable about and experienced in working with children who have VI/B. In addition, formal evaluation methods should be supplemented by informal, often more subjective methods (e.g., student work samples, personal interactions with the child, and interviews with parents and teachers), in order to gain a more complete picture of the child's neuropsychological strengths and weaknesses than may not be obtainable from test scores alone. According to the Texas School for the Blind and Visually Impaired (2016), intellectual assessment of these students should also include direct observations of the child.

Instruments for Assessing Individuals with D/HI

Although the earlier-discussed issues of mode of communication, use of interpreters, and other standardized test problems still make cognitive and neuropsychological assessment of individuals with D/HI complex, there is considerably more research on this topic (for overviews, see Braden, 2001; Maller, 2003; Reesman et al., 2014) than is available for the population with VI/B. Braden (1994) reviewed over 300 studies on IQ and deafness, and found that IQ distributions for deaf people without comorbid conditions and for hearing people are nearly identical; this has also helped to raise awareness of the intellectual potential of students with D/HI. Again, test selection and interpretation should include awareness of individual variables—and, as noted above for individuals with VI/B, it should be a multidisciplinary team effort involving hearing teachers and other special professionals.

Reesman and colleagues (2014) reviewed 13 cognitive assessment measures available at that time for available information on using the tests with students who have D/HI. Many measures offered guidance in their test manuals regarding administration accommodations, including the Stanford–Binet Intelligence Scales, Fifth Edition (SB5); the Leiter International Performance Scale—Revised (Leiter-R) and Leiter-3; the WJ III COG; the Differential Ability Scales—Second Edition (DAS-II); the Kaufman Assessment Battery for Children—Second Edition (KABC-II); the Wechsler Nonverbal Scale of Ability (WNV); the WISC-IV; the Comprehensive Test of Nonverbal Intelligence—Second Edition (CTONI-2); and the Universal Nonverbal Intelligence Test (UNIT). The authors noted that the DAS-II offered particularly specific information regarding translation and back-translation (for fidelity) of administration instructions to and from ASL. Additionally, the ASL translation of the DAS-II instructions is available on DVD. Outside of a reliable ASL test translation, Reesman and colleagues pointed out that tests that rely solely on gesture for instruction presentation and response minimize translation issues for individuals who use different communication modalities from those of the examiner. For example, Maller (2000) examined the UNIT and concluded that no items in the four subtests studied showed differential item functioning across typically developing and D/HI samples.

In terms of test interpretation, although some measures included a small D/HI clinical norm sample, no test manual offered an adequate description of the sample in terms of communication modality, use of assistive listening devices, degree of hearing loss, etc. to allow for confident comparison (Reesman et al., 2014). Furthermore, the test manuals offered little guidance in appropriate interpretation of scores for this population. The manuals for the DAS-II, KABC-II, WNV, WISC-IV, and UNIT, however, did include discussion of comparing these students to the normative sample. Reesman and colleagues (2014) concluded that there is no one superior measure for all individuals with D/HI, and they recommended considering child and testing factors in selecting, administering, and interpreting intellectual assessment measures for these students. They called for future empirical research on the appropriateness and statistical bias of various tests with this population (Reesman et al., 2014). Notably, they also highlighted the possible future impact of technological developments that may allow for greater test accessibility, such as video-assisted administration. It remains to be seen how advances like tablet and computerized administration and scoring will affect the validity and reliability of testing with this group.

When examiners are assessing fluid reasoning skills, Hill-Briggs and colleagues (2007) suggest emphasizing spatial reasoning tasks, while keeping in mind that language differences and cognitive processing delays or use of an interpreter may affect the delivery of instructions. On Digit Span working memory tasks, they also remind clinicians that the average forward span for deaf signers is approximately five digits, and that the backward span is usually equivalent. Differences between signing system and English grammatical structure or syntax must be considered in using subtests such as the WJ IV COG Visual–Auditory Learning test and other sentence or story recall tasks.

Notably, the SB5 test manual includes an entire appendix titled "Use of the Stanford–Binet Intelligence Scales, Fifth Edition, with Deaf and Hard of Hearing Individuals: General Considerations and Tailored Administration" (Roid, 2003, Appendix E). The appendix divides individuals with D/HI into four categories: those who use sign language, simultaneous communication, cued speech, or auditory verbal–oral communication. The categories are based on the SB5 standardization administration to a special sample of individuals with D/HI, and suggestions for the appropriateness of administering each SB5 subtest to each of these groups are provided. In addition, both the WISC-V and WJ IV COG manuals provide suggested accommodations specifically for individuals with D/HI, as well as a table of appropriate subtests for use with such children by primary mode of communication (see Table 24.1).

Miller (2007) suggests including the SB5 in CHC-based cognitive assessment of children with D/HI, as the SB5 provides scores for the five broad abilities Gf, Gc, Gq, Gsm, and Gv, and has adequate information in Appendix E of the test manual (Roid, 2003). In the KABC-II (Kaufman & Kaufman, 2004) manual, the authors report no significant differences between a subgroup (n = 18) of students with D/HI and the hearing standardization group. Therefore, the KABC-II can also be used as part of a CHC-based cognitive assessment for children with D/HI. The KABC-II manual (Kaufman & Kaufman, 2004) reports that hearing and D/HI norms for the following factors were not significantly different, suggesting that the KABC-II is a stable measure of these abilities in children with D/HI: long-term retrieval (Glr), Gf, and Gv. Lastly, the DAS-II manual (Elliott, 2007) reports equivalent performances for children with D/HI and a matched hearing control group on subtests composing the test's Gf and Gv cluster scores. The DAS-II manual (Elliott, 2007) also includes tables of suggestions for administration to children with D/HI by subtest and communication mode, as well as a sign language CD and norming information from a clinical sample with D/HI. Instead of using one of these batteries alone to assess cognitive functioning in a child with D/HI, clinicians should consider employing the XBA approach (Flanagan et al., 2013; see also Flanagan et al., Chapter 27, this volume), based on Miller's suggestions for this special population. It is worth noting, however, that subtests contributing to cluster scores for broad-stratum abilities such as auditory processing (Ga) may be inappropriate to administer to these children or may have problematic interpretations.

With perceptual–motor tests, clinicians can usually use standard tasks from neuropsychological batteries such as the HRB and LNB, with careful attention to how directions are given and processed by the examinee. Again, the NEPSY-II is perhaps more appropriate for use with school-age children but is not yet empirically validated for individuals with D/HI, although it offers D/HI norms from a special D/HI clinical sample. As mentioned previously, examiners should provide an alternative to blindfolding or to asking individuals with D/HI to close their eyes during some neuropsychological tasks (e.g., balance tasks), as these modifications restrict their use of vision, an important mode of communication for individuals with D/HI. Hill-Briggs and colleagues (2007) note that standard visual–motor tasks in neuropsychological batteries are suitable for individuals with D/HI, but that caution should be exercised in interpreting the Rey Complex Figure task because children with D/HI may have different organizational strategies from those of typically developing children.

PHYSICAL DISABILITIES

According to the U.S. Census Bureau (2010), an estimated 8.2% of the general population had a physical disability (i.e., some limitation in basic physical activities), amounting to 19.9 million people in the United States. Difficulties surrounding the assessment of individuals with physical disabilities have recently become more prevalent—due in part to an increase in the number of individuals with such disabilities, as well as the proliferation of health services for these individuals (O'Keefe, 1994), and the increasing presence of these individuals in both post-secondary education (Fleischer, Zames, & Zames, 2012) and the workforce (Nevala, Pehkonen, Koskela, Ruusuvuori, & Anttila, 2015; Verhoef, Miedema, Meeteren, Stam, & Roebroeck, 2013). Many of these difficulties stem from an inability to adequately measure abilities or to interpret test scores on various assessments when an individual's disability inhibits him or her in some way. The difficulties specific to assessing individuals with physical disabilities, as well as the potential testing accommodations, are dependent on the type and severity of the disabilities.

Typically, children with some sort of physical impairment (PI) are served under the IDEA 2004 category of *orthopedic impairment*. Federal guidelines (U.S. Department of Education, 2006) define this as follows:

> *Orthopedic impairment* means a severe orthopedic impairment that adversely affects a child's educational performance. The term includes impairments caused by a congenital anomaly, impairments caused by disease (e.g., poliomyelitis, bone tuberculosis), and impairments from other causes (e.g., cerebral palsy, amputations, and fractures or burns that cause contractures). (pp. 46756–46757)

According to the U.S. Department of Education (n.d.), approximately 6,325 children ages 3–5 and an additional 46,268 children and youth ages 6–21 in the United States were served under this category in the fall of 2015. Occasionally, children with a PI may qualify for special education services under the category of *other health impairment* (OHI). According to the IDEA regulations, OHI is defined as

> having limited strength, vitality, or alertness, including a heightened alertness to environmental stimuli, that results in limited alertness with respect to the educational environment, that—
>
> (i) Is due to chronic or acute health problems such as asthma, attention deficit disorder or attention deficit hyperactivity disorder, diabetes, epilepsy, a heart condition, hemophilia, lead poisoning, leukemia, nephritis, rheumatic fever, sickle cell anemia, and Tourette syndrome; and
>
> (ii) Adversely affects a child's educational performance. (U.S. Department of Education, 2006, p. 46757)

According to the U.S. Department of Education (n.d.), 23,652 children ages 3–5 and 857,544 children and youth ages 6–21 in the United States were served under this category in the fall of 2015. However, OHI covers more than just PI, so these figures are overestimates. Nevertheless, each individual with PI, whether qualifying for school services under orthopedic impairment or OHI, or under a federal Section 504 Accommodation plan, requires some unique testing considerations at both the individual and test level.

Etiologies: A Brief Overview

Motor skill development occurs in stages, preceded and accompanied by the growth and development of the endocrine and nervous systems. The typically developing child passes through a sequence of milestones, such as holding his or her own head up at about 2 months of age, crawling at about 6 months, and walking at about 1 year of age (Sigelman & Rider, 2009). It is during this time of rapid development that some types of PI become apparent. Specifically, delayed walking is often indicative of a motor impairment. PI can be either acquired (e.g., spinal cord injury [SCI] and amputation) or congenital. Congenital disabilities become apparent either at birth or during infancy and toddlerhood when a child fails to reach one or more typical milestones. Cerebral palsy (CP), for example—the most common motor disability in childhood, occurring 2.0–3.6 times in every 1,000 live births (Accardo, 2008; Arneson et al., 2009; Bhasin, Brocksen, Avchen, & Van Naarden Braun, 2006; Cans, 2000)—has many potential causes, including genetic abnormality, intrauterine infection, complications during labor or birth, or experiencing a TBI before the age of 5. CP generally causes problems associated with movement and posture, although comorbid disorders (e.g., intellectual disabilities, sensory disorders, seizures, and growth abnormalities) often occur (Pellegrino, 1997).

Special Considerations for Assessment: Child Factors

There is little literature on the neuropsychological testing of children with PI, and much of what exists is very specific to one disorder or another. Two of the most researched conditions are CP and SCI. For example, a recent study examined the performance of children with CP on the Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III) (Sherwell et al., 2014). Results indicated that a proportion of the children (20.5%) were unable to complete the full test battery, due to insufficient pointing or verbal responses. Of those who were able to complete the WPPSI-III, fine/gross motor weaknesses depressed their overall scores (i.e., Full Scale IQ) by approximately 5 points (i.e., one-third of a standard deviation). Another variable contributing to the heterogeneity of this population is a child's age at onset of a disability. Many physical disabilities are the result of congenital defects or diseases (e.g., CP, muscular dystrophy [MD], and multiple sclerosis). The age of onset can significantly affect one's ability to compensate for mild to severe PI. Individuals with congenital disabilities learn to cope with their limitations from birth. Those who acquire impairments at a relatively young age are less likely to have multiple deficits as adults. Li and Moore (1998) found that both younger participants and those with congenital disabilities had higher levels of disability acceptance. Those with PI acquired as the result of disease, amputation, or accidents have the added difficulty of relearning many tasks without the full use of a particular body part. Research conducted on motor recovery following stroke, SCI, and CP has indicated that motor recovery, if it occurs, can be incomplete or deficient (Krishnan, 2006).

To qualify for services under federal programs (IDEA or Section 504), an individual's disability must be severe. The type and severity of disability will greatly affect test selection and interpretation. For individuals who fatigue easily as a result of their disability, for example, it may be appropriate to extend the time to allow for additional breaks (AERA et al., 2014). Individuals with PI may also suffer from comorbid conditions. For example, in assessing individuals with CP, associated difficulties—such as learning difficulties, attention-deficit/hyperactivity disorder, intellectual disability, and sensory impairment—warrant consideration. It is these problems that put students at the greatest disadvantage when contrasted with their peers (Pellegrino, 1997). In the case of students with chronic disease or those with PI, distinguishing gaps in achievement resulting from the illness or impairment from indirect effects, such as those caused by school absence, pain or fatigue, depression, and low self-efficacy, is imperative (Donnelly, 2005). Comorbidity is also a possible source of construct-irrelevant variance,[1] which should be considered in interpreting the results of assessments conducted on individuals with PI.

Special Considerations for Assessment: Test Factors

Historically, accommodations for PI have been considered less salient than those for sensory disabilities, such as VI/B or D/HI (Lezak, Howieson, & Loring, 2004). In the past, clinical judgment and prior experience alone were considered acceptable standards for modifying test administration (Pratt & Moreland, 1998). Currently, many assessments suggest accommodations in the administration or technical manuals, in an attempt to retain standardization even in work with special populations (see Table 24.1). Common test accommodations include modifications in presentation, response format, or timing, as well as in test selection (AERA et al., 2014). Test setting and environment can also be modified for accessibility and physical comfort due to problems associated with a disability (AERA et al., 2014; Pratt & Moreland, 1998). A few test manuals provide normative information for clinical populations such as those with PI (see Table 24.1), but those that do are typically based on small samples that are not necessarily representative of the greater population with PI.

Motor functioning is a primary consideration for those assessing individuals with PI to bear in mind (Hill-Briggs et al., 2007), as motor functioning dictates the use of various response options. Specifically, impairment or loss of motor function in the upper extremities is of utmost concern, as it can affect fine motor control and/or dexterity. Impairments in this area will significantly reduce response format options because such individuals will not be able to complete tasks such as those requiring the manipulation of stimuli, or holding or utilizing a writing utensil (e.g., writing, drawing, or copying a figure). If standardized test administration requires such motor skills, then test accommodations for these individuals are necessary (Nester, 1994). The only exception would be when the goal of testing is to determine the extent of limitation caused by the PI.

Awareness of the individual's level of discomfort is another important variable for consideration and may require continual evaluation throughout the duration of testing. Individuals with PI may require repositioning or additional breaks to combat discomfort or fatigue, and some individuals (e.g., those with CP or MD) may additionally experience pain or cramping. Studies conducted on secondary pain (often the result of treatment) in adults with CP illustrated the immense impact of pain. From 67 to 84% of participants across three studies reported chronic pain (Engel, Jensen, Hoffman, & Kartin, 2003; Schwartz, Engel, & Jensen, 1999; Turk, Geremski, Rosenbaum, & Weber, 1997), and 18–56% reported daily incidents of pain (Andersson & Mattsson, 2001; Engel et al., 2003; Schwartz et al., 1999). Similarly, Zebracki and Drotar (2008) found that 54% of those with Duchenne MD and 80% of those with Becker MD reported experiencing pain associated with their disease. Also, according to the Child Activity Limitations Interview (CALI; Palermo, Witherspoon, Valenzuela, & Drotar, 2004), sitting was endorsed most frequently as causing discomfort in daily life, further justifying the need to monitor discomfort during an assessment (Zebracki & Drotar, 2008). Accessibility is also a concern. For example, if an individual uses a wheelchair, the testing location needs to be wheelchair-accessible.

Instruments for Assessing Individuals with PI

Since specialized cognitive tests do not exist for persons with PI in the same way that they do for individuals with VI/B or D/HI, great care must be taken to accommodate varying degrees of impairment within the framework of existing measures such as the WISC-V, WJ IV COG, or SB5. Many popular tests like these provide recommended test accommodations/modifications and norming information (see Table 24.1). Although the norms exist, they are unlikely to be representative; resultant scores should therefore be interpreted with caution.

Within SCI research, certain measures have been noted as being particularly useful in cases of paraplegia and quadriplegia. Others should be avoided; for example, the Trail Making Tests would not be an ideal selection for someone with a fine motor impairment because of construct-irrelevant variance. The following are some of the most commonly used measures within a motor-free neuropsychological assessment battery: the Wechsler Memory Scale (minus visual reproduction); the verbal subtests of the WAIS; the Symbol Digit Modalities Test; the Stroop Test; the Rey Auditory Verbal Learning Test; the Hooper Visual Organization Test; the Halstead Category Test; and the California Oral Word Association Test (Davidoff, Roth, & Richards, 1992; Dowler et al., 1997; Richards, Bown, Hagglund, Bua, & Reeder, 1988; Roth et al., 1989). Using various tests in this way is reminiscent of the XBA cognitive assessment approach (Flanagan et al., 2013), in which subtests from a variety of batteries can be combined in order to provide a comprehensive assessment of an individual's cognitive abilities. A mixed approach, like XBA, can provide the opportunity to assess individuals with PI more fully. Using such an approach could allow for subtest substitutions, which could assess an individual's skills in a different way, independent of his or her PI.

TRAUMATIC BRAIN INJURY

Approximately 1.7 million individuals suffer from a TBI every year, and children and youth between the ages of 0–4 and 15–19 are in the highest risk categories for sustaining a TBI (Brain Injury Association of America, 2012). Approximately 24% of children with severe head injuries do not survive (White et al., 2001). Comprehensive evaluations that include neuropsychological assessment are critical in assisting neurologists, determining the degree and severity of functional impairment resulting from the injury, and documenting changes in impairment over time (American Academy of Neurology, 1996). Schools and educational institutions are obligated under federal law to identify and provide special educational assistance to children with TBI. TBI is specified in IDEA 2004 as one of 13 categories under which students may receive special education services. The IDEA 2004 regulations define TBI as

> an acquired injury to the brain caused by an external physical force, resulting in total or partial functional disability or psychosocial impairment, or both, that adversely affects a child's educational performance. The term applies to open or closed head injuries resulting in impairments in one or more areas, such as cognition; language; memory; attention; reasoning; abstract thinking; judgment; problem-solving; sensory, perceptual, and motor abilities; psychosocial behavior; physical functions; information processing; and speech. [The term] does not apply to brain injuries that are congenital or degenerative, or to brain injuries induced by birth trauma. (U.S. Department of Education, 2006, p. 46757)

According to the U.S. Department of Education (n.d.), approximately 1,106 children ages 3–5 and an additional 25,419 children and youth ages 6–21 in the United States were served under this category during the 2014–2015 academic year. Although the federal definition explicitly excludes children whose brains may be affected by congenital problems or injury at birth, these children may still be eligible for special education under Section 504 of the Americans with Disabilities Act.

TBI may result from penetrating or nonpenetrating wounds. Impairment may range from none to severe, and injuries to the brain may result from direct damage to the brain or from indirect damage caused by brain swelling or bruising. Because falls and motor vehicle accidents are the most common causes of TBI (Faul, Xu, Wald, & Coronado, 2010) and usually result in diffuse injury, the deficits and symptoms associated with TBI vary widely. Early research on the developmental sequelae of brain injuries in childhood suggested that most children fully recover from the injury by adulthood, due to brain plasticity. However, more contemporary research has suggested that TBI sustained during childhood—even a mild brain injury—may in fact have lasting deficits in adulthood that include not only cognitive but psychiatric problems (Konrad et al., 2011; Pirozzolo & Papanicolaou, 1986), as well as lower quality of life (Anderson, Brown, Newitt, & Hoile, 2011).

The specific location of injury also influences long-term outcomes. Counterintuitively, right-hemispheric injuries have been shown to result in more long-term impairments than left-hemispheric injuries, even though the left hemisphere is typically the dominant hemisphere and highly involved in language functions. However, the right hemisphere seems to deal with new learning experiences (novelty) more proficiently than the left hemisphere—and, for children, almost all experiences are new learning experiences.

Neuropsychological Assessment

No specific cognitive profile or cluster of symptoms has been considered diagnostic of TBI; this is not surprising, considering that injuries may be focal (one specific area) or diffuse (general). Focal neurological damage results in more specific behavioral dysfunction than does diffuse damage (Berninger & Richards, 2002; Halstead, 1947; Kolb & Whishaw, 2003; Luria, 1980). Clinical hypothesis testing within an individualized assessment plan is important, to match the clinical symptoms and measured neuropsychological deficits to the type and location of the injury. However, two common cognitive deficits of children with TBI are slowed processing speed and attention problems (Rutter, 1981). Recent literature has identified these deficits by using a variety of measures and batteries (Allen, Thaler, Donohue, & Mayfield, 2010; Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Frencham, Fox, & Maybery, 2005; Mathias & Wheaton, 2007). Cognitive deficits from TBI are direct causes of academic and educational problems, which require specialized attention from educational professionals (Blosser & DePompei, 1991; D'Amato & Rothlisberg, 1996).

Although the connection between TBI and cognitive deficits is well known, the neurological mechanisms have only recently begun to be understood, primarily due to advances in brain imaging technology. Thatcher, Walker, Gerson, and Geisler (1989) used quantitative electroencephalography (QEEG) to compare 264 patients with mild head injuries to a control group, and found differences in the frontal and fronto-temporal regions of the brain. Results from this study suggested that QEEG measures of brain activity could distinguish patients with TBI from controls with 93% accuracy, which clearly indicates that TBI causes changes in brain activity. Using functional magnetic resonance imaging, Kraus and colleagues (2007) found that TBI resulted in white matter disruption, and that the degree of disruption was correlated with neuropsychological impairment, particularly with measures of executive functions and attention. White matter disruptions were found in children with TBI who had persistent motor control deficits. Compared to controls, these children showed greater brain activation in the parietal and cerebellar areas, which resulted from greater attentional resources needed for motor action (Caeyenberghs, Wenderoth, Smits-Engelsman, Sunaert, & Swinnen, 2009). Finally, a study by Wilde and colleagues (2011) investigated working memory tasks in children with TBI. Compared to an orthopedic control group, children with TBI displayed disrupted attentional and working memory functioning in fronto-parietal regions. Additionally, reduced white matter integrity in the cingulum bundle and frontal lobes was associated with slower reaction times. Together, these studies suggest that white matter disruption caused by TBI is a core underlying cause of both cognitive and motor deficits in children with TBI. Cognitive deficits including slowed processing speed are directly linked to white matter disruption, which interferes with automatized behaviors. Compensation with effortful behaviors ultimately results in slowed performance, detected as slowed processing speed and attentional problems on standardized tests.

Neuropsychological consultation has been recommended as a core educational service to be provided by school psychologists (D'Amato, 1990; Decker, 2008; Hale & Fiorello, 2004; Miller, 2007). School psychologists who serve children with TBI, but who lack neuropsychological training, may consider referring the children to or directly consulting with a neuropsychologist. Neuropsychological consultation involves not only direct service delivery, but also an understanding of recovery patterns, rehabilitation and effective interventions, and school reentry assistance. Critically, families commonly experience an often overlooked emotional reaction to loss of a "normal" child. Consultation with families is necessary to help set appropriate expectations for academic and behavioral performance.

Special Considerations for Assessment: Child and Test Factors

One primary function of the school psychologist is to conduct a comprehensive, broad-based assessment of a child with TBI, to ensure that a valid picture of the child can be communicated to parents and teachers and incorporated into intervention (Havey, 2002). Test accommodations are most needed for children with TBI when sensory impairments or PI are evident. When such impairments are not evident, most measures frequently used by school psychologists are appropriate for assessing the underlying impairment that has resulted from the TBI. However, it is important for clinicians to be aware of how core deficits caused by a TBI may inadvertently affect a variety of test constructs, thereby creating construct-irrelevant variance. For example, attention problems and slowed processing speed may have an impact on any variety of cognitive measures, not just those subtests explicitly targeting attention and processing speed. To minimize construct-irrelevant factors, it may be necessary to administer directions at a slower pace to children with TBI to ensure comprehension. In addition, children with TBI may need examiners to repeat directions to ensure proper encoding. When engaging in longer tasks, children with TBI may require more frequent prompting to prevent lapses in attention. More frequent breaks may be necessary, too, as children with TBI may become fatigued more quickly than other children of similar developmental age without a TBI. Most adaptations needed for assessment of children with mild TBI should not seriously jeopardize test validity.

As previously mentioned, age is an important variable for several reasons. First, age is a risk factor in two ways: Younger children (ages 0–4) are at higher risk for TBI due to falls, and older adolescents (ages 15–20) are at higher risk for TBI due to automobile accidents (Faul et al., 2010). Age also moderates the sequelae of behavioral impairment: TBI in younger children may interfere with the later development of skills, whereas TBI in adolescents may result in the loss of learned skills.

Estimating preinjury functional status is important to determine the degree to which cognition has changed as a result of the injury. Premorbid intelligence is typically estimated by a qualitative review of historical records (past educational achievement) and quantitative performance on cognitive measures typically resistant to neurological injury. Measures of crystallized intelligence (e.g., vocabulary) are used to measure abilities most resistant to neurological injury (Königs, Engenhorst, & Oosterlaan, 2016; Lezak et al., 2004; Russell, 1980).

Instruments for Assessing Individuals with TBI

As with standard school psychology assessment procedures, no one test or test battery can comprehensively assess the emotional, cognitive, and behavioral status of a child with TBI. Thus various measures from different domains are required. The Glasgow Coma Scale is widely used in making the initial determination of injury severity (mild, moderate, or severe). This is a short rating scale of eye, verbal, and motor responses that is used at or near the time of injury. Scores range from 3 to 15, with 3 representing no eye, verbal, or motor response and 15 representing a fully awake and functioning person. Similarly, a structured interview specifically tailored to important factors related to neurological injury is recommended (Dean, Woodcock, Decker, & Schrank, 2003). Sensory–motor measures are useful adjuncts to intellectual assessment, to rule out basic input–output deficits as a source of impairment on cognitive measures and to validate clinical hypotheses of cortical injury. Residual sensory–motor impairment from a brain injury has been demonstrated in children several weeks after release from the hospital (Gagnon, Forget, Sullivan, & Friedman, 1998).

Telzrow (1991) recommends assessing a variety of domains of intellectual functioning in children with TBI. Although a variety of assessment models may work for this purpose (Hale & Fiorello, 2004; Miller, 2007), a cross-battery model (e.g., the XBA approach) provides ample coverage to assess comprehensively different domains of cognitive functioning (see Flanagan et al., Chapter 27, this volume). A cross-battery model not only provides a close correspondence between cognitive functions of interest to neuropsychologists, but also enjoys wide familiarity among school psychologists (Decker, 2008). The cognitive domains incorporated in the XBA approach (Flanagan et al., 2013) include processing speed, a crucial area of assessment for children with TBI. Attention is not an explicit cognitive domain within the model, and so clinicians should be familiar with test-specific attentional demands and how attention may influence underlying test performance.

SUMMARY

This chapter has discussed issues related to sensory and physical disabilities as well as TBI within a neuropsychological framework. A neuropsychological framework is necessary because the causes of VI/B, D/HI, PI, and TBI all have some basis in the physiological disruption of typical sensory, motor, and cognitive pathways in the brain and body. In a school setting, individuals with all of these impairments may be eligible for services under IDEA 2004. We have reviewed the ways in which individual child factors—such as age at onset, nature and severity of the disability, etiology, comorbidity, primary modes of communication, and reading and language ability—can affect test selection, administration, and interpretation. We have also offered examples of the ways in which individual differences within each disability category may manifest themselves during academic and testing situations. Some important considerations noted for all disability categories include attention to within-category heterogeneity; the child's communicative style and linguistic skill level; the fit between educational environment and child; and the necessity of collaborating with both families and a multidisciplinary team to achieve systematic data-based decision making on a case-by-case basis.

This chapter has highlighted special considerations required for the assessment and evaluation of children with neurological conditions involving sensory or cortical impairment. Specifically, we have reviewed the availability of cognitive and neuropsychological test batteries with special standardized procedures for test accommodations, appropriate subtests, and normative information in special populations. Clinical judgment is required for modifying standardized tests and interpreting test results in light of a child's particular sensory or motor deficits. With careful attention to both child and test factors, and the help of a multidisciplinary team, many currently available cognitive and neuropsychological test batteries can be used effectively for the assessment of children with VI/B, D/HI, PI, and TBI. Cautious interpretation of the results, along with multisource information from systematic observations and from professional colleagues and specialists, can be integrated to provide a sound psychoeducational plan for every child, regardless of sensory or physical disability.

NOTE

1. The term *construct-irrelevant variance* is used to describe instances in which variability in responses is affected by something unrelated to the interpreted construct (Messick, 1995). This can occur as a result of changes in methodology, and as such it is especially important for practitioners to keep in mind when testing individuals who require modifications to standardized procedures.

REFERENCES

Accardo, P. J. (2008). *Cupute and Accardo's neurodevelopmental disabilities in infancy and childhood* (Vols. 1–2). Baltimore: Brookes.

Allen, D. N., Thaler, N. S., Donohue, B., & Mayfield, J. (2010). WISC-IV profiles in children with traumatic brain injury: Similarities to and differences from the WISC-III. *Psychological Assessment, 22*(1), 57–64.

American Academy of Neurology. (1996). Assessment: Neuropsychological testing in adults: Considerations for neurologists. *Neurology, 47,* 592–599.

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Speech–Language–Hearing Association (ASHA). (2010). Type, degree, and configuration of hearing loss. Retrieved from *www.asha.org/public/hearing/disorders/types.htm.*

Anderson, D. (2006). Lexical development of deaf children acquiring signed languages. In B. Schick, M. Marschark, & P. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 135–160). New York: Oxford University Press.

Anderson, V., Brown, S., Newitt, H., & Hoile, H. (2011). Long-term outcome from childhood traumatic brain injury: Intellectual ability, personality, and quality of life. *Neuropsychology, 25*(2), 176.

Andersson, C., & Mattsson, F. (2001). Adults with cerebral palsy: A survey describing problems, needs, and resources with special emphasis on locomotion. *Developmental Medicine and Child Neurology, 43*(2), 76–82.

Arneson, C. L., Durkin, M. S., Benedict, R. E., Kirby, R. S., Yeargin-Allsopp, M., Van Naarden Braun, K., et al. (2009). Prevalence of cerebral palsy: Autism and Developmental Disabilities monitoring network, three sites, United States, 2004. *Disability and Health Journal, 2*(1), 45–48.

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215–227.

Berninger, V. W., & Richards, T. L. (2002). *Brain literacy for educators and psychologists*. San Diego, CA: Academic Press.

Bhasin, T. K., Brocksen, S., Avchen, R. N., & Van Naarden Braun, K. (2006). Prevalence of four developmental disabilities among children aged 8 years—Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1996 and 2000. *Morbidity and Mortality Weekly Report Surveillance Summaries, 55*(1), 1–9.

Blosser, J. L., & DePompei, R. (1991). Preparing education professionals for meeting the needs of students with traumatic brain injury. *Journal of Head Trauma Rehabilitation* 6(1), 73–82.

Bosworth, R. G., & Dobkins, K. R. (1999). Left-hemisphere dominance for motion processing in deaf signers. *Psychological Science, 10*(3), 256–262.

Braden, J. P. (1994). *Deafness, deprivation, and IQ*. New York: Plenum Press.

Braden, J. P. (2001). The clinical assessment of deaf people's cognitive abilities. In M. D. Clark, M. Marschark, & M. Karchmer (Eds.), *Context, cognition, and deafness* (pp. 14–37). Washington, DC: Gallaudet University Press.

Bradley-Johnson, S., & Morgan, S. K. (2008). Best practices for instructing students who are visually impaired or blind. In A. Thomas & T. Grimes (Eds.), *Best practices in school psychology V* (pp. 1833–1846). Bethesda, MD: National Association of School Psychologists.

Brain Injury Association of America. (2012). Facts about traumatic brain injury. Retrieved from *www.biausa.org/glossary.htm*.

Brambring, M., & Tröster, H. (1994). The assessment of cognitive development in blind infants and preschoolers. *Journal of Visual Impairment and Blindness, 88*(1), 9–18.

Caeyenberghs, K., Wenderoth, N., Smits-Engelsman, B. C. M., Sunaert, S., & Swinnen, S. P. (2009). Neural correlates of motor dysfunction in children with traumatic brain injury: Exploration of compensatory recruitment patterns. *Brain, 132*, 684–694.

Cans, C. (2000). Surveillance of cerebral palsy in Europe: A collaboration of cerebral palsy surveys and registers. *Developmental Medicine and Child Neurology, 42*(12), 816–824.

Cattani, A., & Clibbens, J. (2005). Atypical lateralization of memory for location: Effects of deafness and sign language use. *Brain and Cognition, 58*(2), 226–239.

Chan, F., Lynch, R. T., Dial, J. G., Wong, D. W., & Kates, D. (1993). Applications of the McCarron–Dial System in vocational evaluation: An overview of its operational framework and empirical findings. *Vocational Evaluation and Work Adjustment Bulletin, 26*(2), 57–65.

Charlier, B. L., & Leybaert, J. (2000). The rhyming skills of deaf children educated with phonetically augmented speechreading. *Quarterly Journal of Experimental Psychology, 53a*, 349–375.

Dalby, D. M., Hirdes, J. P., Stolee, P., Strong, J. G., Poss, J., Tjam, E. Y., et al. (2009). Characteristics of individuals with congenital and acquired deaf-blindness. *Journal of Visual Impairment and Blindness, 103*(2), 93.

D'Amato, R. C. (1990). A neuropsychological approach to school psychology. *School Psychology Quarterly, 5*, 141–160.

D'Amato, R. C., & Rothlisberg, B. A. (1996). How education should respond to students with traumatic brain injury. *Journal of Learning Disabilities, 29*, 670–683.

Davidoff, G. N., Roth, E. J., & Richards, J. S. (1992). Cognitive deficits in spinal cord injury: Epidemiology and outcome. *Archives of Physical Medicine and Rehabilitation, 73*, 275–284.

Day, L. A., Adams Costa, E. B., & Raiford, S. E. (2015). WISC-V technical report #2: Testing children who are deaf or hard of hearing. Retrieved from *http://downloads.pearsonclinical.com/images/Assets/WISC-V/WISC-V-TechReport_2_FNL.pdf*.

Dean, R. S., Woodcock, R. W., Decker, S. L., & Schrank, F. A. (2003). A cognitive neuropsychological assessment system. In F. L. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation* (pp. 345–375). Amsterdam: Elsevier.

Decker, S. L. (2002). Confirmatory models of sensory/motor and cognitive constructs. *Dissertation Abstracts International, 63*, 1083B.

Decker, S. L. (2004). Incremental validity of tactile measures in predicting letter–word identification skills. *Archives of Clinical Neuropsychology, 19*(7), 907–908.

Decker, S. L. (2008). School neuropsychology consultation in neurodevelopmental disorders. *Psychology in the Schools, 45*(9), 799–811.

Dial, J. G., Chan, F., Mezger, C., Parker, H. J., Zangla, K., Wong, D. W., et al. (1991). Comprehensive Vocational Education System for visually impaired and blind persons. *Journal of Visual Impairment and Blindness, 85*, 153–157.

Dial, J. G., Chan, F., Tunick, R., Gray, S. G., & Marme, M. (1991). Neuropsychological evaluation: A functional and behavioral approach. In B. T. McMahon & L. R. Shaw (Eds.), *Work worth doing: Advances in brain injury rehabilitation* (pp. 47–76). Orlando, FL: Deutsch.

Donnelly, J. P. (2005). Providing neuropsychological services to learners with chronic illnesses. In R. C. D'Amato, E. Fletcher-Jensen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 511–532). Hoboken, NJ: Wiley.

Dowler, R. N., Harrington, D. L., Haaland, K. Y., Swanda, R. M., Fee, F., & Fiedler, K. (1997). Profiles of cognitive functioning in chronic spinal cord injury and the role of moderating variables. *Journal of the International Neuropsychological Society, 3,* 464–472.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Engel, J. M., Jensen, M. P., Hoffman, A. J., & Kartin, D. (2003). Pain in persons with cerebral palsy: Extension and cross validation. *Archives of Physical Medicine and Rehabilitation, 84,* 1125–1128.

Faul, M., Xu, L., Wald, M. M., & Coronado, V. G. (2010). *Traumatic brain injury in the United States.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Finger, S. (1994). *Origins of neuroscience: A history of explorations into brain functions.* New York: Oxford University Press.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Fleischer, D. Z., Zames, F. D., & Zames, F. (2012). *The disability rights movement: From charity to confrontation.* Philadelphia: Temple University Press.

Frencham, K. A., Fox, A. M., & Maybery, M. T. (2005). Neuropsychological studies of mildtraumatic brain injury: A meta-analytic review of research since 1995. *Journal of Clinical and Experimental Neuropsychology, 27*(3), 334–351.

Gagnon, I., Forget, R., Sullivan, S. J., & Friedman, D. (1998). Motor performances following a mild traumatic brain injury in children: An exploratory study. *Brain Injury, 12,* 843–853.

Gallaudet Research Institute. (2007). *State summary report of data from the 2006–2007 Annual Survey of Deaf or Hard-of-Hearing Children and Youth.* Washington, DC: Gallaudet University. Retrieved from *http://research.gallaudet.edu/Demographics.*

Goodman, S. A., Evans, C., & Loftin, M. (2011). Full position paper: Intelligence testing of individuals who are blind or visually impaired. Retrieved from *www.aph.org/accessible-tests/position-papers/intelligence-testing/full.*

Gutterman, J., Ward, M., & Genshaft, J. (1985). Correlations of scores of low vision children on the Perkins–Binet tests of intelligence for the blind, the WISC-R, and the WRAT. *Journal of Visual Impairment and Blindness, 79*(2), 55–58.

Haapala, E. A. (2013). Cardiorespiratory fitness and motor skills in relation to cognition and academic performance in children: A review. *Journal of Human Kinetics, 36*(1), 55–68.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Halstead, W. C. (1947). *Brain and intelligence: A quantitative study of the frontal lobes.* Chicago: University of Chicago Press.

Havey, J. M. (2002). Working with students with traumatic brain injury. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1433–1446). Bethesda, MD: National Association of School Psychology.

Hill-Briggs, F., Dial, J., Morere, D., & Joyce, A. (2007). Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness. *Archives of Clinical Neuropsychology, 22*(3), 389–404.

Holt, J. A. (2005). Stanford Achievement Test—8th Edition for deaf and hard of hearing students: Reading comprehension subgroup results. *American Annals of the Deaf, 138*(2), 172–175.

Humes, L. E., Busey, T. A., Craig, J., & Kewley-Port, D. (2013). Are age-related changes in cognitive function driven by age-related changes in sensory processing? *Attention, Perception, and Psychophysics, 75*(3), 508–524

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No., 108-446, 118 Stat. 2647 (2004).

Joyce, A., Isom, R., Dial, J. G., & Sandel, M. G. (2004). Implications of perceptual–motor differences within blind populations. *Journal of Applied Rehabilitation Counseling, 35*(3), 3–7.

Kaskel, L. M., Dial, J. G., Chan, F., & Roldan, G. (1991). Evaluating work potential of persons with visual impairments: The Comprehensive Vocational Evaluation System (CVES) approach. In R. Fry (Ed.), *The issue papers: Fifth national forum on issues in vocational assessment* (pp. 167–173). Menomonie: University of Wisconsin–Stout, Materials Development Center.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Kim, H., Carlson, A. G., Curby, T. W., & Winsler, A. (2016). Relations among motor, social, and cognitive skills in pre-kindergarten children with developmental disabilities. *Research in Developmental Disabilities, 53,* 43–60.

Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of human neuropsychology.* New York: Worth.

Königs, M., Engenhorst, P. J., & Oosterlaan, J. (2016). Intelligence after traumatic brain injury: Meta-analysis of outcomes and prognosis. *European Journal of Neurology, 23*(1), 21–29.

Konrad, C., Geburek, A. J., Rist, F., Blumenroth, H., Fischer, B., Husstedt, I., et al. (2011). Long-term cognitive and emotional consequences of mild traumatic brain injury. *Psychological Medicine, 41*(6), 1197–1211.

Kraus, M. F., Susmaras, T., Caughlin, B. P., Walker, C. J., Sweeney, J. A., & Little, D. M. (2007). White matter integrity and cognition in chronic traumatic brain injury: A diffusion tensor imaging study. *Brain, 130*(10), 2508–2519.

Krishnan, R. V. (2006). Relearning toward motor recovery in stroke, spinal cord injury, and cerebral palsy: A cognitive neural systems perspective. *International Journal of Neuroscience, 116,* 127–140.

LaSasso, C., Crain, K., & Leybaert, J. (2003). Rhyme generation in deaf students: The effect of exposure to cued speech. *Journal of Deaf Studies and Deaf Education, 8,* 250–270.

Leigh, I. W., Marcus, A. L., Dobosh, P. K., & Allen, T. E. (1998). Deaf/hearing cultural identity paradigms: Modification of the Deaf Identity Development Scale. *Journal of Deaf Studies and Deaf Education, 3*(4), 329–338.

Lezak, M., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Li, L., & Moore, D. (1998). Acceptance of disability and its correlates. *Journal of Social Psychology, 138*(1), 13–25.

Lukomski, J. (2008). Best practices in planning effective instruction for children who are deaf/hard of hearing. In A. Thomas & T. Grimes (Eds.), *Best practices in school psychology V* (pp. 1819–1832). Bethesda, MD: National Association of School Psychologists.

Lund, E. M., Miller, K. B., & Ganz, J. B. (2014). Access to assessment?: Legal and practical issues regarding psychoeducational assessment in children with sensory disabilities. *Journal of Disability Policy Studies, 25*(3), 135–145.

Luria, A. R. (1980). *Higher cortical functions in man.* New York: Basic Books.

MacCluskie, K., Tunick, R., Dial, J., & Paul, D. (1998). The role of vision in the development of abstraction. *Journal of Visual Impairment and Blindness, 92*(3), 189.

Maller, S. J. (2000). Item invariance in four subtests of the Universal Test of Nonverbal Intelligence (UNIT) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18,* 240–254.

Maller, S. J. (2003). Intellectual assessment of deaf people: A critical review of core concepts and issues. In M. Marschark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (pp. 451–463). New York: Oxford University Press.

Marschark, M., & Clark, M. D. (Eds.). (1993). *Psychological perspectives on deafness.* Hillsdale, NJ: Erlbaum.

Mather, N., & Wendling, B. J. (2014). *Examiner's manual: Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Mathias, J. L., & Wheaton, P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: A meta-analytic review. *Neuropsychology, 21*(2), 212.

McGee, J. M. (1994). *Neuropsychological functioning of adult subjects with diabetic retinopathy compared to a normal blind population.* Unpublished doctoral dissertation, University of North Texas, Denton.

Meadow, K. P. (1980). *Deafness and child development.* Berkeley: University of California Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment.* Hoboken, NJ: Wiley.

Miller, K. R., & Vernon, M. (2001). Linguistic diversity in deaf defendants and due process rights. *Journal of Deaf Studies and Deaf Education, 6*(3), 226–234.

Mitchell, R. E. (2006). How many deaf people are there in the United States?: Estimates from the survey of income and program participation. *Journal of Deaf Studies and Deaf Education, 11,* 112–119.

Nelson, P. A., Dial, J. G., & Joyce, A. (2002). Validation of the Cognitive Test for the Blind as an assessment of intellectual functioning. *Rehabilitation Psychology, 47*(2), 184–193.

Nelson, P. A., O'Brien, E. P., Dial, J. G., & Joyce, A. (2001). Neuropsychological correlates of retinopathy of prematurity (ROP) compares to other causes of blindness [Abstract]. *Archives of Clinical Neuropsychology, 16,* 766–767.

Nester, M. A. (1994). Psychometric testing and reasonable accommodation for persons with disabilities. In S. M. Bruyere & J. O'Keefe (Eds.), *Implications of the Americans with Disabilities Act for psychology* (pp. 25–36). New York: Springer/American Psychological Association.

Nevala, N., Pehkonen, I., Koskela, I., Ruusuvuori, J., & Anttila, H. (2015). Workplace accommodation among persons with disabilities: A systematic review of its effectiveness and barriers or facilitators. *Journal of Occupational Rehabilitation, 25*(2), 432–448.

O'Keefe, J. (1994). Disability, discrimination, and the Americans with Disabilities Act. In S. M. Bruyere & J. O'Keefe (Eds.), *Implications of the Americans with Disabilities Act for psychology* (pp. 1–14). New York: Springer/American Psychological Association.

Palermo, T. M., Witherspoon, D., Valenzuela, D., & Drotar, D. (2004). Development and validation of the Child Activity Limitations Interview: A measure of pain-related functional impairment in school-age children and adolescents. *Pain, 109,* 461–464.

Pellegrino, L. (1997). Cerebral palsy. In M. Batshaw (Ed.), *Children with disabilities* (4th ed., pp. 499–528). Baltimore: Brookes.

Pirozzolo, F. J., & Papanicolaou, A. C. (1986). Plasticity and recovery of function in the central nervous system. In J. E. Obrzut & G. W. Hynd (Eds.), *Child* *neuropsychology: Vol. 1. Theory and research* (pp. 141–154). Orlando, FL: Academic Press.

Polat, F. (2003). Factors affecting psychosocial adjustment of deaf students. *Journal of Deaf Studies and Deaf Education, 8*(3), 325–339.

Powell-Smith, K. A., Stoner, G., Bilter, K. J., & Sansosti, F. J. (2008). Best practices in supporting the education of students with severe and low-incidence disabilities. In A. Thomas & T. Grimes (Eds.), *Best practices in school psychology V* (pp. 1233–1248). Bethesda, MD: National Association of School Psychologists.

Pratt, S. I., & Moreland, K. L. (1998). Individuals with other characteristics. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity* (pp. 349–371). Washington, DC: American Psychological Association.

Reesman, J. H., Day, L. A., Szymanski, C. A., Hughes-Wheatland, R., Witkin, G. A., Kalback, S. R., et al. (2014). Review of intellectual assessment measures for children who are deaf or hard of hearing. *Rehabilitation Psychology, 59*(1), 99.

Rich, C. C., & Anderson, R. P. (1965). A tactual form of the Progressive Matrices for use with blind children. *Personnel and Guidance Journal, 43*(9), 912–919.

Richards, J. S., Brown, L., Hagglund, K., Bua, G., & Reeder, K. (1988). Spinal cord injury and concomitant traumatic brain injury: Results of a longitudinal investigation. *American Journal of Physical Medicine and Rehabilitation, 67,* 211–216.

Roberts, R. D., Pallier, G., & Goff, G. N. (1999). Sensory processes within the structure of human cognitive abilities. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 339–370). Washington, DC: American Psychological Association.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside.

Roth, E., Davidoff, G., Thomas, P., Doljanac, R., Dijkers, M., Berent, S., et al. (1989). A controlled study of neuropsychological deficits in acute spinal cord injury patients. *Spinal Cord, 27*(6), 480–489.

Russell, E. W. (1980). Fluid and crystallized intelligence: Effects of diffuse brain damage on the WAIS. *Perceptual and Motor Skills, 51,* 121–122.

Rutter, M. (1981). Psychological sequelae of brain damage in children. *American Journal of Psychiatry, 138*(12), 1533–1544.

Schwartz, L. S., Engel, J. M., & Jensen, M. P. (1999). Pain in persons with cerebral palsy. *Archives of Physical Medicine and Rehabilitation, 80,* 1243–1246.

Sherwell, S., Reid, S. M., Reddihough, D. S., Wrennall, J., Ong, B., & Stargatt, R. (2014). Measuring intellectual ability in children with cerebral palsy: Can we do better? *Research in Developmental Disabilities, 35*(10), 2558–2567.

Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: A meta-analysis. *Pediatric Exercise Science, 15*(3), 243–256.

Sigelman, C. K., & Rider, E. A. (2009). *Life-span human development* (6th ed.). Belmont, CA: Wadsworth Cengage Learning.

Stankov, L., Seizova-Cajic, T., & Roberts, R. D. (2001). Tactile and kinesthetic perceptual processes within the taxonomy of human cognitive abilities. *Intelligence, 29,* 1–29.

Stöckel, T., & Hughes, C. M. (2016). The relation between measures of cognitive and motor functioning in 5- to 6-year-old children. *Psychological Research, 80*(4), 543–554.

Szymanski, C. A., Brice, P. J., Lam, K. H., & Hotto, S. A. (2012). Deaf children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42,* 2027–2037.

Telzrow, C. F. (1991). The school psychologist's perspective on testing students with traumatic brain injury. *Journal of Head Trauma Rehabilitation, 6,* 23–34.

Texas School for the Blind and Visually Impaired. (2010). Resources. Retrieved from *www.tsbvi.edu/resources.*

Texas School for the Blind and Visually Impaired. (2016). Assessment of learning disabilities in students with visual impairment. Retrieved from *www.tsbvi.edu/instructional-resources/105-learning-disabilities-and-visual-impairments-workshop.*

Thatcher, R. W., Walker, R. A., Gerson, I., & Geisler, F. H. (1989). EEG discriminant analyses of mild head trauma. *Electroencephalography and Clinical Neurophysiology, 73,* 94–106.

Thomas, A., & Grimes, T. (Eds.). (2008). *Best practices in school psychology V*. Bethesda, MD: National Association of School Psychologists.

Tomporowski, P. D., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Psychology Review, 20*(2), 111–131.

Turk, M. A., Geremski, C. A., Rosenbaum, P. F., & Weber, R. J., (1997). The health status of women with cerebral palsy. *Archives of Physical Medicine and Rehabilitation, 78,* 810–817.

U.S. Census Bureau. (2010). Americans with disabilities: 2010. Retrieved from *www.census.gov/people/disability.*

U.S. Department of Education. (2006, August 14). 34 C.F.R. Parts 300 and 301: Assistance to states for the education of children with disabilities and preschool grants for children with disabilities: Final rule. *Federal Register, 71*(156), 46539–46845.

U.S. Department of Education. (n.d.). IDEA Section 618 data products: Static tables. Retrieved from *www2.ed.gov/programs/osepidea/618-data/static-tables/index.html#partc-cc.*

Verhoef, J. A., Miedema, H. S., Meeteren, J., Stam, H.

J., & Roebroeck, M. E. (2013). A new intervention to improve work participation of young adults with physical disabilities: A feasibility study. *Developmental Medicine and Child Neurology, 55*(8), 722–728.

Vernon, M. (2005). Fifty years of research on intelligence of deaf and hard-of-hearing children: A review of literature and discussion of implications. *Journal of Deaf Studies and Deaf Education, 10*(3), 225–231.

White, J. R., Farukhi, Z., Bull, C., Christensen, J., Gordon, T., Paidas, C., & Nichols, D. G. (2001). Predictors of outcome in severely head-injured children. *Critical Care Medicine, 29*(3), 534–540.

Wilde, E. A., Newsome, M. R., Bigler, E. D., Pertab, J., Merkley, T. L., Hanten, G., et al. (2011). Brain imaging correlates of verbal working memory in children following traumatic brain injury. *International Journal of Psychophysiology, 82*(1), 86–96.

Zebracki, K., & Drotar, D. (2008). Pain and activity in children with Duchenne or Becker muscular dystrophy. *Developmental Medicine and Child Neurology, 50*, 546–552.

CHAPTER 25

Testing with Culturally and Linguistically Diverse Populations

New Directions in Fairness and Validity

Samuel O. Ortiz
Nicole Piazza
Salvador Hector Ochoa
Agnieszka M. Dynda

As the racial and ethnic composition of the U.S. population continues to diversify at a rapid pace, practitioners are increasingly likely to find themselves evaluating individuals who come from culturally and linguistically diverse backgrounds. Among children of school age across the United States, ethnic minorities now constitute the majority group in various areas (e.g., New Jersey, Washington, D.C.; see Llorente & Sheingold, 2010; Smith, 2008), and the precollege grade-level population of Hispanics currently exceeds that of white non-Hispanics in states such as Arizona, California, and Nevada (U.S. Census Bureau, 2017). Reports from the U.S. Census Bureau's (2010, 2014) American Community Survey also indicate that one-fourth of children in U.S. public school kindergarten classes are Hispanic, and that ethnic minority children will become the majority by 2023. The time when practitioners might only encounter ethnic minority children on rare occasions has long since passed. The data make it clear that diversity within the population is increasing quickly and rapidly becoming a fact in everyday life and clinical practice.

In a survey of current practices in psychological assessment of students with limited English proficiency (LEP), Ochoa, Riccio, Jimenez, Garcia de Alba, and Sines (2004) found that practitioners engaged in the assessment of diverse individuals utilized a variety of different instruments. The most commonly used combination included an intelligence scale (most often a Wechsler scale administered completely in English), a test of visual–motor integration, an informal projective drawing test, and a nonverbal intelligence instrument. Interestingly, Ochoa and colleagues noted that this makeshift battery is not substantially different from the instruments employed in the evaluation of mainstream, monolingual English speakers. Indeed, even when a practitioner had professional bilingual capability, the battery remained largely the same; the only notable changes occurred in the language version or language of administration of the tests. This finding was replicated to some extent more recently by Sotelo-Dynega and Dixon (2014). Whether this practice is driven by staunch adherence to training or by mere inertia, the similarity in test usage and its continued ubiquity in current practice suggest that practitioners still largely ignore major factors that differ markedly in the testing of one population versus another. These factors include such things as inadequately representative norms and inappropriate comparison groups, as well as linguistic and cultural confounds that are manifested in unequal levels of age and grade-expected language and acculturative knowledge acquisition—and that occur primarily as a function of limited experience and exposure, not actual lack of ability.

Perhaps one of the reasons why testing practices with diverse populations have remained rather stagnant for so long is the fact that there are very few researchers whose theoretical or empirical work centers on this topic. Although it is not an area of study that is in any way limited to those who are themselves from diverse backgrounds, there remains a significant lack of knowledgeable and committed researchers interested enough to advance the field. When research is conducted along these lines, it is often carried out by researchers and academicians with only a passing interest in the subject, whose limited understanding of the many facets and variables involved tends to undermine the quality of their results and the conclusions they draw from them (cf., Kranzler, Flores, & Coady, 2010; Styck & Watkins, 2013, 2014). Another problem that stifles advances in this area has more to do with a general failure to look beyond the traditional reliability-based view of test bias; there has been little focus on the concept of validity, particularly construct validity. When notions of bias were first leveled at tests and testing applied to diverse populations, there was a concerted effort to investigate the claims and provide some sort of scientific analysis of the truth. At this early stage, it was believed that factors such as race and ethnicity posed threats to the inherent reliability of tests—threats that were manifested in supposed bias in regard to factor structure, difficulty level, or prediction. Yet such bias was rarely found (Figueroa, 1983; Reynolds & Ramsay, 2003; Sandoval, Frisby, Geisinger, Scheuneman, & Grenier, 1998). This is hardly surprising, especially when the issue is considered in the bluntest sense: That is, why should people's skin color dictate their level of performance on a test of intelligence or cognitive ability, any more than would their favorite flavor of ice cream or their first pet's name? For these reasons, many early researchers adopted the notion that the differences in performance must be genetically based. In other words, they concluded that if these differences exist (they do), and they are real and measurable (they are), and the only other difference between populations is skin color (this is where the error occurs because it isn't), then the mean differences must be rooted in some sort of genetically driven explanation (Gottfredson, 1997; Jensen, 1969).

These arguments have felt quite convincing for a long time, and with respect to the lack of *psychometric* bias as related to its typical definition, they are correct. What is wrong is that there has been a tendency to overlook the manner in which factors other than skin color or cultural heritage actually affect test performance. Indeed, it is better understood now, but not widely known, that simple racial or cultural differences are not what create the noted mean differences in performance; rather, it is the degree to which individuals' racial or cultural characteristics might have affected two particular developmental processes that are the foundations for measurement comparisons and believed to be inherently tied to physical maturation and age. These variables are language and cultural knowledge. The color of an individual's skin is not what matters. Rather, what matters is the degree to which any individual, regardless of race or cultural heritage, has had the same opportunity to learn and develop in the language of the test being administered and the same opportunity to learn, acquire, and automatize the cultural knowledge that is invariably built into the test's tasks (Ortiz, 2014; Rhodes, Ochoa, & Ortiz, 2005; Salvia & Ysseldyke, 1991; Valdés & Figueroa, 1994). These factors, however, do not in any real way directly affect the reliability or accuracy of measurements taken with psychometrically sound instruments. Rather, bias comes only as an issue of construct validity. More pointedly, it stems from interpretation of a measurement's meaning when it is based on comparisons among individuals who, despite being of the same age or grade, nevertheless possess differential rates of language and cultural knowledge acquisition. For example, a 10-year-old native English speaker raised in the U.S. mainstream by fully acculturated and native-English-speaking parents, in a home and community in which English is the cultural language, is not comparable to—and cannot be used as the standard for expectations of performance for—a 10-year-old whose parents speak limited English, who is raised in a non-mainstream culture, and who must learn to speak English independently or with limited experiential opportunities (such as at school).

The survey by Sotelo-Dynega and Dixon (2014) highlights another important point: Practitioners are not simply choosing the "wrong" test because the reality is that there is no "right" test, at least not yet. Rather, regardless of what test is selected, practitioners appear to administer and interpret it in much the same way with examinees who have LEP as they do with any other individuals, regardless of whether the test is given in English or the examinees' native language. Thus, although tests continue to need significant improvement, advances in practice will not stem only from better psychometrics. Rather, there is likely to be more value for practice in being able to assist in rendering valid interpretation of results. For example, it has been reported that 98% of school psychologists who evaluate culturally and linguistically diverse individuals stated that they consider both the "validity of test scores" and "an examinee's level of English language proficiency before selecting a test and interpreting test scores" (Sotelo-Dynega, Cuskley, Geddes, McSwiggan & Soldano, 2011). But when asked how they actually evaluate such individuals so that the relevant issues are properly considered, particularly the validity of the test scores, the vast majority simply reported using different types of tests; there was no mention of how any of them actually addressed test score validity. Approximately 88% reported using nonverbal tests, 40% reported use of an interpreter to administer a test, 20% reported use of a native-language test, and 2% reported "on-the-fly" translation of a test. Practitioners thus appear to be presuming that validity issues are resolved merely by decisions regarding test selection. In short, there seem to be few indications that practitioners are actually examining issues of validity directly or attempting to integrate current research regarding the manner in which cultural and linguistic factors are known to influence test performance. As a consequence, there is often no defensible or scientific basis for the types of interpretations, conclusions, and diagnoses stemming from the use of tests in such practice.

The first purpose of this chapter is to describe the current issues pertaining to the use of standardized, norm-referenced tests in attempts to evaluate the abilities (intellectual, cognitive, academic, and neuropsychological) of individuals from culturally and linguistically diverse backgrounds and provide direction for future test development. The second purpose is to discuss the advances that have been made in providing guidance for evidence-based practice in this area, as well as technical and psychometric innovations that can increase practitioners' ability to more directly ameliorate concerns over test score validity. Specific sections of this chapter are devoted to (1) evaluating current standards of practice; (2) evaluating the advantages and disadvantages of current practices; (3) integrating research into defensible practice; and (4) describing new directions in test development.

EVALUATING CONTEMPORARY STANDARDS OF PRACTICE

Changing demographics will continue to place pressure on graduate training programs and their trainees to acquire the requisite multicultural competency necessary for conducting fair and equitable evaluations of individuals from culturally and linguistically diverse backgrounds (Ortiz, 2014). It is unclear at present whether sufficient faculty and supervisors exist, or sufficient resources are available, to ensure that all graduate students will gain adequate experience in formulating appropriate knowledge and skills, but it seems safe to say that supervised experience with multicultural populations probably remains somewhat limited. When practitioners find themselves at a loss or in need of guidance on the matter, they are likely to refer to the usual sources of information. For example, over 20 years ago the American Psychological Association (1993) published *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations,* which emphasizes the need for psychologists to acknowledge the influences of language and culture on behavior, and to consider those factors when working with diverse groups. The guidelines also include admonitions regarding appraisal of the validity of the methods and procedures used for assessment and interpretation. But exactly what should or can be done when validity is deemed questionable, or what specific steps exist to reduce bias and maintain validity in the first place, are unfortunately absent from this rather dated set of guidelines.

The search for more definitive answers invariably leads practitioners to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014), known simply as "the *Standards*." Through revised editions with contributions from leading experts in the field, the authors of the *Standards* have made progress in providing more definitive guidance for practitioners faced with evaluating the knowledge and abilities of individuals from culturally and linguistically diverse backgrounds.

The 2014 edition provides a comprehensive chapter (Chapter 3) centered on elements of testing that support various facets of fairness, protection against measurement bias, and validity for testing culturally and linguistically diverse individuals. Specifically, Chapter 3 outlines fairness provisions to be considered in (1) test design, development, administration, and scoring procedures that minimize barriers to valid score interpretations for the widest possible range of individuals and relevant subgroups; (2) validity of test score interpretations for intended uses for the intended examinee population; (3) accommodations to remove construct-irrelevant barriers and support valid interpretations of scores for their intended uses; and (4) safeguards against inappropriate score interpretations for intended uses. The *Standards* volume also posits two distinct, novel concepts to be adopted by test makers and administers to minimize bias and promote fairness. The first is *accessibility* to all test takers—that is, allowing examinees "an unobstructed opportunity to demonstrate their standing on the construct(s) being measured." The second is *universal design,* or testing that ensures "clarity surrounding construct(s) to be measured," including the target, purpose, inferences to be made, and characteristics of examinees for a given test. However, while the new *Standards* volume is improved in organization and comprehension, it remains nondefinitive in providing concrete answers for practitioners related to the complex issues presented.

The opening description of purpose and background in Chapter 3 of the 2014 *Standards* provides a broad definition of *fairness*:

> A test that is fair within the meaning of the Standards reflects the same construct(s) for all test takers, and scores from it have the same meaning for all individuals in the intended populations; a fair test does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct . . . characteristics of all individuals in the intended test population, including those associated with race, ethnicity, gender, age, socioeconomic status, or linguistic or cultural background, must be considered throughout all stages of development, administration, scoring, interpretation, and use so that barriers to fair assessment can be reduced. (p. 50)

Chapter 3 also clearly identifies the measurement bias that is rampant in traditional methods of testing the ability of culturally and linguistically diverse individuals:

> Individuals who differ culturally and linguistically from the majority of the test takers are at risk for inaccurate score interpretations because of multiple factors associated with the assumption that, absent language proficiency issues, these individuals have developmental trajectories comparable to those of individuals who have been raised in an environment mediated by a single language and culture. (p. 53)

These themes serve as the common thread woven throughout the *Standards* in discussing the multifaceted issues surrounding testing the abilities of culturally and linguistically diverse individuals and provisions to address those issues.

Perhaps the most fundamental issue that cuts across measurement bias is that socially constructed categories, such as race and ethnicity, do little in achieving accurate representation along the relevant and important dimensions that actually affect test performance. Neither skin color nor an individual's ethnic heritage affects test performance directly. It is the individual's developmental background, particularly with respect to linguistic and acculturative experiences, that influences the manner in which the individual responds on standardized, norm-referenced tests. Salvia and Ysseldyke (1991) make this point clear in their statement regarding the *assumption of comparability*:

> When we test students using a standardized device and compare them to a set of norms to gain an index of their relative standing, we assume that the students we test are similar to those on whom the test was standardized; that is, we assume their acculturation is comparable, but not necessarily identical, to that of the students who made up the normative sample for the test. When a child's general background experiences differ from those of the children on whom a test was standardized, then the use of the norms of that test as an index for evaluating that child's current performance or for predicting future performances may be inappropriate. (p. 18)

Controlling for racial or ethnic differences via stratified random sampling may provide a desired measure of "face validity" for the norm sample, but in fact it only ensures that proportionate numbers of individuals from such backgrounds are included in the norm sample. It does not ensure that individuals from such backgrounds have comparable experiences, particularly as related to linguistic development and acculturative experiences. The 2014 *Standards* volume does clearly acknowledge this point: "In attempting to ensure fairness, we often generalize across groups of test takers . . . however, this is for convenience and not to imply that subgroups themselves are homogenous, or that, consequently, all members of a group should be treated similarly when making interpretations of test scores for individuals" (p. 53). But guidelines on *how* to treat culturally and linguistically diverse subgroups as individuals, and *how* to account for individual characteristics that may interact with different testing situations, are still lacking. It is unfortunate that current researchers continue to examine a multitude of issues related to test performance in studies that use groups formed exclusively on racial or ethnic categories (e.g., Hispanic). When constructed in such a manner, such studies lack generalizability and do not inform practice in any substantive way and are effectively flawed in methodology. As described by Lohman, Korb, and Lakin (2008),

> most studies compare the performance of students from different ethnic groups . . . rather than ELL [English-language learner] and non-ELL children within those ethnic groups. . . . A major difficulty with all of these studies is that the category Hispanic includes students from diverse cultural backgrounds with markedly different English-language skills. . . . This reinforces the need to separate the influences of ethnicity and ELL status on observed score differences. (pp. 276–278)

In a manner quite similar to differences in language skills, differences in test performance related to acquisition of cultural knowledge also preclude valid comparisons of performance. Validity can only be inferred when an examinee's background and experiences are comparable to those of the individuals who make up the norm sample. Whenever this assumption is not met, conclusions regarding the meaning of test results become dubious at best. This notion has also been described by Salvia and Ysseldyke:

> Incorrect educational decisions may well be made. It must be pointed out that acculturation is a matter of experiential background rather than of gender, skin color, race, or ethnic background. When we say that a child's acculturation differs from that of the group used as a norm, we are saying that the *experiential background* differs, not simply that the child is of different ethnic origin, for example, from the children on whom the test was standardized. (1991, p. 18; original emphasis)

It would seem that if they are to advance our scientific understanding, empirical investigations that make their way into academic journals in the future need to begin acknowledging this issue and incorporate participant groupings more representative of the true linguistic and cultural differences that play pivotal roles in how an individual performs on a test.

Perhaps a more utilitarian discussion for practitioners can be found in Chapter 3 of the *Standards,* which deals with fairness in testing and test use, as indicated above. In addition to the ubiquitous discussions regarding psychometric bias, the chapter provides an important definition and analysis of fairness as "opportunity to learn," as well as of content-, response-, and context-related sources of test bias. It is noted that "the extent to which individuals have had exposure to instruction or knowledge that affords them the opportunity to learn the content and skills targeted by the test has several implications for the fair and valid interpretation of test scores" (p. 56). Individuals who move between a home environment where the language and content emanate from their heritage culture, and a school environment where the language and content emanate from the majority, mainstream culture, are likely to fall under this definition of having limited opportunity to learn. Parents who do not speak English at all or well, and who were not raised or educated in the cultural mainstream and content of the school, can do little to transmit to their children the incidental knowledge or language skills that often accompany the requirements of academic and cognitive tests. When an individual comes from such circumstances, test performance will be adversely affected. But how does this help the practitioner? The *Standards* volume addresses problems that arise when a test requires an ability that is not the intended construct under measurement, or a construct-irrelevant component: "A prime threat to fair and valid interpretation of test scores comes from aspects of the test or testing process that may produce construct-irrelevant variance in scores that systematically lowers or raises scores for one identifiable group of test takers and results in inappropriate score interpretations for intended users" (p. 54). Examples of potential construct-irrelevant components that can create bias include "inappropriate sampling of test content, aspects of the test context such as lack of clarity in test instructions, item complexities that are unrelated to the construct being measured and/or test response expectations or scoring criteria that may favor one group over another" (p. 54).

As mentioned earlier, the authors of the *Standards* recommend that test developers and users rely on standardized tests with universal design, to "facilitate accessibility and minimize construct-irrelevant barriers." Universal design "emphasizes the need to develop tests that are usable as possible for all test takers in the intended test population, regardless of characteristics such as gender, age, language background, culture, socioeconomic status, or disability" (p. 57). Yet later the authors note that researchers are still gathering evidence to support the principles of universal design, and that even when principles of universal design are employed, there will still be situations where the test is not appropriate for all in the intended population. Moreover, in recognizing that the concept of degree is critical in these matters, a curious contradiction remains embedded in the *Standards*. For example, in discussing context-related sources of test bias, the *Standards* authors identify that "Construct-irrelevant variance may result from . . . unrelated complexity or language demands in test tasks" (p. 55). In Standard 3.2, it is recommended that

> unnecessary linguistic, communicative, cognitive, cultural, physical, and/or other characteristics in test item stimulus and/or response requirements can impede some individuals in demonstrating their standing on intended constructs . . . the level of language proficiency, physical response, or other demands required by the test should be kept to the minimum required. (p. 64)

Just what constitutes an appropriate level of response demand is not explained, and adopting a stance that there is even a minimal level of language that would not in some way affect performance is inconsistent with the view of language as a developmental process. Such a stance seems to view language proficiency as a threshold ability, beyond which further development provides no benefit to test performance. The well-known correlation between level of education and IQ contradicts such a view (Brody, 1997; Gustafsson & Undheim, 1996; Neisser et al., 1996; Ormond, 2008; Sattler, 2001). Indeed, the fact is that all tests, including nonverbal ones, require some level of communication between the examiner and examinee; thus some tests will indeed be subject to heavy linguistic demands (e.g., expressive vocabulary), and some tests will be less subject to these (e.g., block construction). Test performance is therefore likely to be affected continuously and linearly, and not in an either–or manner (i.e., it is affected or not affected), as is implied by the *Standards*.

The *Standards* volume does pay attention to minimizing construct-irrelevant components through increasing accessibility for specific individuals via test adaptations, accommodations, and modifications. Chapter 3 provides an excellent overview of the salient issues that may lead to significant problems in reliability and validity, including test content, context, response format, and opportunity to learn, as just described. A common theme emerging throughout the discussion is the notion that differences in language, culture, and development must be attended to throughout all aspects of test development, administration, and interpretation, to avoid measurement bias in test content, context, and response format. This is best captured in the opening overview of Chapter 3, which we have quoted above. Unfortunately, the associated practice guidelines are noteworthy in their lack of definitive guidance for practitioners. For example, Standard 3.9 states: "Test developers and/or test users are responsible for developing and providing test accommodations when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target construct" (p. 67). However, the commentary that follows provides no concrete guidance or examples for addressing construct-irrelevant components that may arise in assessing culturally and/or linguistically diverse individuals. Moreover, little attention is paid to providing guidance on how and when test adaptations meet the threshold of changing the intended construct being measured, thereby also hindering validity.

Standard 3.9 continues: "For example, individuals who are not fully proficient in English may need linguistic accommodations that address their language status" (p. 67). This statement implies that some tests are appropriate for use with individuals whose knowledge of the language of the test is questionable; perhaps it alludes to nonverbal instruments or ones that are otherwise less verbally demanding. Beyond the use of nonverbal tests, however, what assessment methods are likely to comply with this standard? What factors should professionals take into account when making a decision about language differences and whether they are relevant? Won't language differences of any kind always be relevant, especially in connection with potential limited opportunity for learning? Another point is that for practitioners who are interested in measuring visual processing, memory, or reasoning abilities, this statement might be sufficient guidance. But what about evaluations where the referral concern involves reading difficulties, which create a need to measure phonological processing skills, extent of vocabulary, and facility with language (e.g., with verbal analogies)? Such abilities cannot be measured in a nonverbal way. They are verbal by definition, and language development and proficiency are integral aspects of what they intend to measure.

For practitioners who are looking for specific advice about the testing of culturally and linguistically diverse individuals, there remains a need to look beyond the *Standards*. Without question, the *Standards* volume is an important and significant work. But although it provides significant direction in the understanding and accommodation of experiential differences in the assessment of culturally and linguistically diverse populations, and the degree to which these differences will affect nearly every aspect of test development, construction, administration, and use, practitioners still must review and evaluate empirical evidence independently and find a manner in which such evidence can be used to make critical decisions in testing and interpretation. A later section of this chapter provides a discussion along these lines, but first, a critical examination of current testing and interpretive practices is necessary.

EVALUATING CURRENT METHODS OF PRACTICE

The problems inherent in applying tests that were developed in the U.S. cultural milieu and normed primarily on monolingual English speakers to examinees with LEP were noted at the very advent of psychological testing (Brigham, 1923; Goddard, 1913; Yerkes, 1921), but they appear to have been either ignored or simply dismissed in the face of prevailing beliefs and arguments to the contrary (Sanchez, 1934). Despite a persistent pattern of lower performance among culturally and linguistically diverse individuals as compared to native English speakers, the matter of differential performance remained largely confined to notions regarding genetic differences (Jensen, 1974, 1976). Perhaps spurred by the spirit of social justice and civil rights that permeated the prior decades, other researchers in the 1970s began to reexamine the issue (e.g., Oakland & Laosa, 1976)—particularly in light of the 1975 passage of Public Law 94-192, the original Education for All Handicapped Children Act, later renamed and currently known as the Individuals with Disabilities Education Improvement Act of 2004 (IDEA, 2004). In the early forays into "nondiscriminatory assessment," the crux of the problem was broadly defined as "one dimension of the more general problem of valid assessment of any child" (Oakland, 1976, p. 1).

Despite the fact that examinations of test bias had been largely focused on issues of reliability, Oakland (1976) was one of the first to note that the most important aspect in testing of diverse individuals was validity. The issues described in the preceding section regarding the recommendations outlined in the *Standards* all reinforce the importance and centrality to the concept of validity in fairness, far more so than the concept of reliability. But as we have also noted previously, the authors of the *Standards* were rather silent until the 2014 edition regarding the manner in which test results are to be examined for evidence of validity, and practitioners have historically been forced to resort to a variety of methods in attempts to address it—often without any empirical support to defend their use. Indeed, many of the methods in current use appear to be based on the mistaken idea that validity, like reliability, is a continuous variable. From a strict psychometric standpoint, a measure or scale can be considered valid or not, depending on whether there is sufficient evidence to support it one way or the other. Viewing validity as shades of gray instead of as a dichotomous concept is easily forgiven in light of the fact that practitioners are simply trying to do whatever can be done to ensure the validity of their obtained test results. Thus doing more should be better than doing less or nothing at all in accomplishing the goal. Unfortunately, more often than not, practitioners cannot actually evaluate the success of their efforts in establishing validity. That is, the extent to which factors such as LEP or differences in opportunity for learning cultural knowledge actually affect the results of testing is rarely addressed in reports, apart from the ubiquitous but hollow warnings that "results should be interpreted with extreme caution." Use of a particular method or strategy may lead practitioners to assume that validity has been "increased" or "maintained," but typical methods simply do not permit independent verification that this is in fact the case. As will be discussed, such unverified assumptions (i.e., that validity has been achieved when in fact it has not) are common to all current approaches and limits their utility as avenues for achieving fairness.

In general, a review of the literature reveals four basic approaches that have been touted as viable methods for dealing with validity issues that stem from cultural and linguistic differences. Each approach is intended in some manner to address questions of fairness, so that test results emerge as valid—and each method has its own particular advantages and disadvantages, many of which do not appear to be acknowledged or recognized by those who employ them. It is especially important that the limitations of each approach be well understood by any practitioner with a desire to implement any of them in actual testing practice.

Modified or Adapted Testing

Perhaps some of the first attempts to address the various problems inherent in the evaluation of culturally and linguistically diverse individuals with standardized tests involved modifications or adaptations of the tests or testing protocols themselves. In this approach, tests are administered primarily in English, but are modified in some way so as to increase their fairness. Among the various adaptations are that have been suggested, the most common include eliminating or not administering certain test items with presumed culturally biased content; mediating culturally based task concepts prior to administration; repeating verbal instructions to ensure full comprehension; accepting responses in either the native language or the language of the test; administering only the subtests that do not rely on oral expression; and eliminating or modifying time constraints (Figueroa, 1983, 1990a, 1990b; Sattler, 1992, 2001). Such procedures are extensions of what is often referred to as "testing the limits" and represent a clinical approach to evaluating diverse individuals. These procedures are designed to aid examinees in performing to the true extent of their actual ability by reducing aspects of the testing process that might attenuate the scores. Unfortunately, any time a test is administered with such alterations, by definition it no longer remains standardized. Unknown amounts of error are introduced into the testing situation, resulting in a loss of confidence in the test's psychometric properties, especially those that determine its validity. Despite the benevolent intent of such procedures, any results derived from their application are rendered suspect at best, and thus they effectively preclude valid and defensible interpretation.

Another common testing adaptation involves attempts to overcome the language barrier via use of a translator/interpreter. Up to 20% of practitioners working with culturally and linguistically diverse children employ this method (Sotelo-Dynega et al., 2011). The presumption that testing will be valid as long as an individual comprehends what is being said or asked has intuitive appeal; however, it neglects the culturally based aspects of the testing process itself, as well as the fact that the test remains culturally bound. More importantly, even if we ignore the significant problems in translating tests "on the fly" with or without the aid of trained and untrained interpreters and the presence of third-party observers, tests have yet to be standardized with the use of a translator/interpreter. That is, the use of a translator/interpreter in the testing process represents another violation of standardized procedures, which again by itself undermines the reliability and validity of the results and continues to prevent interpretation.

Beyond issues related to test administration and modification, it is important to note that such procedures do nothing to address problems related to norm sample representation. Even if modification of the test or its administration protocol did not invalidate the process, could the test scores be interpreted fairly? Even if threats to validity are controlled in some areas, it does not mean that validity has been addressed in all areas. This is particularly true with respect to the adequacy of the norm sample against whom the test scores will be compared. Test developers often attempt to control for cultural or linguistic differences by including individuals from diverse racial and ethnic backgrounds. But race and ethnicity are not the same as culture or cultural differences and do not directly account for differences in experience that affect language or acculturative knowledge development, as noted by Salvia and Ysseldyke (1991) and discussed in the preceding section. Representation within a test's norm sample on the basis of racial or ethnic categories is simply not a sufficient proxy for the degree to which an individual is or is not familiar with the culture of the test. Likewise, neither race nor ethnicity provides specific information about the extent of an individual's proficiency in English. Despite demonstration of high-quality technical characteristics and the use of sophisticated sampling techniques, norm samples that are stratified on the basis of race or ethnicity, but that contain individuals who are predominantly or exclusively monolingual English speakers, are unlikely to meet the necessary standards for adequate representation of genuinely "bilingual" and "bicultural" individuals. For the most part, test developers and researchers have not addressed or recognized this issue. Until norm samples for tests are based upon stratification variables that matter to culturally and linguistically diverse individuals, the results simply cannot be construed as valid, even when tests are carefully modified, adapted, or translated.

Because altering the standardized requirements of the testing process in any manner effectively precludes the assignment of meaning to the collected data, modifications or adaptations in testing are of limited utility. Even if such adaptations could be seen as valid, the significant problems with norm sample adequacy would still preclude validity of any conclusions regarding comparative differences. In practice, such procedures may be most useful in allowing practitioners to derive qualitative information—that is, by observing behavior, evaluating learning propensity, evaluating developmental capabilities, analyzing errors, and so forth. Perhaps the best recommendation for practitioners who are considering use of these types of methods would be to administer tests in a standardized manner first, and then retest with any modifications or adaptations that might help illuminate the actual or true level of the individual's ability. In this way, it may be possible to evaluate the issue of validity by limiting the threats rather than adding to them.

Nonverbal Testing

Much as in the development of the Beta version of the Army examination (Yerkes, 1921), the use of nonverbal methods and tests in the evaluation of English learners has been predicated on a simple notion: Eliminate the language barrier, and testing can proceed as usual. Nonverbal tests have in fact become quite popular in psychological practice, and a variety of tools have been published expressly for this purpose. According to the Sotelo-Dynega and colleagues (2011) survey, when evaluating the intelligence of culturally and linguistically diverse individuals, 88% of all practitioners choose to administer a nonverbal test. Similar to the claims originally put forth by Brigham (1923), these tests offer the promise of validity based on the idea that language has been effectively removed from the testing equation. For example, according to Weiss and colleagues (2006), administration of a nonverbal cognitive assessment is still promoted as "an acceptable answer to this problem" (p. 49). This appears, however, to be an overly optimistic view.

The phrase *nonverbal testing* is itself a bit of a misnomer; what is meant is probably better characterized as *language-reduced testing/assessment*. This is because no matter what a test is like, its use in any evaluation requires that the examiner and examinee be able to communicate with each other. Even tests whose developers claim that they can be administered in a completely nonverbal manner (i.e., via gestures or pantomime) first require that the examinee understand and comprehend the meaning of the gestures. This meaning must necessarily include instructions on when to start, when to stop, what is a right answer, and when to work quickly. Other testing issues (establishing rapport, explaining the purpose of testing, etc.) also need to be conveyed to the examinee. How all this is to be communicated in the absence of any verbal interaction is not clear. Even if it were possible to do so, the fact remains that the teaching of gestures is akin to the teaching of a new, albeit very brief and limited, "language." Thus, whether spoken language is used or not, administration of a test always requires some type of communication between examinee and examiner. Nonverbal testing may well reduce the language barrier, but it clearly does not eliminate it.

In a similar manner, the claim that a test's cultural fairness is increased because it is nonverbal does not mean that cultural content embedded in the test is eliminated. Given the emphasis on abilities that are less verbal, there may be some reduction in cultural content unless the use of visual stimuli includes pictures of actual objects and artifacts, which continue to embed culture even with the reduction in language. Many nonverbal tests continue to rely on visual images that remain culturally bound (Sattler, 1992). In addition, nonverbal tests are often used to derive a score that will serve as an indicator of an individual's general intelligence. Such practice, especially in the context of evaluation for specific learning disabilities (SLD), is problematic for several reasons. First, it has been demonstrated that nonverbal estimates of intelligence may be no more fair or valid than those that include verbal abilities (Figueroa, 1989). Second, the range of abilities measured by a nonverbal composite is by definition likely to be narrower than that measured by verbal batteries, despite correlations with broader measures of intelligence (Flanagan, Ortiz, & Alfonso, 2007; Ortiz, 2014). Third, the majority of referrals for SLD evaluation are based on problems in language arts, particularly reading. This means that in terms of evaluating the cognitive deficits most likely responsible for reading difficulties, an assessment for SLD will need to include testing for those abilities most related to reading, including auditory processing (Ga) and crystallized knowledge (Gc) (Flanagan et al., 2007; Flanagan, Ortiz, Alfonso, & Mascolo, 2006). These abilities cannot be easily measured or measured at all with nonverbal tests, and such tests are therefore not useful for evaluation of SLD in a large majority of cases. Finally, nonverbal tests are also subject to the same problems with norm sample representation as those that exist for verbal tests, as described earlier. That is, neither type of tests has norm samples that systematically and adequately control for differences in acculturative experiences or language development characterizing bilingual and bicultural individuals. In sum, language-reduced tests are not as helpful in the evaluation of the abilities of individuals from diverse cultural and linguistic backgrounds as their developers often claim. Although such tests may provide better estimates of true functioning in certain areas, they do not represent a satisfactory solution with respect to validity and fairness in testing, and in the majority of cases they will be inadequate to serve the purpose of SLD identification.

It seems likely that these problems may help explain why the empirical evidence for the predictive validity of nonverbal tests tends to be rather dubious (Figueroa, 1989; Lohman et al., 2008). In an examination of three different nonverbal tests often used to identify gifted children from culturally and linguistically diverse backgrounds, Lohman and colleagues (2008) noted that "one cannot assume that nonverbal tests level the playing field for children who come from different cultures or who have had different educational opportunities" (p. 292). For example, in contrast to claims of reduced "ethnic" score differences for many nonverbal measures, Lohman and colleagues found "large differences between the scores of ELL and non-ELL children on the three nonverbal tests"; these findings indicated that practitioners "must consider opportunity to learn not only for tests that measure verbal and quantitative abilities and achievements but also for those abilities measured by nonverbal tests" (p. 292).

Despite their widespread popularity, immense intuitive appeal, and long history of clinical use with culturally and linguistically diverse individuals, nonverbal instruments simply do not fulfill their developers' vision of them as tests of innate ability unaffected by culture, education, or experience. This is not to say that such tests are not helpful or valuable in evaluating diverse individuals, but only that they are not the sole or definitive solutions to issues of fairness and validity, as they are often purported to be. The best recommendation that can be made regarding the evidence base for the use of nonverbal tests is that they should be viewed as only one component of a broader, comprehensive evaluation—a component that assists in examining functioning in the particular areas such tests measure. Whether the obtained results are valid remains a question not adequately addressed merely by their use in practice.

Native-Language Testing

The relatively recent development of psychometrically sound, standardized tests of intelligence and cognitive abilities in languages other than English, coupled with a slight increase in the number of psychologists with sufficient competency in evaluations conducted in languages other than English, has led to a growth in approaches based on the use of examinees' native languages. Unfortunately, such practice has become identified with the inaccurate label of *bilingual assessment*. Bilingual assessment implies evaluation that is to be conducted bilingually—that is, with the concurrent use of two languages as the situation may dictate, or desired by the individuals, as is the custom when bilingual persons speak to each other. Native-language tests, however, are not standardized using two languages, but only one. Of course, it would probably be impossible to standardize rote transitions from one language to another because artificial and arbitrary changes by an individual would lead to considerable awkwardness in communication. Except on some tests where responses are accepted when given in either language, code switching (into or out of English) is not specified or standardized. Thus *bilingual assessment* is better described as *monolingual testing,* even in those situations where a test is given in one language followed by retesting in another language.

Regardless of how it may be best characterized, use of a native-language test requires that psychologists speak the language of the test (i.e., they need to be bilingual themselves). The ability to communicate with an examinee directly is an important and significant benefit to this approach; it places a psychologist in a position to conduct assessment activities in a manner (i.e., bilingually) not available to a monolingual psychologist even with the aid of a translator/interpreter. This notion may partly explain why the simple hiring of a bilingual practitioner is often seen as an acceptable solution to the problem of evaluating diverse individuals. However, "mere possession of the capacity to communicate in an individual's native language does not ensure appropriate, nondiscriminatory assessment of that individual. Traditional assessment practices and their inherent biases can be easily replicated in any number of languages" (Flanagan, McGrew, & Ortiz, 2000, p. 291). Speaking the same language as the examinee and utilizing a test available in the language which the examinee speaks do not, by themselves, resolve issues of fairness or validity. Indeed, they may actually lead practitioners, and those for whom they work, into a false sense of security about the meaning of their obtained test results. In addition, not only are there no truly "bilingual" tests or assessment protocols, but very little is currently known about the performance of bilingual individuals on monolingual tests administered in the primary language. Compared to the body of research on the use of tests administered in English, research on testing in the native language is a relatively new tradition, with very little empirical evidence to guide appropriate activities or use as a basis for standards for practice. The general question regarding how a bilingual individual born in or recently moved to the United States would be expected to perform on a test administered in the native language has yet to be answered. Such a question is bound to be complicated by various factors, such as the individual's age, his or her level and type of prior education, the current language of instruction, and the type of instructional program (Goldenberg, 2008).

In addition, when native-language testing is accomplished in the United States, an examinee cannot rightly be viewed as a monolingual speaker or from a monocultural background. Because the norms of native-language tests often utilize monolingual speakers from other countries who are being raised by parents who speak the language and who are being educated in the native language, they do not form an adequately representative norm sample for comparison of performance to individuals now residing in the United States. In such cases, the experiential backgrounds of these two populations are no more similar than they are to the backgrounds of monolinguals. As noted by Harris and Llorente (2005), "these children indeed represent a proportion of U.S. school children who are ELLs. Realistically, however, little is known about the language abilities of these learners and the degree to which they are bilingual" (pp. 392–393). Even when test developers attempt to include bilingual speakers, they are not sampled systematically with respect to the two major variables (current proficiency in both languages and level of acculturation) that would be necessary to create representative groups. For example, despite inclusion of bilinguals in the developmental sample of the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) Spanish version (Wechsler, 2005), they are grouped primarily by country of origin, length of time in the United States, or length of schooling in the United States, all of which fail to account for the influence of cultural and linguistic differences (Harris & Llorente, 2005). In addition, it should be noted that the actual WISC-IV Spanish norms are equated to the WISC-IV English norms, and thus the Spanish version does not have actual, separate norms (Braden & Iribarren, 2005).

It would appear that until such time as a sufficient body of knowledge exists on which bilingual psychologists can base expectations of performance on native-language tests, their use will continue to provide results that are of dubious validity and extremely difficult to interpret. Accordingly, the most prudent recommendation for practice involving the use of native-language tests is similar to those offered previously: Administer tests in a standardized manner first, and then apply whatever modifications or adaptations that may be appropriate and may inform the referral questions. Well-trained bilingual examiners may be limited by the tools and practices available to them, but they remain the best choice for conducting evaluations of culturally and linguistically diverse individuals. Despite the limitations inherent in all approaches, bilingual examiners have one unique and significant advantage over other examiners employing other approaches—their ability to directly communicate with, and observe the behaviors of, diverse examinees in a controlled setting. This advantage alone places native-language testing in front of all other approaches and represents the current "ideal" situation—not because the test results are necessarily more fair or valid, but because of the examiners' ability to engage in direct interaction with the examinees and to utilize a wide variety of procedures (including authentic assessment, informal measures, error analyses, and whatever other methods may be helpful and informative in understanding and gauging the examinees' true abilities).

English-Language Testing

Given the increasingly large numbers of culturally and linguistically diverse individuals in the U.S. population, coupled with the fraction of professionals with sufficient competency to conduct evaluations in such persons' native languages, it is not likely that all such individuals will be evaluated in the native language or by bilingual professionals. Of the 480 practitioners surveyed by Sotelo-Dynega and colleagues (2011), only about 12% identified themselves as "bilingual/multicultural school psychologists," yet 86% reported that they evaluated students who were culturally and linguistically diverse. The reality is that the majority of diverse individuals are being evaluated by monolingual English-speaking practitioners, and that these evaluations will be conducted primarily (if not exclusively) in English. As compared to the three prior methods, this particular approach would seem to be the most biased and least fair. In many ways it echoes Brigham's (1923) comments about handling a "typically American situation," because it makes no concessions to the fact that such a child is not a native English speaker; it does not alter the content or administration of the test; and it does not investigate the abilities the child may be able to demonstrate in the native language but not in English. On the other hand, if we dispense with Brigham's mistaken notions about personal character, we can in fact recognize that this is also the only approach where there exists a great deal of scientific research regarding how culturally and linguistically diverse individuals actually perform on tests—tests given to them in English.

Although it certainly was not intentional, the field of psychometrics has nevertheless provided perhaps the most empirically supportable basis for evaluating the validity of bilingual individuals' test performance. The development of standardized procedures, coupled with repeated evaluation of participants proficient enough in English to reasonably comprehend test instructions, has established a rather extensive and cohesive database regarding the manner in which such examinees perform on tests administered to them in English (Brigham, 1923; Cummins, 1984; Figueroa, 1989; Goddard, 1913; Jensen, 1974, 1976; Mercer, 1979; Sanchez, 1934; Valdés & Figueroa, 1994; Vukovich & Figueroa, 1982; Yerkes, 1921). A review of this research indicates that non-native speakers of English consistently perform more poorly (about one full standard deviation or more below average) than native English speakers on tasks that rely on English-language development, skills, or proficiency, and that they perform comparably to them (at or near the normative mean) on tasks that do not require such verbal or language-based development or skill (Cummins, 1984; Figueroa, 1989; McShane, 1980; Mercer, 1979; Naglieri, 1982; Valdés & Figueroa, 1994; Vukovich & Figueroa, 1982).

On the surface, assessing diverse individuals with English-language tests appears to result in highly biased estimates of ability, particularly those abilities that may rely heavily on language and acculturative knowledge development. On the other hand, if this research is viewed as illuminating the magnitude and degree to which such factors as differences in language and acculturative development actually affect test performance, it seems reasonable that it may be an effective way to develop an empirically based approach for evaluating the validity of obtained test results. Available research provides an estimate of the degree of attenuation that may have occurred in testing as a function of the presence of the main operating variables—namely, English-language proficiency and acculturative knowledge. Whether or not the results are valid can thus be examined by comparing diverse examinees' performance to that of a group far more similar to the examinees in developmental background and experience than existing norm samples are. Ideally, individuals from diverse cultural and linguistic backgrounds should rightly be evaluated by qualified, competent professionals with specific expertise in nondiscriminatory assessment and knowledge of the manner in which such differences influence test performance (Ortiz, 2014). There is nothing, however, that prevents any practitioner from evaluating such an individual in both the native language and English. And when there is no other option available but to evaluate in English, the same type of expertise, knowledge, and research base may well be applied to assist in evaluating validity—a process that is often couched in terms of determining "difference versus disorder." The potential application of this body of literature in support of an evidence-based approach using English test administration is explored further in the next section regarding integration of research into practice. For the moment, testing in English only without regard for the inherent problems in fairness and validity will certainly lead to extremely discriminatory interpretations and conclusions.

INTEGRATING RESEARCH INTO PRACTICE

Fallacies and Misconceptions in Early Research

Although tests can be quite reliable, even perfect reliability does not guarantee validity. It is possible that something can be measured very accurately and consistently over time, but this does not ensure that the construct one *believes* is being measured is actually the one that is. The difference between reliability and validity is central to the issue of testing with culturally and linguistically diverse individuals. Although this notion was certainly well understood by early psychometricians and psychologists, preconceptions and the early cultural *zeitgeist* may have prevented their recognition.

A particularly stark example can be seen in the work of Henry Herbert Goddard, a leading American psychologist who translated the Binet–Simon Scale into English and promptly set about working with bilingual examinees and individuals from diverse cultural backgrounds. Goddard's intent was not to study bilingualism or its effect on test performance, but rather to prove an a priori conviction—that IQ was an innate, inherited trait, and that it explained the division between the "haves" and "have-nots." To examine the issue, he went to Ellis Island in New York and began testing immigrants—not randomly selected ones, but those whom he already believed were intellectually inferior, primarily on the basis of their appearance. In the process of his investigation, Goddard (1913) eventually came to conclude that approximately 80% of all immigrants arriving from Eastern Europe were mentally defective or feeble-minded, a level of functioning. He even developed a specific word to define their level of functioning—*moron*. The degree to which Goddard was working from preconceptions that made him oblivious to patterns in his own data can be seen in his explicit comments, such as "We picked out one young man whom we suspected was defective, and, through the interpreter, proceeded to give him the test" (p. 105); in his inability to entertain any hypothesis contrary to his own when the interpreter questioned the fairness of the testing by stating, "I could not have done that when I came to this country" (p. 105); and in his curt and decisive reply to the interpreter, "We convinced him that the boy was defective" (p. 105).

Not much later, Robert Yerkes, already an eminent American psychologist, developed the Army mental tests in response to the Department of the Army's request to help in distinguishing men who would be good officer candidates from those who could serve merely as infantrymen. In the course of his work, Yerkes (1921) did in fact recognize and acknowledge the problem posed by individuals from diverse cultural and linguistic backgrounds, which he viewed mostly as an issue of illiteracy. Because his initial (Alpha) test had components that required reading, he could not administer the test to anyone (regardless of country of origin) who could not read English. He therefore set about creating an alternative (Beta) version, which eliminated any test requiring reading and which utilized "nonverbal" demonstrations via blackboard to provide presumably comprehensible instructions. He believed that the Beta test represented a valid method for evaluating the abilities of the millions of men he needed to test who were illiterate, at least in English, or who did not speak English. In examining his data on nearly 1.75 million American men, Yerkes noted that the average raw score on the Beta for native English speakers (even those who could not read at all in English) was a stout 101.6, which classified them as Very Superior (Grade A). In contrast, the average raw score for non-native speakers of English (who also could not read in English) came in at only 77.8, which placed these individuals in the Average (Grade C) classification. For Yerkes, and for the contingent of other notable psychologists working with him (e.g., Carl Brigham, David Wechsler, and Lewis Terman), the results confirmed their own beliefs that immigrants—particularly those from certain countries and from lower classes—were merely displaying their inherited lack of intelligence. Questions regarding the potential issues involved in testing individuals with LEP or no English at all, and with limited opportunity for learning about the cultural content embedded in the test, were not addressed, even if they hovered in the air when the pattern of results was examined. Given the atmosphere of the era, none of these psychologists were inclined to seriously abandon their causal explanations related to intelligence as an innate quality even when the data strongly suggested otherwise.

Figure 25.1 provides an example of this tendency using Yerkes's data, which were later reanalyzed by Brigham in his 1923 monograph *A Study of American Intelligence*. The increase in mental age as time of residence in the United States increased is particularly striking. Nevertheless, this particular finding seemed to bother Brigham—perhaps subconsciously, as he first admitted—but then he dismissed the obviously correct interpretation and instead provided a convoluted hypothesis consistent with his and the prevailing beliefs. In examining the data in Figure 25.1, Brigham gave the following explanation:

> Instead of considering that our curve indicates a growth of intelligence with increasing length of residence, we are forced to take the reverse of the picture and accept the hypothesis that the curve indicates a gradual deterioration in the class of immigrants examined in the army, who came to this country in each succeeding 5 year period since 1902.

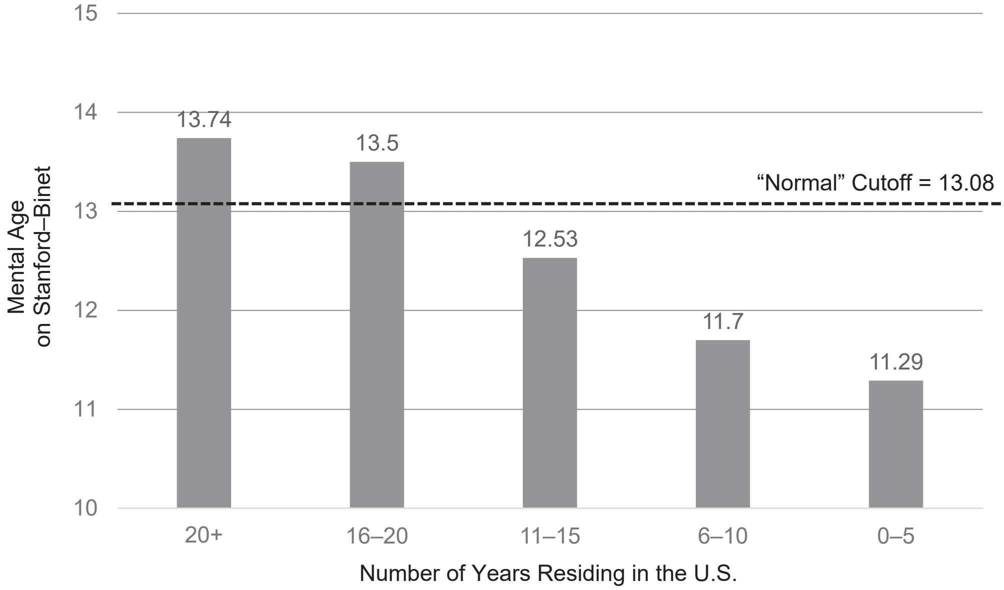


FIGURE 25.1. Mean mental age on the Binet–Simon Scale for immigrants tested with the Army Beta examination. Average raw score for native English speakers on the Beta test = 101.6 (Very Superior; Grade A). Average raw score for non-native English speakers on the Beta test = 77.8 (Average; Grade C). Data from Yerkes (1921).

> (pp. 110–111) . . . The average intelligence of succeeding waves of immigration has become progressively lower. (p. 155)

When combined with the data showing significantly lower raw scores on the Army Beta test for non-native than for native English speakers, to which Brigham had access, it seems reasonable that the far more plausible interpretation as previously noted should also have been reached by Brigham: that the longer one lives in the United States, the more English one learns (which increases comprehension), and the more familiar one becomes with U.S. cultural artifacts (including those that appeared on the test). That such an obvious conclusion was not offered by Brigham again emphasizes the degree to which evidence indicating that testing was being affected by differences in the cultural and linguistic backgrounds of individuals went largely unappreciated. Even when deliberate attempts were made to bring the issue to light, it was often buried in obscure journals or flatly dismissed as inconsistent with scientific fact, as is evident in Terman's (1916) comment:

> The common opinion that the child from a cultured home does better in tests solely by reason of his superior home advantages is an entirely gratuitous assumption. Practically all of the investigations which have been made of the influence of nature and nurture on mental performance agree in attributing far more to original endowment than to environment. (p. 115)

George Sanchez, a Mexican American psychologist, was one of the few who did manage to publish his research on bilingual individuals' performance in scholarly journals (Sanchez, 1932, 1934). He explicitly outlined the idea that differences in language development and proficiency, as well as differing levels of experience with the cultural content embedded in the test, were in fact the variables responsible for the observed difference in test scores between native and non-native English speakers. According to Sanchez (1934), "as long as tests do not at least sample in equal degree a state of saturation [i.e., assimilation of fundamental experiences and activities] that is equal for the 'norm children' and the particular bilingual child it cannot be assumed that the test is a valid one for the child" (p. 770). By the time the issue was being given serious considerations in the 1970s, the legacy of these early studies remained embedded in psychological science, and genetic explanations of test performance remained dominant or were replaced by ones suggesting that bilingualism itself was a handicap (Valdés & Figueroa, 1994).

One of the individuals who helped to promote and maintain notions regarding genetic explanations was Arthur Jensen (1974, 1976, 1980). Ironically, Jensen was also one of the first researchers to admit to the existence of bias (as related to validity) through an experiment of his own, which, unlike his more controversial assertions, garnered very little attention. In an investigation of the convergence of two separate measures of intelligence with Mexican and European American ("white") groups, one using a verbal modality and one using a nonverbal modality, he expected to find similar patterns of performance on the tests and equivalent degrees of difficulty among items for both groups. In fact he did find equivalent degrees of difficulty among items for both groups, but he did not find similar patterns of performance. Instead, he found that whereas both groups did show similar score patterns on the nonverbal task, the Mexican group had significantly lower scores on the verbal task than the European American group. The lack of concurrence between test scores on the two tasks for the two groups led Jensen (1974) to the following conclusion:

> The fact that the Mexican group is very similar to the white in rank order of *p* values and *p* decrements on both the PPVT [Peabody Picture Vocabulary Test] and the Raven, yet has lower scores on the PPVT than on the RAVEN, suggest that some factor is operating to depress the PPVT performance more or less uniformly for all items and that this factor does not depress Raven performance, at least to the same degree. It seems plausible to suggest that this factor is verbal and may be associated with bilingualism in the Mexican group. (pp. 239–240)

Two years later, in a separate publication, Jensen (1976) offered the following comment on the result of his earlier study: "Thus, there is some evidence that a vocabulary test in English may be a biased test of intelligence for Mexican-Americans" (p. 342). Shortly thereafter, other researchers reached much the same conclusion. Although Jensen and others kept pointing to a verbal–nonverbal dichotomy in performance, the evidence began to demonstrate a pattern that was much more consistent with what Yerkes and Brigham had found originally with the Binet Scales—a more or less continuous, linear variation in decline. Table 25.1 provides a comparison of means from ten of the subtests from the WISC-R/WISC-III among "Hispanic" (Mercer, 1972, cited in Vukovich & Figueroa, 1983), "ESL" (Cummins, 1984), and "bilingual" (Nieves-Brull, 2006) groups as compared to the norm sample mean (ScS = 10). Two things are particularly evident in the data: (1) The Verbal tests show significantly more attenuation than the Performance tests; and (2) the degree of attenua-

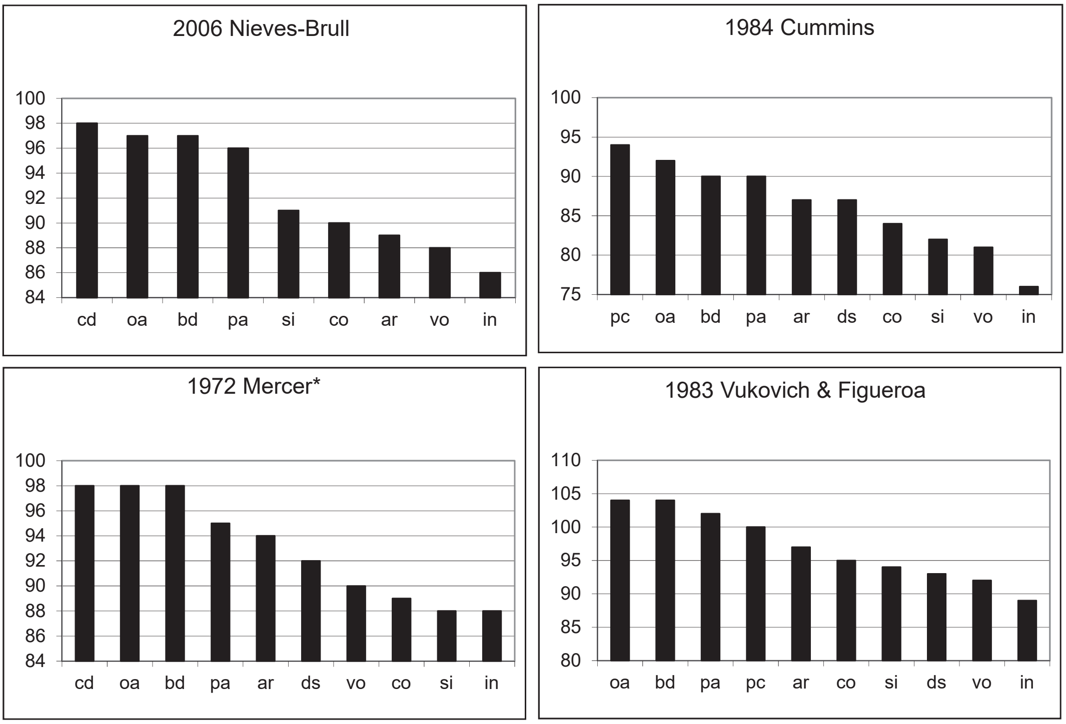
TABLE 25.1. Results of Testing among Four Different Non-native English-Speaking Groups

WISC-R Subtest Name	Hispanic Group (Mercer, 1972, cited in Vukovich & Figueroa, 1983) Mean scaled score	Hispanic Group (Vukovich & Figueroa, 1983) Mean scaled score	ESL Group (Cummins, 1982) Mean scaled score	Bilingual Group (Nieves-Brull, 2006) Mean scaled score
Information	7.5	7.8	5.1	7.2
Vocabulary	8.0	8.3	6.1	7.5
Similarities	7.6	8.8	6.4	8.2
Comprehension	7.8	9.0	6.7	8.0
Digit Span	8.3	8.5	7.3	*
Arithmetic	8.7	9.4	7.4	7.8
Picture Arrangement	9.0	10.3	8.0	9.2
Block Design	9.5	10.8	8.0	9.4
Object Assembly	9.6	10.7	8.4	9.3
Picture Completion	9.7	9.9	8.7	9.5
Coding	9.6	10.9	8.9	9.6

**Data for this subtest were not reported in the study.*

tion lessens across the tests and does not depress performance in a manner that can be considered equal within either the Verbal or the Performance domain. In general, it is clear that all four groups show a variable but systematic decline in performance, and not merely a difference between verbal and nonverbal. Some tests that are considered verbal (e.g., Arithmetic) do not attenuate performance as much as others that are very verbal in nature (e.g., Information). Likewise, some tests that are considered nonverbal and have very low linguistic demands (e.g., Coding) have smaller adverse effects on test performance than other nonverbal ones with low linguistic demands (e.g., Picture Arrangement).

This type of declining pattern is particularly evident when scores are graphed and arranged in terms of highest mean to lowest mean. Figure 25.2 provides such an illustration and demonstrates several important dimensions relevant to the present discussion, including the fact that the change in scores does not appear to be an either–or proposition, but rather a smooth, gradual decline as the tasks increase in their measurement of language skills and cultural knowledge. These graphs are particularly striking in their similarity to Yerkes's data shown in Figure 25.1—data that are from 50 to 80 years older than those illustrated. Another pattern in the data can be seen in the difference in the magnitude of the means between the 1972 Mercer and 1983 Vukovich and Figueroa groups. Although both groups show a clear and similar decline in test performance relative to increasing verbal and knowledge demands, the means for the latter group are consistently higher than those for the former and would suggest the existence of differences between them in the two principal variables (language proficiency and opportunity for learning about the culture). Personal communication with Vukovich (September 29, 2008) has confirmed the difference in these factors and highlights the necessity to account for them in evaluations, as well as for the mistakes that can occur if all "bilingual" individuals are treated as equal or as a monolithic


2006 Nieves-Brull

1984 Cummins

1972 Mercer*

1983 Vukovich & Figueroa

FIGURE 25.2. Comparison of mean WISC-R/WISC-III subtest scores from four investigations with "Hispanic," "ELL," and "bilingual" populations (see Table 25.1). Subtest abbreviations: pc, Picture Completion; cd, Coding; oa, Object Assembly; bd, Block Design; pa, Picture Arrangement; si, Similarities; co, Comprehension; ar, Arithmetic; vo, Vocabulary; in, Information; ds, Digit Span. *The 1972 Mercer sample is reported in Vukovich and Figueroa (1983).

group. In sum, these studies all point to the notion that adherence to a strict verbal–nonverbal conceptualization of test performance for culturally and linguistically diverse individuals is much too simplistic and not entirely supported by the evidence. To understand performance and the extent to which it is affected by cultural and linguistic influences instead appears to require paying attention to the unique characteristics of each subtest, as well as the construct it purports to measure.

Despite research spanning over a century thus far, as well as the robust and persistent finding regarding the pattern of performance for bilingual groups on tests given to them in English, very little of this information has found its way into actual practice. This question was raised some time ago by Valdés and Figueroa (1994), who wondered why "these and other anomalous psychometric outcomes associated with bilingual populations (such as the ubiquitous and intractable low-VIQ, high-PIQ bilingual profile) are curious in and of themselves, but not quite as perplexing as psychometricians' lack of interest about why such outcomes occur" (p. 108). Practitioners do not have the same luxury as academicians in being able to ignore such issues. Perhaps the seductive nature of the verbal–nonverbal duality has resulted in its unquestioned use among practitioners—or, more likely, there simply has not been much of an alternative. As we have noted earlier, there are few general approaches a practitioner can employ in conducting an evaluation of an individual from a diverse cultural and linguistic background, and none of the methods we have discussed provide a truly satisfactory or evidence-based solution, particularly in addressing the issue of validity. Moreover, it cannot be said that any of them are based on any substantive body of research, although nonverbal testing has relied on the general verbal/nonverbal pattern. Nevertheless, even when nonverbal methods are employed, several significant problems remain (Figueroa, 1989, 1990b; Flanagan & Ortiz, 2001; Lohman et al., 2008; Ortiz, 2014), especially in cases where there is a need to measure the full range of cognitive abilities (Carroll, 1993; Horn & Blankson, 2005; McGrew & Flanagan, 1998; Woodcock, 1990).

Moving toward Research-Based Practice

If the currently available instruments have limitations in terms of demonstrating adequate validity, and if such tools lack evidence to support their use with various linguistic or cultural subgroups (i.e., groups that vary in terms of proficiency and experience within a larger ethnic group), and if the notion of verbal–nonverbal views of performance is highly overgeneralized, then what is the average practitioner to do in terms of engaging in evidence-based practice? At this time, the answer to this question may be rather surprising. Despite the more detailed and comprehensive practice guidelines outlined in the *Standards,* if evidence-based practice must still rest on just that—evidence—then the only sufficient body of empirical research currently available to guide the practice of evaluating a wide range of cognitive abilities in individuals from diverse cultural and linguistic backgrounds is the use of standardized tests administered in English. It is perhaps ironic, but the fact remains that there is a considerable amount of information (much of which has been discussed previously) regarding the performance of bilingual individuals on tests given to them in English. If this literature is combined with defensible psychometric procedures (particularly maintenance of standardization without alteration), and if the main focus in evaluation is placed on examining the validity of the obtained test data (leaving interpretation to occur only if the data are deemed valid), it would be possible to conduct evaluations on diverse individuals that begin to meet the criteria for being both evidence-based and defensible. At present, few such methods that are expressly designed to deal with the matter of validity have been proposed in the literature. Those that do exist are discussed below.

The Culture–Language Test Classifications and Interpretive Matrix

Regardless of the type, language, or number of tests selected for an evaluation, the mere inclusion of a standardized, norm-referenced test in a battery means that practitioners must then contend with the need to establish the construct validity of the obtained test scores; otherwise, interpretation is likely to remain inequitable, discriminatory, and purely speculative. If practitioners are to comply with relevant legal and professional standards, there must be evidence that the validity of test scores has been directly examined, and (at a minimum) a description of how the determination of validity was made. In other words, practitioners are required to provide some sort of convincing evidence that measurement of an ability domain that gave rise to a standard score or percentile rank is, in fact, a valid estimate of that specific ability and is not instead a reflection of the influence of some other construct (e.g., LEP). The familiar refrain "difference versus disorder" elegantly captures this dilemma regarding test score validity, which is the heart of the matter in the evaluation of culturally and linguistically diverse individuals. If test scores are believed to be valid, low performance may then be interpreted as possibly reflecting the lack of an ability or attribute. If test scores are not believed to be valid, low performance may then be interpreted as a reflection of the influence of cultural and linguistic differences. So how do practitioners accurately account for linguistic and cultural differences, to avoid misinterpretation of test results? The manner in which this fundamental question may be addressed is the purpose and intent of the Culture–Language Test Classifications (C-LTC) and the Culture–Language Interpretive Matrix (C-LIM).

The Culture–Language Test Classifications

Development of the C-LTC (Flanagan et al., 2000, 2007; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998) and its companion the C-LIM (Flanagan & Ortiz, 2001, Flanagan et al., 2007; Ortiz, 2001, 2004; Ortiz & Flanagan, 1998) was spurred by the need to consider the "difference versus disorder" question, as well as by the wealth of research available on the performance of bilinguals tested in English. The C-LTC was initially developed as an extension of the Cattell–Horn–Carroll (CHC) theoretical classifications presented as the basis of the CHC cross-battery assessment and interpretive approach (Flanagan et al., 2000; McGrew & Flanagan, 1998). The C-LIM evolved shortly afterward as a refinement of the C-LTC, designed specifically to aid in interpretation by allowing practitioners to assess whether or not what is measured is due primarily to the influence of cultural or linguistic variables (Flanagan & Ortiz, 2001; Flanagan et al., 2007; Mpofu & Ortiz, 2009; Ortiz, 2001, 2004; Ortiz & Dynda, 2010). Although the C-LTC and the C-LIM were initially linked to the CHC cross-battery approach, they can be used independently, and their utility does not depend on the use or application of any particular assessment procedure. The C-LTC and C-LIM are designed to evaluate whether obtained test results are either valid (permitting interpretation) or invalid (thereby precluding interpretation).

In an appeal for less discriminatory practices, Figueroa (1990a, 1990b) suggested that application of defensible theoretical frameworks in the assessment of culturally and linguistically diverse individuals was an important avenue to explore. In addition, he admonished practitioners to pay particular attention to the cultural and linguistic dimensions of tests that were often ignored or misunderstood in evaluation. In response to such issues, Ortiz and Flanagan (1998), Flanagan and Ortiz (2001), and Flanagan and colleagues (2000) developed the C-LTC, essentially a classification system for cognitive ability tests based on two critical test dimensions: degree of cultural loading and degree of linguistic demand. These two dimensions were deliberately selected because they have been identified as factors that have a significant and powerful relationship to test performance, and can render results invalid for individuals who are culturally and linguistically diverse (Figueroa, 1990a, 1990b; Sandoval et al., 1998; Valdés & Figueroa, 1994). What establishes the C-LTC as an evidence-based practice is the fact that the initial and some of the subsequent test classifications were and continue to be drawn directly from actual research on bilingual persons tested in English. By using the comparative subtest means available for such groups (see Table 25.1), one can easily sort tests into categories that correspond to the three basic classification levels (low, moderate, high) used for both dimensions of the C-LTC framework. It bears repeating that the classifications are data-driven, organized by the available empirical studies on the testing of bilingual individuals in English. In cases where no such data exist, classifications have been made via an expert consensus procedure as well as by examination of task characteristics, manner of administration, and construct that the subtest was designed to measure. Given the extent to which ability tests establish validity via correlations with other ability tests and via factor-analytic methods, test classifications based on this information also represent an application of research.

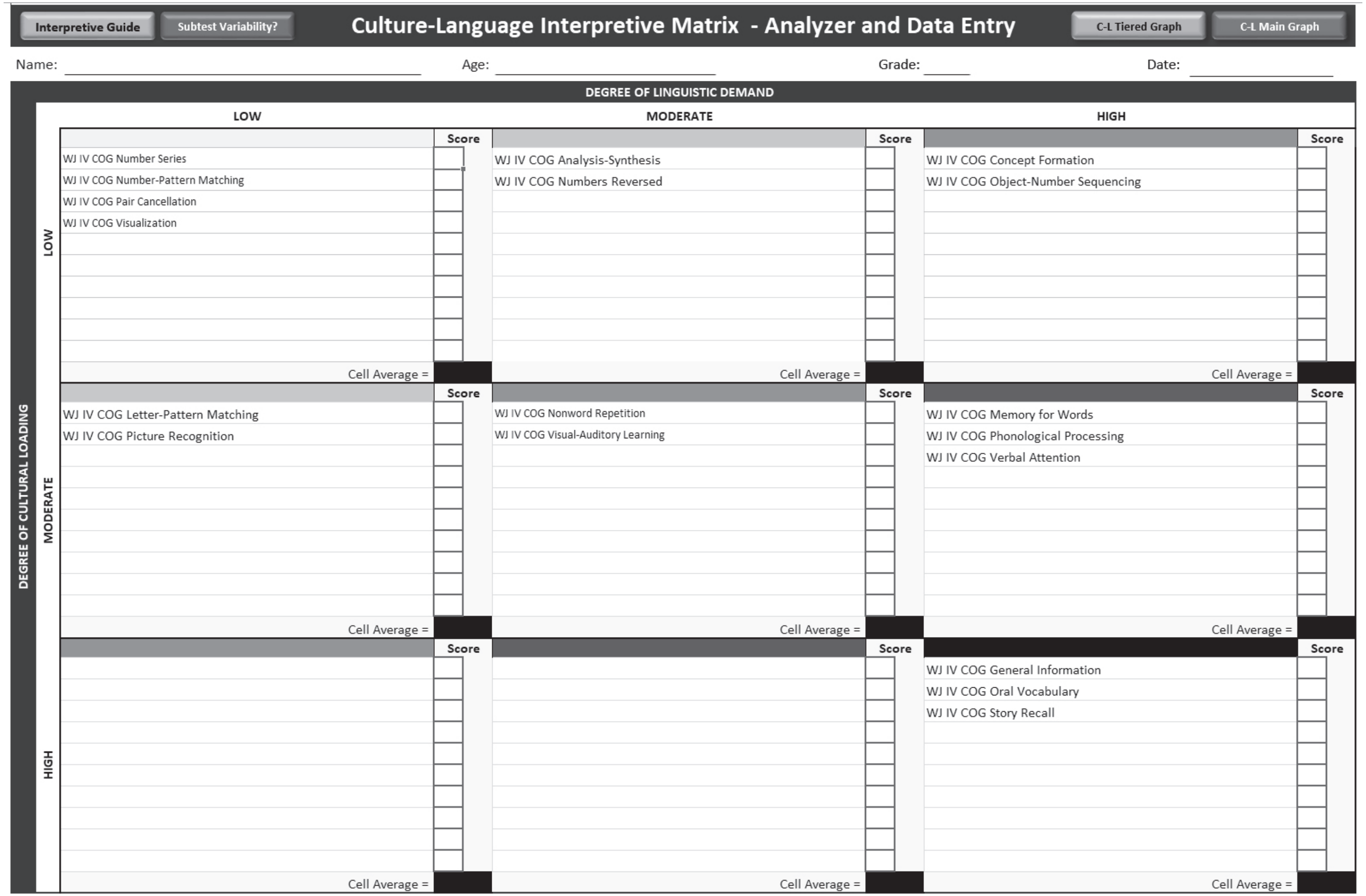
The manner in which standardized, norm-referenced tests included in the C-LTC are organized represents a departure from the more common organization related to the theoretical construct to be measured. The C-LTC categorizes tests only on the basis of the degree to which subtest means indicate that bilingual examinees' performance is attenuated. High attenuation earns classification in the high category; moderate attenuation suggests the moderate category; and little attenuation points to the low category. Classification of tests in this manner is meant to reflect the degree of cultural loading and the extent of linguistic demand that are embedded in a particular subtest and that are responsible for the degree of test score attenuation. In effect, the organization of the C-LTC provides a unique frame of reference from which to view test performance. An example of the C-LTC for various subtests of the Woodcock–Johnson IV Tests of Cognitive Ability (WJ IV COG; Woodcock, McGrew, & Mather, 2001) is presented in Figure 25.3.

As is evident in Figure 25.3, the C-LTC is organized as a matrix, with degree of cultural loading as the variable along the vertical axis and degree of linguistic demand along the horizontal axis. Each variable is subdivided into three levels (low, moderate, and high) that are intended to distinguish the classifications further. In the resulting 3 × 3 matrix, some of the nine cells contain tests that share a particular combination of cultural loading and linguistic demand. The classifications are not based on cognitive ability constructs; that is, two tests within the same cell are not there because they measure the same thing (e.g., visual processing), but rather because research has indicated that they appear to share similar levels of cultural loading and linguistic demand, as manifested in comparable subtest means. Subtests classified as high along both dimensions have relatively lower means, and those classified as low along both dimensions have relatively higher means. A notable feature of the C-LTC is that the arrangement of the tests is dynamic and easily altered to be consistent with new research on the performance of bilingual individuals as it may emerge as well as adaptations in the subtests that compose the tests themselves. The C-LTC classifications do not imply significant differences between subtests classified in one cell or in one category versus another, but rather provides a guide to the order in which performance may be expected to decline as a function of increasing cultural loading and linguistic demands of the tests. The C-LTC for the WJ IV COG, as illustrated in Figure 25.3, thus suggests the following order from highest to lowest value for the first seven subtests of the Standard Battery: Number Series, Visualization, Letter-Pattern Matching, Phonological Processing, Verbal Attention, Oral Vocabulary, and Story Recall.

It was initially thought that the C-LTC would allow practitioners to select tests measuring a particular construct that would have the best chance of producing valid data. Naturally, this meant selecting tests that were classified as low in cultural loading and linguistic demand. As discussed previously, individuals who have had less opportunity for learning about mainstream U.S. culture, or who have a level of English-language proficiency different from that of same-age or same-grade peers who are native speakers, tend to obtain lower scores than the individuals on whom virtually all tests are typically normed (Aguera, 2006; Dynda, 2008; Figueroa, 1990a, 1990b; Hamayan & Damico, 1991; Jensen, 1974; Mercer, 1979; Nieves-Brull, 2006; Sotelo-Dynega, 2007; Tychanska, 2009; Valdés & Figueroa, 1994). Consequently, scores for diverse individuals are expected to be better approximations of true ability on tests that are lower in cultural loading and linguistic demand, and poorer estimates of true ability on tests that are higher in cultural loading and linguistic demand. Unfortunately, use of the C-LTC in selecting tests for administration runs up against some problems. Despite the use of research to guide the classifications, the presumption of validity remains a question not answerable by simply selecting tests that are low in cultural content and linguistic demands. In addition, it quickly became apparent that some abilities, particularly those related to language skills and verbal ability, simply could not be measured through tests that were culturally or linguistically reduced or classified as low on both dimensions. The problem of "difference versus disorder" remained, and it was not until the development of the C-LIM (Flanagan & Ortiz, 2001; Flanagan et al., 2007) that this issue was more fully addressed.

The Culture–Language Interpretive Matrix

The classification of tests on the C-LIM according to shared levels of cultural loading and linguistic demand helped to identify tests that might result in the fairest estimates of true ability for culturally and linguistically diverse individuals. But this turned out to be only one benefit of the manner in which the tests were organized. In reviewing the decades of research on the test performance of bilingual individuals, Flanagan and colleagues (2007) realized that the arrangement of the classifications meant that tests in the upper left cell of the C-LIM (low cultural loading/low linguistic demand) would collectively produce a much higher aggregate score than tests in the lower right cell (high cultural loading/high linguistic demand). Data from numerous studies supported not only the classifications themselves, but also the nature of expected patterns of performance for diverse individuals (Aguera, 2006; Brigham, 1923, 1930; Cummins, 1984; Dynda, 2008; Figueroa, 1990a; Gould, 1996; Jensen, 1974, 1976, 1980; Nieves-



Culture-Language Interpretive Matrix - Analyzer and Data Entry

Name: ______ Age: ______ Grade: ______ Date: ______

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING	LOW	Score	MODERATE	Score	HIGH	Score
LOW	WJ IV COG Number Series WJ IV COG Number-Pattern Matching WJ IV COG Pair Cancellation WJ IV COG Visualization Cell Average =		WJ IV COG Analysis-Synthesis WJ IV COG Numbers Reversed Cell Average =		WJ IV COG Concept Formation WJ IV COG Object-Number Sequencing Cell Average =	
MODERATE	WJ IV COG Letter-Pattern Matching WJ IV COG Picture Recognition Cell Average =		WJ IV COG Nonword Repetition WJ IV COG Visual-Auditory Learning Cell Average =		WJ IV COG Memory for Words WJ IV COG Phonological Processing WJ IV COG Verbal Attention Cell Average =	
HIGH	Cell Average =		Cell Average =		WJ IV COG General Information WJ IV COG Oral Vocabulary WJ IV COG Story Recall Cell Average =	

FIGURE 25.3. The Culture–Language Test Classifications (C-LTC) matrix for the WJ IV Tests of Cognitive Abilities, depicted in a screen capture of the Culture–Language Interpretive Matrix (C-LIM) from the Cross-Battery Assessment Software System (X-BASS; Flanagan, Ortiz, & Alfonso, 2017).

Brull, 2006; Sanchez, 1932, 1934; Sotelo-Dynega, 2007; Tychanska, 2009; Valdés & Figueroa, 1994; Vukovich & Figueroa, 1983; Yerkes, 1921). This pattern is illustrated in Figure 25.4, which depicts the C-LIM.

Although placed in an orthogonal arrangement, the two dimensions in Figure 25.4 are in fact highly correlated because it is not entirely appropriate to separate the effects of culture from language or vice versa. Nevertheless, the arrows in the illustration depict the three possible ways in which the test results of diverse individuals may be attenuated. First, test performance may decrease primarily as a function of the increasing cultural loading of tests. Similarly, test performance may decrease largely as a function of the increasing linguistic demands of tests. And finally, test performance may decrease as a function of the combination of cultural loading and linguistic demand. In practice and research, however, there has not been significant evidence of a singular effect for either culture or language, with the exception that strong, primary language effects have been seen in culturally and linguistically diverse children who also have significant speech–language disorders (Aziz, 2010; Lella, 2010; Tychanska, 2009). Therefore, except for some specific occasions, interpretation of the pattern of test performance via the C-LIM should be limited to examination of the combined effect of both dimensions, and should not be focused on the singular influence of either one alone. This information, coupled with knowledge regarding an individual's cultural and linguistic experience, makes it possible to accomplish defensible interpretation through analysis of the patterns formed by test data collected over the past century.

The value of understanding this declining pattern of performance lies in its empirical base, which provides predictability not only for diverse groups, but also for diverse individuals. The research on the test performance of bilingual individuals reflects a linear and continuous decline in performance on tests as a function of their cultural content and linguistic demands. For example, apart from their examination of the classification of the WJ III COG subtests, Kranzler and colleagues (2010) conducted additional analyses and concluded that a "statistically significant (decreasing) trend was observed for the effect of linguistic demand and cultural loading combined" (p. 431). Despite their concerns stemming primarily from expectations of statistically significant differences, their investigation provides an independent replication of the robust, linearly declining performance of bilingual examinees as subtests increase

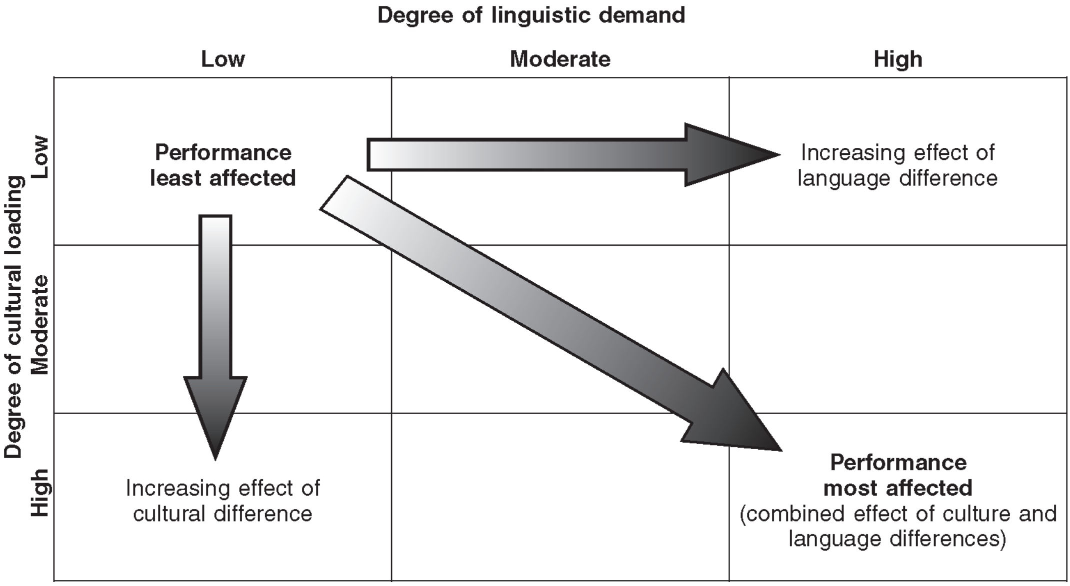


FIGURE 25.4. The Culture–Language Interpretive Matrix (C-LIM): General pattern of expected performance for culturally and linguistically diverse children. From Flanagan and Ortiz (2001). Copyright © John Wiley & Sons, Inc. Reprinted by permission.

their demands for acquired cultural knowledge and developmental language proficiency.

In short, the Kranzler and colleagues (2010) study reinforces the very foundations upon which the C-LIM is built and the principles that guide examination of test score validity. To evaluate issues of validity, an individual's obtained subtest scores are classified within the cells specified by the C-LTC and are then aggregated to create values across the matrix. These mean values then permit closer examination and more importantly, comparison against the means obtained for other bilingual examinees as reported in the literature. For example, if the pattern of aggregate scores within the matrix approximates the declining pattern of scores (in terms of both magnitude and rate of decline) derived from the literature, the results can be said to be invalid, in that they are reflecting primarily the effects of cultural and linguistic influences—not the constructs the subtests were intended to measure. Conversely, if the pattern of aggregate scores within the matrix does not approximate the pattern of scores derived from the literature (i.e., the magnitude of scores is lower than the range predicted, or there is an absence of systematic decline as linguistic and cultural demands increase), then the results can be said to be valid, in that they are reflecting the primary influence of a variable or variables other than those related to cultural or linguistic differences. Use of the term *primary* is important here because cultural and linguistic differences may never be completely absent in such cases, and may well be contributory factors in most all situations where bilingual examinees are concerned. However, any potential deficits must be identified on the basis of performance that cannot be attributed primarily to cultural or linguistic difference. This is where the C-LIM provides significant utility to practitioners. Failure to identify a clear pattern of decline within the expected range for bilingual individuals implies that cultural or linguistic factors cannot be viewed as the primary or only factors affecting the results (although they may be contributing to some part of the pattern), and this strongly suggests that the test results are valid.

However, an extremely important caveat in using the C-LIM is that there are many other variables that might have affected test score performance apart from the possibility of cognitive deficits or disorder (e.g., lack of motivation, emotional disturbance, incorrect scoring or administration). As such, the lack of a declining pattern does not automatically indicate disorder, and any diagnosis involving deficient ability must be made by excluding other potential explanations and making use of corroborating data. In any event, the C-LIM appears to provide a solid, evidence-based method for establishing test score validity and for helping practitioners ask and answer the question of "difference versus disorder." Moreover, two particular advantages of the C-LIM make it exceedingly practical. First, the use of testing in English allows it to remain accessible to all practitioners; second, once applied to evaluate the validity of obtained test scores, it permits use of any interpretive method, schema, or framework with which a practitioner may already be familiar or comfortable. Thus, apart from learning how to use and apply it, the C-LIM does not require alteration of procedures that virtually every practitioner already knows and uses.

A complete discussion of the C-LTC and C-LIM is beyond the scope of this chapter. The reader is referred to the original sources for better and more detailed guidance on their use and application in testing practices. In addition, despite the research base upon which these approaches have been developed, neither of them should be relied upon as the only method for establishing validity and making defensible interpretations. Rather, both of these approaches are intended to supplement those assessment and evaluation practices already in use by practitioners. They are designed and intended to bring more rigor and defensibility to current testing practices, not to replace them entirely. When used in conjunction with other relevant assessment data and information (e.g., direct observations, review of records, interviews, language proficiency testing, socioeconomic status, developmental data, family history), these methods should assist in bringing assessment and testing procedures into accordance with current calls for evidence-based practice.

INNOVATIONS IN TEST DEVELOPMENT

As emphasized throughout this chapter, significant measurement biases arise when traditional testing and assessment batteries are applied to culturally and linguistically diverse individuals. Construct-irrelevant components emerge, fostering unfair testing conditions and producing invalid results. However, key practice guidelines have been established and detailed in the publication of the 2014 *Standards*, and these guidelines offer the beginnings of a roadmap toward creating and implementing more reliable and valid testing procedures. Moreover, as discussed in the previous section, the research-supported C-LTC and C-LIM now provide perhaps the most concrete, data-driven recommendations for choosing cognitive subtests that offer the greatest chance of producing valid test results for individuals from linguistically and culturally diverse backgrounds. These research-based approaches highlight that recognizing the source of potential bias and adopting tests that seek to minimize potential bias are the first steps in ensuring validity. Although the C-LTC and C-LIM lack sufficient empirical support to stand alone as viable or defensible evidence-based practices, they can serve as an excellent foundation for the progression of the field and the push toward evidence-based assessment for ELLs.

Efforts to Overcome Language-Based Limitations

Contemporary assessment practices have emerged that build on key components of the C-LTC and C-LIM. An example of these practices is allowing assessment of vocabulary acquisition, for which reliability and validity can be maintained for culturally and linguistically diverse individuals, particularly ELLs. It is not surprising that many of the most widely used batteries of intelligence testing incorporate measures of vocabulary acquisition. Vocabulary acquisition is a key component of verbal ability, which has been analytically proven to account for 20–50% of variance in measures of general intelligence or ability, depending on the number of factors in the model being tested (Flanagan, Ortiz, & Alfonso, 2017; Schrank, Decker, & Garruto, 2016); it also relates to reading and writing skills, receptive and expressive language ability, general academic attainment, broad linguistic proficiency, and competence (Cattell, 1943; Schneider & McGrew, 2012; Thorndike, 1914). Aside from associations with intelligence, vocabulary is a foundational and inherent aspect of language, necessary in perhaps all assessments and evaluations, given the need for communication in even so-called "nonverbal" tests. It is therefore not surprising that measurement bias and fairness concerns arise when traditional tests for native English speakers are used to assess ELLs. It is also not surprising that attempts to rid intelligence tests of language components (e.g., the Army Beta test [Yerkes, 1921]; Performance IQ [Wechsler, 1939]), and those aimed at managing language through the creation of instruments in languages other than English (Schlueter, Carlson, Geisinger, & Murphy, 2013), have historically fallen short.

Perhaps the most compelling attempt to overcome language-based limitations is the creation of true "bilingual" tests (e.g., the Bilingual English–Spanish Assessment [BESA]; Pena, Gutierrez-Clellan, Iglesias, Goldstein, & Bedore, 2014). The BESA employs a unique normative sample based fully on bilingual English–Spanish speakers, and it provides guidelines on language use for administration in either language or only one. Although the BESA is an innovative and concentrated effort to reduce common measurement and testing biases, its testing format is still not without significant shortcomings. For example, while individuals are granted the opportunity to demonstrate both their English- and native-language proficiency, the age range of individuals for whom the test is appropriate is narrow; the evaluator is required to be bilingual, or a trained interpreter is necessary, which removes the evaluator from the clinical aspects of testing; and the test is limited to Spanish–English learners. Although the BESA is thus a step in the right direction, fairness and validity concerns remain. Combined, the BESA and C-LTC/C-LIM highlight the need for test procedures to employ appropriate peer normative samples and to control for construct-irrelevant factors, particularly amount of English-language exposure.

Group-Specific Norms

The development of novel, valid testing methods striving for fairness for non-native speakers of English is complex and requires careful consideration of test innovations and associated limitations. Dual norming may be the perfect answer to achieve a reliable, valid test assessment that not only is consistent with the *Standards* as outlined earlier in this chapter, but also builds on effective data-driven practices such as the C-LTC and C-LIM. The use of more than one normative sample for a particular test may be a key to producing the fairest and most valid test results possible for culturally and linguistically diverse individuals, in that it allows such individuals to be compared to a "true peer" group. More specifically, the creation of one standardized test with accompanying administration procedures that include two distinct normative samples—one for native English speakers and a separate one for ELLs—begins to address some of the psychometric issues discussed throughout this chapter that are inherent in obtaining valid test results for ELLs. Specifically, it begins to account for differences in English exposure and other unique aspects of the linguistic and cultural backgrounds of ELLs that too often contaminate testing results. Additionally, adoption of dual-norming approaches has the significant advantage of providing diagnostic capability for ELLs. Traditional single-norm tests limit practitioners' capacity to diagnose disability in English learners, as comparison to normative samples of native English speakers is inherently biased and discriminatory for the reasons previously discussed. Dual-norming techniques thus free practitioners from the confines of evaluating solely for instructional need or intervention intensity. The value of providing ELL-specific norms, and the utility for practitioners in being able to draw valid interpretations from test results even when these are obtained by using English, are illustrated in Figures 25.5 and 25.6.

Figure 25.5 depicts the predicted performance on the Ortiz Picture Vocabulary Acquisition Test (PVAT; Ortiz, 2017) of individuals grouped by level of English exposure (i.e., ELLs with low, moderate, and high exposure, and native English speakers). As is evident from the curves, the predicted mean scores (based on the performance of each group, taken as an average) increase not only with age from 2 to 22 years, but also with level of English exposure. Native English speakers are predicted to have the highest scores at each age, followed by the high-, moderate-, and low-exposure groups, respectively and without variance. English exposure thus emerges as a construct-irrelevant factor implicated in test score achievement. Likewise, Figure 25.6 illustrates the differences between scores similarly based on level of English exposure. At first glance at the graph on the left, the figure appears to illustrate disordered performance of non-native speakers of English. However, this is primarily due to the ELLs' performance being compared to native English speakers' performance

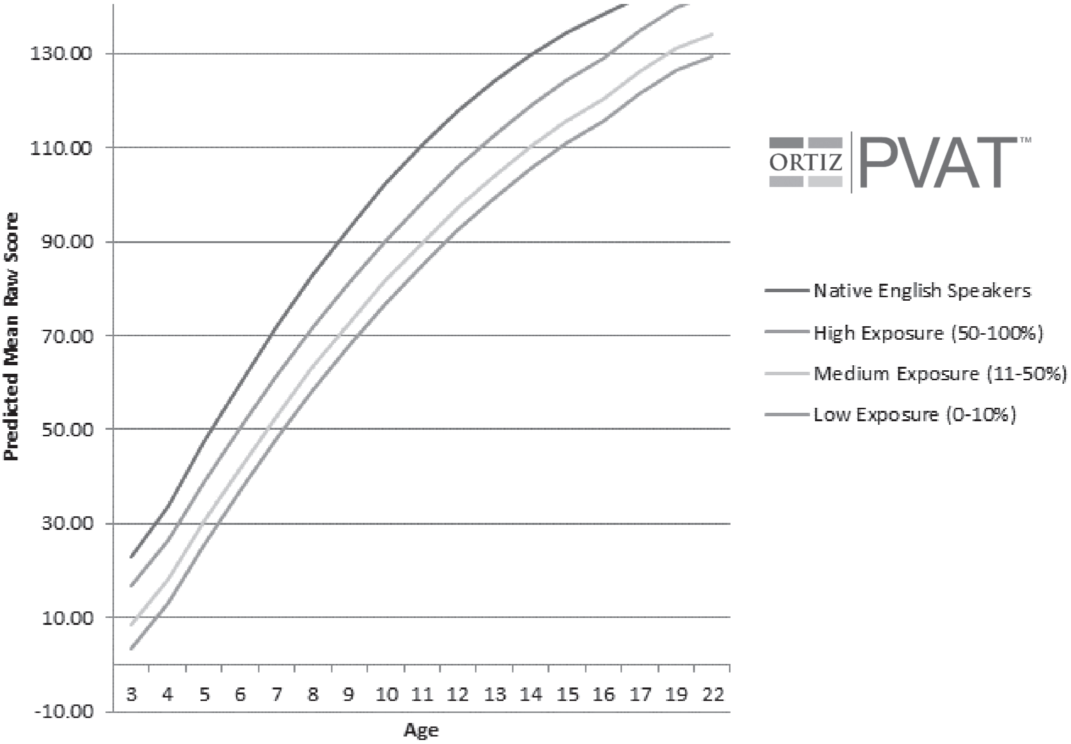


FIGURE 25.5. Comparison of performance on the Ortiz Picture Vocabulary Acquisition Test (PVAT) by samples of English learners with low, medium, and high exposure to English, and of native English speakers. From Ortiz (2017). Copyright © Multi-Health Systems. Used by permission.

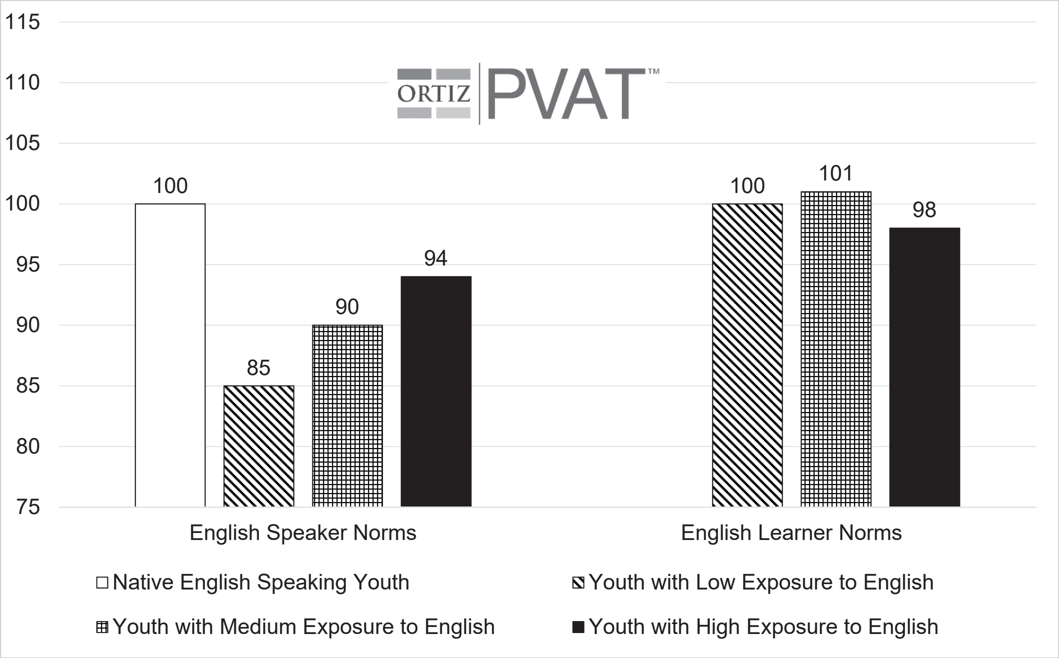


FIGURE 25.6. (Left) Comparison of performance on the Ortiz PVAT by the same four samples as in Figure 25.5, based on norms for native English speakers. (Right) Comparison of the three groups of English learners, based on norms for English learners (controlled for length of English exposure). From Ortiz (2017). Copyright © Multi-Health Systems. Used by permission.

without controls for developmental differences in English-language exposure and opportunity for acculturative knowledge acquisition. The graph on the right shows the performance of the three ELL groups according to norms for English learners, with the application of such controls. It becomes evident that comparison of non-native speakers of English to native speakers is discriminatory, and that the potential for misidentification of a disorder is stark. Native English speakers simply do not represent a "true peer" comparison group for ELLs. The absence of controls for differential linguistic and developmental experiences typical of ELLs in the United States permeates even native-language tests.

By controlling for differential amounts of English exposure, the creation of group-specific norms for ELLs effectively provides the necessary "true peer" foundation for making valid comparisons of performance. In this way, the use of a single test that can be administered to any individual who speaks English (as a native or otherwise) not only yields efficiency in the evaluation process, but directly addresses the "difference versus disorder" dilemma in a manner that is easily accessible and consistent with current practices and procedures. Indeed, the use of dual norms is an innovative approach to the issue of test score validity and may serve as a model for future test development. Considering that more than 350 languages are spoken in the United States alone (U.S. Census Bureau, 2015), producing adaptations or translations of tests into all these other languages is simply untenable and impractical. Rather, creating tests that serve dual purposes, and that remain accessible to and usable by every practitioner (whether bilingual or not), is a promising avenue for test development—one that seeks to embrace the concept of universal design as recommended in the *Standards,* as well as to meet the need for ensuring test score validity.

SUMMARY

The purpose of this chapter has been to present a discussion of the main issues facing practitioners seeking to evaluate the cognitive abilities of diverse individuals in a manner that integrates research and results in evidence-based practice. Practitioners must have a solid understanding of these issues, in order to integrate them into current practice in a manner that is not only comprehensive and systematic but guided by the available scientific literature. Fair and equitable assessment is accomplished via recognition of the nature and sources of potential bias, and by application of methods and procedures that are specifically designed to ensure validity. Toward that end, cognitive assessment of culturally and linguistically diverse individuals must move beyond traditional practices that lack defensibility or utility in addressing problems with validity, and must dispense with simplistic notions that recognition of a verbal–nonverbal pattern of differences is sufficient. Integration of research into practice appears to necessitate a good grasp of the basic influences on test performance including developmentally based cultural and linguistic differences and their relation to tests and testing, such as norm sample representation, degree and patterns of score attenuation, and test characteristics and demands. Existing approaches lack sufficient empirical support to stand alone as viable or defensible evidence-based practice and require supplementation via emerging methods that are designed specifically to examine issues of validity. Without application of frameworks that permit direct inspection of score validity, use of any particular approach—whether testing in an examinee's native language, testing in English, nonverbal testing, modified/adapted testing, or any combination of them—will remain limited in the extent to which the obtained results can be interpreted validly as measures or estimates of true ability. Success in these endeavors may be facilitated by emerging research-based techniques, including the C-LTC and C-LIM. Such approaches, as well as other contemporary testing approaches (e.g., dual norming), appear to hold significant promise for elevating current assessment practices to meet the call for evidence-based assessment.

Whether current practitioners and researchers realize it or not, differences in the cultural and linguistic backgrounds and experiences of diverse individuals have always posed and will continue to pose serious threats to the validity and meaning of test results. As we have discussed, the type of bias that stems from such differences is not related to any technical or psychometric flaws within the tests themselves, but rather primarily to violations of the assumption of comparability. Individuals who are culturally and linguistically different cannot be held to the same expectations as those for their same-age or same-grade peers when their linguistic or acculturative development is not comparable to theirs; unfortunately, this is what is done when performance standards are based on mainstream, monolingual English speakers. Questions regarding validity are crucial, if not central, to the task of evaluating individuals from diverse cultural and linguistic backgrounds. And such complicated questions will not be answered via simple prescriptions. Bilingualism and level of acculturation are such complex concepts that practitioners should not expect that the question of how to conduct nondiscriminatory assessment can ever be reduced to the question of what is the "right" test to use. Sattler (1992) notes:

> Probably no test can be created that will entirely eliminate the influence of learning and cultural experiences. The test content and materials, the language in which the questions are phrased, the test directions, the categories for classifying the responses, the scoring criteria, and the validity criteria are all culture bound. (p. 579)

At the core of any evaluation lies the issue of validity, and this issue, more than any other, will require significant attention and scrutiny because it alone provides the defensible foundation upon which meaning can be ascribed confidently to obtained results. In the measurement of cognitive abilities in individuals from culturally and linguistically diverse backgrounds, the question of validity will always be tied directly to the success of the chosen methods and procedures in discerning the influence of cultural and linguistic differences on test performance.

REFERENCES

Aguera, F. (2006). *How language and culture impact test performance on the Differential Ability Scales in a preschool population.* Unpublished doctoral dissertation, St. John's University, Jamaica, NY.

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychological Association. (1990). *Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations.* Washington, DC: Author.

Aziz, N. (2010). *English language learners with global cognitive impairment: Evaluation of patterns within the*

Culture–Language Interpretive Matrix. Unpublished doctoral dissertation, St. John's University, Jamaica, NY.

Braden, J. P., & Iribarren, J. A. (2007). Test review: Wechsler, D. (2005). Wechsler Intelligence Scale for Children—Fourth Edition Spanish. *Journal of Psychoeducational Assessment, 25,* 292–299.

Brigham, C. C. (1923). *A study of American intelligence.* Princeton, NJ: Princeton University Press.

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist, 52,* 1046–1050.

Carroll, J. B. (1993). *Human cognitive abilities.* New York: Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cummins, J. C. (1984). *Bilingual and special education: Issues in assessment and pedagogy.* Austin, TX: PRO-ED.

Dynda, A. M. (2008). *The relation between language proficiency and IQ test performance*. Unpublished manuscript, St. John's University, Jamaica, NY.

Figueroa, R. A. (1983). Test bias and Hispanic children. *Journal of Special Education, 17,* 431–440.

Figueroa, R. A. (1989). Psychological testing of linguistic-minority students: Knowledge gaps and regulations. *Exceptional Children, 56,* 145–152.

Figueroa, R. A. (1990a). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 691–696). New York: Guilford Press.

Figueroa, R. A. (1990b). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology II* (pp. 93–106) Washington, DC: National Association of School Psychologists.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment.* New York: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2017). *The cross-battery assessment software system, version 2.0* (X-BASS v2.0). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2006). *The achievement test desk reference (ATDR): A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Goddard, H. H. (1913). The Binet tests in relation to immigration. *Journal of Psycho-Asthenics, 18,* 105–107.

Goldenberg, C. (2008, Summer). Teaching English language learners what the research does—and does not—say. *American Education*, pp. 8–44.

Gottfredson, L. S. (1997). Mainstream science on intelligence [Editorial]. *Intelligence, 24,* 13–23.

Gould, S. J. (1996). *The mismeasure of man* (rev. ed.). New York: Norton.

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Macmillan.

Hamayan, E. V., & Damico, J. S. (1991). *Limiting bias in the assessment of bilingual students.* Austin, TX: PRO-ED.

Harris, J. G., & Llorente, A. M. (2005). Cultural considerations in the use of the Wechsler Intelligence Scale—Fourth Edition (WISC-IV). In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 381–413). Burlington, MA: Elsevier.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 41–68). New York: Guilford Press.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review, 39*(1), 1–123.

Jensen, A. R. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs, 90,* 185–244.

Jensen, A. R. (1976). Construct validity and test bias. *Phi Delta Kappan, 58,* 340–346.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Kranzler, J. H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review, 39*(3), 431–446.

Lella, S. (2010). *Evaluating speech–language and cognitive impairment patterns via the Culture-Language Interpretive Matrix.* Unpublished doctoral dissertation. St. John's University, Jamaica, NY.

Llorente, E., & Sheingold, D. (2010, September 15). Minorities now nearly half of Bergen County children, and just over half in the state. *Passaic County News.*

Lohman, D. F., Korb, K., & Lakin, J. (2008). Identifying academically gifted English language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly, 52,* 275–296.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

McShane, D. (1980). A review of scores of American Indian children on the Wechsler intelligence scales. *White Cloud Journal, 1,* 3–10.

Mercer, J. R. (1979). *System of multicultural pluralistic assessment: Technical manual.* New York: Psychological Corporation.

Mpofu, E., & Ortiz, S. O. (2009). Equitable assessment practices in diverse contexts. In E. L. Grigorenko (Ed.), *Assessment of abilities and competencies in the era of globalization* (pp. 41–76). New York: Springer.

Naglieri, J. A. (1982). Does the WISC-R measure verbal intelligence for non-English speaking children? *Psychology in the Schools, 19,* 478–479.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nieves-Brull, A. (2006). *Evaluation of the Culture–Language Matrix: A validation study of test performance in monolingual English speaking and bilingual English/Spanish speaking populations.* Unpublished manuscript, St. John's University, Jamaica, NY.

Oakland, T. (Ed.). (1976). *Nonbiased assessment of minority group children.* Lexington, KY: Coordinating Office for Regional Resource Centers.

Oakland, T., & Laosa, L. (1976). Professional, legislative, and judicial influences on psychoeducational assessment practices in schools. In T. Oakland (Ed.), *Nonbiased assessment of minority group children.* Lexington, KY: Coordinating Office for Regional Resource Centers.

Ochoa, S. H., Riccio, C. A., Jimenez, S., Garcia de Alba, R., & Sines, M. (2004). Psychological assessment of limited English proficient and/or bilingual students: An investigation of school psychologists' current practices. *Journal of Psychoeducational Assessment, 22,* 93–105.

Ormrod, J. E. (2008). *Educational psychology: Developing learners.* Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.

Ortiz, S. O. (2001). Assessment of cognitive abilities in Hispanic children. *Seminars in Speech and Language, 22*(1), 17–37.

Ortiz, S. O. (2004). Use of the WISC-IV with culturally and linguistically diverse populations. In D. P. Flanagan & A. S. Kaufman, *Essentials of WISC-IV assessment* (pp. 183–215). Hoboken, NJ: Wiley.

Ortiz, S. O. (2014). Best practices in nondiscriminatory assessment. In P. Harrison & A. Thomas (Eds.), *Best practices in school psychology VI: Foundations* (pp. 61–74). Bethesda, MD: National Association of School Psychologists.

Ortiz, S. O. (2017). *Technical manual: Ortiz Picture Vocabulary Acquisition Test (PVAT).* North Tonawanda, NY: Multi-Health Systems.

Ortiz, S. O., & Dynda, A. M. (2010). Diversity, fairness, utility and social issues. In E. Mpofu & T. Oakland (Eds.), *Assessment in rehabilitation and health* (pp. 37–55). Upper Saddle River, NJ: Merrill.

Ortiz, S. O., & Flanagan, D. P. (1998). Gf-Gc cross-battery interpretation and selective cross-battery assessment: Referral concerns and the needs of culturally and linguistically diverse populations. In K. S. McGrew & D. P. Flanagan, *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment* (pp. 401–444). Boston: Allyn & Bacon.

Pena, E. D., Gutierrez-Clellan, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *BESA: Bilingual English-Spanish Assessment manual.* San Rafael, CA: AR-Clinical Publications.

Reynolds, C. R., & Ramsay, M. C. (2003). Bias in psychological assessment: An empirical review and recommendations. In I. B. Weiner (Series Ed.) & J. R. Graham & J. A. Naglieri (Vol. Eds.), *Handbook of psychology: Vol. 10. Assessment psychology* (pp. 67–94). Hoboken, NJ: Wiley.

Rhodes, R., Ochoa, H. S., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide.* New York: Guilford Press.

Salvia, J., & Ysseldyke, J. (1991). *Assessment in special and remedial education* (5th ed.). Boston: Houghton Mifflin.

Sanchez, G. I. (1932). Group differences and Spanish-speaking children: A critical review. *Journal of Applied Psychology, 16,* 549–558.

Sanchez, G. I. (1934). Bilingualism and mental measures: A word of caution. *Journal of Applied Psychology, 18,* 756–772.

Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J. D., & Grenier, J. R. (Eds.). (1998). *Test interpretation and diversity: Achieving equity in assessment.* Washington, DC: American Psychological Association.

Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego, CA: Author.

Sattler, J. M. (2001). *Assessment of children: Behavioral and clinical applications* (4th ed.). San Diego: Author.

Schlueter, J. E., Carlson, J. F., Geisinger, K. F., & Murphy, L. M. (Eds). (2013) *Pruebas Publicadas en Espanol: An index of Spanish tests in print.* Lincoln: University of Nebraska Press.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of the WJ IV Cognitive Abilities Assessment.* New York: Wiley.

Smith, D. (2008, August 7). In the under-5 set, minority becoming the majority. *Washington Post.* Retrieved from *www.washingtonpost.com/wp-dyn/content/article/2008/08/06/AR2008080603683.html.*

Sotelo-Dynega, M. (2007). *Cognitive performance and the development of English language proficiency.* Unpublished doctoral dissertation, St. John's University, Jamaica, NY.

Sotelo-Dynega, M., Cuskley, T., Geddes, L., McSwiggan, K., & Soldano, A. (2011). *Cognitive Assessment: A survey of current school psychologists' practices.* Poster presented at the annual conference of the National Association of School Psychologists, San Francisco, CA.

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*(10), 1031–1045.

Styck, K. M., & Watkins, M. W. (2013). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities, 49*(2), 216–224.

Styck, K. M., & Watkins, M. W. (2014). Structural validity of the WISC-IV for students with ADHD. *Journal of Attention Disorders, 21*(11), 921–928.

Terman, L. M. (1916). *The uses of intelligence tests.* Boston: Houghton Mifflin.

Thorndike, E. L. (1914). *The measurement of ability in reading, preliminary scales and tests.* New York: Teachers College, Columbia University.

Tychanska, J. (2009). *Evaluation of speech and language impairment using the Culture–Language Test Classifications and Interpretive Matrix.* Unpublished doctoral dissertation, St. John's University, Jamaica, NY.

U.S. Census Bureau. (2010). American community survey 2010. Washington, DC: U.S. Government Printing Office. Retrieved April 7, 2014, from *www.census.gov/acs/www/data/data-tables-and-tools/american-factfinder.*

U.S. Census Bureau. (2014). *American community survey 2014: 1-year estimates.* Washington, DC: U.S. Government Printing Office. Retrieved May 6, 2015, from *http://factfinder2.census.gov.*

U.S. Census Bureau. (2015). Census Bureau reports at least 350 languages spoken in U.S. homes. Retrieved from *www.census.gov/newsroom/press-releases/2015/cb15-185.html.*

U.S. Census Bureau. (2017, August 28). More than 77 million people enrolled in U.S. schools. Retrieved September 11, 2017, from *www.census.gov/newsroom/press-releases/2017/school-enrollment.html.*

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias.* Norwood, NJ: Ablex.

Vukovich, D., & Figueroa, R. A. (1983). *The validation of the system of multicultural pluralistic assessment: 1980–1982.* Unpublished manuscript, University of California at Davis.

Wechsler, D. (1939). *Wechsler Bellevue Intelligence Scale.* New York: Psychological Corporation.

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children—Fourth Edition Spanish.* San Antonio, TX: Psychological Corporation.

Weiss, L. G., Harris, J. G., Prifitera, A., Courville, T., Rolfhus, E., Saklofske, D. H., et al. (2006). WISC-IV interpretation in societal context. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 1–58). Burlington, MA: Academic Press.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

Yerkes, R. M. (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences, 15,* 1–890.

PART V

Linking Assessment Data to Intervention

CHAPTER 26

Cognitive Hypothesis Testing

Linking Test Results to the Real World

Catherine A. Fiorello
Kirby L. Wycoff

IMPORTANCE OF COMPREHENSIVE COGNITIVE ASSESSMENT

This chapter presents a model, *cognitive hypothesis testing* (CHT), which is a balanced approach to assessment and intervention (Hale & Fiorello, 2004). Cognitive assessment rests on a foundation of cognitive, neuropsychological, and neuroscientific research, and can inform our interventions with children displaying academic and behavioral difficulties. Research into the neuropsychology of specific learning disabilities (SLD) has made tremendous strides over the last 30 years (Hale, Fiorello, Miller, et al., 2008; Hale, Wilcox, & Reddy, 2016), and provides support for cognitive and neuropsychological assessment that can lead to the development of truly individualized education programs.

The CHT model described in this chapter encourages the use of response to intervention (RTI) as a *prevention* and *intervention* program, although not a *diagnostic* one (Hale, Wycoff, & Fiorello, 2010). We hope that a balanced practice model that includes CHT will provide the basis for a new paradigm shift—a multi-tiered model that combines academic and behavioral interventions for all children who need them, with comprehensive psychoeducational evaluations for those who do not respond to those interventions (Decker, Hale, & Flanagan, 2013; Hale, 2006; Hale, Chen, et al., 2016). For students who do not respond to increasingly intensive intervention approaches, a comprehensive evaluation that includes assessment of cognitive processing will allow us to examine their learning and psychosocial needs, as well as to determine whether or not they have SLD (Backenson et al., 2015; Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012; Fiorello, Hale, Snyder, Forrest, & Teodori, 2008; Fuchs & Deshler, 2007; Hale, Fiorello, Dumont, et al., 2008; Hain, Hale, & Glass-Kendorski, 2009; Miller & Hale, 2008). This model acknowledges that students may fail to respond to interventions for a variety of reasons—one of which is SLD, a set of neuropsychological disabilities requiring individualized instruction designed to meet students' unique needs (Fenwick et al., 2016; Hale, Flanagan, & Naglieri, 2008). Students who do not respond to RTI service delivery approaches are a heterogeneous group, and a comprehensive CHT evaluation of neuropsychological relationships will aid in both differential diagnosis and intervention for these children.

Diagnosis

The Education for All Handicapped Children Act (Pub. L. No. 94-142, 1975) defined SLD as a deficit in the basic psychological processes that adversely affects academic achievement, and its regulatory language required the presence of a severe discrepancy between ability and achievement. The formal definition was deemphasized, and instead the regulatory language of "severe discrepancy" was emphasized (e.g., Mercer, Jordan, Alsopp, & Mercer, 1996), as it was in most subsequent research and practice. Unfortunately, the process deficits presumed to underlie the academic difficulties were not given the same attention as the ability–achievement discrepancy (AAD)—partly because there was little consensus as to what processes were involved in learning, and partly because good measures of basic psychological processes were not available at the time (Hale & Fiorello, 2004). Much has changed since then, as there are now many well-standardized tools with excellent technical quality to assess psychological processes (Hale & Fiorello, 2004; Hale, Wycoff, & Fiorello, 2010).

As the SLD definition and implementation of the AAD method differed among districts and states (Reschly & Hosp, 2004) and identification of processing disorders was not required, SLD became anything but "specific," and there was an enormous increase in prevalence (MacMillan, Gresham, Lopez, & Bocian, 1996; MacMillan & Speece, 1999). Students with SLD eventually accounted for about half of all special education students (Kavale, Holdnack, & Mostert, 2005). In response, there were calls for a paradigm shift (e.g., Reschly & Ysseldyke, 2002) that would focus on intervention for learning problems rather than diagnosis of them. There are unquestionably problems with the AAD; in particular, the lack of specificity and sensitivity of measures leads to overclassification (e.g., MacMillan, Siperstein, & Gresham, 1996). We do not believe that this condemns all cognitive assessment, however. Instead, we argue that we need to use our well-standardized cognitive assessment tools more wisely, within the context of a comprehensive CHT approach (Fiorello, Hale, & Snyder, 2006; Hale & Fiorello, 2004).

RTI advocates suggest that many learning problems are due to delays that can be remedied by more intensive instruction (e.g., Barnett, Daly, Jones, & Lentz, 2004). We agree; however, there are many reasons why children might not respond to intervention, only one of which is SLD (Hale, Kaufman, Naglieri, & Kavale, 2006). SLD and other neuropsychological disorders are characterized by specific deficits, not just delays (Berninger & Richards, 2002; Collins & Rourke, 2003; Compton et al., 2012; Fenwick et al., 2016; Fiez & Petersen, 1998; Filipek, 1999; Fine, Semrud-Clikeman, Keith, Stapleton, & Hynd, 2007; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Geary, Hoard, & Hamson, 1999; Hale, Chen, et al., 2016; Hale & Fiorello, 2004; Kubas et al., 2014; Naglieri & Bornstein, 2003; Nicholson & Fawcett, 2001; Pugh et al., 2000; Shaywitz, Lyon, & Shaywitz, 2006; Simos et al., 2005; Stein, 2001). These specific deficits are markers that differentiate SLD from simple delays or lack of instruction, and can also identify subtypes of SLD. In a study of 684 students across grades 3–5, Compton and colleagues (2012) found distinct patterns of strengths and weaknesses; they thus showed that the individual differences found in SLD, which have clinical utility, are lost when disparate individuals are collapsed into a single heterogeneous SLD sample. Meta-analysis also supports the deficit model, with phonological awareness, processing speed, working memory, and executive function deficits being the strongest indicators of SLD status (Johnson, Humphrey, Mellard, Woods, & Swanson, 2010). These deficits require comprehensive evaluation, according to the law (e.g., Dixon, Eusebio, Turton, Wright, & Hale, 2011), and a survey of renowned SLD experts interprets this to mean use of a *pattern of strengths and weaknesses* (PSW) approach to identification of SLD (Hale, Alfonso, et al., 2010).

Children with SLD have specific cognitive deficits in the presence of cognitive integrities that can be ascertained during comprehensive evaluations with well-standardized cognitive and neuropsychological tests (Fiorello, Hale, & Snyder, 2006; Hale, 2006). Research by Hale and colleagues has demonstrated significant profile variability, and more predictive validity for subtests over global composites like IQ, for students with SLD, attention-deficit/hyperactivity disorder (ADHD), and traumatic brain injury (Elliott, Hale, Fiorello, Dorvil, & Moldovan, 2010; Fiorello, Hale, McGrath, Ryan, & Quinn, 2001; Fiorello et al., 2006, 2007; Hale, Fiorello, Dumont, et al., 2008; Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007), indicating that empirically based interpretation of strengths and weaknesses is necessary for these students. Research on the diagnostic role of cognitive and neuropsychological assessment will continue to be important, especially for students who do not respond to intervention (Berninger, 2006; Hale et al., 2006; Hale, Fiorello, Dumont, et al., 2008; Kavale et al., 2005; Semrud-Clikeman, 2005; Willis & Dumont, 2006).

As researchers have moved away from AAD, and have started to develop a better understanding of PSW in this heterogeneous population, studies have shown that children with reading SLD may have similar deficits but multiple causes, with subtypes linked to specific PSW (Feifer & Della Tofallo, 2007). These deficits include auditory speech processing (Boets et al., 2011), phonological processing (Cho et al., 2015; Fletcher et al., 2011; Morris et al., 1998; Ramus et al., 2003), visual orthographic processing (Facoetti et al., 2009), integration/mapping of sounds and letters (the alphabetic principle; Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009), rapid automatic naming (Cho et al., 2015; Fletcher et al., 2011; Morris et al., 1998; Norton & Wolf, 2012), processing speed (Morris et al., 1998), working memory (Morris et al., 1998; Swanson, 2011), fluid reasoning (Fletcher et al., 2011), and receptive and expressive language, verbal knowledge, and following directions (Cho et al., 2015; Fletcher et al., 2011; Hulme & Snowling, 2011). Similarly, various subtypes of SLD in written expression have been identified (Fenwick et al., 2016), including ones characterized by weaknesses in working memory, processing speed, left-hemisphere (crystallized) functions, right-hemisphere (fluid) functions, and executive functions.

Students with math SLD have not been studied as extensively, but there too research has identified a number of processing weaknesses in these students compared to low achievers. Students with math SLD show weaknesses in memory retrieval, rapid automatized naming, and processing speed (Cowan & Powell, 2014; Geary, 2011), working memory (Cowan & Powell, 2014; Geary, 2011), fluid reasoning (Cowan & Powell, 2014), number sense and estimating (subitizing and associating numerals with sets; Geary, 2011), oral language (Cowan & Powell, 2014), procedural knowledge (Geary, 2011), and number systems knowledge (Cowan & Powell, 2014).

The Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004; Pub. L. No. 108-446) no longer requires an AAD for identification of SLD, allows the use of an RTI methodology, and also allows what has become known as a *third-method* approach. The final IDEA 2004 regulations (34 C.F.R. Parts 300 and 301; U.S. Department of Education, 2006) indicate that states "may permit the use of other alternative research-based procedures" (p. 46786) for identifying SLD. Although the language is nonspecific, it has typically been interpreted to refer to the PSW approach to identification (Flanagan, Fiorello, & Ortiz, 2010; Hale, Flanagan, & Naglieri, 2008) and is considered an appropriate alternative for SLD classification by several state boards of education (Zirkel & Thomas, 2010). Evaluation of cognitive and neuropsychological processing, as well as academic achievement, for PSW that are diagnostic of SLD is therefore consistent with the IDEA 2004 SLD definition (Fiorello et al., 2008; Hale, Fiorello, Dumont, et al., 2008; Kavale et al., 2005).

When we look more closely at specific types of SLD, we see clear differences in subtype profiles, which further supports the need for an assessment method that examines processing strengths and weaknesses. Students with different strengths and weaknesses do not perform comparably on measures, and make different cognitive and academic errors based on their PSW (Flanagan & Mascolo, 2016; Koriakin et al., 2017). Unlike AAD or RTI outcome variables, understanding PSW requires a careful examination of within- and between-subtest differences. In particular, understanding the pattern of response and error analysis holds a critical key for understanding the disability and guiding intervention (Hale, Wilcox, & Reddy, 2016). The analysis of the pattern of performance within and across measures, validated by concurrent data sources, helps us understand each student's individual profile and thus provide the most appropriate intervention to address the student's needs. This type of comprehensive evaluation is also helpful in identifying other possible disorders or difficulties that may be interfering with academic achievement because there are many reasons for nonresponse in an RTI model, not just SLD (Hale et al., 2006).

A large group of experts came together to publish a white paper on the identification of SLD (Hale, Alfonso, et al., 2010). This group offered five major recommendations:

1. The SLD statutory requirements that include processing deficits in the definition of SLD should be strengthened and adhered to in practice.
2. Neither RTI nor AAD is sufficient for SLD identification.
3. A PSW approach to the identification of SLD has the most empirical and clinical support.
4. RTI is appropriate as a prevention approach, but should be combined with a PSW approach to identify and serve children who fail to respond to interventions.
5. Identification and intervention development for students with SLD should include neuropsychological assessment (process) approaches.

Although an RTI approach can improve outcomes for many children (e.g., Jimerson, Burns, & VanDerHeyden, 2016), lack of response to interventions is not effective in identifying SLD (e.g., Barth et al., 2008; Brown-Waesche, Schatschneider, Maner, Ahmed, & Wagner, 2011). In addition, one large-scale study, conducted under the auspices of the U.S. Department of Education, using hundreds of children and dozens of schools, found that using a multi-tiered RTI approach in the absence of differentiated instruction does not lead to better academic outcomes for struggling learners (Balu et al., 2015). In fact, when a standard instructional approach was compared to the multi-tiered system of support approach using RTI, children in the standard approach did better than those in RTI, with results suggesting little to no gains in the undifferentiated RTI approach (Balu et al., 2015). We believe that RTI is an excellent first step to helping children learn and overcome challenges, but is not alone sufficient to diagnose SLD (Hale et al., 2006) or guide individualized interventions to help children overcome their learning difficulties (e.g., Balu et al., 2015).

Work by Hale and colleagues (Hale, Betts, Morley, & Chambers, 2010; Hale & Morley, 2009) has demonstrated that not all nonresponders in an RTI model have SLD. These researchers successfully used a combined RTI–CHT approach to provide RTI services to all children, and this approach dramatically reduced referrals for special education evaluation. Although all of the nonresponders in their multi-tiered RTI approach did meet IDEA 2004 criteria for a disability, the comprehensive CHT evaluation revealed that these children had many different types of disorders (e.g., ADHD, anxiety disorder) or different subtypes of SLD (e.g., orthographic SLD, working memory SLD). Not only did the CHT evaluation help identify each child's unique PSW and disabling condition, but it also led to targeted interventions designed to meet each child's unique needs, and ultimately to treatment efficacy.

The third-method PSW approach to identifying SLD is supported not only by SLD experts (Hale, Alfonso, et al., 2010), but also by samples of school psychology practitioners (Caterino, Sullivan, Long, & Bacal, 2008; Machek & Nelson, 2007) and several national organizations, including the National Association of School Psychologists (2007), the American Academy of School Psychology (Schrank, Miller, Caterino, & Desrochers, 2006), and the Learning Disabilities Association of America (Hale, Alfonso, et al., 2010).

Different approaches to SLD identification lead to differing outcomes, depending on the methodology and data examined (Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016; Maki, Burns, & Sullivan, 2016), so we must return to a critical defining element of SLD to guide our methodology. The statutory requirement of IDEA specifies that children with SLD have a deficit in the basic psychological processes. Focusing on both the SLD construct and measurement issues of SLD identification, instead of solely considering the methodology, will enable us to develop a more nuanced and sophisticated approach to serve this diverse population for both diagnostic and intervention purposes. The idea that children with SLD are different from their neurotypical peers, both in their processing characteristics and in their brain functioning, is confirmed time and again in the neuroscience literature (Hale, Chen, et al., 2016)—wherein students with different PSW may have similar achievement deficits, but profile and error pattern differences attest to just how they are different (Koriakin et al., 2017). The point has been supported by meta-analysis (Johnson et al., 2010).

Several studies of the PSW model have reported negative results, using simulated and actual data (e.g., Kranzler et al., 2016; Miciak et al., 2016; Stuebing, Fletcher, Branum-Martin, & Francis, 2012). Although these studies question the utility of the PSW approach, we argue that the negative results are related to the study methodology, rather than constituting a critique of the model itself. Specifically, studies that examine the PSW model often identify the lowest score relative to the highest score and make few actual attempts to *verify* that the weakness is related to a disability, as we (Hale, Wycoff, & Fiorello, 2010) suggest is necessary before identification should occur (see also Flanagan, Alfonso, Costa, Palma, & Leahy, Chapter 22, this volume). In addition, those students are essentially collapsed into one "big bucket" for single-heterogeneous-group analysis, which misses the essence of the model entirely (Kranzler et al., 2016; Taylor, Miciak, Fletcher, & Francis, 2017; see Flanagan & Schneider, 2016, for commentary on this type of analysis). Researchers' assumptions have played a large role in these analyses, and are thus likely to have undermined potential findings, with designs likely to maximize the likelihood of Type II errors. For instance, Kranzler and colleagues (2016) only used the cognitive–achievement relationships determined for a typical population (i.e., they used McGrew & Wendling, 2010), not those with a SLD or subtypes of disability. Although we agree that the relationships found between Cattell–Horn–Carroll (CHC) factors and achievement measures is a good place to start in understanding these relationships, results from our work and others suggest that there is not just one cause of a particular academic disability, but several. As a result, collapsing disparate scores from disparate subpopulations (subtypes) into a single group for subsequent analysis is akin to the measurement problems seen in both AAD and undifferentiated RTI.

As Decker and colleagues (2013) note, it is easy to create studies that lead to Type II errors when individual cognitive–achievement differences are minimized, and then to ignore the literature that could guide more accurate analyses. In fact, it is interesting to note that many of these anti-PSW papers ignore the positive PSW findings in the literature. For instance, the Miciak and colleagues (2016) and Kranzler and colleagues (2016) studies do not cite one positive PSW study, as if the research was never conducted. These studies also avoid reviewing any neuropsychology or neuroscience literature on SLD. A quick review of their reference lists shows little attention to this broad and telling literature, which attests to both the importance of individual processing differences for identification of SLD and targeted intervention specific to child needs (see Hale, Chen, et al., 2016). Ignoring the literature and creating designs that maximize Type II errors do not bode well for the conclusions rendered by these researchers. Regardless, the great asset of PSW is its specificity, but this specificity is lost as we climb the nomothetic ladder or treat all children with SLD as if they have the same processing cause for their disorder.

In education, we often hear of the problems associated with a "one-size-fits-all" orientation. It fails us with AAD. It also fails us in an RTI approach. What makes researchers think that PSW would be any different? As long as the individual processing differences are ignored in researcher designs, PSW will suffer the same fate as AAD and RTI. In their comprehensive analyses of the topic, Compton, Fuchs, and others have demonstrated that when these students are collapsed into a single group, the effects disappear. The sensitivity of the PSW model is lost when we do not attempt to verify that the weakness may be related to a disability (Flanagan & Schneider, 2016; Fuchs et al., 2012; Fuchs, Fuchs, & Compton, 2010; McMaster, Fuchs, Fuchs, Compton, 2005), and when students are collapsed into a heterogeneous SLD group.

Intervention Development

Evaluations for students who do not respond to intervention must be comprehensive, and we define *comprehensive* as including cognitive and neuropsychological measures in addition to other data sources. The goal of the comprehensive evaluation before tier 3 service delivery is not just differential diagnosis or determination of special education eligibility, but also development of targeted interventions designed to meet a child's unique needs. Most eligibility assessment procedures have not been directly linked to intervention and have poor validity for this purpose (Bocian, Beebe, MacMillan, & Gresham, 1999). Although early special education research suggested that there were no "aptitude–treatment interactions" (Reschly & Ysseldyke, 2002), neuropsychological research has been more fruitful in linking assessment to intervention (see a summary in Hale, Fiorello, Miller, et al., 2008). In fact, a special issue of the *Journal of Learning Disabilities*, titled "Cognitive and Neuropsychological Assessment Data That Inform Educational Intervention" (see Fuchs, Hale, & Kearns, 2011), attests to recent advances in the area of linking these types of data to intervention. Of course, there is no easy "cookbook" approach for linking assessment results to interventions. Our CHT model calls for a complete problem-solving process: collaborative development of interventions, regular monitoring of progress/results, and recycling of the process until a successful intervention is found for each individual child.

MODELS OF COGNITIVE FUNCTIONING: CHC THEORY AND NEUROPSYCHOLOGY

CHC theory, based in psychometrics and large-scale factor analyses, is the predominant model of cognitive functioning today (Newton & McGrew, 2010). The broad and narrow cognitive abilities identified in CHC theory have been linked with a variety of educational outcomes (Elliott et al., 2010; Hale, Fiorello, Dumont, et al., 2008; McGrew & Wendling, 2010). Most measures of cognitive functioning today are designed to measure a variety of constructs, rather than simply to provide a single IQ or *g* measure (Elliott et al., 2010; Fiorello et al., 2001, 2007; Flanagan, Ortiz, & Alfonso, 2013; Hale et al., 2006, 2007; Hale, Fiorello, Dumont, et al., 2008; Hale, Fiorello, Miller, et al., 2008; McGrew & Wendling, 2010), and CHC theory is often the model used for that development (Keith & Reynolds, 2010).

At the same time as CHC has been coming into prominence as a psychometric theory of cognitive functioning, neuropsychological research has burgeoned; much of it has examined learning disabilities, ADHD, and other high-prevalence disorders (e.g., D'Amato, Fletcher-Janzen, & Reynolds, 2005; Feifer & Rattan, 2009; Hale & Fiorello, 2004; Miller, 2010). The convergence of evidence that cognitive and neuropsychological processes can be reliably and validly measured and linked to outcomes is leading to a synthesis of the psychometric and neuropsychological approaches to assessment (Fiorello et al., 2008). This knowledge can be used during CHT evaluations for nonresponders in a multi-tiered RTI model—an approach that includes the strengths of both CHT and RTI models.

COGNITIVE HYPOTHESIS TESTING

Overview of the Model

The CHT model is based on the idea that professionals must *intervene* to *assess* (Hale & Fiorello, 2004). A multi-tiered RTI model, implemented with fidelity, will ensure that the number of students referred for comprehensive evaluations will be relatively small. When professionals are completing an assessment on every child who is at risk for school failure, they do not have time to complete the kind of comprehensive assessment that we recommend. However, if most students are served through an RTI process, the number who require an assessment for diagnosis and placement eligibility should be manageable. School psychologists operating in typical public schools are often faced with an overwhelming number of students who are referred for full psychoeducational evaluations. Many of these clinicians cringe at the thought of having to do additional testing on top of what is already a very demanding caseload. However, the CHT model does not simply advocate for "more testing." The model is not intended to place more responsibility on school psychologists, but rather to help redistribute that responsibility back to general education teachers, special educators, and others who are intervening at tiers 1 and 2. When RTI is correctly implemented, all team members are responsible for data collection. When other team members are responsible for interventions and data collection at tiers 1 and 2, the number of referrals for full evaluations will decrease, and school psychologists will have more time to assess the small subset of the student population with true, neurologically based learning disabilities. An RTI model that includes both standard protocol and collaborative problem-solving approaches (that also include participation from special educators across tier levels) will ultimately maximize external and internal validity in the decision-making process (Hale, Flanagan, & Naglieri, 2008). This position is advocated by a majority of leading researchers in SLD (Hale, Alfonso, et al., 2010).

The CHT model (Hale & Fiorello, 2004; see Figure 26.1) uses cognitive and neuropsychological measures to assess students who do not respond to intervention. Based on a scientist-practitioner approach, the CHT model uses the scientific method to assess children over time; this mitigates some of the difficulties with one-shot assessments, and establishes concurrent and ecological validity in the process. Any hypotheses about processing weaknesses derived from the initial cognitive battery or other data sources are tested further with more specific measures, and are evaluated to ensure their ecological and treatment validity (Hale & Fiorello, 2004). Although we do recommend empirical profile analysis (Elliott et al., 2010), CHT avoids many of the difficulties of that process by confirming or disconfirming hypotheses with further data collection, including further testing of psychological processes beyond a single cognitive test.

The beginning stages of CHT are similar to typical assessment practices. A student who has failed to respond to instruction and intervention is referred for formal evaluation. The referral question is considered together with historical records, classroom permanent products, and RTI data to develop a theory of the problem. Hypotheses are proposed to explain the academic or behavioral deficits, and if the hypothesis revolves around a question of cognitive functioning, a cognitive/intelligence test is used during the first round of data collection. The initial battery should be chosen to cover a broad range of cognitive processes, and to be a fair measure of the student's functioning based on his or her cultural and linguistic background. That measure should be scored and interpreted to identify possible processing strengths and weaknesses—using clinical references to generate initial hypotheses (such as Dehn, 2013; Flanagan et al., 2013; Hale & Fiorello, 2004; Miller, 2013; Naglieri & Goldstein, 2009; Sattler, 2008), and using *demands analysis* (see below) as needed

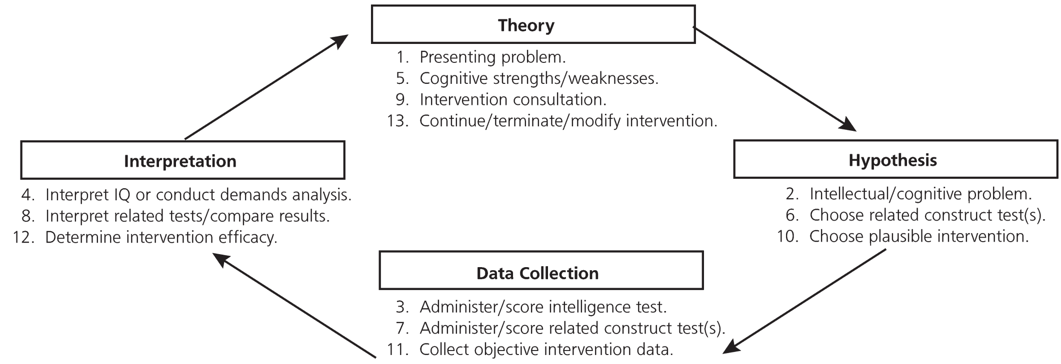


FIGURE 26.1. The cognitive hypothesis testing (CHT) model. From Hale and Fiorello (2004). Copyright © The Guilford Press. Reprinted by permission.

to determine what cognitive processes are being assessed by a given measure.

Although large-group studies can provide valuable information about what particular tests measure, they are often based on the standardization samples of cognitive batteries (e.g., McGrew & Wendling, 2010), and therefore do not capture the differences in what is measured that may be found in individuals or subgroups. Students with disabilities, in particular, may use different cognitive processes to complete complex tasks from those used by the typically developing majority (e.g., Elliott et al., 2010; Hale, Fiorello, Dumont, et al., 2008; Shaywitz et al., 1998). Neuroimaging research confirms that most students use multiple brain areas to solve complex tasks, but that different brain areas are primarily responsible for different components of cognitive tests (Glascher et al., 2009). In addition, most complex cognitive tasks can be completed in a variety of ways, so using a "cookbook" approach by listing all potential processing possibilities is not helpful in intervention development. For example, poor Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) Processing Speed performance may result from many different problems, including attention, visual acuity, visual scanning, visual–spatial functioning, visual memory, associative learning, somatosensory functioning, processing speed, psychomotor speed, fine motor coordination, and/or graphomotor skills. Further testing and comparison to ecological information should evaluate these hypotheses to allow development of appropriate interventions.

Practically speaking, this process may unfold in the following way: The school psychologist generates hypotheses about the potential reasons for a student's poor WISC-V Processing Speed score. Next, the school psychologist consults with the special educator who was responsible for the tier 2 interventions. The special educator provides work samples, classroom examples, and progress-monitoring data to suggest that the student's visual memory and visual–spatial functioning are poor, but that graphomotor skills appear intact. The school psychologist can then choose specific subtests (e.g., Wide Range Assessment of Memory and Learning, Second Edition subtests for Visual Memory; NEPSY-II Arrows, Design Copy) from different neuropsychological and cognitive measures to test the specific hypotheses about the underlying processes—visual memory, visual–spatial processing, and graphomotor skills—that may be responsible for poor Processing Speed scores on the WISC-V. If the visual memory hypothesis is confirmed, then the school psychologist is in a better position to recommend very specific, clear recommendations and accommodations focusing on poor visual memory. On the other hand, if the school psychologist stops the process prematurely and simply reports back to the team that this child has "poor processing speed," little headway will be made with regard to providing appropriate, practical suggestions to help this child learn.

Clinical interpretation that takes into account a variety of empirical information about what tests measure, together with deep knowledge of the tests themselves and close observation of *how* a student performs the required tasks, can identify potential strengths and weaknesses. Both level (i.e., nomothetic) and pattern (i.e., idiographic) of performance are examined to determine the student's cognitive, neuropsychological, academic, and behavioral state at the time of the evaluation (Hale

& Fiorello, 2004). Note that it is not assumed that the psychologist is assessing unchanging *traits* of the student, but obtaining a picture of the student's current *state* of functioning. Since cognitive states are measured, it is important to administer measures over more than one session to confirm or refute the hypotheses derived from any given session, which is why CHT is so critical for interpretation. However, at this point in a typical psychoeducational assessment, most clinicians write a report describing the purported strengths and weaknesses and make recommendations for placement and interventions. But this is only the beginning of the CHT process because any hypotheses developed need to be confirmed or disconfirmed by using additional data sources and conducting additional testing (Hale & Fiorello, 2004).

Many supplemental tests are brief neuropsychological processing measures with adequate sensitivity and more specificity, so this additional testing need not take an inordinate amount of time. The results of the additional testing are examined in light of all the data collected about the child, and a theory about a likely intervention approach is developed (Hale & Fiorello, 2004). Through a process of collaborative consultation with the teacher and/or parent, an intervention plan is devised. The intervention is implemented with regular progress monitoring, and evaluated to determine efficacy. If the intervention is not effective, the psychologist revises the plan or recycles through the process until a successful intervention is found (Hale & Fiorello, 2004). In this way, CHT combines information about cognitive and neuropsychological functioning within a collaborative problem-solving approach, and uses single-subject methodology to evaluate the effectiveness of interventions (Fiorello et al., 2006; Hale et al., 2006; Reddy & Hale, 2007).

CHT can be used to link assessment to intervention for students with difficulties in various areas, including reading (Fiorello et al., 2006), math (Hale et al., 2006), and attention (Reddy & Hale, 2007). CHT has also been recommended for use in neuropsychological settings (Fletcher-Janzen, 2005; Miller, Getz, & Leffard, 2006), as well as in schools (Elliott et al., 2010; Hale, Fiorello, Dumont, et al., 2008). Because CHT is incorporated within the context of a collaborative problem-solving approach, it is inherently self-correcting and leads to successful intervention for children with a number of disabilities (Fiorello et al., 2006; Hale, Fiorello, Miller, et al., 2008).

Demands Analysis

Interpreting an IQ score is simple, but it seldom is reflective of a child's ability, nor is it the best predictor of academic achievement (Fiorello et al., 2007; Hale, Fiorello, Miller, et al., 2008). Instead, we must acknowledge that interpretation of intelligence/cognitive subcomponent scores is necessary, and that it is not a simple or straightforward process. Various intelligence tests involve different tasks, different cognitive demands, different cultural and linguistic loading, and different administration procedures. In addition, every test measures, to a greater or lesser extent, a combination of ability and achievement; therefore, an examinee's background and exposure to similar tasks must be taken into account during interpretation. Scores are not interchangeable and should not be interpreted as a measure of a student's *trait* of intelligence, but as a measure of the student's current *state* of cognitive functioning (Hale, Wycoff, & Fiorello, 2010). The choice of a battery should be based on a priori information about the student's prior RTI data, prior experience and education, cultural and linguistic background, and any sensory or motor difficulties, in order to minimize the construct-irrelevant variance that these factors can introduce.

After administration and scoring, the profile should be examined for significant variability. If there is significant subtest or factor variability, the global IQ score should not be interpreted. But even factor scores are complex, and should not be interpreted as unitary clusters if they too show significant subtest variability within the factors. Factor scores should only be interpreted if they hold together, or are reliable, for the child being evaluated. Of course, we do not mean to imply that variability implies a disability, as the majority of people show significant variability on complex batteries like our current IQ tests (Fiorello et al., 2007; Hale, Fiorello, Dumont, et al., 2008). We simply mean that profiles should be interpreted as indicating cognitive strengths and weaknesses when there is significant variability present. In fact, it may even be necessary to examine differences *within* a subtest if it contains disparate tasks, such as Digits Forward and Digits Backward on the WISC-V Digit Span subtest. For instance, research has shown that the difference between Digits Forward and Backward can be useful in identification of children with SLD (e.g., Hain et al., 2009) and attention problems (Hale, Hoeppner, & Fiorello, 2002), suggesting that interpretation of a Digit Span score may not accurately reflect a child's functioning. The newly added process scores may help in evaluating a child's performance.

Many school psychologists report using factor scores, profile analysis, or both to examine cognitive strengths and weaknesses in practice (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000). However, we need a consistent methodology for interpreting the pattern of strengths and weaknesses to increase the reliability and validity of those conclusions. We have derived our CHT model in such a way as to increase the reliability and validity by systematically testing our hypotheses, and evaluating ecological and treatment validity. *Demands analysis* (see Figure 26.2) systematizes the process of identifying the input, processing, and output demands of individual tasks. Rather than just basing interpretation on a score and the test maker's description of what a subtest measures, demands analysis provides a wide range of possible interpretive factors to be considered in evaluating a student's strengths and weaknesses. After testing these hypotheses with further evaluation, demands analysis allows professionals to develop an individualized education program that will truly meet a student's individual needs.

Demands analysis is a combination of the "intelligent testing" approach begun by Alan Kaufman (1979) and in widespread use in school psychology (e.g., Sattler, 2008), along with the CHC cross-battery approach advocated by McGrew, Flanagan, and colleagues (e.g., Flanagan et al., 2013; McGrew & Wendling, 2010) and a neuropsychological assessment "process approach" (e.g., Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Hebben & Milberg, 2002). This emphasis on the neuropsychological processes underlying task completion has become more widespread in school psychology since the introduction of the process in *School Neuropsychology: A Practitioner's Handbook* (Hale & Fiorello, 2004; see, e.g., D'Amato et al., 2005; Dehn, 2013; Miller, 2010, 2013).

To complete a demands analysis, a professional must first consider the *input* demands of the task. This refers to the directions and stimulus materials—what modalities are used; the presence of pictures, manipulatives, oral directions, or written materials; how abstract or meaningful the content is; how much English-language mastery is called for; and how much cultural loading there is in the task. Next, the *processing* demands must be considered. It is important not to depend solely on the primary process suggested in the test manual or the loading on CHC abilities indicated by factor-analytic research, but to consider other neurological processing demands, such as executive and working memory skills. As Goldberg (2001) suggests, several brain processes are typically involved in any given task, so it is important to interpret

Student's Name: ______________________ Age: ________ Grade: ________

	Test/subtest	Input	Process	Output
Strengths				
Weaknesses				

FIGURE 26.2. Demands analysis. From Hale and Fiorello (2004). Copyright © The Guilford Press. Reprinted by permission.

this interrelationship among cognitive processes when examining any given test result. Also, it is important to consider the different ways that examinees can solve a given task. Some students use the visual gestalt to solve Block Design on the Wechsler scales; others use a trial-and-error approach; still others talk their way through the problem in a linear, sequential manner and instead focus on stimulus details.

Finally, it is important to consider the *output* demands of the task. What modalities and skills are required to complete the task—simple pointing, a complex motoric task, a complex verbal explanation? Taking copious notes on the student's actual behavior on the task, even at the level of individual items, is required for this fine-grained interpretation. A poor score on the Wechsler Information subtest may be due to memory retrieval problems or lack of general knowledge—but it may also be due to lack of knowledge in one specific area, like science. Once a professional has noted the demands of the tasks on which a student exhibits strengths and weaknesses, it is important to examine the notes for commonalities and contradictions. The practitioner must keep in mind that input and output difficulties are most likely due to sensory or motor difficulties or even to cultural or linguistic differences, whereas processing difficulties are likely to be characteristic of neuropsychological disorders (e.g., SLD, ADHD, or depression).

After a theory of the student's strengths and weaknesses is developed, the practitioner will then determine what measures are necessary to confirm or refute this theory during subsequent testing. Tasks that assess these specific hypotheses are more sensitive and specific for the processes in question. Then these results are compared to data from history, work samples, behavior observations, or rating forms to confirm or refute initial findings. Comparing the conclusions about input, processing, and output demands that are drawn from a child's test performance to indicators of classroom performance and behavior will help establish the accuracy of interpretation and establish ecological validity of findings. Tests that are specifically designed to cleanly assess narrow abilities (like the Woodcock–Johnson IV [WJ IV] Tests of Cognitive Abilities and Oral Language), or neuropsychological instruments that yield a variety of process scores (like the WISC-V Integrated, the NEPSY-II, and the Delis–Kaplan Executive Function System), are good places to look for tasks that can test hypotheses.

Concordance–Discordance Model

In order to identify SLD, our concordance–discordance model (C-DM; Hale & Fiorello, 2004) (see Figure 26.3) provides practitioners with an empirical approach to examining patterns of performance on cognitive and academic measures to establish that a processing deficit is the cause of the SLD, and therefore meets the IDEA 2004 statutory definition for the disorder (Hale et al., 2006). Establishing this PSW, and associated achievement deficit(s), appears to be the preferred method among those advocating third-method approaches to SLD identification (Flanagan et al., 2010; Flanagan, Ortiz, & Alfonso, 2015; Hale, Alfonso, et al., 2010; Hale, Flanagan et al., 2008), and is growing in popularity across states (Zirkel & Thomas, 2010). By identifying a pattern of cognitive and

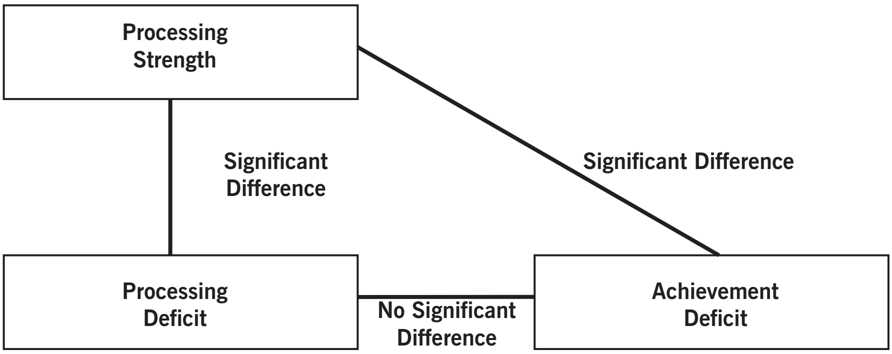


FIGURE 26.3. The concordance–discordance model (C-DM) of SLD identification. Based on Hale and Fiorello (2004).

academic strengths and weaknesses, and evaluating whether they are statistically and clinically different from each other, practitioners can determine the presence of SLD as part of the CHT model.

A step-by-step approach for using the C-DM model can be found in Hale, Wycoff, and Fiorello (2010). The C-DM approach establishes a pattern of cognitive strength(s), together with cognitive weakness(es) associated with an academic weakness. Each component should be composed of a test or cluster score that is reliable and valid for individual, high-stakes decision making. A cluster score may be provided by the test itself, like a General Ability Index score from the WISC-V or a Nonverbal Index score from the Kaufman Assessment Battery for Children—Second Edition in the cognitive area, or practitioners may have to construct a cluster themselves. In most cases, if professionals do this, they will have to calculate the mean cluster score and reliability coefficient themselves, using Fisher's z' transformation (Hale, Fiorello, Miller, et al., 2008).

It is important that practitioners not merely use the highest and lowest cognitive scores and lowest academic scores to calculate the C-DM differences. The cognitive strength should be one that the literature indicates is seldom related to the academic weakness (e.g., visual processing and reading comprehension), while the cognitive weakness should be empirically linked to the academic weakness (e.g., working memory and reading comprehension). Following the standard error of the difference (*SED*) formula (Anastasi & Urbina, 1997), $SED = SD\sqrt{2 - r_{xx} - r_{xy}}$, the practitioner then calculates the *SED* between the cognitive strength and the academic weakness, using the reliability of those cluster scores. This value is then multiplied by 1.96 to obtain the $p < .05$ difference score (or 2.58 for $p < .01$). If the cognitive strength score minus the academic weakness score is equal to or greater than that number, there is a significant difference. Then the same *SED* formula is applied to the cognitive strength and the cognitive weakness to see whether there is a significant difference. Then the calculation is performed again using the cognitive weakness and the academic weakness, and these scores should *not* be significantly different. If the pattern of results fits the pattern shown in Figure 26.3, this is evidence for the presence of SLD. The null hypothesis is that this pattern will not be found, indicating that something other than SLD is responsible for the learning difficulties.

Establishing Ecological and Treatment Validity

We recommend direct behavior observations in natural environments (e.g., classroom) and behavior rating scales to help establish the ecological validity of assessment findings. If a student has a true neuropsychological disorder, it should manifest itself in some form across settings. Of course, behavior is interactional, so a student's behavior will vary in environments with differing demands; however, if difficulties observed during individual testing are not present in the classroom environment, findings need to be evaluated more closely. Sattler (2014) presents a variety of methods for systematically observing student behavior in the classroom; there are formal coding systems available as well, such as the Achenbach System of Empirically Based Assessment (ASEBA) Direct Observation Form (McConaughy & Achenbach, 2009) and the Behavior Assessment System for Children, Third Edition (BASC-3) Student Observation System (Reynolds & Kamphaus, 2015). Behavior rating scales sample behavior across settings and over a period of time, and so add important data to an evaluation. We recommend starting with a general rating scale so as to evaluate a broad range of behaviors, such as the ASEBA Teacher Report Form (Achenbach & Rescorla, 2001), the BASC-3 Teacher Rating Scale (Reynolds & Kamphaus, 2015), or the Clinical Assessment of Behavior (Bracken & Keith, 2004). If practitioners later want more detailed information about a specific class of behaviors, they can follow up with a more focused rating scale. However, it is important to note that the indirect data gained through behavior ratings are quite different from those obtained through direct assessment, so clinical judgment will be necessary to reconcile differences obtained on these measures (e.g., Hale et al., 2009).

Intervention must occur within the context of a collaborative problem-solving approach, and therefore the importance of ongoing progress monitoring to evaluate treatment efficacy cannot be overstated. Gone are the days when a school psychologist would file an evaluation report and not check back for 1 or even 3 years. Effective practitioners must be involved in treatment implementation and monitoring to ensure that what was recommended is actually implemented with fidelity and is effective; if not, consultation skills are necessary to retool the plan as necessary until treatment efficacy is obtained (Hale & Fiorello, 2004).

Using the CHT model to link cognitive and neuropsychological assessment data to intervention on an individual level can ensure that the students served will obtain the individualized services they need. But more group and single-subject research is needed to establish the utility of the CHT approach, and to identify intervention approaches that are effective with students displaying specific patterns of SLD, especially since there are numerous SLD subtypes (e.g., Compton et al., 2012; Fiorello et al., 2006; Hain et al., 2009; Hale, Fiorello, Dumont, et al., 2008; Kubas et al., 2014; Tolar, Fuchs, Fletcher, Fuchs, & Hamlett, 2016). The evidence is just now emerging that cognitive and neuropsychological processes are relevant for academic and behavioral intervention, and considerable empirical work is needed to establish the validity of these relationships (e.g., Fuchs et al., 2011). Only then will the true promise of special education be realized.

ACKNOWLEDGMENT

We acknowledge James B. Hale for his contribution to the chapter in the third edition of this volume, on which the present chapter is based.

REFERENCES

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles.* Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Backenson, E. M., Holland, S. C., Kubas, H. A., Fitzer, K. R., Wilcox, G., Carmichael, J. A., et al. (2015). Psychosocial and adaptive deficits associated with learning disability subtypes. *Journal of Learning Disabilities, 48*(5), 511–522.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE Publication No. 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *Journal of Special Education, 38,* 66–79.

Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., et al. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences, 18,* 296–307.

Berninger, V. W. (2006). Research-supported ideas for implementing reauthorized IDEA with intelligent professional psychological services. *Psychology in the Schools, 43,* 781–796.

Berninger, V. W., & Richards, T. L. (2002). *Brain literacy for educators and psychologists.* Boston: Academic Press.

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology, 19*(6), 503–508.

Bocian, K., Beebe, M., MacMillan, D., & Gresham, F. M. (1999). Competing paradigms in learning disabilities classification by schools and the variations in the meaning of discrepant achievement. *Learning Disabilities Research and Practice, 14,* 1–14.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities, 32*(2), 560–570.

Bracken, B. A., & Keith, L. K. (2004). *Examiner's manual: Clinical Assessment of Behavior.* Lutz, FL: Psychological Assessment Resources.

Brown-Waesche, J. S., Schatschneider, C., Maner, J. K., Ahmed, Y., & Wagner, R. K. (2011). Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities, 44,* 296–307.

Caterino, L., Sullivan, A., Long, L., & Bacal, E. (2008). A survey of school psychologists' perceptions of the reauthorization of IDEA 2004. *The School Psychologist, 62*(2), 45–49.

Cho, E., Roberts, G. J., Capin, P., Roberts, G., Miciak, J., & Vaughn, S. (2015). Cognitive attributes, attention, and self-efficacy of adequate and inadequate responders in a fourth grade reading intervention. *Learning Disabilities Research and Practice, 30*(4), 159–170.

Collins, D. W., & Rourke, B. P. (2003). Learning-disabled brains: A review of the literature. *Journal of Clinical and Experimental Neuropsychology, 25,* 1011–1034.

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45,* 79–95.

Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology, 106*(1), 214–229.

D'Amato, R. C., Fletcher-Janzen, E., & Reynolds, C. R. (Eds.). (2005). *Handbook of school neuropsychology.* Hoboken, NJ: Wiley.

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools, 50*(3), 300–313.

Dehn, M. J. (2013). *Essentials of processing assessment* (2nd ed.). Hoboken, NJ: Wiley.

Dixon, S. G., Eusebio, E. C., Turton, W. J., Wright, P. W. D., & Hale, J. B. (2011). *Forest Grove School District v. T. A.* Supreme Court case: Implications for school psychology practice. *Journal of Psychoeducational Assessment, 29*(2), 103–113.

Education for All Handicapped Children Act, Pub. L. No. 94-142, 20 U.S.C. 1401 (1975).

Elliott, C. D., Hale, J. B., Fiorello, C. A., Dorvil, C., & Moldovan, J. (2010). Differential Ability Scales—Second Edition prediction of reading performance: Global scales are not enough. *Psychology in the Schools, 47,* 698–720.

Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., et al. (2009). Multisensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of Cognitive Neuroscience, 22,* 1011–1025.

Feifer, S. G., & Della Toffalo, D. (2007). *Integrating RTI with cognitive neuropsychology: A scientific approach to reading.* Middletown, MD: School Neuropsych Press.

Feifer, S. G., & Rattan, G. (Eds.). (2009). *Emotional disorders: A neuropsychological, psychopharmacological, and educational perspective*. Middletown, MD: School Neuropsych Press.

Fenwick, M. E., Kubas, K. A., Witzke, J. W., Fitzer, K. R., Miller, D. C., Maricle, D. E., et al. (2016). Neuropsychological profiles of written expression learning disabilities determined by concordance–discordance model criteria. *Applied Neuropsychology: Child, 5*(2), 83–96.

Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences USA, 95,* 914–921.

Filipek, P. A. (1999). Neuroimaging in the developmental disorders: The state of the science. *Journal of Child Psychology and Psychiatry, 40,* 113–128.

Fine, J. G., Semrud-Clikeman, M., Keith, T. Z., Stapleton, L. M., & Hynd, G. W. (2007). Reading and the corpus callosum: An MRI family study of volume and area. *Neuropsychology, 21,* 235–241.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14,* 2–12.

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2001). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13,* 115–125.

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading disabilities. *Psychology in the Schools, 43,* 835–854.

Fiorello, C. A., Hale, J. B., Snyder, L. E., Forrest, E., & Teodori, A. (2008). Validating individual differences through examination of converging psychometric and neuropsychological models of cognitive functioning. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K–3 classrooms* (pp. 151–186). New York: Routledge.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of CHC theory and research: A "third method" approach to SLD identification. *Psychology in the Schools, 47,* 739–760.

Flanagan, D. P., & Mascolo, J. T. (2016). Utility of KTEA-3 error analysis for the diagnosis of specific learning disabilities. *Journal of Psychoeducational Assessment. 35*(1–2), 226–241.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2015). *Cross-battery assessment software system (X-BASS).* Hoboken, NJ: Wiley.

Flanagan, D. P., & Schneider, W. J. (2016). Cross-battery assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues' "Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth." *International Journal of School and Educational Psychology, 4,* 137–145.

Fletcher, J. M., Stuebing, K. K., Barth, A., Denton, C. A., Cirino, P. T., Francis, D. J., et al. (2011). Cognitive correlates of inadequate response to intervention. *School Psychology Review, 40*(1), 3.

Fletcher-Janzen, E. (2005). The school neuropsychological examination. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 172–212). Hoboken, NJ: Wiley.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental delay versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology, 88,* 3–17.

Fuchs, D., Compton, D. L., Fuchs, L., Bryant, V. J., Hamlett, C. L., & Lambert, W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*(3), 217–231.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice, 22,* 129–136.

Fuchs, D., Hale, J. B., & Kearns, D. (2011). On the importance of a cognitive processing perspective: An introduction. *Journal of Learning Disabilities, 44,* 99–104.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle and high school. *School Psychology Review, 39*(1), 22–28.

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250.

Geary, D. C., Hoard, M. K., & Hamsom, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74,* 213–239.

Glascher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, A., et al. (2009). Lesion mapping of cognitive abilities linked to intelligence. *Neuron, 61,* 681–691.

Goldberg, E. H. (2001). *The executive brain: Frontal lobes and the civilized mind.* New York: Oxford University Press.

Groth-Marnat, G., Gallagher, R. E., Hale, J. B., & Kaplan, E. (2000). The Wechsler intelligence scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 129–194). New York: Wiley.

Hain, L. A., Hale, J. B., & Glass-Kendorski, J. (2009). Comorbidity of psychopathology in cognitive and academic SLD subtypes. In S. G. Feifer & G. Rattan (Eds.), *Emotional disorders: A neuropsychological, psychopharmacological, and educational perspective* (pp. 199–226). Middletown, MD: School Neuropsych Press.

Hale, J. B. (2006). Implementing IDEA with a three–tier model that includes response to intervention and cognitive assessment methods. *School Psychology Forum: Research and Practice, 1,* 16–27.

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223–236.

Hale, J. B., Betts, E. C., Morley, J., & Chambers, C. (2010, March). *SLD third method approaches for combining RTI and comprehensive evaluation.* Mini Skills Workshop presented at the annual convention of the National Association of School Psychologists, Chicago, IL.

Hale, J. B., Chen, S. H. A., Tan, S. C., Poon, K., Fitzer, K. R., & Boyd, L. A. (2016). Reconciling individual differences with collective needs: The juxtaposition of sociopolitical and neuroscience perspectives on remediation and compensation of student skill deficits. *Trends in Neuroscience and Education, 5*(2), 41–51.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales—Second Edition (neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools, 45,* 838–858.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified?: A response to special issue authors. *Applied Neuropsychology, 14,* 37–51.

Hale, J. B., Fiorello, C. A., Miller, J. A., Wenrich, K., Teodori, A. M., & Henzel, J. (2008). WISC-IV assessment and intervention strategies for children with specific learning disabilities. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 109–171). Amsterdam: Elsevier.

Hale, J. B., Flanagan, D. P., & Naglieri, J. A. (2008). Alternative research-based methods for IDEA (2004) identification of children with specific learning disabilities. *Communiqué, 36*(8), 1, 14–17.

Hale, J. B., Hoeppner, J. B., & Fiorello, C. A. (2002). Analyzing Digit Span components for assessment of attention processes. *Journal of Psychoeducational Assessment, 20,* 128–143.

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools, 43,* 753–770.

Hale, J. B., & Morley, J. (2009, February). *Combining RTI with cognitive hypothesis testing for effective classroom instruction.* Invited workshop at the annual convention of the National Association of School Psychologists, Boston, MA.

Hale, J. B., Reddy, L. A., Decker, S. L., Thompson, R., Henzel, J., Teodori, A., et al. (2009). Development and validation of an executive function and behavior rating screening battery sensitive to ADHD. *Journal of Clinical and Experimental Neuropsychology, 31,* 897–912.

Hale, J. B., Wilcox, G., & Reddy, L. (2016). Neuropsychological assessment. In J. C. Norcross, G. R. VandenBos, & D. K. Freedheim (Eds.), *APA handbook of clinical psychology: Vol. 3. Applications and methods* (pp. 139–165). Washington, DC: American Psychological Association.

Hale, J. B., Wycoff, K. L., & Fiorello, C. A. (2010). RTI and cognitive hypothesis testing for specific learning disabilities identification and intervention: The best of both worlds. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 173–202). Hoboken, NJ: Wiley.

Hebben, N., & Milberg, W. (2002). *Essentials of neuropsychological assessment.* New York: Wiley.

Hulme, C., & Snowling, M. (2011). Children's reading comprehension difficulties. *Current Directions in Psychological Science, 20*(3), 139–142.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1400 (2004).

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.). New York: Springer.

Johnson, E., Humphrey, M., Mellard, D., Woods, K., &

Swanson, L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disability Quarterly, 33*(1), 3–18.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

Kavale, K. A., Holdnack, J. A., & Mostert, M. P. (2005). Responsiveness to intervention and the identification of specific learning disability: A critique and alternative proposal. *Learning Disability Quarterly, 28,* 2–16.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650.

Koriakin, T., White, E., Breaux, K. C., DeBiase, E., O'Brien, R., Howell, M., et al. (2017). Patterns of cognitive strengths and weaknesses and relationships to math errors. *Journal of Psychoeducational Assessment, 35*(1–2), 155–167.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School and Educational Psychology, 4,* 124–136.

Kubas, H. A., Drefs, M. A., Poole, J. M., Schmid, A. D., Holland, S. C., & Fiorello, C. A. (2014). Cognitive and academic profiles associated with math disability subtypes. *Learning Disabilities: A Multidisciplinary Journal, 20*(1), 31–44.

Machek, G. R., & Nelson, J. M. (2007). How should reading disabilities be operationalized?: A survey of practicing school psychologists. *Learning Disabilities Research and Practice, 22,* 147–157.

MacMillan, D. L., Gresham, F. M., Lopez, M. F., & Bocian, K. M. (1996). Comparison of students nominated for prereferral interventions by ethnicity and gender. *Journal of Special Education, 30,* 133–151.

MacMillan, D. L., Siperstein, G., & Gresham, F. M. (1996). Mild mental retardation: A challenge to its viability as a diagnostic category. *Exceptional Children, 62,* 356–371.

MacMillan, D. L., & Speece, D. (1999). Utility of current diagnostic categories for research and practice. In R. Gallimore, L. Hernheimer, D. MacMillan, D. Speece, & S. Vaughn (Eds.), *Developmental perspectives on children with high-incidence disabilities* (pp. 111–133). Mahwah, NJ: Erlbaum.

Maki, K. E., Burns, M. K., & Sullivan, A. (2016). Learning disability identification consistency: The impact of methodology and student evaluation data. *School Psychology Quarterly, 32*(2), 254–267.

McConaughy, S. H., & Achenbach, T. M. (2009). *Manual for the ASEBA Direct Observation Form.* Burlington: University of Vermont, Research Center for Children, Youth, and Families.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651–675.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445–463.

Mercer, C. D., Jordan, L., Alsopp, D. H., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. *Learning Disability Quarterly, 19,* 217–232.

Miciak, J., Williams, J. L., Taylor, W. P., Cirino, P. T., Fletcher, J. M., & Vaughn, S. (2016). Do processing patterns of strengths and weaknesses predict differential treatment response? *Journal of Educational Psychology, 108*(6), 898–909.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention.* Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D. C., & Hale, J. B. (2008). The neuropsychological applications of the Wechsler Intelligence Scale for Children—Fourth Edition. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV advanced clinical interpretation* (2nd ed., pp. 445–495). San Diego, CA: Academic Press.

Miller, J. A., Getz, G., & Leffard, S. A. (2006, February). *Neuropsychology and the diagnosis of learning disabilities under IDEA 2004.* Poster presented at the 34th annual meeting of the International Neuropsychological Society, Boston, MA.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90,* 347–373.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment, 21,* 244–260.

Naglieri, J. A., & Goldstein, S. (Eds.). (2009). *Practitioner's guide to assessing intelligence and achievement.* Hoboken, NJ: Wiley.

National Association of School Psychologists. (2007). *Identification of students with specific learning disabilities (Position statement).* Bethesda, MD: Author.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell–Horn–Carroll–based assessment. *Psychology in the Schools, 47,* 621–634.

Nicholson, R. I., & Fawcett, A. J. (2001). Dyslexia, learning, and the cerebellum. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 159–188). Timonium, MD: York Press.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427–452.

Pfeiffer, S. L., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly, 15,* 376–385.

Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., et al. (2000). The angular gyrus in developmental dyslexia: Task-specific differences in functional connectivity within posterior regions. *Psychological Science, 11,* 51–56.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., et al. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126,* 841–865.

Reddy, L. A., & Hale, J. B. (2007). Inattentiveness. In A. R. Eisen (Ed.), *Clinical handbook of childhood behavior problems: Case formulation and step-by-step treatment programs* (pp. 156–211). New York: Guilford Press.

Reschly, D. J., & Hosp, J. L. (2004). State SLD policies and practices. *Learning Disability Quarterly, 27,* 197–213.

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 3–21). Bethesda, MD: National Association of School Psychologists.

Reynolds, C. R., & Kamphaus, R. (2015). *Behavior Assessment System for Children, Third Edition (BASC-3).* Bloomington, MN: Pearson.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). La Mesa, CA: Author.

Sattler, J. M. (2014). *Assessment of children: Behavioral, social, and clinical foundations* (6th ed.). San Diego, CA: Author.

Schrank, F. A., Miller, J. A., Caterino, L., & Desrochers, J. (2006). American Academy of School Psychology survey on the independent educational evaluation for a specific learning disability: Results and discussion. *Psychology in the Schools, 43,* 771–780.

Semrud-Clikeman, M. (2005). Neuropsychological aspects for evaluating learning disabilities. *Journal of Learning Disabilities, 38,* 563–568.

Shaywitz, B. A., Lyon, G. R., & Shaywitz, S. E. (2006). The role of functional magnetic resonance imaging in understanding reading and dyslexia. *Developmental Neuropsychology, 30,* 613–632.

Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., et al. (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proceedings of the National Academy of Sciences USA, 95,* 2636–2641.

Simos, P. G., Fletcher, J. M., Sarkari, S., Billingsley, R. L., Francis, D. J., Castillo, E. M., et al. (2005). Early reading development of neurophysiological processes involved in normal reading and reading disability: A magnetic source imaging study. *Neuropsychology, 19,* 787–798.

Stein, J. F. (2001). The neurobiology of reading difficulties. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 3–22). Timonium, MD: York Press.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41,* 3–22.

Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities, 44*(4), 358–371.

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment, 29*(4), 446–457.

Tolar, T. D., Fuchs, L., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability. *Journal of Learning Disabilities, 49*(3), 240–256.

U.S. Department of Education. (2006, August 14). 34 C.F.R. Parts 300 and 301: Assistance to states for the education of children with disabilities and preschool grants for children with disabilities: Final rule. *Federal Register, 71*(146), 46539–46845.

Willis, J. O., & Dumont, R. (2006). And never the twain shall meet: Can response to intervention and cognitive assessment be reconciled? *Psychology in the Schools, 43,* 901–908.

Zirkel, P., & Thomas, B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children, 43*(1), 60–73.

CHAPTER 27

Cross-Battery Assessment, the Cross-Battery Assessment Software System, and the Assessment–Intervention Connection

Dawn P. Flanagan
Michael Costa
Katherine Palma
Meghan A. Leahy
Vincent C. Alfonso
Samuel O. Ortiz

AN OVERVIEW OF THE XBA APPROACH

The Cattell–Horn–Carroll (CHC) cross-battery assessment approach (hereafter referred to as the XBA approach) was introduced by Flanagan and her colleagues approximately two decades ago (Flanagan & McGrew, 1997; Flanagan, McGrew, & Ortiz, 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998). The XBA approach is a systematic method of augmenting cognitive, neuropsychological, achievement, and speech-language batteries with tests from other batteries, such as special-purpose tests (e.g., memory batteries, tests of phonological and orthographic processing, tests of executive functions), to gain a more complete understanding of an individual's strengths and weaknesses (Flanagan, Ortiz, & Alfonso, 2007, 2013, 2017). When batteries are crossed, the XBA approach ensures that reliable, psychometrically sound, and theoretically defensible composites are generated. Moving beyond the boundaries of a single battery by adopting the theoretically and psychometrically sound principles and procedures outlined in the XBA approach is a significant improvement over single-battery assessment because it allows practitioners to focus on measurement of the cognitive constructs and neurodevelopmental functions that are most germane to referral concerns (e.g., Carroll, 1998; Cheramie, Schanding, & Streich, 2018; Decker, 2008; Hale et al., 2010; Kaufman, 2000; Wilson, 1992).

According to Carroll (2012), the CHC taxonomy of human cognitive abilities "appears to prescribe that individuals should be assessed with regard to the *total range* of abilities the theory specifies" (p. 889; original emphasis). However, because Carroll recognized that "any such prescription would of course create enormous problems," he indicated that research "is needed to spell out how the assessor can select what abilities need to be tested in particular cases" (p. 889). Flanagan and colleagues' XBA approach was developed to "spell out" how practitioners can conduct assessments that approximate the total range of cognitive and academic abilities and neuropsychological processes more adequately than what is possible with a stand-alone battery, as well as most collections of co-normed tests.

In an early review of the XBA approach, Carroll (1998) stated that it "can be used to develop the most appropriate information about an individual in a given testing situation" (p. xi). In Kaufman's (2000) review of XBA, he stated that the approach is based on sound assessment principles, adds theory to psychometrics, and improves the quality of the assessment and interpretation of cognitive abilities and processes. Moreover, Decker (2008) stated that the XBA approach "may improve school psychology assessment practice and facilitate the integration of neuropsychological methodology in school-based assessments . . . [because it] shift[s] assessment practice from IQ composites to neurodevelopmental functions" (p. 804). Most recently, Schneider and Roman (2017) stated, "The XBA approach to understanding individuals is an excellent way to organize one's thoughts about test scores. It helps practitioners decide which scores need follow-up testing and which scores are likely to be accurate" (p. 50; see also Cheramie et al., 2018).

Interestingly, assessment professionals "crossed" batteries well before Woodcock (1990) recognized the need, and before Flanagan and her colleagues introduced the XBA approach in the late 1990s, in part as a response to Woodcock's suggestion. Neuropsychologists have long adopted the practice of crossing various standardized tests to measure a broader range of brain functions than that offered by any single instrument (Lezak, 1976, 1995; Lezak, Howieson, & Loring, 2004; see Wilson, 1992, for a review). However, several problems with crossing batteries plagued assessment-related fields for years. Many of these problems have been circumvented by Flanagan and colleagues' XBA approach (see Table 27.1 for examples). But unlike the XBA approach, the various so-called "cross-battery" techniques applied within the field of neuropsychological assessment, for example, are not typically grounded in a systematic approach that is theoretically and psychometrically sound. Thus, as Wilson (1992) cogently pointed out, the field of neuropsychological assessment was in need of an approach that would guide practitioners through the selection of measures that would result in more specific and delineated patterns of function and dysfunction—an approach that would provide more clinically useful information than one "wedded to the utilization of subscale scores and IQs" (p. 382). Indeed, all fields involved in the assessment of cognitive and neuropsychological functioning have some need for an approach that would aid practitioners in their attempt to "touch all of the major cognitive areas, with emphasis on those most suspect on the basis of history, observation, and ongoing test findings" (Wilson, 1992, p. 382; see also Carroll, 1998; McCloskey, Slonim, Whitaker, Kaufman, & Nagoshi, 2017). The XBA approach met this need. The definition of and rationale for XBA are presented in this chapter, followed by a description of the XBA method, information on linking XBA data to intervention, and a case example of its use. Figure 27.1 provides an overview of the information presented in this chapter.

DEFINITION

The XBA approach is a method of assessing cognitive abilities, neuropsychological processes, and academic skills that is grounded mainly in CHC theory and research. It allows practitioners to measure reliably a wider range (or a more in-depth but selective range) of psychological constructs than that represented by any given stand-alone assessment battery. The XBA approach is based on three foundational sources of information (Flanagan, Ortiz, & Alfonso, 2013; Flanagan et al., 2017) that together provide the knowledge base necessary to organize theory-driven, comprehensive assessments.

THE FOUNDATION OF THE XBA APPROACH

The foundation of the XBA approach is CHC theory (Schneider & McGrew, 2012; this volume)[1]—specifically, the broad and narrow CHC classifications of all subtests included in current cognitive, neuropsychological, achievement, speech-language, and special-purpose batteries. The CHC theory was selected to guide assessment and interpretation because it is based on a more thorough network of validity evidence than any other contemporary multidimensional model of intelligence within the psychometric tradition (see Carroll, 1993; Horn & Blankson, 2012; McGrew, 2005; Messick, 1992; Reynolds, Keith, Flanagan, & Alfonso, 2013; Schneider & McGrew, Chapter 3, this volume; Sternberg & Kaufman, 1998). Because CHC theory is discussed in detail by Schneider and McGrew in Chapter 3, it is not described here.

TABLE 27.1. Parallel Needs in Assessment-Related Fields Addressed by the Cross-Battery Assessment (XBA) Approach

Need within assessment-related fields[a]	Need addressed by the XBA approach
School psychology, clinical psychology, and neuropsychology have lagged in the development of conceptual models of the assessment of individuals. There is a need for the development of contemporary models.	The XBA approach provides a contemporary model for measurement and interpretation of cognitive and academic abilities and neuropsychological processes.
It is likely that there is a need for events external to a field of endeavor to give impetus to new developments and real advances in that field.	Carroll and Horn's fluid–crystallized theoretical models and systematic programs of research in cognitive psychology provided the impetus for the XBA approach and led to the development of better assessment instruments and interpretive procedures. Research in fields of cognitive psychology, neuropsychology, and neuroscience give impetus to revisions and refinements to the Cattell–Horn–Carroll (CHC) theory, which in turn led to further development of XBA methods.
There is a need for truly unidimensional assessment instruments for children and adults. Without them, valid interpretations of test scores are problematic at best.	Some scale and composite measures on ability batteries are mixed, containing excess reliable variance associated with a construct irrelevant to the one intended for interpretation. The XBA approach ensures that assessments include composites that are *relatively* pure representations of CHC broad and narrow abilities, allowing for valid measurement and interpretation of multiple, relatively distinct abilities and processes.
There is a need to utilize a conceptual framework to direct any approach to assessment. This would aid both in the selection of instruments and methods, and in the interpretation of test findings.	The XBA approach to assessment is based mainly on CHC theory and research, as well as sound measurement and interpretive procedures. At this time, over 135 psychological tests, including over 1,000 subtests, have been linked to contemporary CHC theory. As such, test selection and interpretation can be accomplished within the context of an overarching conceptual framework.
It is necessary for the conceptual framework or model underlying assessment to incorporate various aspects of neuropsychological and cognitive functioning, which can be described in terms of constructs that are recognized in the neuropsychological and cognitive psychology literature.	The XBA approach includes various aspects of neuropsychological and cognitive functioning, which are described in terms of CHC broad- and narrow-ability constructs that are recognized in the neuropsychological and cognitive psychology literature, as well as other related fields.
There is a need to adopt a conceptual framework that allows for the measurement of the full range of behavioral functions subserved by the brain. Unfortunately, in neuropsychological assessment there is no inclusive set of measures that is standardized on a single normative population.	XBA allows for the measurement of a wide range of broad and narrow cognitive abilities and processes specified in CHC theory. Although an XBA norm group does not exist, the method of crossing batteries and the processes of creating cross-battery composites is psychometrically sound and theoretically defensible.
Because there are no truly unidimensional measures in psychological assessment, there is a need to select subtests from standardized instruments that appear to reflect the neurocognitive function of interest. In neuropsychological assessment, therefore, the aim is to select those measures that, on the basis of careful task analysis, appear mainly to tap a given construct.	The XBA approach is defined in part by a CHC classification system. Subtests from the major intelligence/cognitive batteries, academic achievement tests, neuropsychological instruments, and special purpose tests were classified as measures of broad and narrow CHC constructs. Use of these classifications allows practitioners to be reasonably confident that a given test taps a given construct.

(continued)

TABLE 27.1. *(continued)*

Need within assessment-related fields[a]	Need addressed by the XBA approach
It is clear that an eclectic approach is needed in the selection of measures—preferably subtests rather than the omnibus IQs, in order to gain more specificity in the delineation of patterns of function and dysfunction.	The XBA approach ensures that two or more relatively pure, but qualitatively different, indicators of each *broad* cognitive ability are represented in an assessment of broad CHC constructs. Two or more qualitatively similar indicators are necessary to make inferences about specific or *narrow* CHC constructs. The XBA approach is eclectic in its selection of measures, but attempts to represent all broad and narrow abilities and processes of interest by using a subset of measures from one or more batteries to augment another battery.
There is a need to solve the potential problems that can arise from crossing normative groups as well as sets of measures that vary in reliability.	In the XBA approach, one can typically achieve baseline data in cognitive functioning across seven or eight CHC broad abilities and processes through the use of two well-standardized batteries that were normed within a few years of one another; this minimizes the effects of error due to norming differences. Also, since interpretation of both broad and narrow CHC abilities is made at the composite (rather than subtest) level, issues related to low reliability are less problematic in this approach. And, because confidence intervals are used for all broad- and narrow-ability composites, the effects of measurement error are reduced further. Finally, composite that emerges as a weakness or deficit must have ecological validity (see Flanagan, Ortiz, & Alfonso, 2013, for details).

[a] Information obtained in part from Wilson (1992).

RATIONALE FOR THE APPROACH IN PRACTICE

The XBA approach was introduced years ago as providing "a much needed and updated bridge between current intellectual theory and research and practice" (Flanagan & McGrew, 1997, p. 322). The need for the XBA "bridge" became evident after a review of the results of several cross-battery factor analyses conducted prior to 2000. The results demonstrated that none of the intelligence batteries in use at that time contained measures that sufficiently approximated the full range of broad abilities defining the structure of intelligence specified in contemporary psychometric theory (see Table 27.2; see also Alfonso, Flanagan, & Radwan, 2005, for a comprehensive discussion of these findings). Indeed, the joint factor analyses conducted by Woodcock (1990) suggested that it might be necessary to "cross" batteries to measure a broader range of cognitive abilities than that provided by a single intelligence battery.

As may be seen in Table 27.2, most batteries fell far short of measuring all seven of the broad cognitive abilities listed. Of the major intelligence batteries in use prior to 2000, most failed to measure three or more broad CHC abilities (viz., Ga, Glr, Gf, Gs) that were (and are) considered important in understanding and predicting school achievement (Flanagan, Ortiz, Alfonso, & Mascolo, 2006; McDonough, Flanagan, Sy, & Alfonso, 2017; McGrew & Wendling, 2010). In fact, Gf, often considered to be the *essence* of intelligence, was either not measured or not measured adequately by most of the intelligence batteries included in Table 27.2 (i.e., the Wechsler Intelligence Scale for Children—Third Edition [WISC-III], Wechsler Adult Intelligence Scale—Revised [WAIS-R], Wechsler Preschool and Primary Scale of Intelligence—Revised [WPPSI-R], Kaufman Assessment Battery for Children [K-ABC], and Cognitive Assessment System [CAS]) (Alfonso et al., 2005).

The finding that the abilities either *not measured* or *underrepresented* by the intelligence batteries listed in Table 27.2 are important in understanding children's learning difficulties provided much of the impetus for developing the XBA approach (McGrew & Flanagan, 1998). In effect, the XBA approach was developed to systematically augment cognitive batteries, such as those listed in Table

27.2, with tests from other batteries to better represent important cognitive abilities and processes in assessment and to allow for measurement of those that are specific to the referral but that nonetheless are not measured by the practitioner's battery of choice. As such, the XBA approach guides practitioners in the selection of tests that together provide measurement of abilities and processes that can be considered sufficient in both breadth and depth for addressing referral concerns.

Table 27.3 is like Table 27.2, except that it includes current tests of intelligence and cognitive ability. A comparison of the two tables demonstrates that the latest versions of commonly used tests measure a greater breadth of CHC cognitive abilities and processes than their predeces-

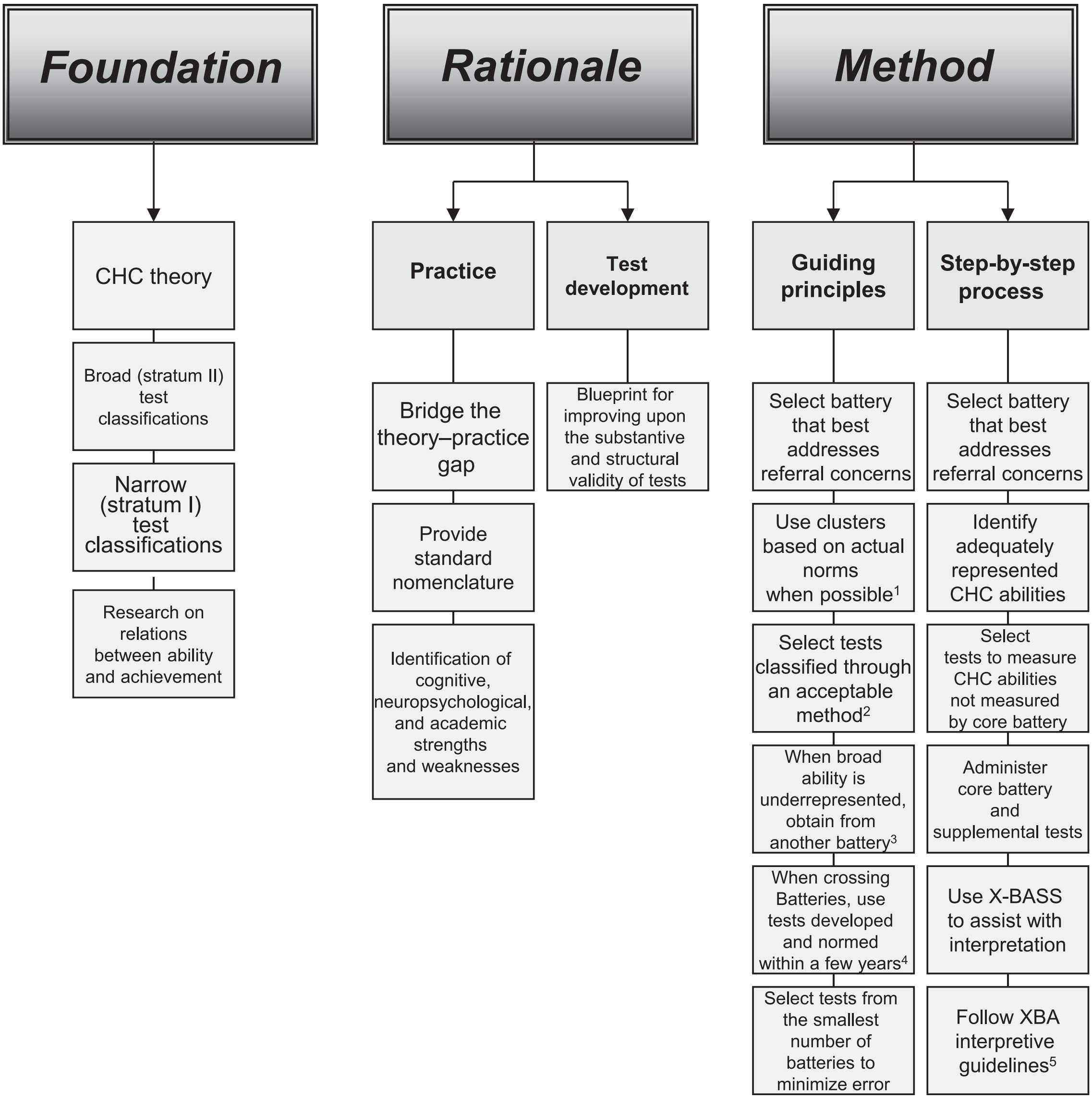


FIGURE 27.1. Overview of the Cattell–Horn–Carroll cross-battery assessment (XBA) approach. CHC, Cattell–Horn–Carroll; X-BASS, Cross-Battery Assessment Software System (Flanagan, Ortiz, & Alfonso, 2017). [1, 3]Facilitated by X-BASS. [2, 4]All tests in X-BASS meet this criterion. [5]Conducted automatically by X-BASS.

TABLE 27.2. Representation of Broad CHC Abilities on Nine Intelligence Batteries Published Prior to 2000

Battery	Gf	Gc	Gv	Gsm	Glr	Ga	Gs
WISC-III	—	Vocabulary Information Similarities Comprehension	Block Design Object Assembly Picture Arrangement Picture Completion Mazes	Digit Span	—	—	Symbol Search Coding
WAIS-R	—	Vocabulary Information Similarities Comprehension	Block Design Object Assembly Picture Completion Picture Arrangement	Digit Span	—	—	Digit–Symbol
WPPSI-R	—	Vocabulary Information Similarities Comprehension	Block Design Object Assembly Picture Completion Mazes Geometric Design	Sentences	—	—	Animal Pegs
KAIT	Mystery Codes Logical Steps	Definitions Famous Faces Auditory Comprehension Double Meanings	Memory for Block Designs	—	Rebus Learning Rebus Delayed Recall Auditory Delayed Recall	—	—
K-ABC	Matrix Analogies	—	Triangles Face Recognition Gestalt Closure Magic Window Hand Movements Spatial Memory Photo Series	Number Recall Word Order	—	—	—

CAS	—	—	Figure Memory Verbal–Spatial Relations Nonverbal Matrices	Word Series Sentence Repetition Sentence Questions	—	—	Matching Numbers Receptive Attention Planned Codes Number Detection Planned Connections Expressive Attention
DAS	Matrices Picture Similarities Sequential and Quantitative Reasoning	Similarities Verbal Comprehension Word Definitions Naming Vocabulary	Pattern Construction Block Building Copying Matching Letter-Like Forms Recall of Designs Recognition of Pictures	Recall of Digits	Recall of Objects	—	Speed of Information Processing
WJ-R	Concept Formation Analysis–Synthesis	Oral Vocabulary Picture Vocabulary Listening Comprehension Verbal Analogies	Spatial Relations Picture Recognition Visual Closure	Memory for Words Memory for Sentences Numbers Reversed	Memory for Names Visual–Auditory Learning Delayed Recall: Memory for Names Delayed Recall: Visual–Auditory Learning	Incomplete Words Sound Blending Sound Patterns	Visual Matching Cross Out
SB-IV	Matrices Equation Building Number Series	Verbal Relations Comprehension Absurdities Vocabulary	Pattern Analysis Bead Memory Copying Memory for Objects Paper Folding and Cutting	Memory for Sentences Memory for Digits	—	—	—

Note. WISC-III, Wechsler Intelligence Scale for Children—Third Edition (Wechsler, 1991); WAIS-R, Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981); WPPSI-R, Wechsler Preschool and Primary Scale of Intelligence—Revised (Wechsler, 1989); KAIT, Kaufman Adolescent and Adult Intelligence Test (Kaufman & Kaufman, 1993); K-ABC, Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983); CAS, Cognitive Assessment System (Naglieri & Das, 1997); DAS, Differential Ability Scales (Elliott, 1990); WJ-R, Woodcock–Johnson Psycho-Educational Battery—Revised (Woodcock & Johnson, 1989); SB-IV, Stanford–Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986).

TABLE 27.3. Representation of Broad and Narrow CHC Abilities on Current Intelligence and Cognitive Batteries

Battery	Gf	Gc	Gv	Gsm	Glr	Ga	Gs
WISC-V	Matrix Reasoning (I) Figure Weights (RG, RQ) Picture Concepts (I) Arithmetic (MW; Gf:RQ)	Similarities (VL, Gf:I) Vocabulary (VL) Information (K0) Comprehension (K0)	Block Design (Vz) Visual Puzzles (Vz) Underrepresented	Digit Span (MW) Picture Span (MS, WM) Letter–Number Sequencing (MW) Arithmetic (MW; Gf:RQ)	Naming Speed Literacy (NA) Naming Speed Quantity (NA; Gs:N) Immediate Symbol Translation (MA) Delayed Symbol Translation (MA) Recognition Symbol Translation (MA)	Not measured	Symbol Search (P) Coding (R9) Cancellation (P)
WAIS-IV	Matrix Reasoning (I) Figure Weights (RQ)	Vocabulary (VL) Information (K0) Similarities (VL, Gf:I) Comprehension (K0)	Block Design (Vz) Picture Completion (CF, Gc:K0) Visual Puzzles (Vz)	Digit Span (MS, MW) Letter–Number Sequencing (MW) Arithmetic (MW; Gf:RQ)	Not measured	Not measured	Symbol Search (P) Coding (R9) Cancellation (P)
WPPSI-IV	Matrix Reasoning (I) Underrepresented	Picture Concepts (Gc:K0, Gf:I) Vocabulary (VL) Information (K0) Similarities (VL, Gf:I) Comprehension (K0) Receptive Vocabulary (VL) Picture Naming (VL)	Block Design (Vz) Object Assembly (CS)	Picture Memory (MS, MW) Zoo Locations (MS, MW)	Not measured	Not measured	Bug Search (P) Cancellation (P)
KABC-II	Pattern Reasoning (I; Gv:Vz) Story Completion (RG, Gc:K0)	Expressive Vocabulary (VL) Verbal Knowledge (VL, K0) Riddles (VL, Gf:RG)	Face Recognition (MV) Triangles (Vz) Gestalt Closure (CS) Rover (SS, Gf:RG)	Number Recall (MS) Word Order (MS, MW) Hand Movements (MS, Gv:MV)	Atlantis (MA) Rebus (MA) Atlantis Delayed (MA)	Not measured	Not measured

			Block Counting (Vz) Conceptual Thinking (Vz; Gf:I)	Underrepresented	Rebus Delayed (MA) Underrepresented		
WJ IV COG	Concept Formation (I) Number Series (RQ) Analysis–Synthesis (RG)	Oral Vocabulary (VL) General Information (K0)	Visualization (Vz) Picture Recognition (MV)	Verbal Attention (MW) Numbers Reversed (MW) Object–Number Sequencing (MW)	Visual–Auditory Learning (MA) Story Recall (MM)	Nonword Repetition (Gsm: MS; Ga:UM) Phonological Processing (PC; Glr: WF) MIXED	Letter–Pattern Matching (P) Pair Cancellation (P) Number–Pattern Matching (P)
SB5	Nonverbal Fluid Reasoning (I; Gv) Verbal Fluid Reasoning (I, RG, Gc:CM) Nonverbal Quantitative Reasoning (RQ, Gq:A3) Verbal Quantitative Reasoning (RQ, Gq:A3)	Nonverbal Knowledge (K0, LS, Gf:RG) Verbal Knowledge (VL,K0)	Nonverbal Visual–Spatial Processing (Vz) Verbal Visual–Spatial Processing (Vz, Gc:VL, K0) Underrepresented	Nonverbal Working Memory (MS, MW) Verbal Working Memory (MS, MW)	Not measured	Not measured	Not measured
DAS-II	Matrices (I) Picture Similarities (I) Sequential and Quantitative Reasoning (RQ)	Early Number Concepts (VL, Gq:A3) Naming Vocabulary (VL) Word Definitions (VL) Verbal Comprehension (LS) Verbal Similarities (VL, Gf:I)	Pattern Construction (Vz) Recall of Designs (MV) Recognition of Pictures (MV) Copying (Vz) Matching Letter-Like Forms (Vz)	Recall of Digits—Forward (MS) Recall of Digits—Backward (MW) Recall of Sequential Order (MW)	Rapid Naming (NA; Gs:R9) Recall of Objects—Immediate (M6) Recall of Objects—Delayed (M6)	Phonological Processing (PC) Underrepresented	Speed of Information Processing (P) Underrepresented

Note. WISC-V, Wechsler Intelligence Scale for Children—Fifth Edition (Wechsler, 2014); WAIS-IV, Wechsler Adult Intelligence Scale—Fourth Edition (Wechsler, 2008); WPPSI-IV, Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (Wechsler, 2012); KABC-II, Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004); WJ IV COG, Woodcock–Johnson IV Tests of Cognitive Abilities (Schrank, McGrew, & Mather, 2014); SB5, Stanford–Binet Intelligence Scales, Fifth Edition (Roid, 2003); DAS-II, Differential Ability Scales—Second Edition (Elliott, 2007).

sors (see also Table 27.4 for brief highlights of the most salient changes across various editions of intelligence and cognitive tests as they pertain to CHC theory). These changes can be attributed to a number of factors, including (1) greater exposure to CHC theory (formerly known as Gf-Gc theory) in the field of school psychology (e.g., Carroll, 1997; Flanagan, Ortiz, & Alfonso, 2007, 2013, 2017; McGrew, 1997, 2005; Schneider & McGrew, 2012; Woodcock, 1990); (2) the development of XBA and its classification system, which shows what intelligence, cognitive ability, achievement, neuropsychological, speech-language, and special purpose tests measure according to CHC theory (Flanagan & McGrew, 1997; Flanagan, Ortiz, & Alfonso, 2017; McGrew, 1997; McGrew & Flanagan, 1998); (3) the publication of CHC-driven confirmatory factor analyses of the standardization data and independent data sets, most notably by Timothy Z. Keith and his colleagues (e.g., Keith, 1997; Keith, Fine, Reynolds, Taub, & Kranzler, 2006; Keith & Reynolds, Chapter 31, this volume; Reynolds & Keith, 2017; Reynolds et al., 2013).

Another contribution of the XBA approach to practice has been that it facilitates communication among professionals. Most scientific disciplines have a standard nomenclature (i.e., a common set of terms and definitions) that facilitates communication and guards against misinterpretation (McGrew & Flanagan, 1998). For example, the standard nomenclature in chemistry is reflected in the periodic table of elements; in biology, it is reflected in the classification of animals according to phyla; in psychology and psychiatry, it is reflected in the *Diagnostic and Statistical Manual of Mental Disorders*; and in medicine, it is reflected in the *International Classification of Diseases*. Underlying the XBA approach is a standard nomenclature or *table of human cognitive abilities* (McGrew & Flanagan, 1998) that includes classifications of hundreds of tests according to the broad and narrow CHC abilities they measure (Flanagan et al., 2017). The XBA classification system has had a positive impact on communication among practitioners; has improved our understanding of and guided the research on the relations between cognitive and academic abilities (McDonough et al., 2017; McGrew & Wendling, 2010; Niileksela, Reynolds, Keith, & McGrew, 2016); and has resulted in improvements in the measurement of cognitive constructs, as may be seen in the design and structure of current cognitive batteries described in this book, most notably the WJ IV (see Schrank & Wendling, Chapter 14, this volume).

Finally, the XBA approach offers practitioners a psychometrically sound means of identifying population-relative (or normative) strengths and weaknesses. The approach focuses interpretation on cognitive ability composites (i.e., via combinations of construct-relevant subtests) that contain either qualitatively different indicators of each CHC broad-ability construct (to represent broad-ability domains) or qualitatively similar indicators of narrow abilities (to represent narrow- or specific-ability domains or processes), thereby making identification of normative strengths and weaknesses reliable. Adhering closely to the guiding principles of the approach (described later) will help to ensure that any identified strengths and weaknesses are interpreted in a theoretically and psychometrically sound manner. In sum, the XBA approach has addressed the long-standing need within assessment-related fields for methods that "provide a greater range of information about the ways individuals learn—the ways individuals receive, store, integrate, and express information" (Brackett & McPherson, 1996, p. 80; see also Decker, 2008; Wilhoit, 2017).

GUIDING PRINCIPLES

To ensure that XBA procedures are theoretically and psychometrically sound, it is recommended that practitioners adhere to seven guiding principles (Flanagan et al., 2013). These principles are listed in Figure 27.1 and are defined briefly below. Note that the use of the Cross-Battery Assessment Software System (X-BASS; Flanagan, Ortiz, & Alfonso, 2017) ensures that guiding principles 5 and 6 are followed.

1. *Select a comprehensive cognitive battery as the core battery in assessment,* ensuring that it is (a) responsive to referral concerns and (b) suitable to meet each examinee's unique needs. Common core batteries include (but are not limited to) the Wechsler scales; the WJ IV; the Stanford–Binet Intelligence Scales, Fifth Edition (SB5); the Differential Ability Scales—Second Editon (DAS-II); the CAS2; and the KABC-II.

2. *Use subtests and composites from a single battery whenever possible to represent broad and narrow CHC abilities.* In other words, best practice involves using actual norms whenever they are available, rather than a formula. However, when it is necessary to follow up on aberrant score per-

TABLE 27.4. Impact of CHC Theory and XBA CHC Classifications on Intelligence Test Development

Test (year of publication) CHC and XBA impact	Revision (year of publication) CHC and XBA impact
WISC-III (1991) No obvious impact. **WISC-IV (2003)** Eliminated Verbal and Performance IQs; replaced the Freedom from Distractibility Index with the Working Memory Index; replaced the Perceptual Organization Index with the Perceptual Reasoning Index; enhanced the measurement of fluid reasoning by adding Matrix Reasoning and Picture Concepts subtests; enhanced measurement of processing speed with the Cancellation subtest.	**WISC-V (2014)** CHC theory was one guiding theory in the development of this battery; the PRI was divided into the Fluid Reasoning Index (Gf) and Visual Spatial Index (Gv); a Glr factor was added that assesses associate memory and retrieval fluency; improvements were made in measurement of working memory and fluid reasoning.
WAIS-III (1997) No obvious impact.	**WAIS-IV (2008)** Minor impact. Eliminated Verbal and Performance IQs; replaced the Perceptual Organization Index with the Perceptual Reasoning Index; enhanced the measurement of fluid reasoning by adding the Figure Weights and Visual Puzzles subtests; enhanced measurement of Processing Speed with the Cancellation subtest; enhanced measurement of memory with the Working Memory Index.
WPPSI-III (2002) Incorporated measures of processing speed that yielded a Processing Speed Quotient, based on recent research indicating the importance of processing speed for early academic success; enhanced the measurement of fluid reasoning by adding the Matrix Reasoning and Picture Concepts subtests.	**WPPSI-IV (2012)** Minor impact. Eliminated Verbal and Performance IQs; enhanced measures of working memory, processing speed, and inhibitory control.
K-ABC (1983) No obvious impact.	**KABC-II (2004)** Provides a second global score that includes fluid and crystallized abilities; includes several new subtests measuring reasoning; interpretation of test performance may be based on CHC theory or Luria's theory; provides assessment of five CHC abilities.
WJ-R (1989) Used modern Gf-Gc theory as the cognitive model for test development; included two measures of each of eight broad abilities. **WJ III NU (2001, 2007)** Used CHC theory as a "blueprint" for test development; included two or three qualitatively different narrow abilities for each broad ability; the combined Cognitive and Achievement batteries of the WJ III NU included 9 of the 10 broad abilities subsumed in CHC theory at the time.	**WJ IV (2014)** Uses expanded CHC theory (Schneider & McGrew, 2012) and introduces cognitive complexity in test development; changes Gsm to Gwm; adds attentional control (AC) under Gwm; added speed of lexical access (LA) under Glr; includes a new Gf-Gc composite; brings back Scholastic Aptitude Clusters; adds an Oral Language battery; provides adequate measurement of 10 broad CHC abilities across batteries.

(continued)

TABLE 27.4. *(continued)*

Test (year of publication) CHC and XBA impact	**Revision (year of publication)** CHC and XBA impact
SB:IV (1986)	**SB5 (2003)**
Used a three-level hierarchical model of the structure of cognitive abilities to guide construction of the test: The top level included a general reasoning factor or g; the middle level included three broad factors called Crystallized Abilities, Fluid-Analytic Abilities, and Short-Term Memory; the third level included more specific factors, including Verbal Reasoning, Quantitative Reasoning, and Abstract/Visual Reasoning.	Uses CHC theory to guide test development; increases the number of broad factors from four to five; includes a Working Memory factor, based on research indicating its importance for academic success.
DAS (1990)	**DAS-II (2007)**
No obvious impact.	Measures seven broad CHC abilities and also includes measures of certain narrow abilities not found on other major cognitive batteries (e.g., F6 or free-recall memory).

Note. WISC-III, Wechsler Intelligence Scale for Children—Third Edition (Wechsler, 1991); WISC-IV, Wechsler Intelligence Scale for Children—Fourth Edition (Wechsler, 2003); WAIS-III, Wechsler Adult Intelligence Scale—Third Edition (Wechsler, 1997); WAIS-IV, Wechsler Adult Intelligence Scale—Fourth Edition (Wechsler, 2008); WPPSI-III, Wechsler Preschool and Primary Scale of Intelligence—Third Edition (Wechsler, 2002); WPPSI-IV, Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (Pearson, 2012); K-ABC, Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983); KABC-II, Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004); WJ-R, Woodcock–Johnson Psycho-Educational Battery—Revised (Woodcock & Johnson, 1989); WJ III NU, Woodcock–Johnson III Tests of Cognitive Abilities Normative Update (Woodcock, McGrew, & Mather, 2001, 2007); SB:IV, Stanford–Binet Intelligence Scale—Fourth Edition (Thorndike, Hagen, & Sattler, 1986); SB5, Stanford–Binet Intelligence Scales, Fifth Edition (Roid, 2003); DAS, Differential Ability Scales (Elliott, 1990); DAS-II, Differential Ability Scales—Second Edition (Elliott, 2007). Adapted from Flanagan, Ortiz, and Alfonso (2005). Copyright © The Guilford Press. Adapted by permission.

formances, to test hypotheses regarding unusual patterns in test data, and so forth, crossing batteries is almost always necessary; therefore, cross-battery composites are helpful in understanding an individual's strengths and weaknesses. When cross-battery composites are necessary to represent broad or narrow abilities and processes, X-BASS provides the most psychometrically defensible approach to creating these composites (see Flanagan, Ortiz, & Alfonso, 2017, for details).

3. When assessment of a broad ability is desired, and the core battery does not contain at least two qualitatively different indicators of that broad ability, *supplement the core battery with at least two qualitatively different indicators of that broad ability from another battery*. For example, if an evaluator is interested in measuring visual processing (Gv), and the core battery has only one or no Gv subtests (or two Gv subtests that measure the same narrow ability), then the evaluator should select two qualitatively different indicators of Gv from another battery and use that Gv composite to supplement the core battery. This practice ensures that actual norms are used for interpreting broad-ability performance whenever they are available. Note that it is not always necessary to assess an ability broadly via two qualitatively different indicators. For example, the WISC-V has a Visual Spatial Index (VSI) that is comprised of two tests of Visualization (Vz). As such, the VSI is better described as a measure of the narrow Vz ability than a measure of the broad Gv ability. Based on the nature of the referral, a Vz composite may be sufficient. Unlike broad abilities, reliable and valid assessment of CHC narrow abilities very often requires crossing batteries. This is because most batteries do not contain two or more measures of the *same* narrow ability. When batteries are crossed to obtain information about narrow abilities, X-BASS may be used to generate theoretically and psychometrically defensible narrow-ability composites.

4. *Select tests from the smallest possible number of batteries* to minimize the effect of spurious differences between test scores that may be attributable to differences in the characteristics of inde-

pendent norm samples (McGrew, 1994). When tests are selected carefully and tailored to address well-defined referral concerns, it is typically not necessary to cross more than two or three batteries. That is, using selected tests from one or two batteries to augment the constructs measured by the core battery is typically sufficient to represent adequately the range of broad and narrow cognitive abilities and processes considered necessary based on referral concerns. However, when there is a need to follow up on aberrant score performances or test hypotheses related to unexpected performances, it is often necessary to select tests from additional batteries.

5. When constructing CHC broad- and narrow-ability clusters, *select tests that have been classified through an acceptable method, such as through CHC theory-driven factor analyses and expert-consensus content validity studies*. The most popular or widely used tests included in X-BASS have been classified through these methods (e.g., Flanagan, Alfonso, & Reynolds, 2013; Flanagan et al., 2006; Flanagan, Ortiz, & Alfonso, 2013) and cross-checked with the construct validity data reported in each test's technical and interpretive manual (Flanagan et al., 2017). Use of X-BASS ensures that this guiding principle is followed.

6. When crossing batteries, *select tests that were developed and normed within a few years of one another,* to minimize the effect of spurious differences between test scores that may be attributable to the so-called "Flynn effect" (Flynn, 1984, 2010). X-BASS includes approximately 135 cognitive, neuropsychological, achievement, language, and special-purpose tests, all of which were normed within 10–12 years of one another. Thus, use of X-BASS ensures that this guiding principle is followed.

7. *Establish ecological validity for test performances that are suggestive of normative weaknesses or deficits*. The finding of a cognitive weakness or deficit is largely meaningless without evidence of how the weakness manifests itself in activities of daily living, including academic achievement (Flanagan, Alfonso, & Mascolo, 2011; Mascolo, Alfonso, & Flanagan, 2014; see also Fiorello & Wycoff, Chapter 26, this volume). The validity of test findings is bolstered when clear connections are made between the cognitive weakness (as measured by standardized tests) and the educational impact of that weakness—for example, as observed in classroom performance and as may be gleaned from a student's work samples. To demonstrate, Table 27.5 includes information about (a) the major cognitive domains of functioning included in CHC theory; (b) the manifestations of deficits in these domains in general, as well as in specific academic areas; and (c) interventions and recommendations that can be tailored to an individual's unique learning needs when such weaknesses are found. In effect, the information in this table depicts the assessment–intervention connection. The key is understanding the manifestations of cognitive deficits for the individual (which provides ecological validity for test scores), which in turn provides a target for intervention. Once the target is identified, then the practitioner is in a better position to select interventions that are likely to minimize the effects of the deficit, allowing the student greater access to the curriculum and instruction.

CHC THEORY, NEUROPSYCHOLOGICAL THEORY, AND RESEARCH KNOWLEDGE BASES

To organize XBAs and interpret XBA data, practitioners should have knowledge of CHC theory or neuropsychological theory (or both) and the expansive research foundations supporting them. Meaningful and defensible interpretation of test data, whether from a single battery or from XBAs, requires knowledge of contemporary theory and research. Such information is critical in the early stages of assessment because it provides the foundation from which to specify the relations between manifest academic performance deficits and suspected underlying cognitive ability and neuropsychological processing deficits. Logical deductions and presuppositions that emanate from current theory and research allow for the formation and subsequent testing of a priori hypotheses.

SPECIFICATION OF A PRIORI HYPOTHESES

The definition of *a priori* as it is used in the XBA approach is found in *The American Heritage Dictionary of the English Language, Fifth Edition* (2018): "Proceeding from a known or assumed cause to a necessarily related effect; deductive." When case history data and current information are coupled with knowledge of contemporary theory and research (and perhaps with information from other fields, such as the literature on specific learning disabilities [SLD]), defensible connections among

TABLE 27.5. CHC Broad Abilities, Their Definitions, Manifestations of CHC Weaknesses, and Suggested Recommendations and Interventions

CHC broad ability	CHC broad-ability definition	General manifestations of the CHC broad-ability weakness	Specific manifestations of the CHC broad-ability weakness in academic areas	Recommendations/interventions
Fluid reasoning (Gf)	• Novel reasoning and problem solving • Processes are minimally dependent on learning and acculturation • Involves manipulating rules, abstracting, generalizing, and identifying logical relationships	***Difficulties with:*** • Higher-level thinking • Transferring or generalizing learning • Deriving solutions for novel problems • Extending knowledge through critical thinking • Perceiving and applying underlying rules or process(es) to solve problems	***Reading difficulties:*** • Inferential reading comprehension • Abstracting main idea(s) • ***Math difficulties:*** • Math reasoning (word problems) • Internalizing procedures and processes used to solve problems • Apprehending relationships between numbers ***Writing difficulties:*** • Essay writing and generalizing concepts • Developing a theme • Comparing and contrasting ideas	• Develop student's skill in categorizing objects and drawing conclusions • Use demonstrations to externalize the reasoning process • Gradually offer guided practice (e.g., guided questions list) to promote internalization of procedures or process(es) • Provide targeted feedback • Use cooperative learning • Use reciprocal teaching • Provide graphic organizers to arrange information in visual format • Use metacognitive strategies • Compare new concepts to previously learned (same vs. different) • Use analogies, similes, and metaphors when presenting tasks
Crystallized intelligence (Gc)	• Breadth and depth of knowledge of a culture • Developed through formal education and general learning experiences • Stores of information and declarative and procedural knowledge • Ability to communicate verbally and reason with previously learned procedures	***Difficulties with:*** • Vocabulary acquisition • Knowledge acquisition • Comprehending language • Fact-based/informational questions • Using prior knowledge to support learning	***Reading difficulties:*** • Decoding and comprehension ***Math difficulties:*** • Understanding math concepts and the "vocabulary of math" ***Writing difficulties:*** • Grammar (syntax) • Bland writing with limited descriptors • Verbose writing with repetitions • Inappropriate word usage ***Language difficulties:*** • Understanding class lessons • Expressive language—"poverty of thought"	• Provide an environment rich in language and experiences • Provide frequent practice with and exposure to words • Read aloud to children • Vary reading purpose (leisure, information) • Work on vocabulary building • Teach morphology • Use text talks
Auditory processing (Ga)	• Ability to analyze and synthesize auditory information	***Difficulties with:*** • Hearing information presented orally, initially processing oral information	***Reading difficulties:*** • Acquiring phonics skills • Decoding and comprehension • Using phonetic strategies	• Provide phonemic awareness activities • Emphasize sight word reading • Teach comprehension monitoring (e.g., "Does the word I heard/read make sense in

		• Paying attention, especially in the presence of background noise • Discerning the direction from which auditory information is coming • Foreign-language acquisition • Acquiring receptive vocabulary	***Math difficulties:*** • Word problems ***Writing difficulties:*** • Spelling • Note taking • Poor quality of writing	context?") • Enunciate sounds in words in an emphatic manner when teaching new words for reading or spelling • Use work preview/text preview to clarify unknown words • Provide guided notes during note-taking activities • Build in time for clarification questions related to "missed" or "misheard" items during lecture • Supplement oral instructions with written instructions • Shorten instructions • Use preferential seating • Localize sound source for student • Minimize background noise
Long-term retrieval (Glr)	• Ability to store information (e.g., concepts, words, facts) and fluently retrieve it later through association	***Difficulties with:*** • Learning new concepts • Retrieving or recalling information by using association • Performing consistently across different task formats (e.g., recognition vs. recall formats) • Speed with which information is retrieved and/or learned • Paired learning (visual–auditory) • Recalling specific information (words, facts)	***Reading difficulties:*** • Accessing background knowledge to support new learning while reading (associative memory deficit) • Slow to access phonological representations during decoding (RAN deficit) ***Math difficulties:*** • Recalling procedures to use for math problems • Memorizing and recalling math facts ***Writing difficulties:*** • Accessing words to use during essay writing • Specific writing tasks (compare and contrast; persuasive writing; conceptual) • Note taking ***Language difficulties:*** • Expressive—circumlocutions, speech fillers, "interrupted" thoughts, pauses • Receptive—making connections throughout oral presentations (e.g., class lectures)	• Provide repeated practice with and review of newly presented information • Teach memory strategies (verbal rehearsal to support encoding, use of mnemonic devices) • Use multiple modalities when teaching new concepts (e.g., pair written with oral information) • Limit the amount of new material to be learned; introduce new concepts gradually and with a lot of context • Be mindful of when new concepts are presented • Make associations between newly learned and prior explicit information • Use lists to facilitate recall (prompts) • Expand vocabulary to minimize impact of word retrieval deficits • Build in waiting time for student when fluency of retrieval is an issue • Use text previews to "prime" knowledge • Provide background knowledge first before asking a question to "prime" student for retrieval

(continued)

TABLE 27.5. *(continued)*

CHC broad ability	CHC broad-ability definition	General manifestations of the CHC broad-ability weakness	Specific manifestations of the CHC broad-ability weakness in academic areas	Recommendations/interventions
Processing speed (Gs)	• Speed of processing, particularly when required to pay focused attention • Usually measured by tasks that require rapid processing, but are relatively easy	***Difficulties with:*** • Efficient processing of information • Quickly perceiving relationships (similarities and differences between stimuli or information) • Working within time parameters • Completing simple, rote tasks quickly	***Reading difficulties:*** • Slow reading speed • Impaired comprehension • Need to reread for understanding • ***Math difficulties:*** • Automatic computations • Computational speed is slow despite accuracy • Slow speed can result in reduced accuracy due to memory decay ***Writing difficulties:*** • Limited output due to time factors • Labored process results in reduced motivation to produce ***Language difficulties:*** • Cannot retrieve information quickly—slow, disrupted speech, as cannot get out thoughts quickly enough • Is slow to process incoming information; puts demands on memory store, which can result in information overload and loss of meaning	• Provide repeated practice • Provide speed drills • Use computer activities that require quick, simple decisions • Provide extended time • Reduce the quantity of work required • Increase waiting times both after questions are asked and after responses are given
Visual processing (Gv)	• Ability to generate, perceive, analyze, synthesize, manipulate, transform, and think with visual patterns and stimuli	***Difficulties with:*** • Recognizing patterns • Reading maps, graphs, and charts • Attending to fine visual detail • Recalling visual information	***Reading difficulties:*** • Orthographic coding (using visual features of letters to decode) • Sight word acquisition • Using charts and graphs within a text in conjunction with reading • Comprehension of text involving spatial	• Capitalize on student's phonemic skills for decoding tasks • Teach orthographic strategies for decoding (e.g., word length, shape of word) • Provide oral explanation for visual concepts • Review spatial concepts and support comprehension through use of hands-on

		• Appreciation of spatial characteristics of objects (e.g., size, length) • Recognition of spatial orientation of objects	concepts (e.g., social studies text describing physical boundaries, movement of troops along a specified route) ***Math difficulties:*** • Number alignment during computations • Reading and interpreting graphs, tables, and charts ***Writing difficulties:*** • Spelling sight words • Spatial planning during writing tasks (e.g., no attention to margins, words that overhang a line) • Inconsistent size, spacing, position, and slant of letters	activities and manipulatives (e.g., using models to demonstrate the moon's orbital path) • Highlight margins during writing tasks • Provide direct handwriting practice • Use graph paper to assist with number alignment
Short-term memory (Gsm)	• Ability to hold information in immediate awareness and use or transform it within a few seconds	***Difficulties with:*** • Following oral and written instructions • Remembering information long enough to apply it Remembering the sequence of information • Rote memorization	***Reading difficulties:*** • Reading comprehension • Decoding multisyllabic words • Orally retelling or paraphrasing what one has read ***Math difficulties:*** • Rote memorization of facts • Remembering mathematical procedures • Multistep problems and regrouping • Extracting information to be used in word problems ***Writing difficulties:*** • Spelling multisyllabic words • Redundancy in writing (word and conceptual levels) • Note taking	• Provide opportunities for repeated practice and review • Provide supports (e.g., lecture notes, study guides, written directions) to supplement oral instruction • Break down instructional steps for student • Provide visual support (e.g., times table) to support acquisition of basic math facts • Outline math procedures for student, and provide procedural guides or flashcards for the student to use when approaching problems • Highlight important information within a word problem • Have student write all steps and show all work for math computations

academic skills, cognitive abilities, and neuropsychological processes can be made. For example, when a student presents with reading difficulties, CHC theory and its research base assist practitioners in identifying the most salient broad and narrow constructs that are related to reading achievement (e.g., short-term working memory, phonetic coding, lexical knowledge, naming facility). The neuropsychological research also supports these same reading ability–processing connections, placing additional focus on the process of attentional capacity as well as specific executive functions. Using this information, a practitioner can logically assume that if the student indeed has cognitive ability or neuropsychological processing deficits that are related to (or that are, in part, the presumptive causes of) reading difficulties, such deficits are likely to be found through an evaluation of the abilities and processes known to explain significant variance in reading achievement (Flanagan & Schneider, 2016). Note that although the practitioner has a suspicion that the individual's reading difficulties may be related to deficits in certain cognitive areas, the a priori hypothesis remains null, indicating that expected performance on any ability or processing test is *within normal limits* (WNL). This is discussed in more detail later.

ORGANIZING AN XBA UTILIZING SPECIFIC REFERRAL INFORMATION

Scenario 1: Cognitive–Achievement Relations

The first scenario relates to the need to evaluate the relationship between an individual's manifest performance (e.g., academic skills) and cognitive abilities and neuropsychological processes. This is often the situation in evaluations conducted in accordance with the Individuals with Disabilities Education Improvement Act of 2004 (IDEA; 2004), which seek to determine the presence of a disability that may be used to establish eligibility for special education programs and services. For example, if there are concerns with reading skills, practitioners should review current research that provides evidence linking cognitive abilities and neuropsychological processes to reading. Practitioners should then ensure that measures of these specific cognitive abilities and processes are included in the initial assessment.

Research on the relationships among cognitive abilities, neuropsychological processes, and specific academic skills has grown over the years (for summaries, see Flanagan, Ortiz, & Alfonso, 2013; Fletcher, Lyon, Fuchs, & Barnes, 2007; McDonough et al., 2017; McGrew & Wendling, 2010; Niileksela et al., 2016). Much of the recent research on cognitive–academic relationships has been interpreted within the context of CHC theory (e.g., Flanagan et al., 2011) and with specific instruments developed from CHC theory (e.g., McGrew & Wendling, 2010; Niileksela et al., 2016). In addition, statistical analyses, such as structural equation modeling, have been used to understand the extent to which specific cognitive abilities explain variance in academic skills above and beyond the variance accounted for by *g* (e.g., Floyd, McGrew, & Evans, 2008; Juarez, 2012; McGrew, Flanagan, Keith, & Vanderwood, 1997; Vanderwood, McGrew, Flanagan, & Keith, 2002). Finally, many valuable resources summarize the research on cognitive and neurobiological processes associated with specific academic skill deficits (e.g., Alfonso & Flanagan, 2018; Feifer & DeFina, 2005; Feifer & Della Toffalo, 2007; Flanagan et al., 2017; Fletcher et al., 2007; Fletcher-Janzen & Reynolds, 2008; Hale & Fiorello, 2004; Miller, 2010, 2013).

Scenario 2: Practical and Legal Considerations

Another scenario that illustrates the effect of referral concerns on test selection and organization in the context of XBA occurs when practical or legal considerations may constrain the evaluation in some way. With respect to practical considerations, it is unreasonable to expect that every practitioner has every published test or has expertise in administering, scoring, and interpreting all available tests. Therefore, decisions regarding test selection and organization will be directly influenced by this reality. For example, of the major cognitive batteries, the KABC-II may be considered the best one for testing a child who, after having exited an English as a second language (ESL) program in fifth grade, is nevertheless falling rapidly behind classmates in most academic areas. However, because the KABC-II does not measure certain abilities and processes important for understanding learning difficulties (e.g., working memory, processing speed, phonological processing), it will need to be supplemented with subtests from another battery (or batteries) with which the practitioner is familiar.

In similar fashion, with respect to legal considerations, there are times when federal or local regulations mandate that certain types of data be collected (e.g., IQ or global ability scores from cognitive batteries). Although this most often occurs in assessments that are conducted to gather data to inform decisions regarding special education eligibility, many states and districts no longer mandate global ability scores for SLD identification. However, in those locations where global ability is still mandated or encouraged, practitioners may find it necessary to obtain the required score even though they may not find it to be particularly informative. For example, instead of administering the WJ IV COG (which measures seven CHC broad cognitive abilities adequately), a practitioner may be required to administer the WISC-V, which measures six, precisely to obtain the Full Scale IQ (FSIQ). Then, if measurement of Ga is warranted, the WISC-V would need to be supplemented with tests from another battery, such as the Comprehensive Test of Phonological Processing—Second Edition (CTOPP-2). Although a WJ IV COG evaluation is more straightforward than the WISC-V/CTOPP-2 cross-battery, the evaluator in this case is constrained by the need to obtain the WISC-V FSIQ.

Scenario 3: Consideration of Examinee Characteristics

The third scenario in which decisions regarding test selection and organization may be highly subject to specific referral concerns involves testing individuals whose characteristics set them apart from the mainstream. For example, practitioners are often called on to assess the abilities of individuals who have sensory or perceptual impairments (e.g., deafness, blindness), who have fine motor impairments (e.g., individuals with cerebral palsy, tremors, seizure activity), or who come from culturally and linguistically diverse backgrounds. Obviously, if an individual is unable to manipulate objects because he or she cannot see or hold them, test selection and organization will be affected. Decisions about test selection and organization are not, of course, specific to conducting XBAs. An individual's unique characteristics must be considered before tests are selected for any evaluation. In the case of individuals who are culturally and linguistically diverse, the Culture–Language Test Classifications (C-LTC) can be utilized to make decisions that respond directly to issues of limited English proficiency or age- or grade-appropriate acculturative knowledge acquisition. This information and the ability to evaluate the validity of obtained results via the Culture–Language Interpretive Matrix (C-LIM) by using X-BASS allows practitioners the opportunity to construct and carry out XBAs that are tailored to specific referral concerns related to individual cultural and linguistic variables. (See Ortiz, Piazza, Ochoa, & Dynda, Chapter 25, this volume, for a discussion of the C-LTC and C-LIM.)

IMPLEMENTING THE XBA APPROACH STEP BY STEP

Practitioners should follow four general steps when conducting XBAs. Each step is described here.

Step 1: Select an Ability Battery

The first step of the XBA approach requires selecting a battery that is appropriate and responsive to various factors: age and developmental level of the examinee; acculturative experiences and background; developmental English-language proficiency of the examinee; the specific referral concerns; and so forth. As such, although a test like the WJ IV COG may be appropriate for a relatively bright and articulate seventh grader who is experiencing difficulties in reading comprehension, it may not be the best instrument of choice for a second grader who is an English-language learner (EL) and who is significantly behind classmates in all academic areas, even though the WJ IV COG provides the most comprehensive coverage of CHC abilities. The WJ IV COG may be less suitable for this scenario because it lacks manipulatives, relies exclusively on verbal instructions, and utilizes some tasks with high receptive language demands (e.g., Concept Formation). In the case of this second grader, a battery such as the KABC-II may be more appropriate because its language demands and cultural loadings are generally lower than those associated with the WJ IV COG.

Step 2: Identify Broad Abilities That Are and Are Not Measured by the Selected Battery; Identify Narrow Abilities and Processes That Need to be Measured

When examiners are interested in a comprehensive evaluation that samples functioning across a wide range of CHC broad abilities (approximately seven or more), nearly all batteries will need to be supplemented via XBA procedures. For example,

Table 27.6 shows that, on average, most major intelligence batteries have adequate representation of five broad CHC abilities. However, in most assessments of suspected SLD, it is desirable to measure abilities and processes across at least seven CHC domains. Therefore, most intelligence and cognitive batteries will need to be supplemented in a comprehensive evaluation for SLD.

Even when seven broad ability areas are measured in an initial assessment, it may be necessary to follow up on significant differences between scores within a broad-ability domain. For example, if an individual scores in the above-average range on a test of general sequential (deductive) reasoning and in the well-below-average range on a test of inductive reasoning, then the resulting Gf composite (reflecting average reasoning ability) is not particularly meaningful from a clinical perspective, even though it is a reliable and valid composite (Schneider & Roman, 2017). That is, the Gf composite in this example masks important information about the individual's reasoning strengths and weaknesses. In this case, it seems necessary to follow up on the inductive reasoning score, since it is suggestive of a weakness or deficit. Because an individual subtest score does not provide a solid basis upon which to render an interpretation of a weakness, another test of inductive reasoning

TABLE 27.6. Representation of Broad CHC Abilities on Selected Cognitive, Achievement, and Neuropsychological Batteries

Battery	Gf	Gc	Gv	Gsm	Glr	Ga	Gs	Grw	Gq	Gkn	Gp	Gh
WISC-V	✓	✓	U	✓	✓	—	✓	—	—	—	—	—
WAIS-IV	✓	✓	✓	✓	—	—	✓	—	—	—	—	—
WPPSI-IV	U	✓	✓	✓	—	—	✓					
WJ IV COG	✓	✓	✓	✓	✓	✓	✓	—	—	—	—	—
SB5	✓	✓	U	✓	—	—	—	—	—	—	—	—
DAS-II	✓	✓	✓	✓	✓	U	U	—	—	—	—	—
KABC-II	✓	✓	✓	U	U	—	—	—	—	—	—	—
KTEA-3	U	✓	—	—	✓	U	✓	✓	U	—	—	—
WIAT-III	U	✓	—	—	U	U	U	✓	U	—	—	—
WJ IV ACH	U	✓	—	—	U	✓	✓	✓	✓	✓	—	—
NEPSY-II	U	✓	✓	✓	✓	U	U	—	—	U	✓	—
D-KEFS	✓	U	U	U	✓	—	✓	—	—	—	U	—
DWNB	—	U	U	U	—	—	—	—	—	—	✓	✓

Note. ✓, *adequately represented* (i.e., the battery contains at least two qualitatively different indicators [subtests] of the broad ability); U, *underrepresented* (i.e., the battery contains only one indicator of the broad ability, or two or more indicators of only one narrow ability subsumed by the broad ability); —, *not measured.* Gf, fluid reasoning; Gc, comprehension–knowledge; Gv, visual processing; Gsm, short-term memory; Glr, long-term storage and retrieval; Ga, auditory processing; Grw, reading and writing; Gq, quantitative knowledge; Gkn, domain-specific knowledge; Gp, psychomotor abilities; Gh, tactile abilities. There are four broad CHC abilities not included in this rapid reference (i.e., olfactory abilities [Go], psychomotor speed [Gps], reaction and decision speed [Gt], and kinesthetic abilities [Gk]). WISC-V, Wechsler Intelligence Scale for Children—Fifth Edition (Wechlser, 2014); WAIS-IV, Wechsler Adult Intelligence Scale—Fourth Edition (Wechsler, 2008); WPPSI-IV, Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (Wechsler, 2012); WJ IV COG and ACH, Woodcock–Johnson IV Tests of Cognitive Abilities and Tests of Achievement (Schrank, McGrew, & Mather, 2014); SB5, Stanford–Binet Intelligence Scales, Fifth Edition (Roid, 2003); DAS-II, Differential Ability Scales—Second Edition (Elliott, 2007); KABC-II, Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004); KTEA-3, Kaufman Test of Educational Achievement, Third Edition (Kaufman & Kaufman, 2014); WIAT-III, Wechsler Individual Achievement Test—Third Edition (Pearson, 2009); NEPSY-II (Korkman, Kirk, & Kemp, 2007); D-KEFS, Delis–Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001); DWNB, Dean–Woodcock Neuropsychological Battery (Dean & Woodcock, 2003).

should be administered. Most batteries do not have two measures of the same narrow abilities (see Table 27.3); therefore, it is often necessary to cross batteries to follow up on below-average score performances.

Finally, even when a battery contains adequate representation of seven main CHC domains and some narrow-ability domains, close inspection of the relations between the cognitive constructs measured by the battery and the referral concern may reveal that the battery does not include specific processes considered important to measure (e.g., orthographic processing in a reading referral or subitizing in a math referral). As such, crossing batteries to obtain information about specific narrow abilities and processes germane to referral concerns is often necessary. Following the XBA approach, X-BASS can be used to examine cross-battery data and create cross-battery composites that are psychometrically defensible, as is demonstrated later in this chapter).

Step 3: Administer and Score Selected Battery and Supplemental Tests

There are no unique administration or scoring instructions associated with XBAs to be followed, apart from those already specified by the test publishers. Practitioners should incorporate general testing and scoring considerations applicable to the use of standardized tests, as well as the specific guidelines provided by test publishers in the manuals of any tests that are used.

Step 4: Enter Scores into X-BASS

After all tests selected for an XBA have been administered, all subtest scaled and standard scores and composite standard scores should be calculated with the appropriate procedures (i.e., manual or hand scoring, computerized scoring). Having a summary of all scores will be very helpful, whether from the original protocol or from a computerized print out, and will facilitate data entry into X-BASS. Once all scores have been calculated, the examiner should open X-BASS, click on the "Start" tab from the "Welcome" screen, and enter demographic information. The individual's age will be calculated automatically; however, his or her grade must be entered manually. Note that it is important to enter the individual's grade for all analyses in X-BASS to run. After demographic data have been entered, click on the "Create New Record" tab and X-BASS will advance automatically to the "Test Index and Main Navigation" tab (or "Index" tab for short). From the index, the examiner clicks on the button corresponding to the battery for which there are scores to be entered. The program will then automatically go to the appropriate test tab, where the examiner may begin to enter data (e.g., WISC-V tab, WJ IV tab, KABC-II tab, WIAT-III tab).

After all initial data have been entered into X-BASS, scores are automatically analyzed to determine whether composites are cohesive or not cohesive and whether there is a need to follow up on any aberrant test performances, regardless of composite cohesion. This is an important point to be noted: Recommendations for follow up are independent of composite cohesion. Follow-up may well be warranted even in cases where the composite is cohesive, and in other cases, no follow up may be necessary even when the composite is not cohesive. Although there is some relation between the two, follow-up recommendations are made independently of cohesion (see Flanagan, Alfonso, Sy, Mascolo, & McDonough, 2018, for a discussion).

INTERPRETING TEST DATA

Test data must be interpreted in a manner that is both theoretically and psychometrically defensible. The XBA approach includes a set of interpretive guidelines that allows practitioners to interpret data from one or more batteries from CHC theory and research (and, in some instances, neuropsychological theory and research) via psychometrically defensible methods. Because the XBA approach represents an advance over traditional assessment practices in terms of both *measurement* and *meaning,* it has informed the interpretive approaches of widely used intelligence batteries (e.g., the WISC-V, WAIS-IV, and KABC-II; see Alfonso et al., 2005; Flanagan & Alfonso, 2017; Flanagan & Kaufman, 2009; Flanagan, Ortiz, & Alfonso, 2013; Lichtenberger & Kaufman, 2009).

Interpretation of XBA data adheres strictly to sound psychometric and statistical precepts that establish the basis for comparative evaluations of test performance, such as interindividual (or population-relative) analysis of cognitive strengths and weaknesses. Interpretation must not, however, be thought of as a separate or distinct endeavor from measurement. Rather, measurement and interpretation are related, and each influences the other in many ways. To interpret data properly, the manner in which measurement and interpretation are related must be specified. To this end, interpretation of test data is embedded in a broader conceptual framework for assessment that relies on the generation and testing of functional assumptions or hypotheses about expected performance. In general, both a priori and a posteriori assumptions are incorporated into the interpretive approach to control for confirmatory bias (explained in the next section). This part of the chapter begins with a discussion of a hypothesis-driven framework and its relationship to the iterative nature of measurement and interpretation. Next, specific guidelines are described that allow practitioners to make defensible interpretations of all data entered into X-BASS.

Hypothesis-Driven Assessment and Interpretation

Inherent in the XBA approach are the emphasis on conducting assessments within a broad, comprehensive framework and the recognition that measurement methods, however precise, might form only a part of the entire scope of assessment-related activities. When standardized testing is to be carried out, practitioners should adhere to guidelines based on a philosophy of hypothesis generation and hypothesis testing. Although psychometric data may seem to be rather objective, interpretation of such data is hardly an unambiguous exercise. Therefore, to reduce the chances of drawing incorrect inferences from test data on the basis of preconceived ideas, hypothesis-generating and hypothesis-testing approaches are necessary and critical components of XBA.

Confirmatory bias occurs when an examiner begins testing with preconceived notions regarding expected performance on a test. After the test is administered and the data are collected, the examiner reviews the data, looking specifically for patterns and results that support the preconception. In other words, the examiner becomes predisposed to seeing only those patterns in the data that support the prevailing assumption, and tends to minimize or reject data that are counter to the assumption (Sandoval, Frisby, Geisinger, Scheuneman, & Grenier, 1998). To reduce the tendency to see patterns of disability and dysfunction in data where in fact no such patterns exist, diagnostic interpretation should not begin with the presumption of preexisting deficits. Rather, interpretation of test data should be guided by the assumption that the examinee is not impaired and that his or her performance on tests (e.g., subtests, composites) will be within normal limits (WNL). The assumption of WNL performance represents the *null hypothesis,* which is evaluated to determine whether it should be retained or rejected in favor of an alternative hypothesis (i.e., performance is not WNL).

Adoption of the stance that performance will be WNL, until and unless convincingly contraindicated by the data, reduces the chance that examiners will view standardized test data only in a manner that corroborates the beliefs they had prior to testing. It is important to note that even when factors other than cognitive weakness or deficiency have been ruled out as the primary cause of observed difficulties (e.g., poor academic performance), it cannot be concluded automatically that an internally based disability is present (e.g., SLD). In every case, the null hypothesis must be expected and retained until the data suggest that such a position is no longer tenable or defensible. Notwithstanding, practitioners can and will entertain thoughts of dysfunction. After all, if standardized testing of a student is being contemplated, it is likely that the examiner has already been made aware of the possibility that a disability exists. However, a clear distinction must be drawn between the specific hypotheses that are to be evaluated and the opinions, conjecture, or suppositions of the examiner. Only the hypotheses specified a priori or a posteriori are actually tested and evaluated directly in light of the data; opinion and conjecture are not. Consequently, unless and until the data strongly suggest otherwise, the null hypothesis that performance is WNL must not be rejected, no matter how strong the examiner's belief, expectation, or desire to the contrary.

When the null hypothesis is rejected in favor of the alternative hypothesis, an examiner can be confident that (1) the data do not support the notion that performance is WNL and (2) performance is likely outside the range of normal limits. Accepting the alternative hypothesis, however, does not provide de facto support for the presence of a disability. The specific reasons for this level of performance should be investigated further and corroborated by additional data sources (e.g., review of school records, work samples, observations, diagnostic interviews). Likewise, failure to reject the null hypothesis does not provide de facto evidence that no disability is present. The range of normal limits encompasses 30 standard score points (SS = 85–115), which spans the below-average, average, and above-average ranges of ability. Performances in the below-average range (e.g., standard scores of 86, 87, 88) may very well be included as part of a constellation of findings that support the presence of a disability. As such, sole reliance on standard scores and cutoff points should be avoided in any diagnostic approach to determining disability, particularly SLD.

Integrating Hypothesis Testing and Data Interpretation

The next discussion is meant to assist practitioners in understanding the various stages of the XBA approach as they apply to interpretation. The assessment and interpretive process requires careful evaluation of case history information (e.g., educational records, progress-monitoring data, authentic measures of achievement, medical records); the inclusion of data from relevant sources (e.g., parents, teachers); and the framing of an individual's difficulties within the context of CHC or neuropsychological theory and research. No matter how compelling the results from the administration of a single ability battery or combination of batteries may appear, test data alone should not be used to make definitive diagnostic decisions.

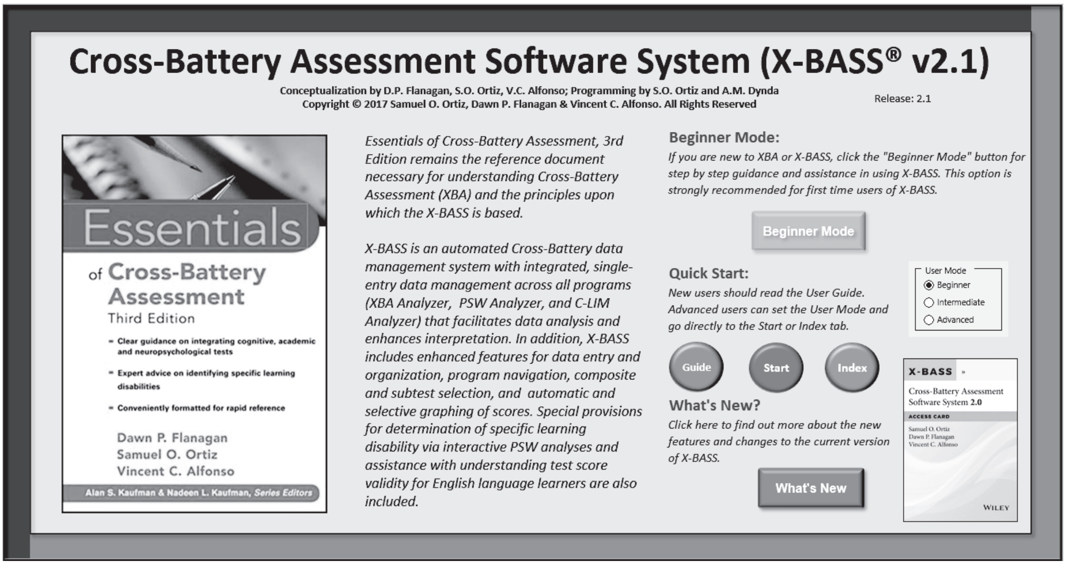
This stage is at the heart of the interpretive process. It is at this point that the examiner can accomplish several different levels of analysis. Such analysis includes evaluation of data yielded from a single battery (e.g., the WISC-V or WJ IV) and from more than one battery (i.e., XBA). Interpretations that are made within the context of the XBA approach are based on *inter* individual comparisons (i.e., population-relative comparison against same-age peers). Thus an individual's performance on both broad and narrow CHC abilities and neuropsychological processes is based on *between*-individual (or normative) comparisons rather than *within*-individual (or person-relative) comparisons, although the latter are certainly important to consider for intervention planning and are incorporated in X-BASS for the WISC-V, in particular.

In general, the current iteration of the XBA approach is based on a model of interpretation that reflects the integration of CHC and neuropsychological theory and research. With the emergence of the field of school neuropsychology (e.g., Fletcher-Janzen & Reynolds, 2008; Hale & Fiorello, 2004; Miller, 2007, 2010, 2013) came a logical extension for linking and integrating CHC theory and neuropsychological theories. Understanding how CHC theory and neuropsychological theories relate to one another expands the options available for interpreting cognitive test performance and improves the quality and clarity of test interpretation, as a much wider research base is available to inform practice (Flanagan, Alfonso, Ortiz, & Dynda, 2010, 2013; Flanagan, Fiorello, & Ortiz, 2010).

Although scientific understanding of the way the brain functions and how mental activity is expressed on psychometric tasks has increased dramatically in recent years, there is still much to be learned. All efforts to create a framework that guides test interpretation benefit from diverse points of view. For example, according to Fiorello, Hale, Snyder, Forrest, and Teodori (2008), "The compatibility of the neuropsychological and psychometric approaches [CHC] to cognitive functioning suggests converging lines of evidence from separate lines of inquiry, a validity dimension essential to the study of individual differences in how children think and learn" (p. 232). Their analysis of the links between the neuropsychological and psychometric approaches not only provides validity for both, but also suggests that each approach may benefit from knowledge of the other. For that reason, a framework that incorporates the neuropsychological and psychometric approaches to cognitive functioning holds the promise of increasing knowledge about the etiology and nature of a variety of disorders (e.g., SLD) and the manner in which such disorders are treated. This type of framework should not only connect the elements and components of both assessment approaches; it should also allow for interpretation of data within the context of either model. In other words, the framework should serve to translate the concepts, nomenclature, and principles of one approach into their counterparts in the other. For a detailed discussion on integrating neuropsychological and CHC theories, see Flanagan, Alfonso, and colleagues (2010).

X-BASS TUTORIAL

Upon opening X-BASS, a clinician can choose his or her User Mode (Beginner, Intermediate, Advanced). This program feature differentiates the amount of pop-up feedback and instruction provided to the clinician by the software. It is recommended that clinicians who are new to XBA or X-BASS choose Beginner Mode for step-by-step guidance (see Figure 27.2). Additionally, it is recommended that new users read the user "Guide" before beginning to analyze test data with X-BASS.



Cross-Battery Assessment Software System (X-BASS® v2.1)

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

Release: 2.1

Essentials of Cross-Battery Assessment, 3rd Edition remains the reference document necessary for understanding Cross-Battery Assessment (XBA) and the principles upon which the X-BASS is based.

X-BASS is an automated Cross-Battery data management system with integrated, single-entry data management across all programs (XBA Analyzer, PSW Analyzer, and C-LIM Analyzer) that facilitates data analysis and enhances interpretation. In addition, X-BASS includes enhanced features for data entry and organization, program navigation, composite and subtest selection, and automatic and selective graphing of scores. Special provisions for determination of specific learning disability via interactive PSW analyses and assistance with understanding test score validity for English language learners are also included.

Beginner Mode:

If you are new to XBA or X-BASS, click the "Beginner Mode" button for step by step guidance and assistance in using X-BASS. This option is strongly recommended for first time users of X-BASS.

Quick Start:

New users should read the User Guide. Advanced users can set the User Mode and go directly to the Start or Index tab.

User Mode
● Beginner
○ Intermediate
○ Advanced

What's New?

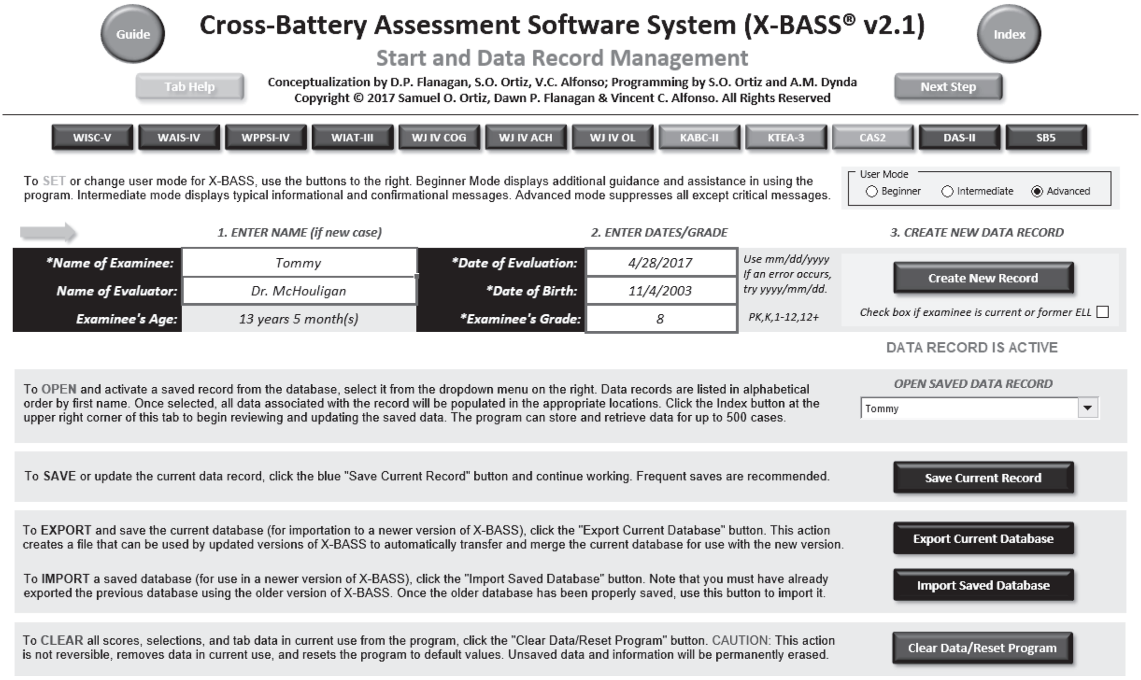
Click here to find out more about the new features and changes to the current version of X-BASS.

FIGURE 27.2. X-BASS Welcome tab.

The "Guide" button is located on the "Welcome" screen as well as subsequent tabs in X-BASS and may be printed by clicking on the "Print" button located in the "Guide" tab. When the user is ready to begin, he or she clicks on the blue Start button, through which the user will be transferred to the Start and Data Record Management tab. Here the student's name, birth date, grade, and date of testing are entered. Additionally, the clinician can indicate if the student is an ELL. When "ELL" is selected, X-BASS is programmed to perform alternative analyses that are specific to ELLs. When the clinician has finished entering the student's identifying information, he or she should click on Create New Record (see Figure 27.3).

The clinician will then be transferred automatically to the Test Index and Main Navigation tab. Here the clinician can choose from any of the major cognitive ability and academic achievement batteries for entering data (see Figure 27.4). From here, the clinician also can access the XBA-CHC Classifications tab, which contains all subtests from all batteries in X-BASS that have been classified according to their CHC broad- and narrow-ability domains, the IDEA-based academic categories related to SLD, and neuropsychological or "other" cognitive domains (see Figure 27.5). Additionally, the index and navigation tab provides access to analysis tabs, graphing tabs, score management tabs, indexes, and references, some of which are discussed below. However, this chapter is only intended to provide a brief overview of the use of X-BASS.

When a clinician is ready to enter test data, he or she clicks on the associated button on the index and navigation tab. This brings the clinician to the Data Analysis tab for the corresponding battery, where the student's scores can be entered. The clinician clicked on the WISC-V tab from the index and navigation tab and entered scores for the Verbal Comprehension Index (see Figure 27.6). If the student's scores are cohesive (as determined by X-BASS) or the composite is judged by the clinician to be a good representation of the examinee's ability in the cognitive domain (regardless of cohesion), the clinician can transfer the composite to the Data Organizer tab. This tab may be thought of as a holding tank for the best estimates of all abilities and processes that will be used in a pattern of strengths and weaknesses analysis (PSW-A). If the examinee's scores are not cohesive, then they may be transferred to the XBA and Test Composite Analyzer tab (XBA Analyzer for short), where they can be combined with a score or scores from another battery for further analysis of performance in that ability or processing domain. The way cohesion is determined in X-BASS is described in the "Guide." When follow-up testing is considered necessary for the lower score in a composite (most notably when the lower score is below average and the higher score is average or better), these scores are typically transferred to



Cross-Battery Assessment Software System (X-BASS® v2.1)

Start and Data Record Management

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

To SET or change user mode for X-BASS, use the buttons to the right. Beginner Mode displays additional guidance and assistance in using the program. Intermediate mode displays typical informational and confirmational messages. Advanced mode suppresses all except critical messages.

User Mode
○ Beginner ○ Intermediate ● Advanced

1. ENTER NAME (if new case)

2. ENTER DATES/GRADE

3. CREATE NEW DATA RECORD

***Name of Examinee:**	Tommy	***Date of Evaluation:**	4/28/2017	Use mm/dd/yyyy If an error occurs, try yyyy/mm/dd.
Name of Evaluator:	Dr. McHouligan	***Date of Birth:**	11/4/2003	
Examinee's Age:	13 years 5 month(s)	***Examinee's Grade:**	8	PK,K,1-12,12+

Check box if examinee is current or former ELL ☐

DATA RECORD IS ACTIVE

To OPEN and activate a saved record from the database, select it from the dropdown menu on the right. Data records are listed in alphabetical order by first name. Once selected, all data associated with the record will be populated in the appropriate locations. Click the Index button at the upper right corner of this tab to begin reviewing and updating the saved data. The program can store and retrieve data for up to 500 cases.

OPEN SAVED DATA RECORD

Tommy

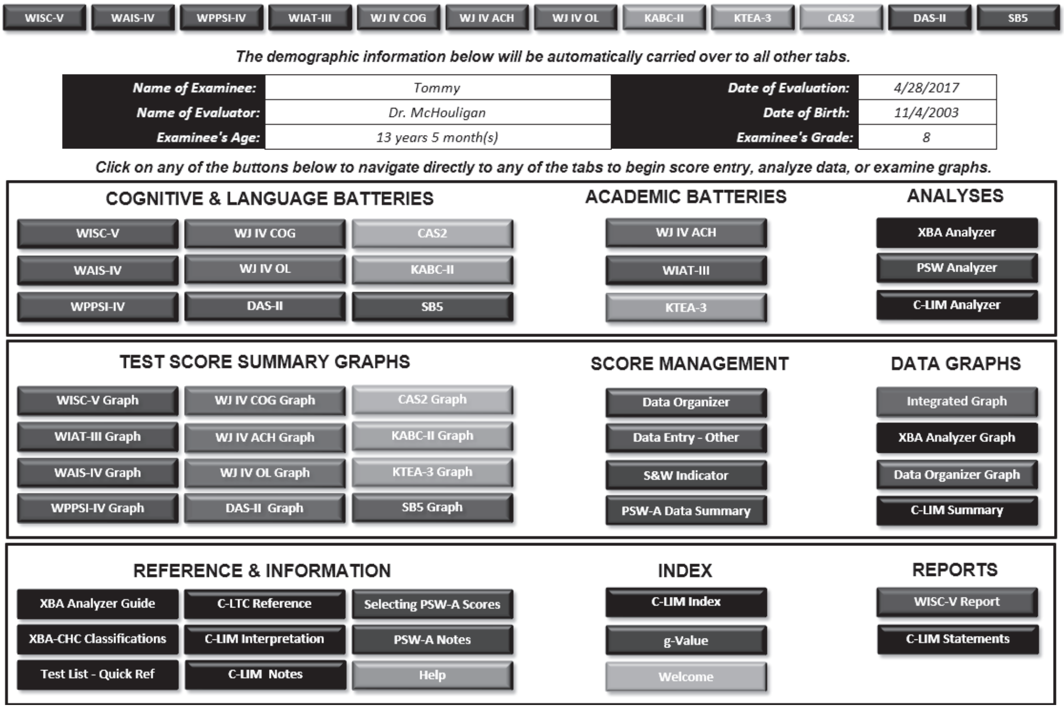
To SAVE or update the current data record, click the blue "Save Current Record" button and continue working. Frequent saves are recommended.

To EXPORT and save the current database (for importation to a newer version of X-BASS), click the "Export Current Database" button. This action creates a file that can be used by updated versions of X-BASS to automatically transfer and merge the current database for use with the new version.

To IMPORT a saved database (for use in a newer version of X-BASS), click the "Import Saved Database" button. Note that you must have already exported the previous database using the older version of X-BASS. Once the older database has been properly saved, use this button to import it.

To CLEAR all scores, selections, and tab data in current use from the program, click the "Clear Data/Reset Program" button. CAUTION: This action is not reversible, removes data in current use, and resets the program to default values. Unsaved data and information will be permanently erased.

FIGURE 27.3. X-BASS Start and Data Record Management tab.

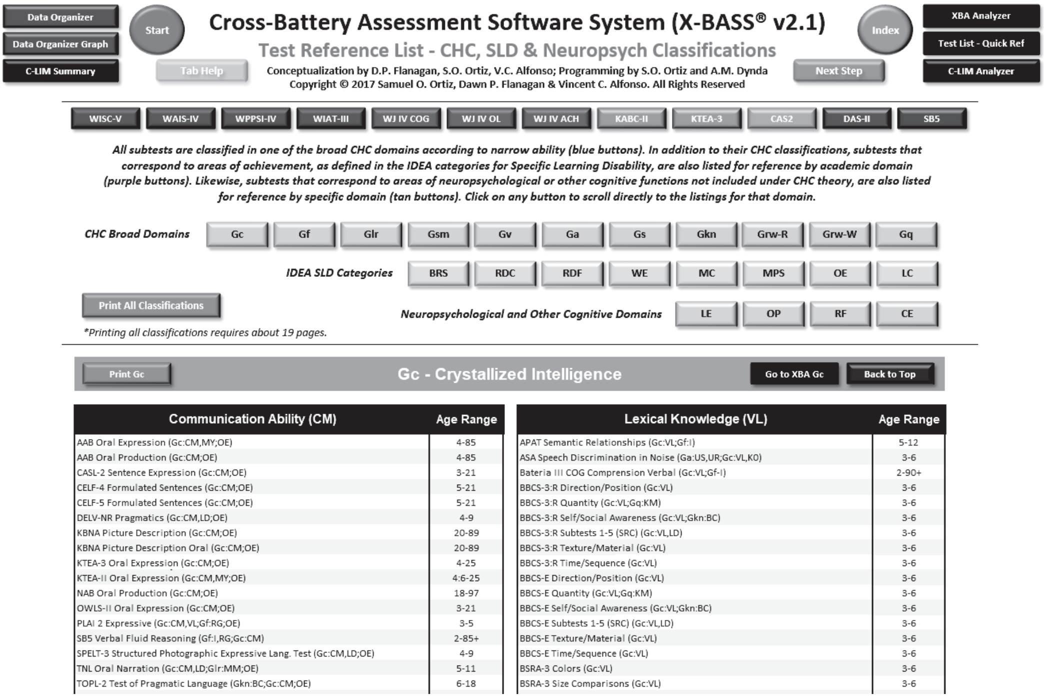


The demographic information below will be automatically carried over to all other tabs.

Name of Examinee:	Tommy	**Date of Evaluation:**	4/28/2017
Name of Evaluator:	Dr. McHouligan	**Date of Birth:**	11/4/2003
Examinee's Age:	13 years 5 month(s)	**Examinee's Grade:**	8

Click on any of the buttons below to navigate directly to any of the tabs to begin score entry, analyze data, or examine graphs.

FIGURE 27.4. Text Index and Main Navigation tab.



Cross-Battery Assessment Software System (X-BASS® v2.1)

Test Reference List - CHC, SLD & Neuropsych Classifications

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

All subtests are classified in one of the broad CHC domains according to narrow ability (blue buttons). In addition to their CHC classifications, subtests that correspond to areas of achievement, as defined in the IDEA categories for Specific Learning Disability, are also listed for reference by academic domain (purple buttons). Likewise, subtests that correspond to areas of neuropsychological or other cognitive functions not included under CHC theory, are also listed for reference by specific domain (tan buttons). Click on any button to scroll directly to the listings for that domain.

CHC Broad Domains

IDEA SLD Categories

Neuropsychological and Other Cognitive Domains

**Printing all classifications requires about 19 pages.*

Gc - Crystallized Intelligence

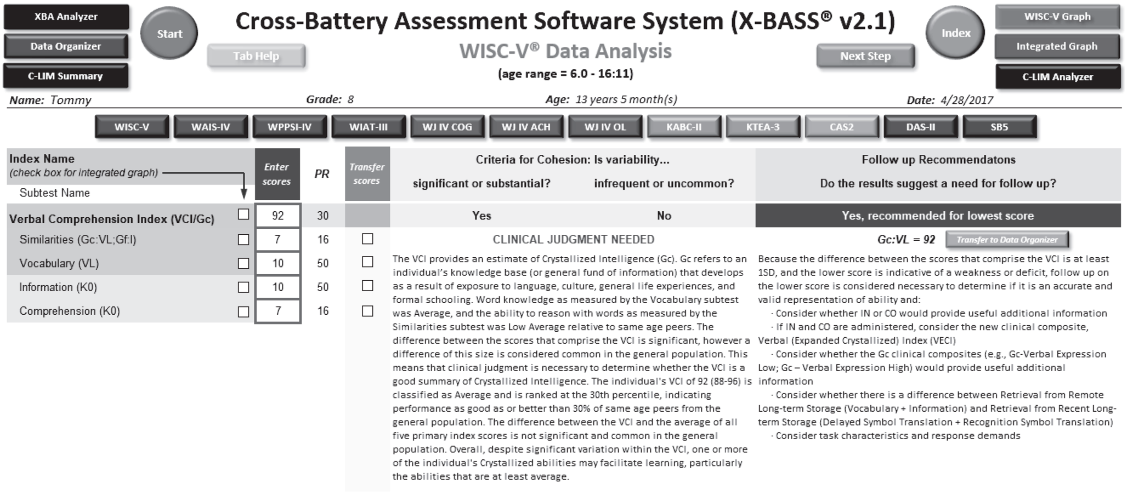
Communication Ability (CM)	Age Range
AAB Oral Expression (Gc:CM,MY;OE)	4-85
AAB Oral Production (Gc:CM;OE)	4-85
CASL-2 Sentence Expression (Gc:CM;OE)	3-21
CELF-4 Formulated Sentences (Gc:CM;OE)	5-21
CELF-5 Formulated Sentences (Gc:CM;OE)	5-21
DELV-NR Pragmatics (Gc:CM,LD;OE)	4-9
KBNA Picture Description (Gc:CM;OE)	20-89
KBNA Picture Description Oral (Gc:CM;OE)	20-89
KTEA-3 Oral Expression (Gc:CM;OE)	4-25
KTEA-II Oral Expression (Gc:CM,MY;OE)	4:6-25
NAB Oral Production (Gc:CM;OE)	18-97
OWLS-II Oral Expression (Gc:CM;OE)	3-21
PLAI 2 Expressive (Gc:CM,VL;Gf:RG;OE)	3-5
SB5 Verbal Fluid Reasoning (Gf:I,RG;Gc:CM)	2-85+
SPELT-3 Structured Photographic Expressive Lang. Test (Gc:CM,LD;OE)	4-9
TNL Oral Narration (Gc:CM,LD;Glr:MM;OE)	5-11
TOPL-2 Test of Pragmatic Language (Gkn:BC;Gc:CM;OE)	6-18

Lexical Knowledge (VL)	Age Range
APAT Semantic Relationships (Gc:VL;Gf:I)	5-12
ASA Speech Discrimination in Noise (Ga:US,UR;Gc:VL,K0)	3-6
Bateria III COG Comprension Verbal (Gc:VL;Gf-I)	2-90+
BBCS-3:R Direction/Position (Gc:VL)	3-6
BBCS-3:R Quantity (Gc:VL;Gq:KM)	3-6
BBCS-3:R Self/Social Awareness (Gc:VL;Gkn:BC)	3-6
BBCS-3:R Subtests 1-5 (SRC) (Gc:VL,LD)	3-6
BBCS-3:R Texture/Material (Gc:VL)	3-6
BBCS-3:R Time/Sequence (Gc:VL)	3-6
BBCS-E Direction/Position (Gc:VL)	3-6
BBCS-E Quantity (Gc:VL;Gq:KM)	3-6
BBCS-E Self/Social Awareness (Gc:VL;Gkn:BC)	3-6
BBCS-E Subtests 1-5 (SRC) (Gc:VL,LD)	3-6
BBCS-E Texture/Material (Gc:VL)	3-6
BBCS-E Time/Sequence (Gc:VL)	3-6
BSRA-3 Colors (Gc:VL)	3-6
BSRA-3 Size Comparisons (Gc:VL)	3-6

FIGURE 27.5. XBA-CHC Classification tab.

the XBA Analyzer where an alternative composite or composites may be calculated that included follow-up testing data. An example of how to conduct follow-up testing is included in the context of the case example below.

When data are transferred to the XBA Analyzer tab, and follow-up testing data are entered into this tab, X-BASS creates XBA composites, which can then be transferred to the Data Organizer and Score Summary tab (or Data Organizer tab for short). For the purpose of PSW analysis, the clinician should have a minimum of seven estimates of cognitive abilities and processes (i.e., referral-related broad and/or narrow abilities



Cross-Battery Assessment Software System (X-BASS® v2.1)

WISC-V® Data Analysis

(age range = 6.0 - 16:11)

Name: *Tommy* ***Grade:*** *8* ***Age:*** *13 years 5 month(s)* ***Date:*** *4/28/2017*

Index Name *(check box for integrated graph)* Subtest Name		*Enter scores*	*PR*	*Transfer scores*
Verbal Comprehension Index (VCI/Gc)	☐	92	30	
Similarities (Gc:VL;Gf:I)	☐	7	16	☐
Vocabulary (VL)	☐	10	50	☐
Information (K0)	☐	10	50	☐
Comprehension (K0)	☐	7	16	☐

Criteria for Cohesion: Is variability... significant or substantial?	infrequent or uncommon?	Follow up Recommendatons Do the results suggest a need for follow up?
Yes	No	Yes, recommended for lowest score

CLINICAL JUDGMENT NEEDED

The VCI provides an estimate of Crystallized Intelligence (Gc). Gc refers to an individual's knowledge base (or general fund of information) that develops as a result of exposure to language, culture, general life experiences, and formal schooling. Word knowledge as measured by the Vocabulary subtest was Average, and the ability to reason with words as measured by the Similarities subtest was Low Average relative to same age peers. The difference between the scores that comprise the VCI is significant, however a difference of this size is considered common in the general population. This means that clinical judgment is necessary to determine whether the VCI is a good summary of Crystallized Intelligence. The individual's VCI of 92 (88-96) is classified as Average and is ranked at the 30th percentile, indicating performance as good as or better than 30% of same age peers from the general population. The difference between the VCI and the average of all five primary index scores is not significant and common in the general population. Overall, despite significant variation within the VCI, one or more of the individual's Crystallized abilities may facilitate learning, particularly the abilities that are at least average.

Gc:VL = 92

Because the difference between the scores that comprise the VCI is at least 1SD, and the lower score is indicative of a weakness or deficit, follow up on the lower score is considered necessary to determine if it is an accurate and valid representation of ability and:

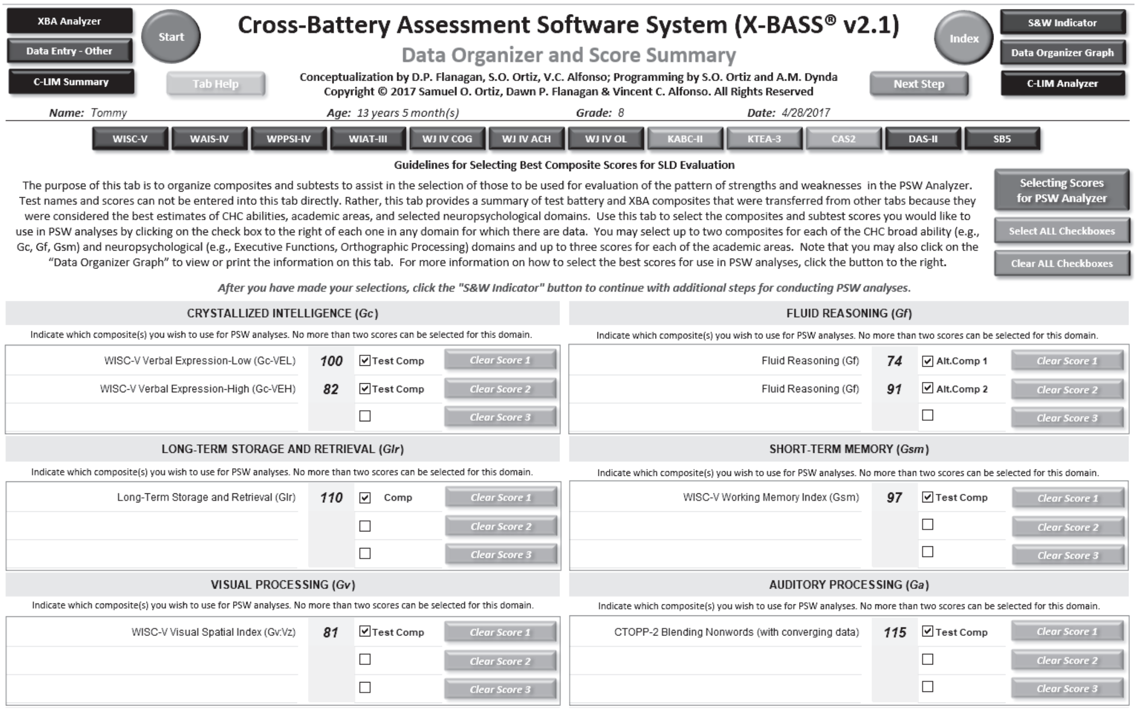
- Consider whether IN or CO would provide useful additional information
- If IN and CO are administered, consider the new clinical composite, Verbal (Expanded Crystallized) Index (VECI)
- Consider whether the Gc clinical composites (e.g., Gc-Verbal Expression Low; Gc – Verbal Expression High) would provide useful additional information
- Consider whether there is a difference between Retrieval from Remote Long-term Storage (Vocabulary + Information) and Retrieval from Recent Long-term Storage (Delayed Symbol Translation + Recognition Symbol Translation)
- Consider task characteristics and response demands

FIGURE 27.6. WISC-V tab with Verbal Comprehension (Gc) Subtest Scaled Scores Entered.

within and across Gf, Gc, Gsm, Glr, Gv, Ga, and Gs) and at least one estimate of achievement (e.g., basic reading skills, reading fluency, math calculation, math problem solving), although the program allows for multiple estimates of achievement across eight areas corresponding to those listed in IDEA. Each estimate of cognitive ability, cognitive processing, and academic achievement should have been transferred to the Data Organizer tab. To conduct a PSW analysis, the clinician clicks the "Select All Checkboxes" button on the Data Organizer tab, which selects all scores that have been transferred to this tab for use in the analysis (see Figure 27.7).

Next, the clinician clicks on the S&W Indicator button, located at in the top right corner of the Data Organizer tab, where he or she is able to select whether a composite is a strength or weakness for the examinee. Typically, scores that are about 90 or higher are considered strengths and scores that fall below 90 are considered weaknesses. However, this is only one guideline that should be considered when designating scores as strengths or weaknesses. For some examinees scores in the upper 80s may be their highest scores and therefore are considered relative strengths for the examinee and should be designated as such. Another guideline for selecting cognitive strengths and weaknesses involves a consideration of whether the score represents an area that is likely facilitating or inhibiting learning and academic performance for the examinee. Thus, clinical judgment is sometimes necessary when determining a strength or weakness, especially when scores fall at or near common cutoff scores, such as 85 or 90. Figure 27.8 shows the S&W tab with all scores marked as either a strength or a weakness, which appear in green and red, respectively, in the actual software, although this and other figures are in black and white. The S&W Indicator tab was purposefully designed to allow for clinical judgment to avoid imposing cutoffs.

When strengths and weaknesses have been determined, the clinician may click on the PSW-A Data Summary button located at the top right of the S&W Indicator tab. The PSW-A Data Summary tab summarizes the data from the S&W Indicator tab (displaying strengths in green and weaknesses in red) and presents the g-value, Facilitating Cognitive Composite (FCC), Alternative Cognitive Composite (ACC), and Inhibiting Cognitive Composite (ICC), terms that are unique to the PSW model operationalized by X-BASS (see Figure 27.9).



Cross-Battery Assessment Software System (X-BASS® v2.1)

Data Organizer and Score Summary

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

Name: *Tommy* **Age:** *13 years 5 month(s)* **Grade:** *8* **Date:** *4/28/2017*

Guidelines for Selecting Best Composite Scores for SLD Evaluation

The purpose of this tab is to organize composites and subtests to assist in the selection of those to be used for evaluation of the pattern of strengths and weaknesses in the PSW Analyzer. Test names and scores can not be entered into this tab directly. Rather, this tab provides a summary of test battery and XBA composites that were transferred from other tabs because they were considered the best estimates of CHC abilities, academic areas, and selected neuropsychological domains. Use this tab to select the composites and subtest scores you would like to use in PSW analyses by clicking on the check box to the right of each one in any domain for which there are data. You may select up to two composites for each of the CHC broad ability (e.g., Gc, Gf, Gsm) and neuropsychological (e.g., Executive Functions, Orthographic Processing) domains and up to three scores for each of the academic areas. Note that you may also click on the "Data Organizer Graph" to view or print the information on this tab. For more information on how to select the best scores for use in PSW analyses, click the button to the right.

After you have made your selections, click the "S&W Indicator" button to continue with additional steps for conducting PSW analyses.

CRYSTALLIZED INTELLIGENCE (*Gc*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

WISC-V Verbal Expression-Low (Gc-VEL)	*100*	☑ Test Comp
WISC-V Verbal Expression-High (Gc-VEH)	*82*	☑ Test Comp
		☐

FLUID REASONING (*Gf*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

Fluid Reasoning (Gf)	*74*	☑ Alt.Comp 1
Fluid Reasoning (Gf)	*91*	☑ Alt.Comp 2
		☐

LONG-TERM STORAGE AND RETRIEVAL (*Glr*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

Long-Term Storage and Retrieval (Glr)	*110*	☑ Comp
		☐
		☐

SHORT-TERM MEMORY (*Gsm*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

WISC-V Working Memory Index (Gsm)	*97*	☑ Test Comp
		☐
		☐

VISUAL PROCESSING (*Gv*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

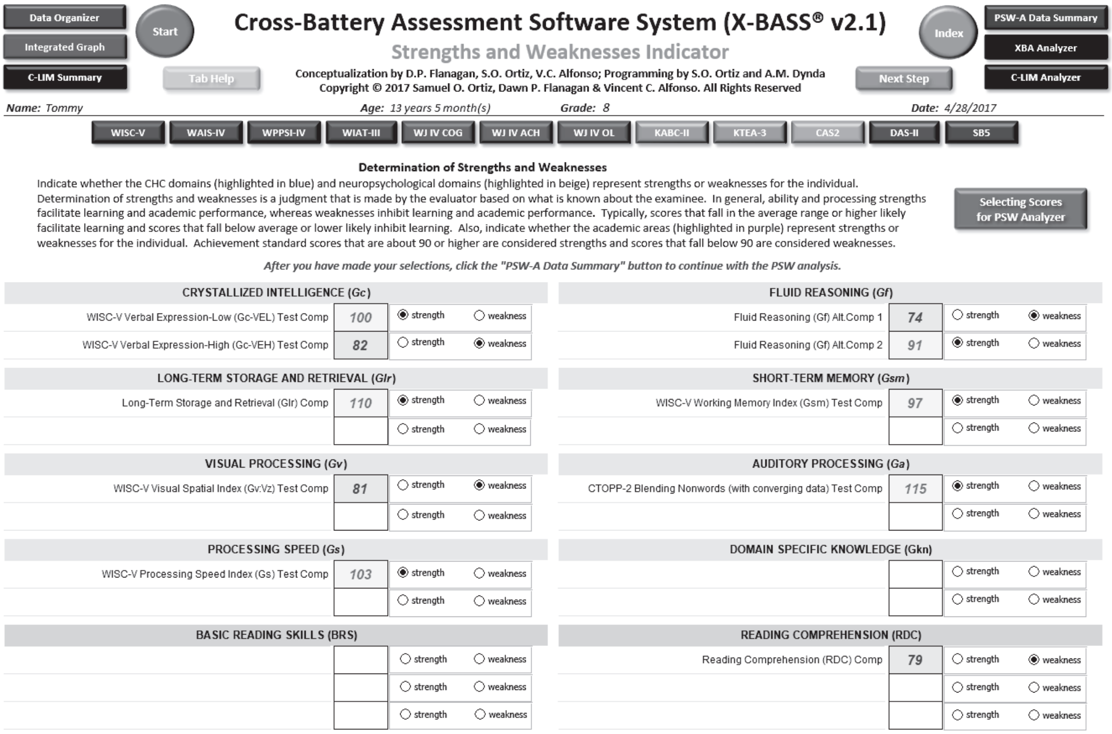
WISC-V Visual Spatial Index (Gv:Vz)	*81*	☑ Test Comp
		☐
		☐

AUDITORY PROCESSING (*Ga*)

Indicate which composite(s) you wish to use for PSW analyses. No more than two scores can be selected for this domain.

CTOPP-2 Blending Nonwords (with converging data)	*115*	☑ Test Comp
		☐
		☐

FIGURE 27.7. Data Organizer tab.



Cross-Battery Assessment Software System (X-BASS® v2.1)

Strengths and Weaknesses Indicator

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

Name: *Tommy* **Age:** *13 years 5 month(s)* **Grade:** *8* **Date:** *4/28/2017*

Determination of Strengths and Weaknesses

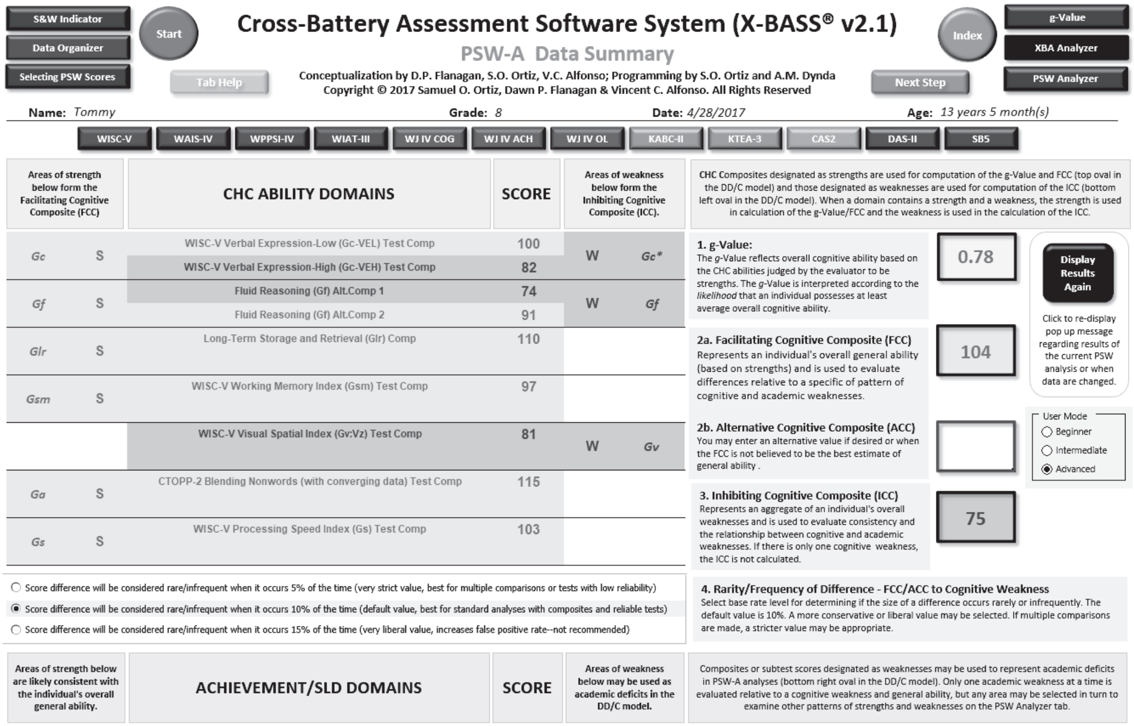
Indicate whether the CHC domains (highlighted in blue) and neuropsychological domains (highlighted in beige) represent strengths or weaknesses for the individual. Determination of strengths and weaknesses is a judgment that is made by the evaluator based on what is known about the examinee. In general, ability and processing strengths facilitate learning and academic performance, whereas weaknesses inhibit learning and academic performance. Typically, scores that fall in the average range or higher likely facilitate learning and scores that fall below average or lower likely inhibit learning. Also, indicate whether the academic areas (highlighted in purple) represent strengths or weaknesses for the individual. Achievement standard scores that are about 90 or higher are considered strengths and scores that fall below 90 are considered weaknesses.

After you have made your selections, click the "PSW-A Data Summary" button to continue with the PSW analysis.

CRYSTALLIZED INTELLIGENCE (Gc)			
WISC-V Verbal Expression-Low (Gc-VEL) Test Comp	100	● strength	○ weakness
WISC-V Verbal Expression-High (Gc-VEH) Test Comp	82	○ strength	● weakness
LONG-TERM STORAGE AND RETRIEVAL (Glr)			
Long-Term Storage and Retrieval (Glr) Comp	110	● strength	○ weakness
		○ strength	○ weakness
VISUAL PROCESSING (Gv)			
WISC-V Visual Spatial Index (Gv:Vz) Test Comp	81	○ strength	● weakness
		○ strength	○ weakness
PROCESSING SPEED (Gs)			
WISC-V Processing Speed Index (Gs) Test Comp	103	● strength	○ weakness
		○ strength	○ weakness
BASIC READING SKILLS (BRS)			
		○ strength	○ weakness
		○ strength	○ weakness
		○ strength	○ weakness

FLUID REASONING (Gf)			
Fluid Reasoning (Gf) Alt.Comp 1	74	○ strength	● weakness
Fluid Reasoning (Gf) Alt.Comp 2	91	● strength	○ weakness
SHORT-TERM MEMORY (Gsm)			
WISC-V Working Memory Index (Gsm) Test Comp	97	● strength	○ weakness
		○ strength	○ weakness
AUDITORY PROCESSING (Ga)			
CTOPP-2 Blending Nonwords (with converging data) Test Comp	115	● strength	○ weakness
		○ strength	○ weakness
DOMAIN SPECIFIC KNOWLEDGE (Gkn)			
		○ strength	○ weakness
		○ strength	○ weakness
READING COMPREHENSION (RDC)			
Reading Comprehension (RDC) Comp	79	○ strength	● weakness
		○ strength	○ weakness
		○ strength	○ weakness

FIGURE 27.8. S&W Indicator tab.



Cross-Battery Assessment Software System (X-BASS® v2.1)

PSW-A Data Summary

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

Name: *Tommy* **Grade:** *8* **Date:** *4/28/2017* **Age:** *13 years 5 month(s)*

Areas of strength below form the Facilitating Cognitive Composite (FCC)		CHC ABILITY DOMAINS	SCORE	Areas of weakness below form the Inhibiting Cognitive Composite (ICC).	
Gc	S	WISC-V Verbal Expression-Low (Gc-VEL) Test Comp	100	W	Gc*
		WISC-V Verbal Expression-High (Gc-VEH) Test Comp	82		
Gf	S	Fluid Reasoning (Gf) Alt.Comp 1	74	W	Gf
		Fluid Reasoning (Gf) Alt.Comp 2	91		
Glr	S	Long-Term Storage and Retrieval (Glr) Comp	110		
Gsm	S	WISC-V Working Memory Index (Gsm) Test Comp	97		
		WISC-V Visual Spatial Index (Gv:Vz) Test Comp	81	W	Gv
Ga	S	CTOPP-2 Blending Nonwords (with converging data) Test Comp	115		
Gs	S	WISC-V Processing Speed Index (Gs) Test Comp	103		

CHC Composites designated as strengths are used for computation of the g-Value and FCC (top oval in the DD/C model) and those designated as weaknesses are used for computation of the ICC (bottom left oval in the DD/C model). When a domain contains a strength and a weakness, the strength is used in calculation of the g-Value/FCC and the weakness is used in the calculation of the ICC.

1. g-Value: 0.78
The *g*-Value reflects overall cognitive ability based on the CHC abilities judged by the evaluator to be strengths. The *g*-Value is interpreted according to the *likelihood* that an individual possesses at least average overall cognitive ability.

2a. Facilitating Cognitive Composite (FCC) 104
Represents an individual's overall general ability (based on strengths) and is used to evaluate differences relative to a specific of pattern of cognitive and academic weaknesses.

2b. Alternative Cognitive Composite (ACC)
You may enter an alternative value if desired or when the FCC is not believed to be the best estimate of general ability .

3. Inhibiting Cognitive Composite (ICC) 75
Represents an aggregate of an individual's overall weaknesses and is used to evaluate consistency and the relationship between cognitive and academic weaknesses. If there is only one cognitive weakness, the ICC is not calculated.

Click to re-display pop up message regarding results of the current PSW analysis or when data are changed.

User Mode
○ Beginner
○ Intermediate
● Advanced

○ Score difference will be considered rare/infrequent when it occurs 5% of the time (very strict value, best for multiple comparisons or tests with low reliability)
● Score difference will be considered rare/infrequent when it occurs 10% of the time (default value, best for standard analyses with composites and reliable tests)
○ Score difference will be considered rare/infrequent when it occurs 15% of the time (very liberal value, increases false positive rate--not recommended)

4. Rarity/Frequency of Difference - FCC/ACC to Cognitive Weakness
Select base rate level for determining if the size of a difference occurs rarely or infrequently. The default value is 10%. A more conservative or liberal value may be selected. If multiple comparisons are made, a stricter value may be appropriate.

Areas of strength below are likely consistent with the individual's overall general ability.	ACHIEVEMENT/SLD DOMAINS	SCORE	Areas of weakness below may be used as academic deficits in the DD/C model.

Composites or subtest scores designated as weaknesses may be used to represent academic deficits in PSW-A analyses (bottom right oval in the DD/C model). Only one academic weakness at a time is evaluated relative to a cognitive weakness and general ability, but any area may be selected in turn to examine other patterns of strengths and weaknesses on the PSW Analyzer tab.

FIGURE 27.9. Pattern of Strengths and Weaknesses Analysis (PSW-A) Data Summary tab.

The *g*-value was developed to estimate the likelihood that the examinee has at least average ability to think and reason (or at least average overall ability) despite specific cognitive weaknesses. The higher the *g*-value, the greater the likelihood of at least average ability to think and reason.

The FCC is an aggregate of the abilities judged by the evaluator to be strengths for the individual, and when the FCC is calculated it means that the examinee has a sufficient breadth of cognitive strengths such that this composite is likely a very good proxy of *g* or general cognitive ability for the examinee. What is unique about the FCC as compared to estimates of overall ability on cognitive and intelligence batteries is that it does not include the examinee's areas of cognitive processing weakness; it is not attenuated by the cognitive processing weaknesses that are presumably interfering with the acquisition and development of the academic skill or skills in question.

The ACC, which is not used often, refers to any cognitive composite derived from an intelligence or cognitive ability battery that is a good estimate of general cognitive ability and considered a better estimate than the FCC. For example, if the *g*-value or FCC is reported by X-BASS in yellow, indicating that the examinee's scores did not demonstrate at least average ability to think and reason despite specific cognitive areas of weakness, then an ACC may be used in place of the FCC if it is judged by the clinician to provide a better estimate of overall cognitive ability as compared to the FCC.

The ICC is an aggregate of the abilities judged by the evaluator to be weaknesses for the individual. In nearly all cases, the PSW-A Data Summary tab does not require data entry. It is simply a summary of strengths and weaknesses for the examinee and reports his or her *g*-value, FCC, and ICC. After reviewing the information on this tab, the clinician clicks on the "*g*-Value tab" button at the top right to advance to this tab. Like the PSW-A Data Summary tab, the *g*-Value tab does not require data entry.

The *g*-Value tab provides information about the likelihood that the individual's pattern of strengths indicates at least average ability to think and reason; it also graphically displays the *g*-value and the scores that contributed to the FCC and ICC (see Figure 27.10). Note that higher *g*-values indicate increased likelihood of at least average ability to think and reason, with *g*-values above .60 indicating "very likely" and that any identified deficits are probably specific in nature. For more detailed information on the development and meaning of the *g*-value, FCC, and ICC, the clinician may go to the Test Index and Main Navigation tab (see Figure 27.4), locate the "Reference & Information" section, and click on the "PSW-A Notes" button.

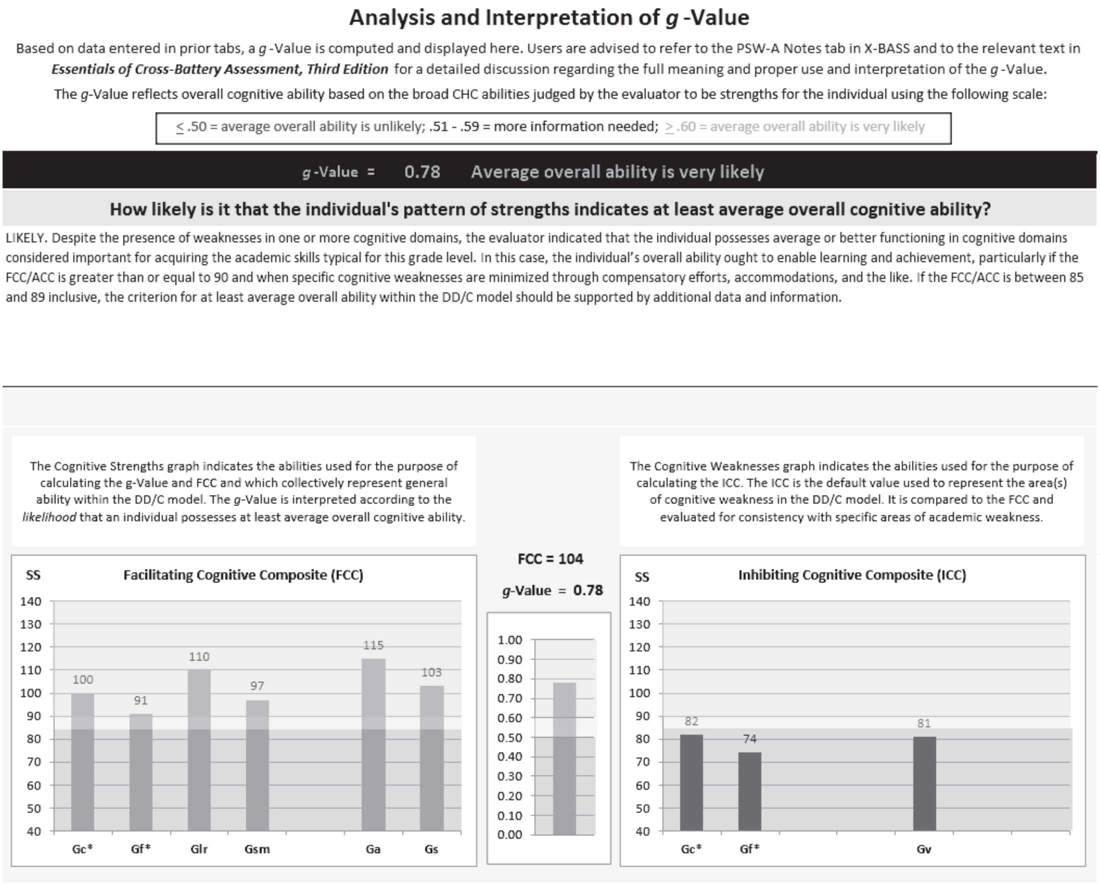
Last, the clinician clicks on the PSW Analyzer button at the top right of the *g*-Value tab, which takes the examiner to a tab labeled "Dual-Discrepancy/Consistency Model: PSW Analyses for SLD" (or PSW Analyzer for short). As shown in Figure 27.11, the tab automatically conducts the PSW analysis according to the DD/C method, which includes (1) a domain-specific cognitive ability weakness with at least average ability to think and reason, (2) unexpected underachievement, and (3) an empirical relationship between the domain-specific cognitive weakness and the academic weakness. The PSW Analyzer tab indicates whether each of these criteria are met through the use of statistical tests, regression analyses, base rate data, and corrections for false negatives. Additionally, the PSW Analyzer tab indicates whether both the cognitive and academic areas are considered weaknesses, and what the strength of the relationship is between the cognitive and academic areas (i.e., low, moderate, high), based on the research. (Again, see Flanagan et al., Chapter 22, this volume, for a more complete explanation of the DD/C method.) The case example below illustrates the use of X-BASS to assist in XBA and in informing SLD identification.

CASE EXAMPLE USING X-BASS FOR XBA AND PSW ANALYSIS

In the following section we summarize a case of a student who is having difficulties in math problem solving and reading comprehension. Use of the XBA and PSW components of X-BASS are highlighted.

Referral Concern

Tommy is a 13-year-old male student in an eighth-grade general education classroom. He was referred by his math teacher due to significant difficulties with math problem solving, such as formulating and reasoning about expressions and equations. In addition, his Language Arts teacher expressed concerns related to Tommy's reading comprehension, such as drawing inferences from text, distinguishing among the connotations of words that are similar in meaning, and understanding word relationships.



Analysis and Interpretation of *g* -Value

Based on data entered in prior tabs, a *g* -Value is computed and displayed here. Users are advised to refer to the PSW-A Notes tab in X-BASS and to the relevant text in ***Essentials of Cross-Battery Assessment, Third Edition*** for a detailed discussion regarding the full meaning and proper use and interpretation of the *g* -Value.

The *g*-Value reflects overall cognitive ability based on the broad CHC abilities judged by the evaluator to be strengths for the individual using the following scale:

≤ .50 = average overall ability is unlikely; .51 - .59 = more information needed; ≥ .60 = average overall ability is very likely

***g* -Value = 0.78 Average overall ability is very likely**

How likely is it that the individual's pattern of strengths indicates at least average overall cognitive ability?

LIKELY. Despite the presence of weaknesses in one or more cognitive domains, the evaluator indicated that the individual possesses average or better functioning in cognitive domains considered important for acquiring the academic skills typical for this grade level. In this case, the individual's overall ability ought to enable learning and achievement, particularly if the FCC/ACC is greater than or equal to 90 and when specific cognitive weaknesses are minimized through compensatory efforts, accommodations, and the like. If the FCC/ACC is between 85 and 89 inclusive, the criterion for at least average overall ability within the DD/C model should be supported by additional data and information.

The Cognitive Strengths graph indicates the abilities used for the purpose of calculating the g-Value and FCC and which collectively represent general ability within the DD/C model. The *g*-Value is interpreted according to the *likelihood* that an individual possesses at least average overall cognitive ability.

The Cognitive Weaknesses graph indicates the abilities used for the purpose of calculating the ICC. The ICC is the default value used to represent the area(s) of cognitive weakness in the DD/C model. It is compared to the FCC and evaluated for consistency with specific areas of academic weakness.

FCC = 104

***g*-Value = 0.78**

FIGURE 27.10. *g*-Value tab.

Choosing a Core Test Battery

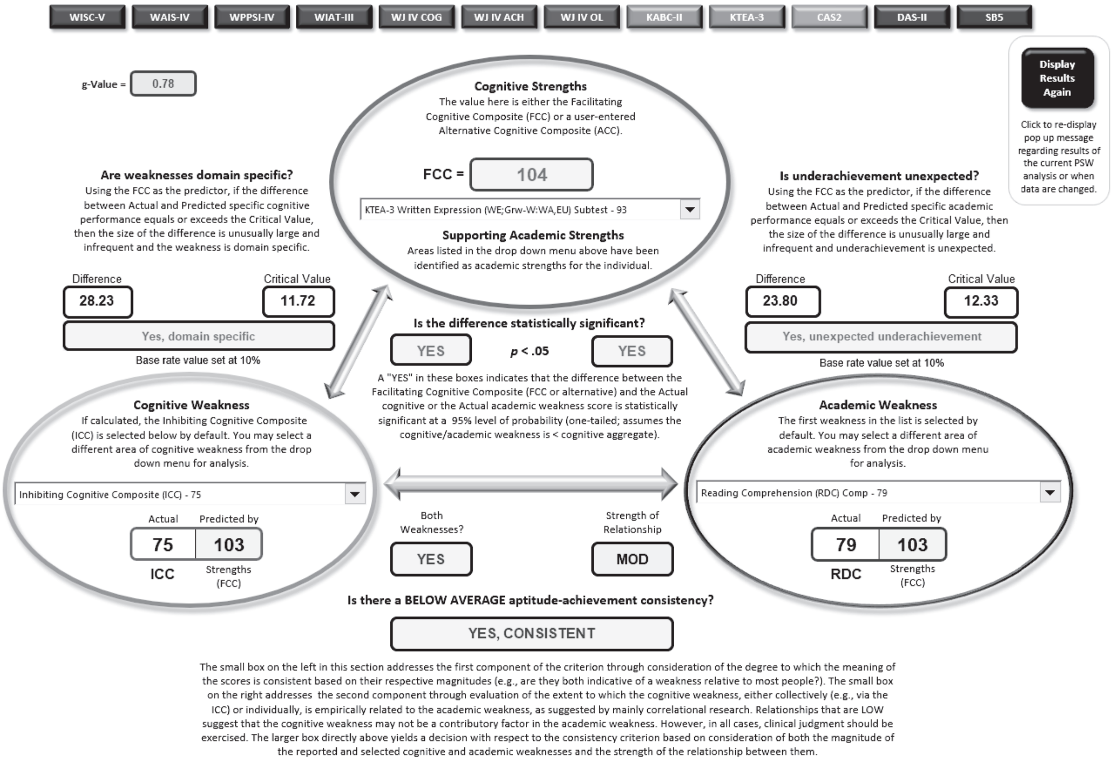
When an examiner is deciding on a core cognitive battery, the following considerations are necessary: referral concern, hypotheses based on the referral concern, age, developmental level, and cultural/linguistic background (e.g., English proficiency, level of acculturation, number of years in the United States). Regarding Tommy's case, there is evidence that fluid reasoning, crystalized intelligence, and short-term working memory are related empirically to both math problem solving and reading comprehension, as well as evidence that visual processing is related to higher-level math. There were no concerns about developmental delays; also, Tommy and his parents were born in the United States and speak only English at home, so cultural/linguistic differences are not involved. The WISC-V was chosen as the core battery for Tommy because it measures the main cognitive abilities and processes that are related empirically to his reported academic difficulties. When choosing an achievement battery, clinicians must consider which batteries most appropriately assess the areas of weakness. In regard to the current case, the Kaufman Test of Educational Achievement, Third Edition (KTEA-3) includes measures of math problem solving and reading comprehension that approximate task demands in the classroom. The WISC-V and KTEA-3 are also statistically linked.

Choosing Supplemental Test Batteries

An examiner who is deciding on supplemental tests must first identify which broad and narrow abilities and processes related to the referral concerns are not measured adequately by the core battery. It is helpful to have an understanding of the abilities and processes most closely associated with areas of academic concern when making this determination and when choosing supplemental tests. The cognitive–achievement relations research was summarized in Tables 22.5 through 22.7 in Chapter 22 (this volume).

In evaluations where the examiner wants to conduct a PSW analysis using X-BASS, a few pointers are in order. First, the PSW analysis requires estimates of performance in seven cognitive areas, namely Gf, Gc, Gwm, Glr, Ga, Gv, and Gs. Estimates in these seven cognitive areas are necessary for the calculation of the *g*-value, FCC, and ICC. Second, other areas of cognitive processing, such as executive functions and orthographic processing, may also be included in the PSW analysis, even though these estimates do not contribute the *g*-value, FCC, and ICC. Third, estimates of cognitive abilities and processes do not need to be *broad* estimates (i.e., comprised of two qualitatively different indicators of the cognitive area). For example, the areas of Ga and Gs are very often *narrow* estimates (i.e., comprised of two similar indicators). In the area of Ga, the estimate typically represents the narrow process of Phonetic Coding (especially in reading referrals); in the area of Gs, the estimate typically represents the narrow process of Perceptual Speed. It is up to the evaluator to determine what estimates are most relevant for the referral concerns—broad or narrow or, in some cases, both. Fourth, when there are strengths and weaknesses within a cognitive area (e.g., average Induction and below average General Sequential Reasoning in the area of Gf), both estimates (i.e., the strength and the weakness) should be used to represent Gf in the PSW analysis. Fifth, a cognitive area may be represented by a single subtest score in some instances, particularly when the subtest score is indicative of average or better performance. Examples are included in this case. Nevertheless, it is important to remember that single subtest scores make for poor measurement (Flanagan & Schneider, 2016). As such, before using a single (*average or better*) subtest score to represent a narrow cognitive area, converging data sources are necessary (e.g., data from previous evaluations, work samples, teacher reports).

Based on a review of the cognitive–achievement relations research and the requirements of a PSW analysis, Tommy's evaluator chose to use



g-Value = 0.78

Click to re-display pop up message regarding results of the current PSW analysis or when data are changed.

Cognitive Strengths

The value here is either the Facilitating Cognitive Composite (FCC) or a user-entered Alternative Cognitive Composite (ACC).

FCC = 104

KTEA-3 Written Expression (WE;Grw-W:WA,EU) Subtest - 93

Supporting Academic Strengths

Areas listed in the drop down menu above have been identified as academic strengths for the individual.

Are weaknesses domain specific?

Using the FCC as the predictor, if the difference between Actual and Predicted specific cognitive performance equals or exceeds the Critical Value, then the size of the difference is unusually large and infrequent and the weakness is domain specific.

Difference 28.23 Critical Value 11.72

Yes, domain specific

Base rate value set at 10%

Is underachievement unexpected?

Using the FCC as the predictor, if the difference between Actual and Predicted specific academic performance equals or exceeds the Critical Value, then the size of the difference is unusually large and infrequent and underachievement is unexpected.

Difference 23.80 Critical Value 12.33

Yes, unexpected underachievement

Base rate value set at 10%

Is the difference statistically significant?

YES $p < .05$ YES

A "YES" in these boxes indicates that the difference between the Facilitating Cognitive Composite (FCC or alternative) and the Actual cognitive or the Actual academic weakness score is statistically significant at a 95% level of probability (one-tailed; assumes the cognitive/academic weakness is < cognitive aggregate).

Cognitive Weakness

If calculated, the Inhibiting Cognitive Composite (ICC) is selected below by default. You may select a different area of cognitive weakness from the drop down menu for analysis.

Inhibiting Cognitive Composite (ICC) - 75

	Actual	Predicted by
	75	103
	ICC	Strengths (FCC)

Academic Weakness

The first weakness in the list is selected by default. You may select a different area of academic weakness from the drop down menu for analysis.

Reading Comprehension (RDC) Comp - 79

	Actual	Predicted by
	79	103
	RDC	Strengths (FCC)

Both Weaknesses? YES

Strength of Relationship MOD

Is there a BELOW AVERAGE aptitude-achievement consistency?

YES, CONSISTENT

The small box on the left in this section addresses the first component of the criterion through consideration of the degree to which the meaning of the scores is consistent based on their respective magnitudes (e.g., are they both indicative of a weakness relative to most people?). The small box on the right addresses the second component through evaluation of the extent to which the cognitive weakness, either collectively (e.g., via the ICC) or individually, is empirically related to the academic weakness, as suggested by mainly correlational research. Relationships that are LOW suggest that the cognitive weakness may not be a contributory factor in the academic weakness. However, in all cases, clinical judgment should be exercised. The larger box directly above yields a decision with respect to the consistency criterion based on consideration of both the magnitude of the reported and selected cognitive and academic weaknesses and the strength of the relationship between them.

FIGURE 27.11. PSW Analyzer tab: Inhibiting Cognitive Composite and Math Problem Solving.

the statistically linked WISC-V and KTEA-3 and supplement these batteries with selected subtests from the CTOPP-2 to assess phonological processing (which is not measured by the WISC-V and not measured well by the KTEA-3) as well as selected subtests from the KeyMath3 to assess math skills that are not measured on the KTEA-3. Testing hypotheses about aberrant score performances may require using subtests from additional batteries. The evaluation results and decisions made throughout the evaluation are described next.

Cognitive Test Data

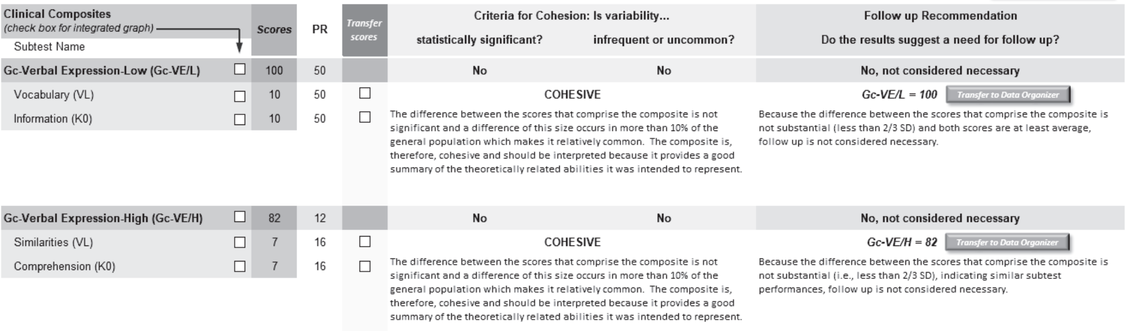
Crystallized Intelligence

Results from the WISC-V revealed an average score on Vocabulary (scaled score = 10) and a below average score on Similarities (scaled score = 8), both of which measure Lexical Knowledge. Together these scores yield a VCI of 92, suggesting that Tommy's Crystallized Intelligence (Gc), particularly his vocabulary knowledge, is average. These data are shown in Figure 27.4. However, Similarities includes a reasoning component, whereas Vocabulary does not. When these subtest scores are entered into X-BASS, the program shows that the difference between the two scores is statistically significant, but not unusual in the general population. This means that clinical judgment is necessary to determine if the VCI of 92 is a good representation of Tommy's Gc. X-BASS assists in determining this by indicating whether follow-up assessment is considered necessary for the lower score in the composite. In this case, the program recommends following up on the Similarities score performance. These decisions regarding composite cohesion and follow-up for the VCI are shown in Figure 27.4.

Tommy's evaluator chose to, but was not required (by XBA procedures) to, administer the remaining two Gc subtests on the WISC-V (i.e., Information and Comprehension), both of which measure the narrow ability of General (verbal) Information. Interestingly, Comprehension has a reasoning component, whereas Information does not. When all four Gc subtest scores are entered on the WISC-V tab, three *clinical composites* are generated based on actual norms provided by Pearson (see Flanagan et al., 2017). The first clinical composite, the Verbal-Expanded Crystallized Index (VECI; originally published in Raiford, Drozdick, Zhang, & Zhou, 2015) is made up of all four Gc subtests. The VECI was 91, which is considered cohesive with no need for follow-up. Psychometrically, the VECI is the best estimate of Tommy's Gc. Clinically, however, Tommy's evaluator believes that the VECI of 91 masks important information about his strengths and weaknesses, suggesting the need to explore why he demonstrated below average performance on two of the four Gc subtests.

Two additional clinical composites revealed that Tommy performed better on the Gc subtests that require low verbal expression (100) as compared to those that required high verbal expression (82) (see Figure 27.12). However, Tommy's expressive language has never been in question. Upon closer examination, it was clear that the subtests that required higher verbal expression were those that also involved reasoning (i.e., Similarities and Comprehension). Thus the composite of 82 could also be interpreted as an indication of Tommy's ability to reason with verbal information, an ability that is consistent with referral concerns.

In the area of Gc, the evaluator chose to report the VECI of 91 as the best estimate of Tommy's Gc. Additionally, the evaluator chose to report



Clinical Composites (check box for integrated graph) Subtest Name		Scores	PR	Transfer scores	Criteria for Cohesion: Is variability... statistically significant?	infrequent or uncommon?	Follow up Recommendation Do the results suggest a need for follow up?
Gc-Verbal Expression-Low (Gc-VE/L)	☐	100	50		**No**	**No**	**No, not considered necessary**
Vocabulary (VL)	☐	10	50	☐	COHESIVE		*Gc-VE/L = 100*
Information (K0)	☐	10	50	☐	The difference between the scores that comprise the composite is not significant and a difference of this size occurs in more than 10% of the general population which makes it relatively common. The composite is, therefore, cohesive and should be interpreted because it provides a good summary of the theoretically related abilities it was intended to represent.		Because the difference between the scores that comprise the composite is not substantial (less than 2/3 SD) and both scores are at least average, follow up is not considered necessary.
Gc-Verbal Expression-High (Gc-VE/H)	☐	82	12		**No**	**No**	**No, not considered necessary**
Similarities (VL)	☐	7	16	☐	COHESIVE		*Gc-VE/H = 82*
Comprehension (K0)	☐	7	16	☐	The difference between the scores that comprise the composite is not significant and a difference of this size occurs in more than 10% of the general population which makes it relatively common. The composite is, therefore, cohesive and should be interpreted because it provides a good summary of the theoretically related abilities it was intended to represent.		Because the difference between the scores that comprise the composite is not substantial (i.e., less than 2/3 SD), indicating similar subtest performances, follow up is not considered necessary.

FIGURE 27.12. WISC-V tab: Gc Clinical Composites.

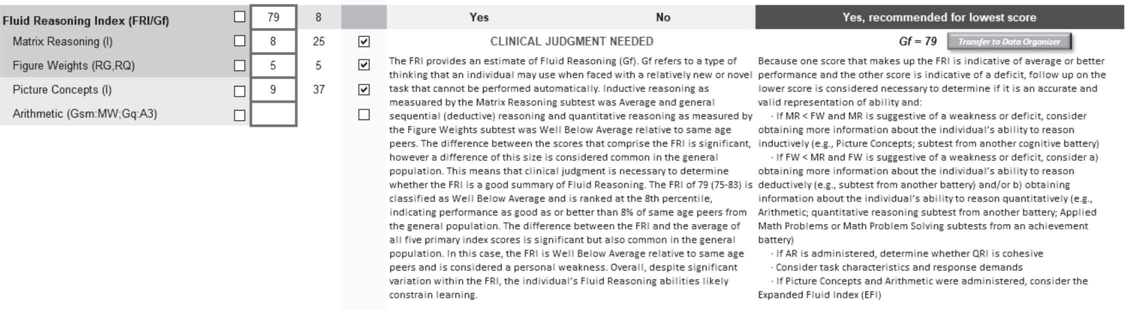
and use 100 and 82 (as depicted in Figure 27.12) to represent his stores of acquired knowledge and his ability to reason with that knowledge, respectively, in the PSW Analysis. As such, the evaluator clicked on the "Transfer to Data Organizer" buttons to the far right of each of these composites (see Figure 27.12) to represent his strength and his weakness in Gc in the PSW analysis.

Fluid Reasoning

Results from the WISC-V revealed an average score on Matrix Reasoning (scaled score = 8), a measure of Induction, and a well below average score on Figure Weights (scaled score = 5), a measure of General Sequential Reasoning and Quantitative Reasoning. Together, these scores yield an FRI of 79, suggesting that Tommy's Fluid Reasoning (Gf) is well below average. Tommy's evaluator was curious about what strategies he used to solve or attempt to solve the items on Figure Weights. She selected items that he got correct and incorrect and asked him how he solved them or attempted to solve them. Tommy's responses suggested that he did not use multiplication or attempt to solve any items mathematically. This information was consistent with the referral, past and current teacher reports, and math problem-solving worksheets. Nevertheless, the evaluator followed the recommendation in X-BASS to follow up on the lower score (Figure Weights), primarily because it was related to the main referral concerns. The UNIT2 Numerical Series subtest was administered to follow-up because it also measures quantitative reasoning but uses a different test format. Additionally, because Similarities was below average, presumably due to reasoning difficulties, the visual analog of this subtest was administered (i.e., Picture Concepts) to determine whether visual information (aided by verbal mediation) would lead to improved concept formation.

To analyze the four reasoning subtests scores, the evaluator checked the boxes next to each Gf subtest on the WISC-V tab, as shown in Figure 27.13, and then clicked on the "Transfer Scores to XBA Analyzer" button at the bottom of this tab (not shown in the figure). Once clicked, the program automatically advances to the XBA Analyzer tab where the Gf scores are displayed in the Gf section, as shown in Figure 27.14. Note that this tab analyzed the three Gf subtest scores from the WISC-V. Figure 27.14 shows that the Matrix Reasoning and Picture Concepts scores formed a cohesive Induction composite of 91 and the Figure Weights score was labeled as "divergent," meaning that it is substantially different from the other scores. Because the evaluator also gave the UNIT2 Numerical Series subtest, it was selected from the drop-down menu in the Gf section of the tab. After Tommy's score of 6 was entered, X-BASS displayed a four-subtest Gf XBA composite of 81 (see Figure 27.15). Psychometrically, 81 is the best estimate of Tommy's Gf. Clinically, the evaluator believes that this composite masks his strengths and weaknesses in the area of Gf; that is, Tommy's inductive reasoning is average, but his quantitative reasoning is well below average and consistent with referral concerns.

To further analyze Tommy's Gf scores, the evaluator clicked on the "Evaluate Score Configuration" button (which is seen in Figure 27.15). This button provides several options for score analysis. In the case of Tommy, the evaluator wanted to create two two-subtest composites: an Induction composite and a General Sequential/Quantitative Reasoning composite. When prompted by X-BASS, the evaluator chose that option. The results are found in Figure 27.16. This figure shows that Tommy's ability to



Fluid Reasoning Index (FRI/Gf)	☐	79	8	
Matrix Reasoning (I)	☐	8	25	☑
Figure Weights (RG,RQ)	☐	5	5	☑
Picture Concepts (I)	☐	9	37	☑
Arithmetic (Gsm:MW;Gq:A3)	☐			☐

CLINICAL JUDGMENT NEEDED

The FRI provides an estimate of Fluid Reasoning (Gf). Gf refers to a type of thinking that an individual may use when faced with a relatively new or novel task that cannot be performed automatically. Inductive reasoning as measuared by the Matrix Reasoning subtest was Average and general sequential (deductive) reasoning and quantitative reasoning as measured by the Figure Weights subtest was Well Below Average relative to same age peers. The difference between the scores that comprise the FRI is significant, however a difference of this size is considered common in the general population. This means that clinical judgment is necessary to determine whether the FRI is a good summary of Fluid Reasoning. The FRI of 79 (75-83) is classified as Well Below Average and is ranked at the 8th percentile, indicating performance as good as or better than 8% of same age peers from the general population. The difference between the FRI and the average of all five primary index scores is significant but also common in the general population. In this case, the FRI is Well Below Average relative to same age peers and is considered a personal weakness. Overall, despite significant variation within the FRI, the individual's Fluid Reasoning abilities likely constrain learning.

Gf = 79

Because one score that makes up the FRI is indicative of average or better performance and the other score is indicative of a deficit, follow up on the lower score is considered necessary to determine if it is an accurate and valid representation of ability and:

· If MR < FW and MR is suggestive of a weakness or deficit, consider obtaining more information about the individual's ability to reason inductively (e.g., Picture Concepts; subtest from another cognitive battery)

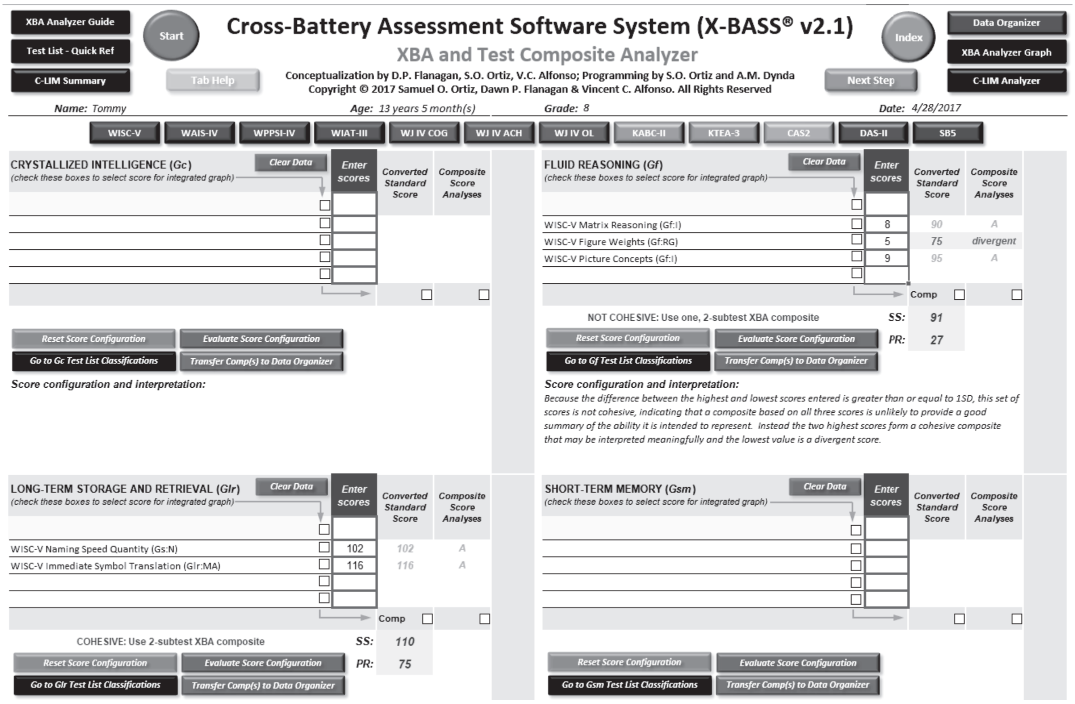
· If FW < MR and FW is suggestive of a weakness or deficit, consider a) obtaining more information about the individual's ability to reason deductively (e.g., subtest from another battery) and/or b) obtaining information about the individual's ability to reason quantitatively (e.g., Arithmetic; quantitative reasoning subtest from another battery; Applied Math Problems or Math Problem Solving subtests from an achievement battery)

· If AR is administered, determine whether QRI is cohesive

· Consider task characteristics and response demands

· If Picture Concepts and Arithmetic were administered, consider the Expanded Fluid Index (EFI)

FIGURE 27.13. WISC-V tab: FRI.



Cross-Battery Assessment Software System (X-BASS® v2.1)

XBA and Test Composite Analyzer

Conceptualization by D.P. Flanagan, S.O. Ortiz, V.C. Alfonso; Programming by S.O. Ortiz and A.M. Dynda
Copyright © 2017 Samuel O. Ortiz, Dawn P. Flanagan & Vincent C. Alfonso. All Rights Reserved

Name: Tommy **Age:** 13 years 5 month(s) **Grade:** 8 **Date:** 4/28/2017

CRYSTALLIZED INTELLIGENCE (*Gc*)
(check these boxes to select score for integrated graph)

	Enter scores	Converted Standard Score	Composite Score Analyses
☐			
☐			
☐			
☐			
☐			
		☐	☐

Score configuration and interpretation:

FLUID REASONING (*Gf*)
(check these boxes to select score for integrated graph)

	Enter scores	Converted Standard Score	Composite Score Analyses
☐			
WISC-V Matrix Reasoning (Gf:I) ☐	8	90	A
WISC-V Figure Weights (Gf:RG) ☐	5	75	divergent
WISC-V Picture Concepts (Gf:I) ☐	9	95	A
☐			
	Comp	☐	☐

NOT COHESIVE: Use one, 2-subtest XBA composite SS: 91 PR: 27

Score configuration and interpretation:
Because the difference between the highest and lowest scores entered is greater than or equal to 1SD, this set of scores is not cohesive, indicating that a composite based on all three scores is unlikely to provide a good summary of the ability it is intended to represent. Instead the two highest scores form a cohesive composite that may be interpreted meaningfully and the lowest value is a divergent score.

LONG-TERM STORAGE AND RETRIEVAL (*Glr*)
(check these boxes to select score for integrated graph)

	Enter scores	Converted Standard Score	Composite Score Analyses
☐			
WISC-V Naming Speed Quantity (Gs:N) ☐	102	102	A
WISC-V Immediate Symbol Translation (Glr:MA) ☐	116	116	A
☐			
☐			
		Comp ☐	☐

COHESIVE: Use 2-subtest XBA composite SS: 110 PR: 75

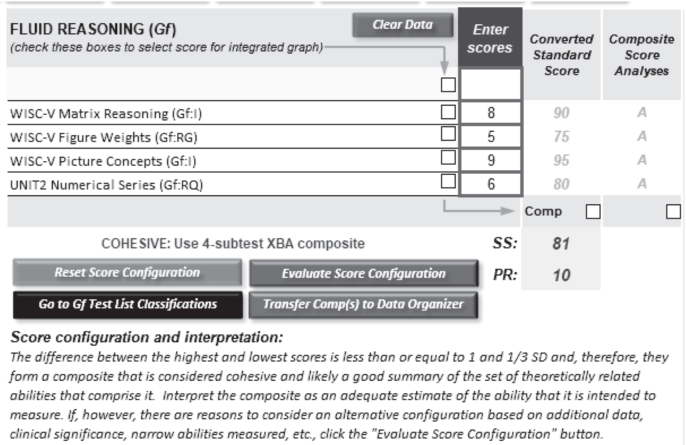
SHORT-TERM MEMORY (*Gsm*)
(check these boxes to select score for integrated graph)

	Enter scores	Converted Standard Score	Composite Score Analyses
☐			
☐			
☐			
☐			
☐			
		☐	☐

FIGURE 27.14. XBA Analyzer tab: Fluid Reasoning.

reason inductively with visual–spatial information is average (standard score = 91), while his ability to use quantitative reasoning to solve problems is well below average (standard score = 74).

Although the best estimate of Tommy's Gf is 81 based on the aggregate of the four Gf subtests, the examiner chose to use 91 and 74 (as depicted in Figure 27.16) to represent his ability to reason inductively with visual–spatial information and his ability to use reasoning processes that can be expressed mathematically, respectively, in the PSW Analysis. As such, the evaluator clicked on the "Transfer Comp(s) to Data Organizer" button in the Gf section on the XBA Analyzer tab (see Figure 27.16) to represent his strength and his weakness in Gf in the PSW analysis.



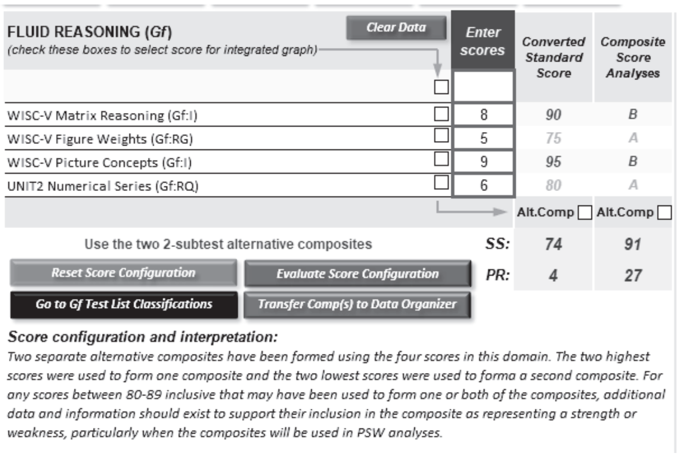
FLUID REASONING (*Gf*)
(check these boxes to select score for integrated graph)

	Enter scores	Converted Standard Score	Composite Score Analyses
☐			
WISC-V Matrix Reasoning (Gf:I) ☐	8	90	A
WISC-V Figure Weights (Gf:RG) ☐	5	75	A
WISC-V Picture Concepts (Gf:I) ☐	9	95	A
UNIT2 Numerical Series (Gf:RQ) ☐	6	80	A
		Comp ☐	☐

COHESIVE: Use 4-subtest XBA composite SS: 81 PR: 10

Score configuration and interpretation:
The difference between the highest and lowest scores is less than or equal to 1 and 1/3 SD and, therefore, they form a composite that is considered cohesive and likely a good summary of the set of theoretically related abilities that comprise it. Interpret the composite as an adequate estimate of the ability that it is intended to measure. If, however, there are reasons to consider an alternative configuration based on additional data, clinical significance, narrow abilities measured, etc., click the "Evaluate Score Configuration" button.

FIGURE 27.15. XBA Analyzer tab: Gf Cross-Battery Data.



FLUID REASONING (*Gf*) *(check these boxes to select score for integrated graph)*		Enter scores	*Converted Standard Score*	*Composite Score Analyses*
	☐			
WISC-V Matrix Reasoning (Gf:I)	☐	8	*90*	*B*
WISC-V Figure Weights (Gf:RG)	☐	5	*75*	*A*
WISC-V Picture Concepts (Gf:I)	☐	9	*95*	*B*
UNIT2 Numerical Series (Gf:RQ)	☐	6	*80*	*A*
			Alt.Comp ☐	Alt.Comp ☐
		SS:	*74*	*91*
		PR:	*4*	*27*

Score configuration and interpretation:

Two separate alternative composites have been formed using the four scores in this domain. The two highest scores were used to form one composite and the two lowest scores were used to forma a second composite. For any scores between 80-89 inclusive that may have been used to form one or both of the composites, additional data and information should exist to support their inclusion in the composite as representing a strength or weakness, particularly when the composites will be used in PSW analyses.

FIGURE 27.16. XBA Analyzer tab: New Score Configuration for Gf.

Visual Processing, Working Memory, and Processing Speed

Tommy's VSI of 81, WMI of 97, and PSI of 103 were cohesive and X-BASS indicated that there was no need to follow up in these areas, to which the examiner agreed. Each of these composites were transferred to the Data Organizer tab.

Long-Term Storage and Retrieval

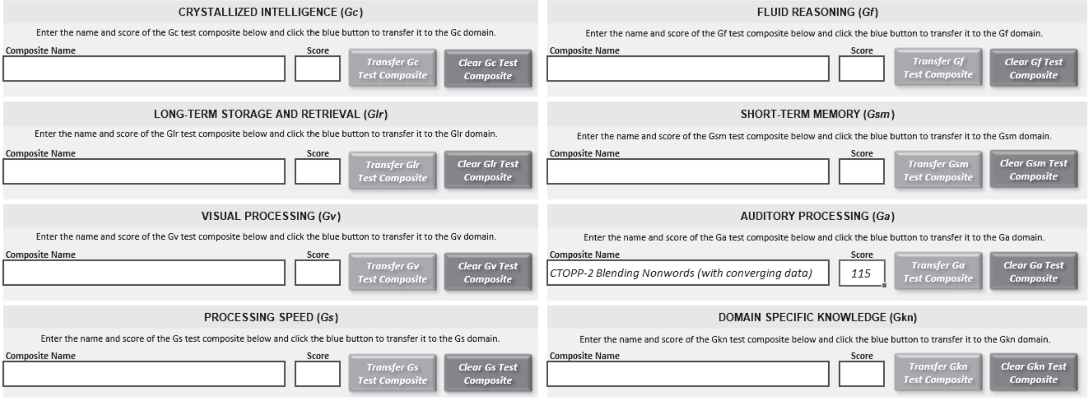
Tommy scored in the average range on Naming Speed Literacy (scaled score = 104), a measure of naming facility, and in the above average range on Immediate Symbol Translation (standard score = 116), a measure of associative memory. Based on initial testing, the evaluator did not believe there was a need for further evaluation in the area of Glr. To create a Glr composite, the evaluator checked the boxes next to each subtest and transferred them to the XBA Analyzer tab. As may be seen in Figure 27.14, a Glr composite of 110 was computed automatically. From this tab (i.e., XBA Analyzer tab), the examiner transferred this score to the Data Organizer tab.

Auditory Processing

Typically, when reading is an area of concern, measures of Phonetic Coding are administered. These measures include tests of phonological processing and phonological memory and are associated with the acquisition and development of basic reading skills. However, Tommy has no difficulty with reading decoding or reading fluency, suggesting that Phonetic Coding is not an area of concern. To test this hypothesis, Blending Nonwords from the CTOPP-2 was administered. Tommy earned an above average score on this subtest (scaled score = 13, which converts to a standard score of 115). This score is consistent with parent and teacher reports, his history of reading achievement, and the evaluator's observations during testing. Therefore, there is no need to conduct additional testing in this area.

XBA has historically required that all cognitive constructs be represented by at least two measures (i.e., subtests). However, there are times when this requirement may lead to unnecessary testing, such as in this situation. To limit the amount of time spent on test administration in the area of Ga, the examiner used the "Other Test Data Entry" tab, as shown in Figure 27.17. On this tab the examiner may enter a single score *only if it is at least average and there are other compelling data sources to support at least average performance,* such as in the case of Tommy's Blending Nonwords score. As seen in Figure 27.17, the evaluator named the score "CTOPP-2 Blending Nonwords (with converging data)" and entered the converted score of 115. From this tab, the Ga score may be transferred to the Data Organizer tab.

In sum, after the assessment of cognitive abilities and processes, including follow-up assessment to explore aberrant score performances and to test specific hypotheses, nine estimates were transferred to the Data Organizer tab, covering the main seven cognitive areas that are necessary to conduct the PSW analysis. Most of these data were displayed earlier in Figure 27.7. The only other information necessary to conduct the PSW analysis



CRYSTALLIZED INTELLIGENCE (*Gc*)

Enter the name and score of the Gc test composite below and click the blue button to transfer it to the Gc domain.

Composite Name | Score

FLUID REASONING (*Gf*)

Enter the name and score of the Gf test composite below and click the blue button to transfer it to the Gf domain.

Composite Name | Score

LONG-TERM STORAGE AND RETRIEVAL (*Glr*)

Enter the name and score of the Glr test composite below and click the blue button to transfer it to the Glr domain.

Composite Name | Score

SHORT-TERM MEMORY (*Gsm*)

Enter the name and score of the Gsm test composite below and click the blue button to transfer it to the Gsm domain.

Composite Name | Score

VISUAL PROCESSING (*Gv*)

Enter the name and score of the Gv test composite below and click the blue button to transfer it to the Gv domain.

Composite Name | Score

AUDITORY PROCESSING (*Ga*)

Enter the name and score of the Ga test composite below and click the blue button to transfer it to the Ga domain.

Composite Name: *CTOPP-2 Blending Nonwords (with converging data)*

Score: 115

PROCESSING SPEED (*Gs*)

Enter the name and score of the Gs test composite below and click the blue button to transfer it to the Gs domain.

Composite Name | Score

DOMAIN SPECIFIC KNOWLEDGE (Gkn)

Enter the name and score of the Gkn test composite below and click the blue button to transfer it to the Gkn domain.

Composite Name | Score

FIGURE 27.17. Data Entry—Other tab: Auditory Processing.

is at least one academic achievement score, although most practitioners gather data across many academic areas. Tommy's achievement data are discussed next.

Achievement Data

Selected subtests from the KTEA-3 were administered, including Reading Comprehension (standard score = 82; below average), Math Computation (standard score = 98; average), Math Concepts and Applications (standard score = 76; well below average), and Written Expression (standard score = 93; average). The KTEA-3 error analysis for Reading Comprehension showed that Tommy had difficulty on items that required drawing inferences from text. As such, the CELF-5 Reading Comprehension subtest was also administered to determine if Tommy had difficulty interpreting inferential information. On this subtest he earned a scaled score of 6 (well below average) and demonstrated much difficulty with inferential information.

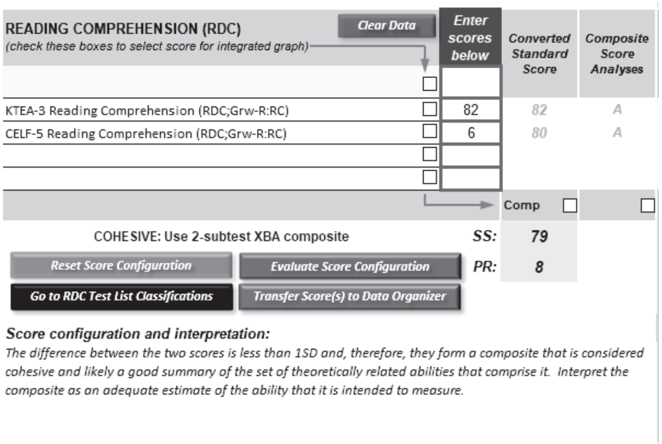
To gather more data in math skills that are not measured (or measured in-depth) on the KTEA-3, the KeyMath3 Applied Problem Solving (APS) subtest was administered. Tommy obtained a scaled score of 7 on this subtest. He appeared to have the most difficulty with drawing inferences and conclusions to solve problems and using charts, tables, graphs, equations, and geometric representations to solve problems, which is consistent with his quantitative reasoning and visual processing weaknesses, respectively.

After the assessment of the academic skills most closely associated with the referral, two subtests were transferred to the Data Organizer tab from the KTEA-3 tab, namely Math Computation and Written Expression. Since the evaluator did more in-depth assessment of Reading Comprehension and Math Problem Solving, the KTEA-3 Reading Comprehension and Math Concepts & Applications subtest scores were transferred to the XBA Analyzer tab from the KTEA-3 tab. On the XBA Analyzer tab the evaluator used the drop-down menu in the Reading Comprehension section to select the CELF-5 Reading Comprehension subtest and then entered the scaled score of 6. X-BASS automatically calculated an XBA Reading Comprehension composite of 79 (see Figure 27.18), which the evaluator transferred to the Data Organizer tab.

In the Math Problem Solving section of the XBA Analyzer tab, the evaluator selected the KeyMath3 APS subtest from the drop-down menu and entered a scaled score of 7. X-BASS automatically calculated an XBA Math Problem Solving composite of 79, which the evaluator transferred to the Data Organizer tab (see Figure 27.19). In all, four achievement scores were transferred to the Data Organizer tab. Now that the evaluation is complete, all score have been entered, XBA composites have been calculated, and all estimates have been transferred to the Data Organizer tab, the evaluator is ready to begin the PSW analysis.

PSW Analysis

On the Data Organizer tab, the evaluator clicked on the "Select All Checkboxes" button to ensure that all cognitive and achievement estimates are used in Tommy's PSW analysis. Next, the evaluator marked each estimate as either a strength or a weakness on the Strengths and Weaknesses Indicator tab. Then X-BASS automatically calculated



READING COMPREHENSION (RDC)
(check these boxes to select score for integrated graph)

		Enter scores below	Converted Standard Score	Composite Score Analyses
	☐			
KTEA-3 Reading Comprehension (RDC;Grw-R:RC)	☐	82	82	A
CELF-5 Reading Comprehension (RDC;Grw-R:RC)	☐	6	80	A
	☐			
	☐			
		Comp ☐		☐

COHESIVE: Use 2-subtest XBA composite

SS: 79

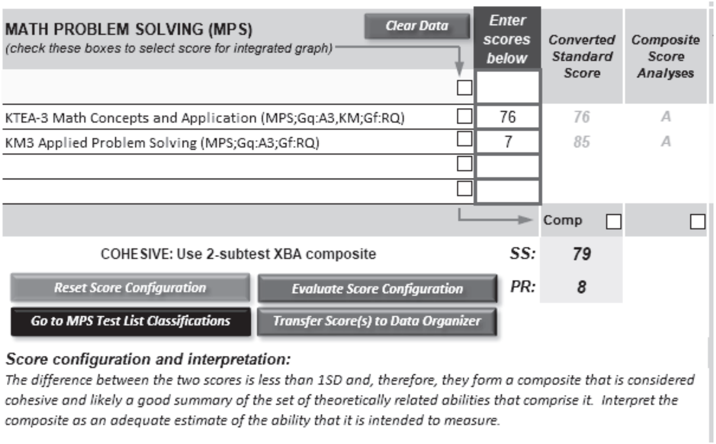
PR: 8

Score configuration and interpretation:
The difference between the two scores is less than 1SD and, therefore, they form a composite that is considered cohesive and likely a good summary of the set of theoretically related abilities that comprise it. Interpret the composite as an adequate estimate of the ability that it is intended to measure.

FIGURE 27.18. XBA Analyzer tab: Reading Comprehension.

the g-Value, FCC, and ICC, as was shown earlier in Figure 27.9. Note that Tommy's WISC-V FSIQ is 87 and his FCC is 104. While both composites provide an estimate of overall cognitive ability, unlike the FSIQ, the FCC is not attenuated by Tommy's processing deficits. Note also that both estimates should be reported by the evaluator and the differences between them should be explained with respect to the composition of each composite. However, the FCC is the value used in the PSW analysis, unless the evaluator purposefully enters an alternative composite to replace the FCC.

Next, the evaluator advanced to the PSW-A tab, which displays the results of a PSW analysis using Tommy's data. Referring back to Figure 27.11, the top oval shows that Tommy's ability to think and reason is at least average (FCC = 104), despite specific cognitive processing weaknesses. The drop-down menu in the top oval lists academic areas that the evaluator indicated were strengths for Tommy. Individuals with SLDs often perform at or close to grade level (or at least average) in one or more academic skills and the top oval reflects this finding. The bottom left oval displays cognitive processing weaknesses. As may be seen in Figure 27.11, the program defaults to the ICC as the cognitive weakness in the bottom left oval. However, the oval contains a drop-down menu, where each area of cognitive processing weakness is listed and therefore any processing weakness may be selected in lieu of the ICC. The bottom right oval represents the academic weakness. This oval also contains a drop-down menu that lists each academic area that the evaluator marked as a weakness. The results in Figure 27.11 show that all criteria were met, indicating that Tommy displays a pattern of



MATH PROBLEM SOLVING (MPS)
(check these boxes to select score for integrated graph)

		Enter scores below	Converted Standard Score	Composite Score Analyses
	☐			
KTEA-3 Math Concepts and Application (MPS;Gq:A3,KM;Gf:RQ)	☐	76	76	A
KM3 Applied Problem Solving (MPS;Gq:A3;Gf:RQ)	☐	7	85	A
	☐			
	☐			
		Comp ☐		☐

COHESIVE: Use 2-subtest XBA composite

SS: 79

PR: 8

Score configuration and interpretation:
The difference between the two scores is less than 1SD and, therefore, they form a composite that is considered cohesive and likely a good summary of the set of theoretically related abilities that comprise it. Interpret the composite as an adequate estimate of the ability that it is intended to measure.

FIGURE 27.19. XBA Analyzer tab: Math Problem Solving.

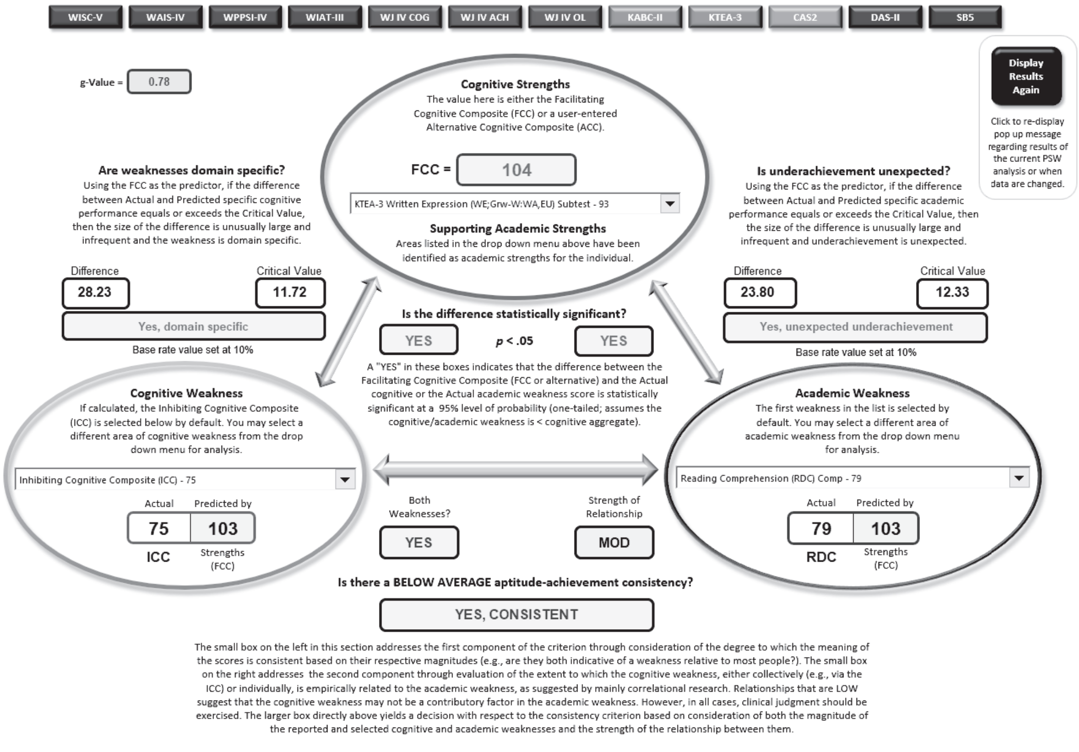
strengths and weaknesses that is consistent with an SLD in math problem solving.

Because reading comprehension was also an area of concern, the evaluator selected the Reading Comprehension composite from the drop-down menu in the bottom right oval and continued to use the ICC in the bottom left oval to represent Tommy's cognitive processing weaknesses (see Figure 27.20). Although X-BASS conducts the PSW analysis automatically each time a different cognitive or academic weakness is displayed in the bottom left or right oval, respectively, it is important to understand how the program conducts the analysis.

First, the program determines whether there is a statistically significant difference between the FCC and the cognitive weakness and the FCC and the academic weakness and reports either "yes" or "no" in the center of the three ovals. As seen in Figure 27.20, "yes" is displayed for both comparisons, meaning that these are reliable differences and are not likely due to chance. However, some statistically significant differences may be common in the general population. Therefore, base rate data are used to determine whether the differences are unusual.

Second, the program uses the FCC in a regression analysis to predict where the individual (in this case Tommy) was expected to perform in the selected areas of cognitive and academic weakness. The bottom ovals contain the actual scores alongside scores that were predicted based on the FCC. For example, the bottom right oval in Figure 27.20 shows that Tommy's "actual" reading comprehension score is 79 and his "predicted" reading comprehension score, based on the results of the regression analysis, is 103. Next, the difference between the actual and predicted scores are compared to a critical value to determine if that difference is unusual in the general population, meaning that the difference occurs in about 10% or less in the population. When the difference between the actual and predicted cognitive scores is unusual, the criterion of a "domain-specific" cognitive weakness is met. When the difference between the actual and predicted achievement scores is unusual, the criterion of "unexpected underachievement" is met. Figure 27.20 shows that both criteria were met.

Third, the program determines whether the pattern of strengths and weaknesses is marked by a "below average aptitude–achievement consistency." Consistency is determined by evaluating whether the cognitive and academic areas identified as weaknesses by the evaluator are actually weaknesses



g-Value = 0.78

Cognitive Strengths
The value here is either the Facilitating Cognitive Composite (FCC) or a user-entered Alternative Cognitive Composite (ACC).

FCC = 104

KTEA-3 Written Expression (WE;Grw-W:WA,EU) Subtest - 93

Supporting Academic Strengths
Areas listed in the drop down menu above have been identified as academic strengths for the individual.

Click to re-display pop up message regarding results of the current PSW analysis or when data are changed.

Are weaknesses domain specific?
Using the FCC as the predictor, if the difference between Actual and Predicted specific cognitive performance equals or exceeds the Critical Value, then the size of the difference is unusually large and infrequent and the weakness is domain specific.

Difference 28.23
Critical Value 11.72
Yes, domain specific
Base rate value set at 10%

Is underachievement unexpected?
Using the FCC as the predictor, if the difference between Actual and Predicted specific academic performance equals or exceeds the Critical Value, then the size of the difference is unusually large and infrequent and underachievement is unexpected.

Difference 23.80
Critical Value 12.33
Yes, unexpected underachievement
Base rate value set at 10%

Is the difference statistically significant?
YES $p < .05$ YES
A "YES" in these boxes indicates that the difference between the Facilitating Cognitive Composite (FCC or alternative) and the Actual cognitive or the Actual academic weakness score is statistically significant at a 95% level of probability (one-tailed; assumes the cognitive/academic weakness is < cognitive aggregate).

Cognitive Weakness
If calculated, the Inhibiting Cognitive Composite (ICC) is selected below by default. You may select a different area of cognitive weakness from the drop down menu for analysis.

Inhibiting Cognitive Composite (ICC) - 75

Actual	Predicted by
75	103
ICC	Strengths (FCC)

Academic Weakness
The first weakness in the list is selected by default. You may select a different area of academic weakness from the drop down menu for analysis.

Reading Comprehension (RDC) Comp - 79

Actual	Predicted by
79	103
RDC	Strengths (FCC)

Both Weaknesses? YES
Strength of Relationship MOD

Is there a BELOW AVERAGE aptitude-achievement consistency?
YES, CONSISTENT

The small box on the left in this section addresses the first component of the criterion through consideration of the degree to which the meaning of the scores is consistent based on their respective magnitudes (e.g., are they both indicative of a weakness relative to most people?). The small box on the right addresses the second component through evaluation of the extent to which the cognitive weakness, either collectively (e.g., via the ICC) or individually, is empirically related to the academic weakness, as suggested by mainly correlational research. Relationships that are LOW suggest that the cognitive weakness may not be a contributory factor in the academic weakness. However, in all cases, clinical judgment should be exercised. The larger box directly above yields a decision with respect to the consistency criterion based on consideration of both the magnitude of the reported and selected cognitive and academic weaknesses and the strength of the relationship between them.

FIGURE 27.20. PSW Analysis tab: Reading Comprehension.

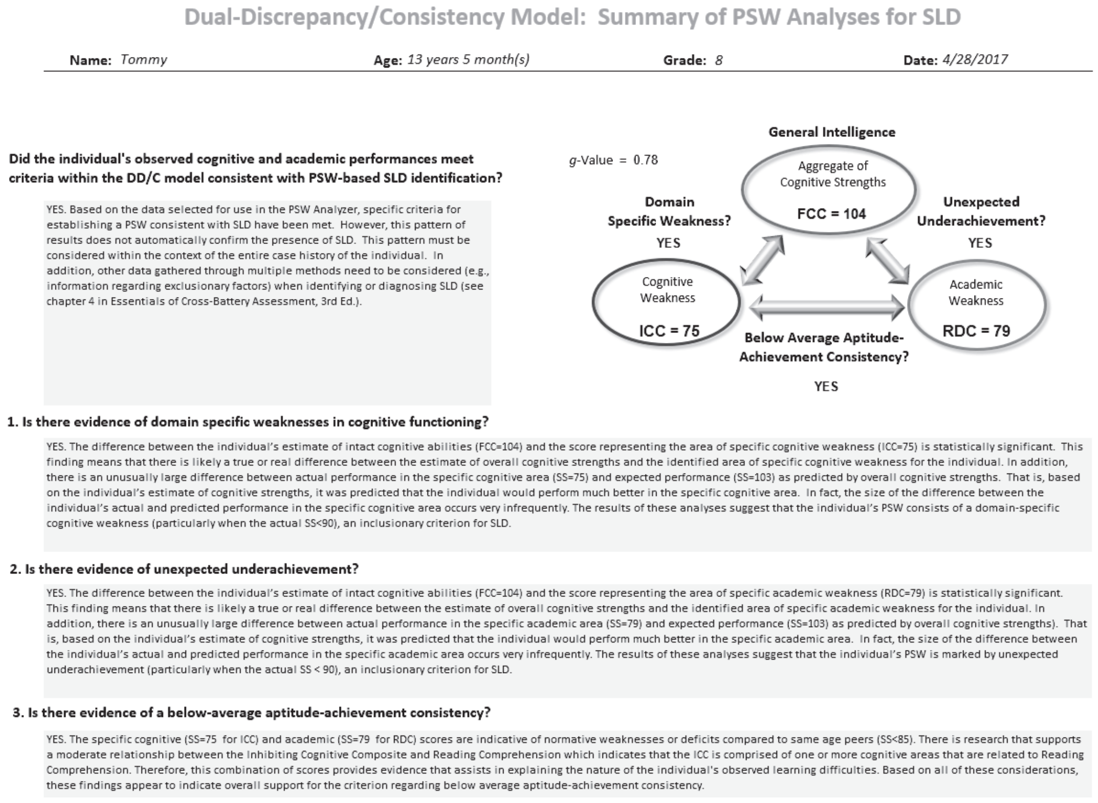
relative to the general population. For example, an evaluator may indicate that a score of 98 is a weakness for a very high-functioning individual. However, a score of 98 is not a weakness relative to most people. Figure 27.20 shows that Tommy's actual cognitive and achievement scores are weaknesses relative to most people. Specifically, the program reported "YES" to the question "Both Weaknesses?" Another component of consistency is an evaluation of the strength of the relationship between the cognitive and academic areas of weakness. The program shows that the combination of cognitive areas that make up the ICC has a moderate relationship to reading comprehension (displayed as "MOD" in Figure 27.20). Moreover, the combination of actual weaknesses and a moderate relationship indicates consistency, which is reported as "YES, CONSISTENT" in the PSW analysis (see Figure 27.20). The results in Figure 27.20 show that all criteria were met, indicating that Tommy displays a pattern of strengths and weaknesses that is consistent with an SLD in reading comprehension.

To assist the evaluator in interpreting the PSW analysis, X-BASS provides answers to the following questions and customizes them according to the individual's scores.

1. Did the individual's observed cognitive and academic performances meet criteria within the DD/C model consistent with PSW-based SLD identification?
2. Is there evidence of domain-specific weaknesses in cognitive functioning?
3. Is there evidence of unexpected underachievement?
4. Is there evidence of a below-average aptitude–achievement consistency?

A customized summary with answers to these questions may be printed and included as an appendix to a psychoeducational report. Figure 27.21 shows responses to these questions for Tommy.

As the case of Tommy reflects, the XBA approach has evolved significantly over the past 20 years. It is more sophisticated theoretically and psychometrically. It is now part of a software system (X-BASS) that provides a means of facilitating the approach and using XBA data to answer



Dual-Discrepancy/Consistency Model: Summary of PSW Analyses for SLD

Name: *Tommy* **Age:** *13 years 5 month(s)* **Grade:** *8* **Date:** *4/28/2017*

Did the individual's observed cognitive and academic performances meet criteria within the DD/C model consistent with PSW-based SLD identification?

YES. Based on the data selected for use in the PSW Analyzer, specific criteria for establishing a PSW consistent with SLD have been met. However, this pattern of results does not automatically confirm the presence of SLD. This pattern must be considered within the context of the entire case history of the individual. In addition, other data gathered through multiple methods need to be considered (e.g., information regarding exclusionary factors) when identifying or diagnosing SLD (see chapter 4 in Essentials of Cross-Battery Assessment, 3rd Ed.).

1. Is there evidence of domain specific weaknesses in cognitive functioning?

YES. The difference between the individual's estimate of intact cognitive abilities (FCC=104) and the score representing the area of specific cognitive weakness (ICC=75) is statistically significant. This finding means that there is likely a true or real difference between the estimate of overall cognitive strengths and the identified area of specific cognitive weakness for the individual. In addition, there is an unusually large difference between actual performance in the specific cognitive area (SS=75) and expected performance (SS=103) as predicted by overall cognitive strengths. That is, based on the individual's estimate of cognitive strengths, it was predicted that the individual would perform much better in the specific cognitive area. In fact, the size of the difference between the individual's actual and predicted performance in the specific cognitive area occurs very infrequently. The results of these analyses suggest that the individual's PSW consists of a domain-specific cognitive weakness (particularly when the actual SS<90), an inclusionary criterion for SLD.

2. Is there evidence of unexpected underachievement?

YES. The difference between the individual's estimate of intact cognitive abilities (FCC=104) and the score representing the area of specific academic weakness (RDC=79) is statistically significant. This finding means that there is likely a true or real difference between the estimate of overall cognitive strengths and the identified area of specific academic weakness for the individual. In addition, there is an unusually large difference between actual performance in the specific academic area (SS=79) and expected performance (SS=103) as predicted by overall cognitive strengths). That is, based on the individual's estimate of cognitive strengths, it was predicted that the individual would perform much better in the specific academic area. In fact, the size of the difference between the individual's actual and predicted performance in the specific academic area occurs very infrequently. The results of these analyses suggest that the individual's PSW is marked by unexpected underachievement (particularly when the actual SS < 90), an inclusionary criterion for SLD.

3. Is there evidence of a below-average aptitude-achievement consistency?

YES. The specific cognitive (SS=75 for ICC) and academic (SS=79 for RDC) scores are indicative of normative weaknesses or deficits compared to same age peers (SS<85). There is research that supports a moderate relationship between the Inhibiting Cognitive Composite and Reading Comprehension which indicates that the ICC is comprised of one or more cognitive areas that are related to Reading Comprehension. Therefore, this combination of scores provides evidence that assists in explaining the nature of the individual's observed learning difficulties. Based on all of these considerations, these findings appear to indicate overall support for the criterion regarding below average aptitude-achievement consistency.

FIGURE 27.21. Summary of PSW Analysis.

questions related to identification of SLD and whether XBA data gathered from cognitive tests yield scores that are likely consistent with an individual's cultural and linguistic background (for details, see Ortiz et al., Chapter 25, this volume). Table 27.7 highlights the many ways in which the XBA approach has evolved and contributed to the practice of conducting reliable, valid, and defensible psychoeducational evaluations.

SUMMARY

In this chapter, we have presented the XBA approach as a method that allows practitioners to augment or supplement any major ability test (e.g., cognitive, neuropsychological, academic, speech–language) to ensure measurement of a wider range of broad and narrow cognitive abilities in a manner that is consistent with contemporary theory and research, and that is predicated upon sound psychometric principles. The foundational sources of information upon which the XBA approach was formulated (i.e., CHC theory and the classifications of ability tests according to this theory), coupled with straightforward, step-by-step procedures, provide a systematic way to conduct a theoretically driven, comprehensive, and valid assessment of a wide range of cognitive abilities and processes. When the XBA approach is applied to the WISC-V, for example, it is possible to measure important abilities that would otherwise not be assessed (e.g., Ga) and to follow up on aberrant score performances and test hypotheses about unexpected variation within cognitive domains.

The XBA approach allows for the measurement of the major cognitive areas specified in CHC theory, with emphasis on those considered most critical on the basis of history, observation, response to intervention, and other available sources of data. The CHC classifications of a multitude of ability tests bring stronger content and construct validity evidence to the evaluation and interpretation process. With a strong research base, the XBA approach can aid practitioners not only in the comprehensive measurement of cognitive abilities and neuropsychological processes, but in the selective measurement of abilities and processes that are deemed to be most important with respect to suspected learning disabilities. Adherence to the guiding principles and steps of the XBA approach, as well as careful attention to specific referral concerns, results in the creation of highly individualized assessment batteries that are ideally suited for the intended purpose of assessment.

Through a case example, this chapter demonstrated the XBA approach and how to use X-BASS to facilitate the approach. Overall, this chapter showed (1) the value of crossing batteries to address referral concerns and inform decisions related to SLD identification; (2) the importance of understanding how identified areas of cognitive weakness manifest in real-world activities (e.g., in the classroom) and how this information informs intervention; and (3) how to use XBA data in a PSW analysis to inform SLD identification using X-BASS.

NOTE

1. CHC theory has been revised and updated in Chapter 3 of this book. Because the information contained in Chapter 3 is new, and because that chapter was written at the same time as this chapter, the classification system underlying the XBA approach as described here does not yet reflect some current refinements to the theory.

REFERENCES

Alfonso, V. C., & Flanagan, D. P. (2006). Best practices in the use of the Stanford–Binet Intelligence Scales, Fifth Edition (SB5) with Preschoolers. In B. A. Bracken & R. Nagle (Eds.), *Psychoeducational assessment of preschoolers* (4th ed., pp. 267– 295). Mahwah, NJ: Erlbaum.

Alfonso, V. C., & Flanagan, D. P. (2018). *Essentials of specific learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185–202). New York: Guilford Press.

The American Heritage Dictionary (5th ed.). (2018). New York: Dell.

Brackett, J., & McPherson, A. (1996). Learning disabilities diagnosis in postsecondary students: A comparison of discrepancy-based diagnostic models. In N. Gregg, C. Hoy, & A. F. Gay (Eds.), *Adults with learning disabilities: Theoretical and practical perspectives* (pp. 68–84). New York: Guilford Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford Press.

Carroll, J. B. (1998). Foreword. In K. S. McGrew & D. P. Flanagan, *The intelligence test desk reference: Gf-Gc cross-battery assessment* (pp. xi–xii). Boston: Allyn & Bacon.

TABLE 27.7. Past and Present Contributions of the XBA Approach to Psychoeducational Evaluation

Source	Contribution
Flanagan, Genshaft, and Harrison (1997)	• First attempt at merging the Cattell–Horn Gf–Gc theory and Carroll's three-stratum theory (McGrew, 1997), which represented the foundation of cross-battery assessment (XBA). • First expert consensus study regarding the narrow abilities measured by intelligence tests (McGrew, 1997), an important component of XBA. • Introduction to the need for XBA and the assumptions and foundations as well as an operationalized set of XPA principles and procedures (Flanagan & McGrew, 1997).
McGrew and Flanagan (1998)	• Introduced a step-by-step approach to XBA in an attempt to improve upon the measurement of cognitive constructs. • Demonstrated how the XBA approach guarded against two ubiquitous sources of invalidity in assessment: construct-irrelevant variance and construct underrepresentation. • Provided worksheets for organizing assessments according to contemporary Gf–Gc theory and for conducting XBA. • Provided a review of the research on the relations between broad and narrow Gf–Gc abilities and academic (reading and math) and occupational outcomes. • Provided a desk reference of all the major intelligence tests, which provided important information for each subtest as a means of informing interpretation of XBA data (e.g., reliability, validity, standardization sample characteristics, test floors and ceilings, item gradients, variables influencing subtest performance, *g* loadings, broad and narrow abilities measured by subtests). • Provided the first comprehensive set of theory-based classifications of tests, in an attempt to further establish a Gf–Gc nomenclature for the field. • Highlighted the importance of joint or cross-battery confirmatory factor-analytic studies for understanding the Gf–Gc broad abilities underlying intelligence tests. • Provided the first set of systematic classifications of ability tests according to degree of cultural loading and degree of linguistic demand.
Flanagan, McGrew, and Ortiz (2000)	• Introduced the *integrated Cattell–Horn and Carroll Gf–Gc model* as the foundation for cross-battery assessment, based on analyses conducted by McGrew (e.g., McGrew, 1997). This integrated model was renamed *Cattell–Horn–Carroll (CHC) theory* shortly thereafter (see McGrew, 2005, for details). • Applied Gf–Gc theory to interpretation of the Wechsler scales. • Demonstrated that the Wechsler scales included redundancy in the assessment of certain constructs (e.g., Gc and Gv) and omitted measurement of other important constructs (e.g., Gf, Ga, and Glr). • Offered step-by-step XBA guidelines for augmenting a Wechsler scale, so that a broader range of cognitive abilities could be measured as deemed relevant and necessary vis à vis referral concerns. • Provided a set of worksheets for conducting XBA with the Wechsler scales.
Flanagan and Ortiz (2001)	• Used CHC theory as the foundation for XBA. • Expanded test classifications to include a variety of special-purpose tests, in addition to the major intelligence tests. • Included more comprehensive coverage of test interpretation. • Provided updated and improved XBA worksheets. • Expert consensus studies provided the basis for narrow-ability classifications of cognitive tests. • Refined classifications of ability tests according to degree of cultural loading and degree of linguistic demand.
Flanagan, Ortiz, Alfonso, and Mascolo (2002)	• Extended the XBA approach to achievement tests. • Included the largest expert consensus study of the narrow abilities underlying ability tests. • Provided an updated review of the literature on the relations between cognitive abilities and reading and math achievement. Review was expanded to include the area of written language. • Demonstrated how to use the XBA approach within the context of a CHC-based operational definition of SLD. • Provided a desk reference of achievement tests, which provided important information for each subtest (e.g., reliability, validity, standardization sample characteristics, test floors and ceilings, broad and narrow abilities measured by each subtest). • Included tables of the qualitative characteristics of individual achievement subtests from 48 batteries—information that informed test selection for XBA as well as interpretation.

(continued)

TABLE 27.7. *(continued)*

Source	Contribution
Flanagan and Kaufman (2004, 2009)	• Provided a CHC interpretive framework for the WISC-IV, thereby facilitating the use of this instrument in the XBA approach. • Included actual norms for seven CHC-based clinical clusters, including narrow-ability clusters that were incorporated into the XBA approach. • Automated the CHC-interpretation method for the WISC-IV (program included on CD that accompanied the book).
Flanagan and Harrison (2005)	• Detailed origins of the XBA approach and the theoretical and research foundation upon which it was based (McGrew, 1997, 2005). • Detailed the manner in which CHC theory and the XBA approach influenced test development (Alfonso, Flanagan, & Radwan, 2005). • Highlighted the XBA approach as an example of the current “wave” of intelligence test interpretation: application of theory (Kamphaus, Winsor, Rowe, & Kim, 2005).
Flanagan, Ortiz, Alfonso, and Mascolo (2006)	• Included variation in task characteristics of the subtests of over 50 achievement batteries—information that informed test selection for XBA as well as interpretation. • Updated CHC-based classifications of achievement tests. • Provided a desk reference of achievement tests, which provided important information for each subtest (e.g., reliability, validity, standardization sample characteristics, test floors and ceilings, broad and narrow abilities measured by each subtest). • Revised and refined the operational definition of SLD, and demonstrated how to use the XBA approach within the context of this definition. • Introduced *academic clinical clusters* according to the eight areas of SLD listed in IDEA 2004.
Flanagan, Ortiz, and Alfonso (2007)	• Introduced automated XBA worksheets in a program called the XBA Data Management and Interpretive Assistant (DMIA). • Introduced an automated Culture–Language Interpretive Matrix (C-LIM) program to evaluate whether test performance systematically declines as a function of increased culture and language demands for English language learners. • Introduced an automated program called the SLD Assistant. This program was intended to assist in determining whether an individual was of at least average overall intellectual ability, despite cognitive deficits in one or more specific areas. • Emphasized use of core tests (and supplemental tests as necessary) from a single battery, rather than selected components of a battery, as part of the assessment, because (1) current intelligence tests have better representation of the broad CHC abilities and use only two or three subtests to represent them; and (2) the broad abilities measured by current intelligence batteries are typically represented by qualitatively different indicators that are relevant only to the broad ability intended to be measured. • Greater emphasis placed on use of actual norms, rather than averages. Averages to be obtained under a selected few circumstances (e.g., narrow-ability level). • Expanded coverage of CHC theory to include abilities typically measured on achievement tests (e.g., Grw, Gq, Ga), providing additional information useful in the identification of SLD. • Addressed the “disorder in a basic psychological process” language of IDEA (2004). • Demonstrated how the XBA approach might be used to operationalize the “pattern of strengths and weaknesses” language of the federal regulations (U.S. Department of Education, 2005).
Flanagan, Alfonso, Ortiz, and Dynda (2010)	• Extended CHC classifications to neuropsychological instruments, thus expanding the range of instruments that might be used in the XBA approach. • Applied neuropsychological domain classifications to cognitive tests; this was intended to expand the interpretive options for XBA data. • Applied XBA principles to neuropsychological evaluation.
Flanagan and Harrison (2012)	• Expanded CHC theory to include 16 broad abilities and over 80 narrow abilities (Schneider & McGrew, Chapter 4). • Emphasized the relevance of the XBA approach for augmenting stand-alone batteries (e.g., McCallum & Bracken, Chapter 14).
Flanagan, Ortiz, and Alfonso (2013)	• Expanded coverage of CHC theory to include abilities not measured by most major intelligence and cognitive batteries (e.g., Gh, tactile abilities; Gk, kinesthetic abilities). • Incorporated and integrated all then-current intelligence batteries (i.e., WJ III, WPPSI-III, WISC-IV, SB5, KABC-II, DAS-II, and WAIS-IV), tests of academic achievement, and selected neuropsychological instruments.

(continued)

TABLE 27.7. *(continued)*

Source	Contribution
Flanagan, Ortiz, and Alfonso (2013) *(continued)*	• Provided a stronger emphasis on using actual norms when available. • Included more stringent guidelines for averaging subtest scores from the same or different batteries under specific circumstances. • Summarized current research on the relations between cognitive abilities and processes and academic skills, and placed even greater emphasis on forming CHC narrow-ability clusters, given their importance in understanding academic outcomes. • Revised the DMIA to incorporate and integrate all features of the XBA approach and to include interpretive statements. It also included tabs for all current intelligence batteries, major achievement tests (e.g., WJ III ACH), and co-normed (e.g., KABC-II and KTEA-II) or linked (WISC-IV and WIAT-III) batteries. Additionally, the DMIA now used a variety of criteria to determine cohesiveness of within-battery clusters. • Revised the SLD Assistant to include a psychometrically defensible analysis of an individual's pattern of strengths and weaknesses. • Revised and updated the C-LIM to include current cognitive tests, special-purpose tests, and selected neuropsychological instruments. The C-LIM now provided additional features for evaluating individuals based on varying levels of language proficiency, acculturative knowledge, and/or giftedness. The C-LIM also allowed for an examination of cognitive performance by the influences of language or culture independently. • Classified current cognitive batteries according to neuropsychological domains of functioning (e.g., sensory–motor, visual–spatial, speed and efficiency, executive). • Included examples of how the XBA approach could be used within the context of various state and district criteria for SLD identification. • Included guidelines for linking findings of cognitive weaknesses or deficits to intervention.
Cross-Battery Assessment Software System (X-BASS; Flanagan, Ortiz, & Alfonso, 2015, 2017)	• Integrated the three software programs from *Essentials of Cross-Battery Assessment*, third edition; improved upon the psychometrics of each program; and provided a means for allowing these programs to work together with the same set of data. • Used over 1,500 coefficients from the technical manuals of ability tests to calculate median reliabilities and intercorrelations that are used in formulas and analyses in the program. • Integrated the WISC-V interpreted system outlined in *Essentials of WISC-V Assessment* (Flanagan & Alfonso, 2017) and allowed for a customized summary report of findings for each examinee. • Incorporated the most psychometrically sophisticated pattern of strengths and weaknesses analysis currently available for assisting in making decisions about SLD. • Allowed for the software to be used with single-battery or cross-battery data. For example, if a single battery provides sufficient data for the PSW analysis, then the analysis will be conducted automatically. In most comprehensive evaluations for SLD, however, cross-battery data are gathered and therefore used in the PSW analysis in the software. • Includes a PSW interpretive summary report.
Flanagan & McDonough (2018)	• Includes the latest revisions and refinements to CHC theory (Schneider & McGrew, Chapter 3, this volume). • Emphasizes the relevance of the XBA approach for augmenting stand-alone batteries (e.g., McCallum & Bracken, Chapter 16, this volume). • Emphasizes the use of XBA data for a PSW analysis following the Dual Discrepancy/Consistency model (Flanagan et al., Chapter 22, this volume). • Demonstrates how to use X-BASS to facilitate use and interpretation of the XBA approach (Flanagan et al., Chapter 27, this volume).

Carroll, J. B. (2012). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 883–890). New York: Guilford Press.

Cheramie, G. M., Schanding, G. T., & Steich, K. (2018). Core selective evaluation process (C-SEP) and dual discrepancy/consistency (DD/C) model for SLD identification: A case study approach. In V. C. Alfonso & D. P. Flanagan (Eds.), *Essentials of specific learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Dean, R. S., & Woodcock, R. W. (2003). *Dean–Woodcock Neuropsychological Battery.* Itasca, IL: Riverside.

Decker, S. L. (2008). School neuropsychology consultation in neurodevelopmental disorders. *Psychology in the Schools, 45,* 799–811.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System.* San Antonio, TX: Psychological Corporation.

Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition*. San Antonio, TX: Harcourt Assessment.

Feifer, S. G., & DeFina, P. D. (2005). *The neuropsychology of mathematics: Diagnosis and intervention*. Middletown, MD: School Neuropsych Press.

Feifer, S. G., & Della Toffalo, D. A. (2007). *Integrating RTI with cognitive neuropsychology: A scientific approach to reading.* Middletown, MD: School Neuropsych Press.

Fiorello, C. A., Hale, J. B., Snyder, L. E., Forrest, E., & Teodori, A. (2008). Validating individual differences through examination of converging psychometric and neuropsychological models of cognitive functioning. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K–3 classrooms* (pp. 151–184). New York: Routledge.

Flanagan, D. P., & Alfonso, V. C. (Eds.). (2011). *Essentials of specific learning disability identification*. Hoboken, NJ: Wiley.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment*. Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2011). A CHC-based operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 233–298). Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459–483). New York: Guilford Press.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. (2013). Cognitive assessment: Progress in psychometric theories of intelligence, the structure of cognitive ability tests, and interpretive approaches to cognitive test performance. In D. Saklofske, C. R. Reynolds, & V. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 239–285). New York: Oxford University Press.

Flanagan, D. P., Alfonso, V. C., & Reynolds, M. R. (2013). Broad and narrow CHC abilities measured and not measured by the Wechsler scales: Moving beyond within-battery factor analysis. *Journal of Psychoeducational Assessment, 31,* 202–223.

Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., & McDonough, E. M. (2018). Dual discrepancy/consistency operational definition of SLD: Integrating multiple data sources and multiple data gathering methods. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Fiorello, C., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools, 47,* 739–760.

Flanagan, D. P., Genshaft, J. L., & Harrison, P. L. (Eds.). (1997). *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford Press.

Flanagan, D. P., & Harrison, P. L. (Eds.). (2005). *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.). New York: Guilford Press.

Flanagan, D. P., & Harrison, P. L. (Eds.). (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York: Guilford Press.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: Wiley.

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314–325). New York: Guilford Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). New York: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2017). *The cross-battery assessment software system, version 2.0* (X-BASS v2.0). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities*. Boston: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *The achievement test desk reference (ADTR): A guide to learning disability identification* (2nd ed.). Boston: Allyn & Bacon.

Flanagan, D. P., & Schneider, W. J. (2016). Cross-Battery Assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues' "Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth." *International Journal of School and Educational Psychology*, *4*(3), 137–145.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford Press.

Fletcher-Janzen, E., & Reynolds, C. R. (Eds.). (2008). *Neuropsychological perspectives on learning disabilities in the era of RTI: Recommendations for diagnosis and intervention*. Hoboken, NJ: Wiley.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell–Horn–Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools, 45,* 132–144.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J. R. (2010). Problems with IQ gains: The huge vocabulary gap. *Journal of Psychoeducational Assessment, 28,* 412–433.

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223–236.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Horn, J. L., & Blankson, N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73–98). New York: Guilford Press.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 118 Stat. 2147 (2004).

Juarez, B. M. (2012). *Effects of general and specific cognitive abilities on reading achievement in a referred sample*. Unpublished doctoral dissertation, St. John's University, Jamaica, NY.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 23–38). New York: Guilford Press.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York, NY: Wiley.

Kaufman, A. S. (2000). Foreword. In D. P. Flanagan, K. S. McGrew, & S. O. Ortiz, *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation* (pp. xiii–xv). Needham Heights, MA: Allyn & Bacon.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement, Third Edition*. Bloomington, MN: Pearson.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373–402). New York: Guilford Press.

Keith, T. Z., Fine, J. G., Reynolds, M. R., Taub, G. E., & Kranzler, J. H. (2006). Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 35,* 108–127.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY—Second Edition*. San Antonio, TX: Harcourt Assessment.

Lezak, M. D. (1976). *Neuropsychological assessment*. New York: Oxford University Press.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Hoboken, NJ: Wiley.

Mascolo, J. T., Alfonso, V. C., & Flanagan, D. P. (2014). *Essentials of planning, selecting, and tailoring interventions for unique learners*. Hoboken, NJ: Wiley.

McCloskey, G., Slonim, J., Whitaker, R., Kaufman, S., & Nagoshi, N. (2017). A neuropsychological approach to interpretation of the WISC-V. In D. P. Flanagan & V. C. Alfonso, *Essentials of WISC-V assessment*. Hoboken, NJ: Wiley.

McDonough, E. M., Flanagan, D. P., Sy, M., & Alfonso, V. C. (2017). Specific learning disorder. In S. Goldstein & M. DeVries (Eds.), *Handbook of DSM-5 disorders in children and adolescents* (pp. 77–104). New York: Springer.

McGrew, K. S. (1994). *Clinical interpretation of the Woodcock–Johnson Tests of Cognitive Ability—Revised*. Boston: Allyn & Bacon.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive CHC framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–180). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–182). New York: Guilford Press.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): CHC cross-battery assessment*. Boston: Allyn & Bacon.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond *g*: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 189—210.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–

Horn–Carroll cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651–675.

Messick, S. (1992). Multiple intelligences or multilevel intelligence?: Selective emphasis on distinctive properties of hierarchy: On Gardner's *Frames of Mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. *Psychological Inquiry, 3,* 365–384.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment.* Hoboken, NJ: Wiley.

Miller, D. C. (Ed.). (2010). *Best practices in school neuropsychology.* Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System.* Itasca, IL: Riverside.

Niileksela, C. R., Reynolds, M. R., Keith, T. Z., & McGrew, K. S. (2016). A special validity study of the WJ IV: Acting on evidence for specific abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 65–106). Amsterdam: Elsevier Academic Press.

Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition.* San Antonio, TX: Psychological Corporation.

Pearson. (2012). *Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition.* San Antonio, TX: Psychological Corporation.

Raiford, S. E., Drozdick, L., Zhang, O., & Zhou, X. (2015). *Expanded index scores* (WISC-V Technical Report No. 1). Bloomington, MN: Pearson.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence, 62,* 31–47.

Reynolds, M., Keith, T., Flanagan, D. & Alfonso, V. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51,* 535–555.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J. D., & Grenier, J. R. (Eds.). (1998). *Test interpretation and diversity: Achieving equality in assessment.* Washington, DC: American Psychological Association.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schneider, W. J., & Roman, Z. (2017). Fine-tuning cross-battery assessment procedures: After follow-up testing, use all valid scores, cohesive or not. *Journal of the Psychoeducational Assessment, 36*(1), 34–54.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479–502.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

U.S. Department of Education. (2005, June 21). 34 C.F.R. Parts 300, 301 and 304. *Federal Register, 70*(118), 35781–35892.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences, 13,* 159–188.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* New York: Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition.* San Antonio, TX: Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition.* San Antonio, TX: Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wilhoit, B. (2017). Best practices in cross-battery assessment of nonverbal cognitive ability. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (2nd ed., pp. 59–76). New York: Springer.

Wilson, B. C. (1992). The neuropsychological assessment of the preschool child: A branching model. In F. Boller & J. Grafman (Series Eds.) & I. Rapin & S. I. Segalowitz (Vol. Eds.), *Handbook of neuropsychology: Vol. 6. Child neuropsychology* (pp. 377–394). Amsterdam: Elsevier.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery—Revised.* Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock–Johnson III Normative Update.* Itasca, IL: Riverside.

CHAPTER 28

Linking Cognitive Abilities to Academic Interventions for Students with Specific Learning Disabilities

Nancy Mather
Barbara J. Wendling

An understanding of the relationships among cognitive abilities and achievement can help an evaluator make informed decisions regarding the choice of assessment tools, the type of diagnosis, the selection of accommodations, and the choice of instructional methodologies. In addition, cognitive assessment data are critical for the identification and understanding of specific learning disabilities (SLD). The definition of SLD indicates that a significant weakness exists in some aspect of basic psychological processing that is affecting learning and interfering with academic progress. Therefore, in this chapter we focus on (1) the cognitive abilities that are most commonly measured by various assessment instruments, (2) the ways these constructs relate to SLD, and (3) the possible instructional implications. We begin by discussing some general issues.

GENERAL ISSUES

In the current educational environment, psychologists and diagnosticians may encounter four general issues related to cognitive assessments of students with SLD: (1) the role of response to intervention (RTI), (2) the rationale for understanding the cognitive correlates of an academic area, (3) the need for and provision of differentiated instruction, and (4) the importance of establishing a *pattern of strengths and weaknesses* (PSW) for the determination of SLD.

The Role of RTI

The reauthorization of the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004) eliminated the requirement of establishing an ability–achievement discrepancy for determining the presence of an SLD and offered other alternative procedures. One alternative approach permits the use of limited response to evidence-based instruction as *part* of the process. This process of documenting progress on interventions has been referred to as *response to intervention* (RTI). For those embracing an RTI model, the results from cognitive assessments may seem to be irrelevant for both identification and instructional planning. Another approach, which is often referred to as the *third method,* allows for the use of alternative research-based methods. This procedure is most often implemented by using a PSW approach to establish the presence of an SLD. Professionals implementing a PSW approach typically use both cognitive and academic assessments.

According to a Response to Intervention Adoption Survey (Spectrum-K12, 2011), 94% of schools surveyed reported implementing some level of RTI back in 2011. This represented an increase from 72% in 2009. Sixty-six percent of schools surveyed reported using RTI as part of the process for determining eligibility for special education—a 25% increase from 2010. As of 2013, 17 states required schools to use an RTI process to help determine which students are eligible for special education, and every state mentions RTI in their regulations (Hauerwas, Brown, & Scott, 2013). With the increase in districts implementing RTI and using RTI as part of the eligibility determination process, evaluation professionals have to consider the role of cognitive assessment. Does cognitive assessment add anything meaningful to the identification process or to instructional planning? Maybe it is not a question of either/or.

Most professionals supporting the role of cognitive assessment also encourage the use of RTI (e.g., Hale, Kaufman, Naglieri, & Kavale, 2006). Most practitioners agree that RTI has value as an instructional model of early intervention and prevention. The disagreement arises when professionals suggest that RTI be used as the sole identification or diagnostic model for determining the presence of SLD. Although an RTI process can help document whether a student is making adequate progress, this information is insufficient for the diagnosis of a disability.

RTI is not a diagnostic model, but rather focuses on analysis of groups or grade levels of students and treats them in the same manner: It is a "one-size-fits-all" approach (Wodrich, Spencer, & Daley, 2006). In contrast, special education is focused on addressing the needs of a specific individual; it is a "one-size-fits-one" approach. To design an individualized approach, an evaluator must first understand an individual's unique strengths and weaknesses. Thus many psychologists and diagnosticians are shifting their focus from the use of full-scale intelligence scores to a more in-depth analysis of an individual's specific abilities. Today's cognitive assessments, as well as the approaches to assessment, are more theory-based and more diagnostic, helping practitioners explore and document an individual's unique strengths and weaknesses.

Each of the methods for SLD identification has both positive and negative aspects, and skilled evaluators consider additional factors beyond test scores and RTI data. As Kovaleski and colleagues (2015) explained, "Knowledgeable practitioners also use clinical judgment to determine which approach is applicable for a given child or in a given school setting. While regulations and policies require school districts to implement a single approach, best practice may reside somewhere in the margins with a hybrid model" (p. 6). Professional judgment is used to (1) investigate a learning problem before testing is conducted; (2) modify a test selection plan during test administration; and (3) determine, after the initial assessment, whether additional testing is needed to clarify the nature of the learning problem (Schrank, Stephens, & Schultz, 2016).

For the results of cognitive tests to be useful, the focus must be on obtaining information that is relevant to academic performance and instruction. Even though a student may be deemed ineligible for certain services, all evaluations need to address the referral concerns and propose solutions. As Cruickshank (1977) noted, "Diagnosis must take second place to instruction, and must be a tool of instruction, not an end in itself" (p. 194). Within the field of SLD, the purpose of testing cognitive abilities is to determine a person's unique abilities and identify the underlying cognitive weaknesses that are often directly linked to the specific weaknesses in achievement. Therefore, it is important to identify the areas of strength and weakness, in order to individualize educational planning for students who do not respond adequately to intervention (Kavale & Flanagan, 2007). From the perspective of making an accurate SLD diagnosis, understanding *why* a student is not responding adequately does indeed matter, and the results from cognitive assessments often help to provide the necessary insights (Mather & Kaufman, 2006). As noted by McCloskey, Whitaker, Murphy, and Rogers (2012), "The nature of a tier 3 referral makes it imperative that the specific cognitive strengths and weaknesses of the student and their impact on learning and production be clearly specified" (p. 870). (See also McCloskey, Slonim, & Rumohr, Chapter 39, this volume).

Although these legal changes in identification procedures have created uncertainty and disequilibrium in the field, they have also led to a greater interest in understanding the relationships among cognitive abilities and achievement. Implementation of a simple discrepancy formula for SLD identification was inadequate for determining eligibility and insufficient for informing instructional planning. Although little research supports the use of IQ–achievement discrepancy, 67% of states still allow practitioners to use an ability–achievement discrepancy in identifying SLD (Maki, Floyd, & Roberson, 2015).

In today's educational world, professionals must assess, explore, and understand an individual's profile of cognitive, linguistic, and academic strengths and weaknesses to inform both eligibility determinations and instructional planning. Individualized education program (IEP) teams can make informed decisions about the educational program for a specific student by linking assessment to interventions (Maricle & Johnson, 2016). So, now more than ever, a need exists for understanding how various cognitive abilities are related to school achievement.

Understanding the Cognitive Correlates of Academic Areas

A growing body of research supports the usefulness of factor or specific ability scores in identifying the cognitive processing problems that specifically inhibit academic development (e.g., Cormier, McGrew, Bulut, & Funamoto, 2017; Fiorello, Hale, & Snyder, 2006; Flanagan, Ortiz, & Alfonso, 2013; McGrew & Wendling, 2010; Moats & Tolman, 2009). As examples in math achievement, weaknesses in number sense characterize many students with mathematical difficulties (Jordan, Fuchs, & Dyson, 2015). The most commonly studied cognitive abilities related to math achievement are working memory, processing speed, and general intelligence (Geary, Hoard, & Bailey, 2011; Peng, Wang, & Namkung, 2018). For example, research indicates that as working memory capacity increases, students become more effective in solving math word problems (e.g., Swanson, 2014; Zheng, Swanson, & Marcoulides, 2011). Also, researchers have identified fluid reasoning as an important correlate of math reasoning (e.g., Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007).

In reading, Adams (1990) found that a child's level of phonemic awareness in kindergarten was the best predictor of reading success in elementary school. Over the last two decades, numerous researchers have confirmed the importance of phonemic awareness to the development of basic reading skills (e.g., Berninger et al., 2006; Cooper, 2006; Cormier et al., 2017; Feifer, 2011; Shaywitz, Morris, & Shaywitz, 2008; Torgesen, 2002; Vellutino, Tunmer, Jaccard, & Chen, 2007). Others have described the relationship between slow rapid automatized naming (RAN) and poor reading (e.g., Berninger et al., 2006; Denckla & Rudel, 1974; Norton & Wolf, 2012; Swanson, Trainin, Necoechea, & Hammill, 2003; Torgesen, 1997; Wolf, Bowers, & Biddle, 2000). Still others have described the significant relationships between working memory and reading comprehension (e.g., Cooper, 2006; Swanson, 1999, 2000; Torgesen, 2002) and between fluid reasoning and reading comprehension (e.g., Cormier et al., 2017). In a recent study that explored the relationships between cognitive abilities and reading comprehension across a large sample of children ($N = 835$) in grades 1–5, the importance of specific cognitive abilities (e.g., auditory processing, long-term retrieval, fluid reasoning) varied across grade levels; however, vocabulary and general knowledge were significant at all levels (Decker, Strait, Roberts, & Wright, 2017).

In regard to reading development, similar difficulties have been observed in both children and adults. For example, in a meta-analysis of reading research, Swanson and Hsieh (2009) found that the cognitive processes of phonological skills, verbal memory, naming speed, and vocabulary all made significant contributions to reading disabilities in adults.

Early identification of students at risk for academic failure continues to be a primary goal of all educators. This desire for early identification fuels the research into the causes of academic failure and seeks to determine the predictive value of specific cognitive and linguistic factors that may be considered "precursors to manifest disabilities" (Fletcher et al., 2002, p. 51). Ideally, a predictor or subset of predictors will accurately differentiate between the children who will struggle with certain academic subjects and those who will not, so that intervention efforts can be initiated early in a timely fashion (Bishop, 2003). Researchers continue to document the relationships among specific cognitive, linguistic, and academic abilities, identifying prerequisite skills and delineating early indicators of risk.

Many investigators have encouraged the early assessment of cognitive abilities to identify children who are at risk for reading disabilities (e.g., Berninger et al., 2006; Fuchs et al., 2005; Hirsch, 2013; Torgesen, 2002). For example, Carroll, Solity, and Shapiro (2016) investigated which cognitive abilities were the best predictors of poor reading in later grades. As was expected, print knowledge, verbal short-term memory, phonological awareness, and rapid naming were all good predictors of later poor reading. In addition, weaknesses in visual search and auditory processing were present in a large minority of the poor readers. Nearly all of the poor readers had a deficit in at least one cognitive area upon school entry, but no single deficit characterized the majority of poor readers. Thus several relevant cognitive abilities can be assessed before students learn to read, such as phonemic awareness, RAN, and short-term memory; at-risk students can thereby be identified at a very young age—even before they experience reading difficulties. Exploring cognitive abilities can also help with establishing a PSW that is relevant to identifying an SLD, trying to understand why a student fails to make adequate progress, and/or planning the most effective instructional programs. The evolution and refinement of theory- and research-based tests measuring multiple abilities have given professionals the opportunity to gain a better understanding of an individual's unique characteristics.

Need for and Provision of Differentiated Instruction

Differentiated instruction is not a program, but a way of teaching. Because all students are not alike, a teacher plans varied approaches to address what individual students need to learn, how they will learn it, and how they will demonstrate what they have learned, thus increasing the likelihood that each student will learn as much as possible (Tomlinson, 2006). Responsive teaching is a simple definition of differentiated, prescriptive instruction. With the growing diversity of the U.S. population and the increased demand for accountability in education, the need to accommodate individual learners has intensified in general education.

Although educators widely embrace the concept of meeting individual students' needs, nearly 40% of new teachers (ones with 4 years or less of teaching experience) responding to the 2012 Schools and Staffing Survey did not feel adequately prepared to implement differentiated instruction (National Center for Education Statistics, 2012). This lack of teacher preparation stems in part from the failure of teacher training programs to integrate recommended methodologies and strategies into the required courses (Elksmin, 2001; Whitaker, 2001). To address these shortcomings, the U.S. Department of Education is drafting regulations for teacher preparation programs, with full implementation called for by 2020. (See *www.ed.gov/teacherprep*.)

Unfortunately, even in special education, many students with IEPs receive the same treatment goals and teaching strategies as their normally achieving peers (Reynolds & Lakin, 1987). To explore the implementation of differentiated instruction, Schumm, Moody, and Vaughn (2000) interviewed third-grade teachers who served students with SLD in inclusive classrooms. Overall, the teachers reported using whole-class instruction that included the same materials for all students, regardless of performance levels. All students were expected to read grade-level materials, even if they could not pronounce the words. Furthermore, students with SLD did not receive instruction directed at improving their word analysis skills. One teacher actually voiced strong opposition to providing instruction in word analysis, stating, "By the time they come to third grade they really should have those skills" (quoted in Schumm et al., 2000, p. 483). With undifferentiated instruction and minimal direct instruction in reading, the students with SLD made little academic improvement, and their attitudes toward reading declined as well.

Historically, the field of SLD has been based on the belief that children differ in their abilities and therefore should be taught and treated differently (Kavale & Forness, 1998; Whitener, 1989); this belief led to the requirement that an IEP must be developed for each student eligible for special education services. As Moats (2015) has noted, "Meeting the needs of students across the spectrum of academic ability . . . requires acceptance of diversity . . . To use one yardstick to measure student growth, one set of standards to drive what is taught, and one view of academic success is indefensible" (p. 22). A student who has difficulty memorizing information requires a different type of instruction than the student who memorizes facts easily. A student who works slowly needs more time than a student who works rapidly. A student who struggles to pay attention requires more novelty and structure than one who attends easily. In some cases, these marked differences in learning and behavior are neurologically based (Semrud-Clikeman et al., 2000; Shaywitz, 1998, 2003; Shaywitz et al., 2003). Clearly, for students with SLD, differentiated instruction that addresses their specific problems will be more effective than global approaches that do not (Aaron, 1997). Some believe that the lack of individualization in special education is a result of the current emphasis on and push for full inclusion. Hallahan and Cohen (2008) cautioned:

> Students with learning disabilities are not receiving special education, which is based on the core principles of intensive, relentless, structured, appropriately paced instruction in small groups with frequent monitoring of each student's progress. We believe

> that the diminution of special education for students with learning disabilities has occurred because of the well-meaning, but misinformed, overly zealous adoption of inclusive educational practices. (p. 3)

Determining a PSW

As mentioned previously, IDEA 2004 states that alternative research-based methods, such as a PSW approach, can be employed to determine eligibility for SLD services. Currently, however, there is debate about the effectiveness of various PSW methods for the identification of SLD. Proponents of using PSW methods state that identifying patterns of strengths and weaknesses helps explain low achievement, can inform instructional planning, and helps differentiate students with SLD from slow learners (Flanagan, Fiorello, & Ortiz, 2010; Naglieri & Otero, 2012; Schultz, Simpson, & Lynch, 2012). Others argue that current PSW methods do not reliably identify SLD (Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016; Stuebing, Fletcher, Branum-Martin, & Francis, 2012; Taylor, Miciak, Fletcher, & Francis, 2017). Although one particular PSW method (e.g., Hale & Fiorello's [2004] concordance–discordance method, Flanagan et al.'s [2013] dual-discrepancy/consistency method, Naglieri's [2011] discrepancy–consistency method) may not always reliably identify individuals with SLD, helping practitioners explore significant intraindividual variations, or the differences among abilities, is precisely how intelligence tests can contribute to SLD determination and educational planning. To be most useful, the data collected from cognitive assessments must be combined with other forms of data, including achievement data, as well as the interpretive insights gained from the evaluator's clinical judgment (Schultz et al., 2012).

Assessments should then focus on understanding a person's information-processing capabilities, including the factors that can facilitate performance. As Gardner (1999) has suggested, in the study of human cognition, awareness of distinctive strengths is of critical importance. Evaluators must also understand the constraints (e.g., limited instruction, specific cognitive or linguistic weaknesses, limited cultural experiences, poor motivation) that affect performance, as well as the multidimensional impact of these constraints (Berninger, 1996). Because the various constraints affect different aspects of academic functioning, they can help inform the type and extent of accommodations and instruction needed. Interpreting intraindividual variations and determining how these differences affect performance then become the cornerstones for linking the results of cognitive ability tests to meaningful interventions. To do this in a valid manner, assessment professionals must know the existing research on the relationship between cognitive abilities and achievement, and must incorporate that knowledge into their decision-making process.

One basic concept underlying identification of SLD is that a student's difficulty does not extend too far into other domains. In other words, the problem is relatively specific, circumscribed, or domain-specific (Stanovich, 1999). This concept of specificity, a main tenet of SLD, is not new. For example, Travis (1935) observed that in some students, a striking disparity exists between achievement in one area and achievement in another. For example, a student may struggle with reading, but can easily comprehend material that is read aloud; or a student excels in reading and writing, but struggles with mathematical concepts and applications. These children, who do not achieve as well as would be expected in one or more areas of performance, may be regarded as having a "special defect or disability" (Travis, 1935, p. 43).

To maintain this concept of specificity, the academic problem is best described as a specific reading disability, math disability, or spelling disability that is associated with weaknesses in specific cognitive or linguistic processes. The first part of an evaluation is then to determine an initial domain-specific classification (Stanovich, 1999); the next part is to identify the specific cognitive processes that underlie the disorder (Robinson, Menchetti, & Torgesen, 2002). An SLD is caused by one or more inherent weaknesses in underlying cognitive processes (Robinson et al., 2002). The assessment process can then be viewed as an ability-oriented evaluation designed to help formulate the problem and then determine specific interventions (Fletcher, Taylor, Levin, & Satz, 1995).

COGNITIVE ABILITIES AND ACADEMIC PERFORMANCE

As noted previously, a growing body of research links certain cognitive and linguistic factors to the various domains of achievement. The relationships between certain cognitive abilities and academic performance are well established (e.g., phonemic awareness and decoding), whereas others are not (e.g., visual processing and reading).

Furthermore, some important cognitive constructs are commonly measured on many intelligence tests (e.g., vocabulary, reasoning, and working memory), whereas others are not (e.g., phonological awareness and RAN). Even though we discuss these cognitive abilities separately, Horn (1991) admonished that attempting to measure cognitive abilities in isolation "is like slicing smoke" (p. 198). Cognitive and academic abilities are interrelated, and various combinations of abilities are employed as an individual completes specific tasks. Therefore, when studying and diagnosing complex learning disorders, evaluators need to embrace a multiple-deficit approach (Carroll et al., 2016; McGrath et al., 2011; Pennington, 2006).

Consider the various skills required to take notes while listening to a lecture. The note taker must pay attention, have knowledge of the topic, understand the vocabulary, use memory to hold on to and paraphrase the important points, and then record these thoughts in writing. A student may struggle with note taking for any or all of these reasons. In addition, the prediction of performance for students with SLD may be improved when several factors are considered (Gregg, Davis, Coleman, Wisenbaker, & Hoy, 2004). For example, when combined, measures of working memory and language comprehension appear to provide the best prediction of reading comprehension ability (Daneman & Carpenter, 1980).

In some circumstances, something influences performance other than what a test was designed to measure. Stern (1938) noted: "It should never, of course, be supposed possible to test a definite, narrowly circumscribed separate capacity of thought with any one of these tests. Other abilities are always involved" (p. 315). He continued:

> Yet this is in no sense to be construed as a defect in the tests. On the contrary, they provide a favorable opportunity for observing the process of thinking in all its complexity. One must not be content to calculate the score for each performance. A completed test, which according to the system of scoring is thrown out as erroneous or deficient in performance, may very frequently result from the fact that *other* kinds of thinking than those expected have intervened, but which may have significance in terms of the subject's particular intellectual approach. (p. 316; original emphasis)

Thus Stern emphasized that intelligence tests are valuable beyond the mere production of scores because careful observation during performance and analysis of the psychological processes that led to the test answers can deepen an evaluator's insight into the structure and functioning of cognitive abilities. Accordingly, we reemphasize the value of forming, exploring, confirming, or rejecting diagnostic hypotheses that are based on test scores; careful, systematic observations of behavior; and prior history.

Vocabulary, Acquired Knowledge, and Language Comprehension

Unless they are designed primarily to measure nonverbal abilities, most intelligence batteries contain measures of vocabulary, acquired knowledge, and language comprehension. Often described as *crystallized intelligence, Gc, verbal or oral language abilities,* or *stores of acquired knowledge,* these abilities are highly correlated with achievement and are good predictors of academic success (Anastasi, 1988; Hirsch, 2013; Niileksela, Reynolds, Keith, & McGrew, 2016; Sinatra, Zygouris-Coe, & Dasinger, 2011). Ample evidence supports the finding that increasing a child's general knowledge and vocabulary before age 6 is the single highest correlate with later school success (Hirsch, 2013). Because most of the measures of crystallized intelligence rely on language, crystallized intelligence can be equated with verbal intelligence (Carroll, 1993; Hunt, 2000) and is often used as a key indicator of giftedness (Benbow & Lubinski, 1996; National Association for Gifted Children, 2010). These types of tests typically measure general aspects of cultural knowledge, rather than specialized knowledge specific to a domain, such as the rules for soccer. Because of the cultural and linguistic content of these tasks, evaluators must use caution when testing and interpreting the performance of individuals from different linguistic or cultural backgrounds, such as English-language learners (ELLs). Poor performance may be the result of limited exposure and/or opportunities, rather than a general weakness in word and world knowledge.

These linguistic abilities are related to most areas of academic performance across the lifespan. Verbal abilities and background knowledge have a strong and consistent relationship with reading (e.g., Beck & McKeown, 2007; Cooper, 2006; Cormier et al., 2017; Decker et al., 2017; Kintsch & Rawson, 2005; Nation, 2007; Perfetti & Stafura, 2014; Reynolds & Turek, 2012; Shaywitz et al., 2008), mathematics (Floyd, Evans, & McGrew, 2003; Gelman & Butterworth, 2005; Swanson & Jerman, 2006), and writing (Berninger, 2009; Cormier, Bulut, McGrew, & Frison, 2016; McCloskey, Perkins, & Van Divner, 2009; Stoeckel et al., 2013).

The most fully substantiated relationship is with reading comprehension and written expression. Decker and colleagues (2017) found that crystallized knowledge was a common cognitive contributor to reading comprehension across all of the elementary grades. Both reading comprehension and written expression depend on background knowledge, which enables a person to understand and create messages, interpret sentence structures, use verbal reasoning abilities, and employ a broad and deep vocabulary (McCardle, Scarborough, & Catts, 2001; Nation, Clarke, & Snowling, 2002). Words and the concepts they represent are thus the building blocks of literacy (Bell & Perfetti, 1994; Cunningham, Stanovich, & Wilson, 1990; Perfetti, Marron, & Foltz, 1996; Sinatra et al., 2011). In addition to vocabulary and background knowledge, researchers have identified other verbal abilities that are frequently weaknesses for individuals with reading comprehension difficulties: listening comprehension (Biemiller, 2003; Nation, Clarke, Marshall, & Durand, 2004), figurative language (Cain, Oakhill, & Lemmon, 2004), grammar (Nation & Snowling, 2000), and oral expression (Nation et al., 2004). Unlike other cognitive abilities, crystallized intelligence has been described as a maintained ability rather than a vulnerable ability because it continues to develop until midlife and does not decline with age as significantly as other cognitive abilities (Horn, 1991). On growth curves, the rate of growth for crystallized intelligence is much greater than for other abilities, and it shows a less rapid rate of decline (McGrew, LaForte, & Schrank, 2014).

Reasons for Differences in Performance

Some people will demonstrate a weakness on verbal ability tests because of language impairments, whereas others will have weaknesses due to limited experiences with language and/or a lack of educational experiences and opportunities (Carlisle & Rice, 2002). In addition, tests of general knowledge, vocabulary, and language comprehension most often reflect the culture and language of the norm group. Therefore, individuals from diverse cultural and/or linguistic backgrounds or from families of low socioeconomic status (SES) often obtain lower scores on measures of acquired knowledge. A child's language acquisition is influenced by the parents' attained level of education and the family's SES, as well as by exposure to early literacy and language activities (Hart & Risley, 1995).

Vocabulary access is influenced by three main factors: (1) familiarity with words, (2) the depth of conceptual understanding of those words, and (3) the ability to retrieve words as needed (Gould, 2001). A student may understand a word's meaning, but have difficulty using the word correctly when speaking. These word-finding or word-retrieving difficulties may also negatively influence performance on the verbal subtests found in many intelligence measures, particularly if the measure is timed. If a student obtains a low score on a vocabulary measure, the evaluator must determine whether that low score is a result of limited verbal knowledge, limited cultural experiences, or difficulty in retrieving verbal labels (a problem more closely linked to associative memory).

Implications for Achievement

Since many academic tasks require linguistic competence, individuals with low verbal abilities are likely to encounter academic difficulties in most areas and will need increased opportunities and experiences to improve linguistic abilities, including increasing the depth and breadth of their vocabulary and world knowledge. Early intervention that develops vocabulary knowledge is crucial to future academic success (Sinatra et al., 2011). Many children, especially those from low-SES communities, enter school with limits in their oral language. In general, students who have difficulty understanding or using spoken language will have difficulty with the aspects of reading, writing, and mathematics that depend on language-specific processes, such as reading comprehension, written expression, and math problem solving. These abilities require the integration of many skills. As an example, Moats and Daken (2008) explain the many components of written expression as follows:

> The ability to compose and transcribe conventional English with accuracy, fluency, and clarity of expression is known as basic writing skills. Writing is dependent on many language skills and processes and is often even more problematic for children than reading. Writing is a language discipline with many component skills that must be directly taught. Because writing demands using different skills at the same time, such as generating language, spelling, handwriting, and using capitalization and punctuation, it puts a significant demand on working memory and attention. Thus, a student may demonstrate mastery of these individual skills, but when asked to integrate

> them all at once, mastery of an individual skill, such as handwriting, often deteriorates. To write on demand, a student has to have mastered, to the point of being automatic, each skill involved. (p. 55)

Because reading comprehension and written expression share many of the same cognitive and linguistic processes, individuals frequently have difficulties in both areas. For example, a study of individuals with writing disabilities found that 75% of the sample also had reading difficulties (Katusic, Colligan, Weaver, & Barbaresi, 2009). As noted previously, both reading comprehension and written expression require vocabulary and background knowledge—in other words, they are built upon a solid foundation of oral language. Early deficits in vocabulary have been identified as a risk factor for later reading problems (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004). A growing body of research illustrates the relationship between academic vocabulary interventions and improvement in comprehension (Lawrence, Rolland, Branum-Martin, & Snow, 2014; Lesaux, Kieffer, Kelley, & Harris, 2014; Townsend & Collins, 2009). Researchers have also established that the primary differences between individuals with good reading comprehension and those with poor reading comprehension are differences in verbal ability (Floyd, Bergeron, & Alfonso, 2006).

In discussing the reasons for reading comprehension failure, Perfetti and colleagues (1996) distinguished between the processes involved in comprehension (e.g., working memory and comprehension monitoring) and knowledge—which includes word meanings or vocabulary, as well as domain knowledge (i.e., the concepts specific to a domain, such as physics, biology, or history). Clearly, knowledge is an important component underlying reading comprehension that contributes to individual differences in reading (Hannon & Daneman, 2001; Oslund, Clemens, Simmons, Smith, & Simmons, 2016). A person's level of acquired knowledge, including domain knowledge obtained through life experiences, school, or work, is highly predictive of academic performance. Breadth and depth of knowledge and a robust oral vocabulary suggest that the person will excel on tasks involving language-learning abilities, whereas limited knowledge and a poor vocabulary suggest that the person will struggle.

In the simple view of reading (Gough & Tunmer, 1986), reading ability equals the product of decoding and linguistic comprehension. The equation used to represent this simple view is $R = D \times C$. Within this model, decoding is measured by the ability to pronounce pseudowords, and linguistic comprehension is assessed by a test of listening comprehension. If either decoding or linguistic comprehension is impaired, reading performance is compromised. Within this framework, Gough and Tunmer proposed three types of reading disabilities: inability to decode (dyslexia), inability to comprehend (hyperlexia), or both ("garden-variety" poor reader). The results of a large-scale study support the simple view of reading with poor reading comprehension associated with inadequate decoding, vocabulary, or both (Spencer, Quinn, & Wagner, 2014). In the absence of poor word recognition skills, a language-based problem is at the core of reading comprehension difficulties (Oakhill & Cain, 2016; Spencer et al., 2014).

Individuals with specific reading disabilities, or dyslexia, typically have verbal abilities that are more advanced than their decoding skills. Essentially, what distinguishes individuals with reading disabilities from other poor readers is that their listening comprehension ability is higher than their ability to decode words (Rack, Snowling, & Olson, 1992). Listening comprehension is frequently cited as a good predictor of reading comprehension (Aaron & Joshi, 1992; Cooper, 2006; Hammill, 2004). Thus measures of verbal abilities, including listening comprehension, can be used to provide the best estimates of how much poor readers would profit from written text if their deficient decoding skills were resolved (Stanovich, 1999). Caution is needed, however, because measures of verbal ability and listening comprehension may underestimate potential for achievement among students with attention or language-processing problems, as well as among students for whom English is a second language (Berninger & Abbott, 1994; Fletcher et al., 1998).

Beyond third grade, individuals with good verbal ability and good reading skills acquire knowledge and new vocabulary primarily through reading. In contrast, individuals with good verbal ability but poor reading skills are much more likely to learn new vocabulary through oral discussions (Carlisle & Rice, 2002). Unfortunately, since reading rather than listening is used to acquire more complex syntax and abstract vocabulary, poor readers tend to fall behind good readers on verbal tasks as they progress through school. Since many intelligence tests include vocabulary measures, a poor reader's relative standing on the verbal scores may decline when compared to that of normally achieving peers. As a result, poor language contributes to a lower ability score, as well as to poor reading (Fletcher et al., 1998; Strang, 1964).

Thus reading experience influences verbal intelligence test scores. In addition, cognitive and academic tests assess many of the same underlying abilities (e.g., vocabulary, general information) (Aaron, 1997). Older students with reading difficulties may have depressed performance on measures of verbal intelligence and reading because of limited experiences with text. Several decades ago, Strang (1964) summarized this problem as follows:

> Intelligence tests are not a sure measure of innate ability to learn. They measure "developed ability," not innate or potential intelligence. Previous achievement affects the test results. The poor reader is penalized on the verbal parts of the test. The fact that his store of information is limited by the small amount of reading he has done also works against him. (p. 212)

For students with reading disabilities, limited print exposure contributes to reduced knowledge and vocabulary, and these deficiencies are likely to increase over time (Vellutino, Scanlon, & Lyon, 2000). This phenomenon has been referred to as the "Matthew effect," which is described in sociology when discussing economic inequality as "the rich get richer and the poor get poorer" (Stanovich, 1986; Walberg & Tsai, 1983). Limited reading alters the course of development in education-related cognitive skills (Stanovich, 1993). In other words, a reading difficulty leads to less time spent reading, which in turn contributes to lower verbal ability and knowledge. The relationship between good reading skills and the development of reading-related cognitive abilities has been described as the "virtuous circle" (Snowling & Hulme, 2011). On the other hand, the "vicious circle" includes students who fail to develop reading skills, tend to avoid reading, and do not develop reading-related and cognitive abilities, all of which limits both cognitive and academic development (Pulido & Hambrick, 2008).

Verbal ability has also been identified as a strong predictor of math performance (Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001) and has been linked to early math achievement, especially to the development of number concepts (Carey, 2004; Gelman & Butterworth, 2005). The importance of verbal ability for math performance appears to increase with age and is most likely related to the increased linguistic demands of complex problem solving (Fuchs et al., 2006, 2008; Geary, 1994; McGrew & Wendling, 2010). Individuals with a math disability tend to score lower on verbal ability measures than typical age peers (Proctor, Floyd, & Shaver, 2005) and have limited oral language abilities (Fuchs et al., 2008). In addition, individuals with low verbal ability often experience problems in both math and reading, whereas those with more intact verbal abilities have problems that are more specific to math.

Interventions for Limited Verbal Ability

An individual with limited knowledge or vocabulary is likely to experience difficulty acquiring new knowledge or vocabulary, unless the new information is connected to prior knowledge (Beck, Perfetti, & McKeown, 1982). Instruction needs to build on prior knowledge and may need to be modified so that it occurs at the individual's language level. One finding from a meta-analysis of the impact of vocabulary instruction was that students with reading difficulties who received vocabulary instruction benefited three times as much as those who did not (Elleman, Lindo, Morphy, & Compton, 2009).

The National Reading Panel (NRP, 2000) found that individuals with limited vocabulary benefit from a variety of approaches, and that no one single approach is best for everyone. Some of the most effective instructional approaches for building vocabulary include both direct and indirect methods, such as reading aloud to students, providing explicit vocabulary instruction, and encouraging wide reading. Ideally, an individual's home and school environments are language-rich, with many opportunities and experiences to reinforce learning. A variety of strategies can be used to help students understand the nature of related words and concepts, such as semantic feature analysis, word webs, and graphic organizers. One important way language develops is through social interactions with more knowledgeable language users (Vygotsky, 1962). As teachers and students work together to attain educational goals, they can model the process of learning by talking about these processes as they perform tasks. Thus modeling and thinking aloud are useful for promoting language development. Research has also focused on the importance of developing academic vocabulary by emphasizing orthographic and morphological instruction (Nagy & Townsend, 2012; Townsend, Bear, Templeton, & Burton, 2016; Uccelli, Phillips Galloway, Barr, Meneses, & Dobbs, 2015). To further aid instruction, Gardner and Davies (2014) have identified the highest-frequency words within specific disciplines (the language of math, history, science, etc.). Swanson, Vaughn, and Wexler (2017) described numerous strategies for enhancing the vocabulary knowledge of adolescents.

A reciprocal relationship exists between learning to read and learning to write (Ehri, 2000), so instruction is more effective when these skills are taught in an integrated manner (Clay, 1982). Reading has been referred to as "language by eye," and writing as "language by hand" (Berninger, 2000; Berninger & Graham, 1998), further connecting the two achievement domains. As discussed, individuals with limited verbal ability will struggle primarily with reading comprehension and written expression. Although these individuals may experience success with lower-level skills, such as spelling or handwriting, their limits in language will interfere with the acquisition of higher-level skills, including vocabulary development, reading comprehension, sentence formulation, and written expression. Thus instruction should focus on increasing vocabulary, as well as both declarative and procedural knowledge (Perfetti & Stafura, 2014).

Although math is considered a less "verbal" achievement area, its demands on language and acquired knowledge are quite significant. Math requires knowledge of content-specific concepts and vocabulary, as well as the ability to understand story or word problems. Supporting the role of verbal ability in math, difficulty with math problem solving has been associated with deficient oral language abilities (Fuchs et al., 2008). In addition, conceptual knowledge of numbers and their relationships is another important correlate of math achievement (Hecht, Close, & Santisi, 2003). Explicit instruction, concrete examples, and guided practice are important for developing math vocabulary and concepts. The concrete–representational–abstract teaching sequence is beneficial for individuals struggling with mathematics. In this teaching sequence, the concrete level involves the use of objects or manipulative devices; the representational level involves visual representations, such as tallies or pictures; and the abstract level involves the use of actual numbers and equations.

Phonological Processing

Phonological awareness, another component of oral language, is important to an understanding of reading, writing, and even math disabilities. *Phonological awareness* refers to the ability to attend to various aspects of the sound structure of speech, whereas *phonemic awareness* refers to the understanding that words can be divided or segmented into phonemes, the individual speech sounds. The importance of phonological processing for promoting reading and spelling achievement has been extensively documented (e.g., Ehri, 1998; Fletcher, Lyon, Fuchs, & Barnes, 2007; Niileksela et al., 2016; Perfetti, 2011; Shaywitz, 2003; Snow, Burns, & Griffin, 1998; Torgesen, 1998; Uhry, 2005). Rack, Snowling, and Olson (1992) hypothesized that phonological awareness underlies the establishment of the graphemic memory store that is required for written language. Supporting this hypothesis, others have found that poor phonological processing interferes with the development of phoneme–grapheme mapping (Kilpatrick, 2015; Noble & McCandliss, 2005). Because phonological awareness abilities are known to be prerequisites for success in word reading and spelling, they should be assessed early, especially in cases where a child is developing slowly in word identification or spelling skill.

The role of phonological processing in math achievement is not as well documented as it is for reading, although phonological processing appears to correlate with math achievement (e.g., Hecht et al., 2003; McGrew & Wendling, 2010; Rasmussen & Bisanz, 2005). Geary (2007) found that individuals with comorbid reading and math difficulties often display phonological processing problems. Some researchers suggest that it plays a role in forming and encoding accurate phonological representations of math facts in working memory (Logie, Gilhooly, & Wynn, 1994; Swanson & Jerman, 2006). Peng and colleagues (2018) found that weaknesses in phonological processing and attention were more severe in younger students than older students with math disabilities.

Reasons for Differences in Performance

Cultural and linguistic differences have an impact on the development of phonological awareness. Individuals who have had limited exposure to the sounds of the English language, have limited oral language, have not been read to during the preschool years, and/or come from low-SES homes may have difficulty discriminating, rhyming, and manipulating speech sounds. They may also have difficulty repeating nonwords, which measures both phonemic sensitivity and phonological short-term working memory capacity (Schrank, Decker,

& Garruto, 2016). Older students may have mastered the more rudimentary skills of rhyming, blending, and segmenting but may still have difficulty with tasks involving the manipulation of sounds (e.g., deleting, substituting; Kilpatrick, 2015).

Implications for Achievement

A weakness in phonological processing as a common factor among individuals with early reading problems has been substantiated by an impressive body of research (e.g., Ehri, 1998; Fletcher & Foorman, 1994; Shaywitz et al., 2008; Stanovich & Siegel, 1994; Torgesen, 2002; Vellutino et al., 2007). Phonological processes are critical for the development of reading and spelling skills (Adams, 1990; Goswami & Bryant, 1990; Gough, 1996; Perfetti, 2011). Results from longitudinal studies suggest that 75% of children who struggle with reading in third grade will still be poor readers at the end of high school, primarily because of problems in phonological awareness (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Lyon, 1998). Individuals with poor phonological abilities typically make less progress in basic word-reading skills and spelling than normally achieving peers. Even spelling problems in young adults often reflect specific problems in the phonological aspects of language (Moats, 2001, 2010).

Numerous findings have documented the neurobiological basis of reading disabilities (Ashkenazi, Black, Abrams, Hoeft, & Menon, 2013; Richlan, Kronbichler, & Wimmer, 2011; Shaywitz, 2003; Wandell, Rauschecker, & Yeatman, 2012). The evolution of functional magnetic resonance imaging technology has made it possible to discover exactly which parts of the brain are engaged during phonological tasks. Good readers engage both the front and back of the left side of the brain as they perform reading tasks, whereas poor readers appear to rely primarily on the front part of the brain. Research has also documented that effective instruction in reading creates changes in brain behavior during reading (Shaywitz, 2003); this finding further emphasizes the extreme importance of implementing high-quality early intervention.

Some children who show phonologically-based reading difficulties also exhibit difficulties in the retrieval of math facts (Ashcraft, 1987, 1992; Geary, 2007; Light & DeFries, 1995). Phonological processing is a persistent weakness in individuals with math fact fluency deficits (Chong & Siegel, 2008). Speech sound processes are used when solving math computations—for example, when counting (Bull & Johnston, 1997; Geary, 1993). Several studies have implicated phonological processing as an underlying cause of individual differences in math problem solving as well (Furst & Hitch, 2000; Gathercole & Pickering, 2000; Geary & Brown, 1991; Swanson & Sachse-Lee, 2001).

Interventions for Limited Phonological Processing

Research results indicate a causal and reciprocal relationship between phonological awareness and reading; gains in one lead to gains in the other (e.g., Castles & Coltheart, 2004; Hulme, Snowling, Caravolas, & Carroll, 2005; Muter, Hulme, Snowling, & Stevenson, 2004). Because children with poor phonological awareness can be identified before they begin learning to read (Hulme & Snowling, 2009; Wise & Snyder, 2001), early intervention is possible. Poor phonological awareness has been described as the single best predictor of risk of early reading failure (Uhry, 2005), so early evaluation is essential. The NRP (2000) identified phonemic awareness as one of the five key components to effective reading instruction. The most important early phonological ability for reading is *blending* (the ability to push together sounds), whereas the most important early ability for spelling is *segmentation* (the ability to break apart the speech sounds in a word). Explicit, sequenced, multisensory instruction at the appropriate level, delivered by highly trained teachers to groups of six or fewer, appears most effective for increasing phonological awareness (Wise & Snyder, 2001). Individuals with limited phonological processing should be exposed to a language-rich environment that includes daily practice with sounds, words, and language. Generally, phonemic awareness instruction should begin with easier tasks such as rhyming, and move to more complex tasks such as segmenting and manipulating sounds (e.g., Anthony & Francis, 2005; Kilpatrick, 2015). Reading aloud to individuals with a weakness in phonological processing and providing audio books can be two beneficial accommodations.

Once a student has learned to blend and segment sounds orally, explicit, systematic, synthetic phonics instruction is needed. This type of instruction involves the direct teaching of the relationships among the phonemes (speech sounds) and the graphemes (letters and letter strings that represent the phonemes). For students who struggle with reading, direct instruction in phoneme–

grapheme connections results in improved word reading (Jenkins & O'Connor, 2001; NRP, 2000).

Short-Term Memory and Working Memory

Two types of memory are discussed briefly in this section: *short-term memory,* or *memory span,* and *working memory.* The relationship between short-term memory and working memory has been described in three different ways: (1) the two as similar constructs; (2) working memory as a subset of short-term memory; and (3) short-term memory as a subset of working memory (Engle, Tuholski, Laughlin, & Conway, 1999). For purposes of this discussion, we address these constructs as being related but distinct.

Short-term memory is a limited-capacity system that requires apprehending and holding information in immediate awareness. It is more narrowly defined as consisting of tasks requiring storage but not a great deal of processing (Gathercole & Alloway, 2008). Most adults can hold seven pieces of information (plus or minus two) at one time. Short-term memory can be thought of as the "use it or lose it" memory. When new information requires a person's short-term memory, the previous information held is either stored or discarded. Common short-term auditory memory span tasks include sentence repetition tasks and repeating digits or words in a serial order. Research has documented the importance of memory span to achievement (Carroll et al., 2016; Flanagan et al., 2013), as well as to the development of verbal abilities (Engle et al., 1999). Memory span also appears to be significantly related to reading recognition (Swanson & Zheng, 2013); basic writing skills, particularly spelling (Berninger, 1996; Lehto, 1996); and math problem solving (Geary, 1993, 2007).

Working memory has been described as a brain-based function in which plans can be retained temporarily as they are being formed, transformed, or executed (Miller, Galanter, & Pribram, 1960). Similarly, Baddeley (1990) described working memory as a system for temporarily storing and manipulating information while executing complex cognitive tasks that involve learning, reasoning, and comprehension. Jensen (1998) described it as the "mind's scratchpad." More recently, working memory has been described as a broader, more complex construct that refers to a dynamic system for temporary storage and manipulation of information and includes the attentional control processes (Unsworth, 2016). An example of a common working memory task is listening to numbers in a forward sequence and then restating the numbers in a reversed order. Working memory shows a strong connection to fluid intelligence and reasoning ability (Chuderski & Necka, 2012; Kane & Engle, 2002; Kyllonen & Christal, 1990), whereas memory span does not (Engle et al., 1999).

Strong connections exist between working memory and most areas of academic performance, making working memory useful in identifying individuals at risk for many types of learning problems (Gathercole & Alloway, 2008). As examples, significant correlations have been found between working memory and reading comprehension (e.g., Berninger, Raskind, Richards, Abbott, & Stock, 2008; Cain, Oakhill, & Bryant, 2004; Cooper, 2006; Fletcher et al., 2007; Shaywitz & Shaywitz, 2008), language comprehension (King & Just, 1991), vocabulary acquisition (Daneman & Green, 1986; Gathercole & Baddeley, 1993), spelling (Ormrod & Cochran, 1988), math computation (Andersson, 2008; Ashcraft & Kirk, 2001; Passolunghi, Mammarella, & Altoè, 2008; Swanson & Jerman, 2006), and math problem solving (Fuchs et al., 2008; Geary, 2007; Logie et al., 1994; Peng et al., 2018). Children who have both reading and math disabilities often have difficulty on tasks involving working memory (Evans, Floyd, McGrew, & Leforgee, 2002; Floyd et al., 2003; Reid, Hresko, & Swanson, 1996; Szucs, 2016; Wilson & Swanson, 2001), as do children who only have difficulties in math. Szucs (2016) found that individuals with reading and math difficulties had verbal working memory deficits, whereas those with difficulty in math only had visual–spatial working memory deficits. Working memory deficits have been identified as the primary characteristic of individuals with a math disability (e.g., Bull, Espy, & Wiebe, 2008; Chong & Siegel, 2008; Geary, 2004; Peng et al., 2018; Rotzer et al., 2009).

Reasons for Differences in Performance

Several factors can influence performance on working memory tasks. Working memory has several different facets: the capacity of primary memory, the small set of items a person is currently working on; attentional control abilities, the set of attentional processes that aid in the ability to maintain focus in the presence of internal and external distractions; and secondary memory abilities, the ability to retrieve and reactivate information that could not be actively maintained in primary memory (Unsworth, 2016). Thus low performance on working memory tasks can arise from a variety of reasons.

If an individual lacks automaticity or efficiency in performing a particular task, or has poor attentional control, performance on memory tasks may be impaired. Language proficiency is a factor for some types of memory span tasks, such as sentence repetition. Knowledge of syntax and vocabulary helps facilitate performance on sentence repetition tasks, placing individuals with different or limited linguistic backgrounds at a distinct disadvantage. Because most memory tasks present the stimulus briefly and only once, performance can also be affected by attention or anxiety. For example, individuals with high math anxiety demonstrate smaller working memory spans when performing math-related tasks (Ashcraft & Kirk, 2001), and stress in general has a negative impact on working memory capacity (Klein & Boals, 2001). Moreover, attention and working memory are closely related; this has been substantiated by a meta-analysis of 26 studies, which concluded that individuals with attentional problems also manifested limits in working memory (Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005). In one study, over 74% of individuals diagnosed with attention-deficit/hyperactivity disorder (ADHD) were found to have working memory deficits (Brown, Reichel, & Quinlan, 2009).

Implications for Achievement

Memory deficits are characteristic of individuals with SLD (Gathercole, Lamont, & Alloway, 2006; Swanson & Zheng, 2013). Individuals with limited memory abilities may (1) appear inattentive, (2) have difficulty following directions or recalling sequences (e.g., months of the year), (3) have trouble memorizing factual information, (4) have difficulty following a lecture or a class discussion, (5) have trouble taking notes, or (6) struggle to comprehend what has been stated or read. For reading comprehension to occur, an individual must decode the words to obtain meaning. If decoding is labored, then fluency is reduced, and greater demands are placed on working memory, diminishing comprehension.

In math, weaknesses in short-term or working memory may contribute to difficulties in retrieving basic facts or solving algorithms. Individuals with math disabilities appear to have difficulty holding information in their minds while completing other processes. Both verbal short-term working memory and visual–spatial short-term working memory have an impact on math performance (Szucs, 2016). Individuals with specific math disabilities often have difficulty in attentional control or logical problem solving (Geary, 2013). These individuals may understand the rules, but forget the numerical information or have trouble following the steps of an algorithm in order. They know fewer facts and forget them more quickly than other children do. Difficulties in learning basic number facts do not necessarily mean that a person has poor memory. Limited knowledge can also result from insufficient exposure, poor instruction, or attentional weaknesses, rather than specific math disabilities (Robinson et al., 2002).

In contrast, above-average performance on memory tasks can indicate good attention. If information can be dealt with quickly, then the limited-capacity system of short-term memory will not be overloaded, and more attention can be directed to higher-level tasks. Good working memory facilitates proficiency in higher-level abilities, such as reading comprehension, math problem solving, and written expression.

Interventions for Limited Short-Term Memory or Working Memory

Individuals with limited short-term memory or working memory often benefit from practice, review, and specific instruction in memory strategies. For example, the use of chunking strategies, mnemonics, and verbal rehearsal of information can help improve performance. The more routines are practiced, the more automatic these tasks become. Automaticity is especially important for activities that require rapid, efficient responses, such as pronouncing words or responding quickly to math facts. For individuals with memory difficulties, explicit instruction in the academic area of concern is essential. For example, in mathematics, explicit, direct instruction of core numerical relations is beneficial (Geary, 2013).

As a basic principle, students who struggle with memory are helped and supported by effective instruction. For all areas of achievement, teachers should review prerequisite information and previously learned skills, provide distributed practice over time, and introduce new skills carefully and systematically. Validated instructional techniques to improve academic performance include (1) providing demonstration and modeling of the skill to be learned, using a think-aloud procedure; (2) incorporating guided practice with immediate corrective feedback; (3) requiring independent practice to promote mastery; (4) setting goals; and (5) monitoring progress.

At times, accommodations may be necessary. To accommodate individuals with memory difficulties, oral directions need to be short—or, better yet, written down. In addition, oral instructions can be supported with visual cues, such as demonstrations, pictures, or graphic representations. Accommodations for memory difficulties often involve reducing the amount of information that must be memorized. For example, a teacher can provide a student with a fact chart or calculator, rather than requiring memorization of math facts; prepare a specific study guide that highlights the information to be learned; or have the student maintain a personal dictionary of words the student commonly misspells.

Some students will require specific accommodations, such as the use of audio books, permission to record lectures, and/or the provision of lecture notes. In addition, individuals who struggle with tasks involving memory need to understand how their difficulties with memory affect their learning, so that they can request specific accommodations when needed.

Long-Term Storage and Retrieval and RAN

Long-term storage and retrieval constitutes another type of memory process that involves associative memory or the process of storing and retrieving information. This aspect of memory would be similar to what Unsworth (2016) has referred to as *secondary memory*. Problems with this process can influence how effectively new information is stored, as well as how efficiently it is retrieved. Long-term storage and retrieval is not to be confused with the actual information being stored or recalled, which is considered to be crystallized or verbal intelligence. Word-finding difficulties (discussed below) are related to problems with the retrieval process.

Associative memory, a narrow ability of long-term storage and retrieval, appears to be an important ability at the early stages of reading (Cormier et al., 2017; Evans et al., 2002; Flanagan et al., 2013; McGrew & Wendling, 2010), as well as math development (Floyd et al., 2003; McGrew & Wendling, 2010). Acquisition of basic reading or math skills requires the individual to associate pairs of information, such as phonemes (speech sounds) and graphemes (a letter or letters that represent the speech sound), and store this information for later use. This ability to form, store, and retrieve sounds and symbols, as well as to store and retrieve lexical knowledge, is important to early reading development (Cooper, 2006; Hammill, 2004; Perfetti, 2007). The acquisition of alphabetic knowledge (phoneme–grapheme correspondence) can be described as a visual–verbal paired-associate learning task (Hulme, 1981; Manis, Seidenberg, Stallings, et al., 1999). Research indicates that paired-associate learning (PAL) accounts for unique variance in reading, independent of the influence of phonological awareness (Hulme, Goetz, Gooch, Adams, & Snowling, 2007; Windfuhr & Snowling, 2001). These findings suggest that difficulties in storing and recalling associations may affect the development of early reading skills. Warmington and Hulme (2012) explain that "recent research suggests that visual-verbal PAL may be a unique cross-modal associative learning mechanism that is specific to the creation of mappings between visual (orthographic) and phonological stimuli" (p. 46).

Both letter sound knowledge and letter name knowledge have also been identified as strong predictors of reading attainment (Adams, 1990; Muter, Hulme, Snowling, & Taylor, 1997). These aspects of literacy development require the ability to form associations between visual and verbal representations, store those associations, and retrieve them later as needed. Knowledge of the letter patterns, or orthography of a language, plays an important role in developing reading skill (Perfetti, 2011). In addition, several studies have reported that individuals with dyslexia have difficulties associating verbal labels with visual stimuli (Holmes & Castles, 2001; Vellutino, 1995).

The same basic memory problem that results in common features of reading disabilities, such as difficulties in retaining phoneme–grapheme correspondences and retrieving words from memory, may also contribute to the fact retrieval problems of many children with math disabilities. Conceivably, a weakness in the long-term storage and retrieval process is a core difficulty that helps explain the high comorbidity of reading and math disabilities (Robinson et al., 2002). Geary (2007) has hypothesized that individuals with both reading and math disabilities have a common memory problem that affects decoding and math fact learning. Associative memory may be that common memory deficit.

Naming facility, another narrow ability of long-term storage and retrieval, has also been identified as a key predictor of early reading achievement (e.g., Berninger et al., 2006; Scarborough, 1998; Wolf et al., 2000). Carroll (1993) classified naming facility as a narrow ability of long-term storage and retrieval that is sometimes referred to as the *speed of lexical access,* or the efficiency with which individuals retrieve and pronounce letters or words.

As noted earlier, this type of naming facility has been referred to as RAN (Denckla & Rudel, 1974). On RAN tasks, a person is typically shown a randomized array of several objects, colors, letters, or digits (6–8 in a row, with a total of 30–50), and is asked to name the stimuli as quickly as possible. Unlike other long-term retrieval tasks, these measures are timed, and the person must name the symbols as quickly as possible. Although RAN has been the focus of extensive research in recent years, use of this type of assessment began with the original work of Geschwind (1965) and Denckla and Rudel (1974). Since these early reports, results from many studies have demonstrated a connection between poor RAN and poor reading skill (e.g., Hammill, 2004; Perfetti, 1994; Torgesen et al., 1999; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993; Wolf, 2007; Wolf & Bowers, 1999). Phonemic awareness and RAN appear to account for independent variance in later reading scores and relate to distinct aspects of reading development (Manis, Seidenberg, & Doi, 1999). RAN is more strongly related to reading fluency than to the reading of isolated words (Protopapas, Altani, & Georgiou, 2013).

To attempt to refine explanations of reading failure, Wolf and Bowers (1999) proposed a theory referred to as the *double-deficit hypothesis.* According to this theory, three major subtypes of poor readers exist: (1) ones with phonological deficits, (2) ones with naming speed deficits, and (3) ones with a combination of the two. Wolf and Bowers have hypothesized that RAN tasks tap nonphonological skills related to reading, such as the processes involved in the serial scanning of print. Presumably, children who are slow to name symbols are slower to form orthographic representations of words (Bowers, Sunseth, & Golden, 1999)—abilities related to the visual aspects of reading. If common letter patterns are not recognized easily and quickly, orthographic pattern knowledge, and subsequently reading rate, will be slow to develop (Bowers & Wolf, 1993; Kilpatrick, 2015).

Some evidence also suggests that RAN differentially predicts reading, based on the level of reading skill. For example, Meyer, Wood, Hart, and Felton (1998) found that RAN tasks had predictive power only for poor readers. Manis, Seidenberg, and Doi (1999) and Abu-Hamour (2009) summarized what existing research suggests about RAN: (1) RAN appears to be independent of phonology and to contribute independent variance to word identification and comprehension; (2) its independent contribution appears larger with younger children and individuals with more severe reading disabilities; (3) RAN is more closely related to reading irregular words than to reading phonically regular nonsense words; (4) it appears to be more closely related to tasks involving orthography than to tasks involving phonology; and (5) RAN is related to both the accuracy and speed of reading words, but the relationship is stronger with speeded measures. In addition, pause time is significantly correlated with both reading accuracy and reading fluency measures, whereas articulation time is not (Georgiou, Parrila, & Kirby, 2006; Georgiou, Parrila, Kirby, & Stephenson, 2008). Thus RAN tasks seem to be measuring the speed in which an individual can retrieve and name visual symbols.

Reasons for Differences in Performance

As with other measures of memory, tasks measuring long-term storage and retrieval may be affected by attention or anxiety. Individuals with math disabilities have difficulty learning basic facts and then, once facts are stored, have difficulty accessing them (Geary, 1993; Miller & Mercer, 1997); these problems suggest difficulties in the storage and retrieval process, which appear to be similar to the word retrieval difficulties common in individuals with reading disabilities. Another problem noted in the retrieval process is the inability to inhibit the recall of related but unnecessary information when one is trying to retrieve a specific answer. For example, an individual may recall that 9 is the answer to 4 + 5, but may also recall 6, the number following the 4-5 sequence, or 20, the product of 4 × 5. Thinking of these extraneous facts slows down the process of getting to the correct answer and increases the chance for error.

Word retrieval difficulties can also impede the effortless retrieval of numbers, letters, and words. German (2001) described three types of word-finding errors as "slip of the tongue" (recalling the wrong word), "tip of the tongue" (unable to recall the word), and "twist of the tongue" (mispronouncing the target word). An individual who manifests word-finding difficulties is not necessarily lacking "knowledge," but may be unable to retrieve and express that knowledge on demand. Higbee (1993) distinguished between *available* and *accessible* information. Available information is known and stored; accessible information is available information that is retrievable. When known information cannot be recalled, a word-finding difficulty is present.

Like word retrieval difficulties, differences in performance on RAN tasks can be attributed to a variety of cognitive and linguistic processes. Wolf and colleagues (2000) describe serial naming speed as similar to reading because it involves a "combination of rapid, serial processing, and integration of attention, perceptual, conceptual, lexical, and motoric subprocesses" (p. 393). A person may have slow naming speed because of any one, or several of, the multiple processes underlying these tasks. Students with reading disabilities as well as those with ADHD can have impairments in RAN (de Jong, Licht, Sergeant, & Oosterlaan, (2012).

Morris and colleagues (1998) have described this specific subtype of reading disability as a "rate deficit." Students are impaired on tasks requiring rapid serial naming, but not on measures of phonological awareness. Conceivably, rapid sequential processing is common to naming speed, processing speed, and reading speed, and slow naming speed reflects a global deficit in the rapid execution of a variety of cognitive and linguistic processes (Kail, Hall, & Caskey, 1999). Whatever RAN measures, it may be partially subsumed under the rubric of processing speed (Denckla & Cutting, 1999).

Implications for Achievement

High performance on associative storage and retrieval tasks suggests that an individual will be successful in learning new information and recalling stored information. Long-term storage and retrieval help an individual store, retrieve, and demonstrate knowledge. Low performance on tasks measuring this ability suggests that the individual will experience difficulty storing new information and recalling previously learned information. These individuals may have difficulty acquiring phoneme–grapheme knowledge, memorizing math facts, and completing fill-in-the-blank tests that require the precise recall of specific information.

Presently, more is known about RAN than about other associative memory abilities. The best predictive measures of early reading achievement appear to be a combination of letter identification, phonological awareness, and RAN (Adams, 1990; Bishop, 2003; NRP, 2000). In addition, children with weaknesses in both RAN and phonemic awareness appear to be the most resistant to reading intervention (Wolf & Bowers, 1999). Results from one study indicated that low RAN scores were the single best predictor of treatment resistance among second-grade students (Vaughn, Linan-Thompson, & Hickman, 2003). Children with only naming speed deficits (no weaknesses in phonological awareness) are characterized by problems in word identification, fluency, and comprehension (Wolf et al., 2000). Although future research is likely to confirm the exact processes involved in RAN tasks, students with naming deficits appear to have a poorer prognosis for reading success than do other subgroups (Korhonen, 1991). Denckla (1979) described these students as a "hard-to-learn" group. Naming speed deficits persist into adolescence and adulthood (e.g., Denckla & Rudel, 1974; Vukovic, Wilson, & Nash, 2004), making them an important marker for identifying reading problems in older individuals.

The double-deficit theory attempts to explain two cognitive correlates of reading failure—poor phonological awareness and slow naming speed—but these are not the only tasks that differentiate poor readers from good readers. For example, Ackerman, Holloway, Youngdahl, and Dykman (2001) found that poor readers scored lower than typically achieving peers on orthographic tasks, attention ratings, and arithmetic achievement. Wolf (1999) also acknowledged the importance of using multidimensional models for explaining reading difficulties, stating that

> this new conceptualization of reading disabilities was ironically, named too quickly. To be sure, double deficit captures the phenomenon of study—that is, the importance of understanding the separate and combined effects of two core deficits—but it fails miserably in redirecting our simultaneous attention as a field to the entire profile of strengths and limitations manifest in children with reading disabilities. Only when we develop truly multi-dimensional models of deficits and strengths will our diagnostic and remedial efforts be best matched to individual children. (p. 23)

Thus more attention has recently been paid to identifying the multiple cognitive deficits that can be found in individuals with reading disabilities (Carroll et al., 2016; McGrath et al., 2011; Pennington, 2006). The research on which and how many cognitive abilities are involved is critical for informing reliable diagnoses of dyslexia (Tamboer, Vorst, & Oort, 2016). In addition, reading disabilities and math disabilities have been associated with different cognitive profiles: In one study, whereas reading was more associated with slow RAN and poor verbal memory, mathematics was more associated with poor temporal processing and visual–spatial memory (Moll, Gobel, Gooch, Landerl, & Snowling, 2016).

Interventions for Deficits in Long-Term Storage and Retrieval and RAN

Individuals with difficulties in associative memory and retrieval will require repeated opportunities and more practice to learn new information. Carroll (1989) suggested that the degree of learning results from the ratio of the time spent learning to the amount of time that was needed for learning. In other words, students who have trouble retaining associations require more time to learn. The strategies that may be most useful include limiting the amount of information presented at one time, and using multisensory and meaning-based instructional approaches that help a student make connections and retain new learning. Examples of approaches include verbal rehearsal, active learning, use of manipulatives, and real-life projects. Smith and Rivera (1998) found that demonstration plus a permanent model was an effective strategy for helping children master computational mathematics, especially learning math skills and organizing and remembering the sequences of multistep algorithms. In addition, techniques that activate the emotional center of the brain by using humor, dramatizations, or movement can enhance learning (Leamnson, 2000). The most effective strategies help a learner to form associations between new and learned information by activating prior knowledge, so that the learning of new information occurs in the context of what the learner already knows (Marzano, 1992).

Another helpful method to facilitate recall is instruction in the use of mnemonic strategies. For example, the *keyword* method involves associating new words with visual images, to help students recall word meanings and learn new vocabulary (Mastropieri, 1988). Three steps are used: *recoding, relating,* and *retrieving.* For recoding, students change the new vocabulary word into a known word, the keyword, which has a similar sound and is easily pictured (e.g., the word *apex* changed to *ape*). For relating, students associate the keyword with the definition of the new vocabulary word through a mental image or a sentence (e.g., a picture of an ape sitting on top of a mountain). For retrieving, students think of the keyword, remember the association, and then retrieve the definition (e.g., the tip or summit). A more specific program that addresses the challenges imposed by RAN deficits is RAVE-O (Reading through Automaticity, Vocabulary, Engagement, and Orthography) (Wolf, 2010). This program emphasizes expansion of vocabulary and building fluency through rapid recognition of the most frequent orthographic patterns in the language.

Visual Processing

Visual–spatial tasks, because they are inherently less verbal in nature, are commonly included in intelligence tests. Carroll (1993) described broad *visual–spatial ability* as including the narrow abilities of *spatial relations, visualization, visual memory, closure speed, spatial scanning,* and a number of others that are not typically included on intelligence tests. Thus a wide range of these abilities exists, and they emphasize the processes of image generation, storage, retrieval, and transformation (Lohman, 1994).

Results from current research do not indicate a strong relationship between visual–spatial abilities and reading or writing (Ackerman et al., 2001; Cormier et al., 2016, 2017; McGrew & Wendling, 2010; Nation et al., 2002; Swanson & Berninger, 1995). This is not to say that such abilities are unimportant to academic success. Clearly, spelling some words involves an ability to retrieve a visual orthographic image of the word to spell it correctly, but the types of visual–spatial tasks on intelligence measures have little relationship with spelling competence (Liberman, Rubin, Duques, & Carlisle, 1985; Sweeney & Rourke, 1985; Vellutino, 1979). This lack of correlation may be due to the types of visual–spatial tasks traditionally included on intelligence tests, such as manipulating patterns, assembling objects, or noting visual details in pictures, that differ from the visual–orthographic processing abilities required for efficient reading and spelling. Visual–spatial abilities are often three-dimensional in nature, and should therefore not be confused with the orthographic processing that includes the visual representations and spelling patterns of the writing system (Berninger, 1990). Orthographic processing is the awareness of the visual representations of language and is one of the two major skills required for printed word recognition; the other being phonological processing (Mather & Wendling, 2012). Orthographic processing is often measured through the reading and spelling of exception or irregular words.

Visual–spatial thinking abilities do appear, however, to be related to math achievement (De Cruz, 2012; Geary, 1994, 2007; Hegarty & Kozhevnikov, 1999; Lubinski, 2010; Moll et al., 2016; Rourke, 1993; Strawser & Miller, 2001). Estimation skills, representations of magnitude, and visualizing a mental number line are dependent on visual–spatial systems (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Zorzi, Priftis, & Umiltá, 2002). Visual processing is frequently a weakness in individuals with math disabilities (Hale et al., 2008; McLean & Hitch, 1999; Proctor et al., 2005), and difficulties in visual memory and visual–spatial working memory have been noted (Fletcher, 1985; McLean & Hitch, 1999; Moll et al., 2016). Geary (1993) identified a visual–spatial disorder subtype of math disability, characterized by difficulties with spatial representations (e.g., alignment) as well as place value errors. Several researchers have found a relationship between math and specific spatial abilities (Assel, Landry, Swank, Smith, & Steelman, 2003; Osmon, Smertz, Braun, & Plambeck, 2006). Still others have suggested that visual–spatial abilities are related to performance on higher-level mathematics, but not to basic math skills (Flanagan et al., 2013). The results from research that explores the relationship between spatial ability and STEM domains (science, technology, engineering, and mathematics) indicates that spatial ability has importance above and beyond the influence of mathematical and verbal abilities in identifying individuals with aptitude for STEM (Wai, Lubinski, & Benbow, 2009).

Reasons for Differences in Performance

Visual processing tasks can measure an array of narrow abilities, so identifying a person's specific weakness(es) is an important prerequisite to understanding performance. For example, an individual may have strengths in visual memory of objects, but weaknesses in spatial relations. Other factors, such as speed of performance on timed visual–spatial tasks, attention, motivation, and working memory, can influence performance. Problems in visual–motor coordination can also affect performance on timed visual–spatial tasks that involve the use of manipulatives, such as moving and assembling blocks or puzzle pieces, or drawing with a pencil.

Implications for Achievement

Except for math, research has not documented a significant relationship between visual processing and reading or writing achievement. Therefore, it would be erroneous to conclude that a student with high scores on visual–spatial tasks would benefit from a sight word approach to teaching reading, or that a student with low scores would benefit from a phonics approach to reading instruction. Many individuals, including those in clinical groups, demonstrate average scores on visual–spatial tasks with simultaneous low achievement. For example, visual processing scores do not differentiate between college students with and without SLD (McGrew & Woodcock, 2001). In many children with reading disabilities, visual–spatial skills are better developed than other abilities (Fletcher et al., 1995). Furthermore, in a review of over 6,000 clinical cases representing 21 different clinical groups, visual processing was generally not impaired (Schrank, Miller, Wendling, & Woodcock, 2010). For example, individuals with autism were found to have little difficulty with visual–spatial tasks, which is consistent with other research (Corbett, Carmean, & Fein, 2009). These findings provide further evidence that visual processing abilities remain relatively intact in clinical groups, and therefore are not good predictors of academic performance in reading or writing.

Interventions for Limited Visual Processing

Research indicates that spatial reasoning skills can be improved with training (Uttal et al., 2013). Although more research is needed on whether improved spatial skills lead to improved achievement, some studies indicate that this is the case. For example, one study of 6- to 8-year-olds found improvement in calculation accuracy immediately after 40 minutes of spatial training (Cheng & Mix, 2014). Another study found improved math achievement in first graders after 6 months of pattern training (Kidd et al., 2013).

In general, students with visual processing deficits benefit when interventions are highly concrete and as verbal as possible. The most effective methods are highly structured and provide external guidance—methods employed in explicit instruction. Expectations may need to be simplified, broken down, or modified. Teaching a student how to use self-talk to reinforce routines or procedures can help with the completion of simple as well as more complex tasks. Rourke (1995) recommended using a "part-to-whole" verbal teaching approach by presenting information in a logical sequence, one step at a time, so that the student can pay attention to details.

Teaching specific learning strategies, such as the use of imagery, graphic organizers, and puzzles, may significantly improve less skilled individuals' performances on visual–spatial tasks. Another strategy, verbal labeling, uses language to describe visual forms as they are manipulated and represented spatially (Kibel, 1992). For individuals with strengths in visual–spatial abilities, teachers may enhance the students' performance by instruction with pictures, diagrams, or graphic organizers. These individuals often excel in tasks such as visualizing and drawing three-dimensional objects.

Processing Speed

Schneider and McGrew (2012) have defined *processing speed* "as the ability to perform simple, repetitive cognitive tasks quickly and fluently" (p. 119). They indicate that once a person knows how to do a task, processing speed becomes an important predictor of skilled performance. From an information-processing perspective, speediness and automaticity of processing underlie efficient performance (Kail, 1991; Lohman, 1989). Processing information quickly frees up limited resources so that higher-level thinking can occur. Slow processing speed may be characterized as a domain-general deficit because it underlies performance in many areas and is not specific to any one area or disability. For example, slow processing speed characterizes individuals with dyslexia, math disabilities, as well as those with ADHD (Eden & Vaidya, 2008; Peng et al., 2018), although children with reading disabilities appear to have more significant weaknesses than those with ADHD (Shanahan et al., 2006).

Perceptual speed is a narrow ability of processing speed; Carroll (1993) described it as the ability to search for and compare visual symbols. Perceptual speed is strongly related to reading achievement (Cormier et al., 2017; McGrew, Flanagan, Keith, & Vanderwood, 1997; McGrew & Wendling, 2010), math achievement (Fiorello & Primerano, 2005; Floyd et al., 2003; McGrew & Hessler, 1995; McGrew & Wendling, 2010; Peng et al., 2018), and writing achievement (Cormier et al., 2016; Floyd, McGrew, & Evans, 2008; McGrew & Knopik, 1993; Williams, Zolten, Rickert, Spence, & Ashcraft, 1993). Thus the ability to process symbols rapidly is strongly related to academic performance and success.

Speed of processing has been identified as a primary process in reading (Joshi & Aaron, 2000), and many researchers have emphasized the importance of speed constructs in early reading acquisition (e.g., Berninger et al., 2006; Kintsch & Rawson, 2005; Shaywitz et al., 2008; Torgesen et al., 1999). In studies investigating the differences between typically achieving readers and those with reading disabilities, processing speed was slower for the students with reading disabilities (Kruk & Willows, 2001; Pammer, Lavis, Hansen, & Cornelissen, 2004) on both linguistic and nonlinguistic tasks (Shanahan et al., 2006). Research results have indicated a relationship between perceptual speed and word reading (e.g., Apel, 2009; Berninger, 1990; Urso, 2008), as well as basic writing skills and composition (Berninger, 2009; Cormier et al., 2016; Floyd et al., 2008).

In the area of math, processing speed was found to be the best predictor of arithmetic competence in 7-year-olds (Bull & Johnston, 1997). In addition to reading, various researchers have identified speed-related issues for individuals with math disabilities: counting speed (Geary, 1993, 2007), numerical processing fluency (Swanson & Jerman, 2006), and efficiency in executing simple cognitive tasks during math fact tasks (Fuchs et al., 2006).

Reasons for Differences in Performance

When an evaluator is examining a person's performance on processing speed tasks, several additional factors need to be considered. Because most tasks used to measure processing speed are visual in nature (often involving rapid searching of symbols or shapes), the individual's vision and visual processing may be a factor. Motivation and attention are factors to consider as well. Processing speed tasks are typically timed and clerical in nature. Some individuals may have difficulty maintaining attention, and some may not be motivated to complete a relatively simple task quickly. Personality style can also affect performance on speeded tasks, as can cultural differences. Some cultures do not value speeded performance as an important behavioral attribute. Age can also influence performance; some younger students do not understand the concept of working quickly for a certain amount of time, whereas older adults may not care. In general, individuals who are reflective will work more slowly, carefully reviewing their options before responding. Some gifted individuals exhibit a relative weakness on speeded tasks because they reflect and check answers before making a decision. In contrast, individuals who are impulsive may work quickly and carelessly.

Implications for Achievement

Limited processing speed suggests that a person may process information slowly, thus creating a "bottleneck" that affects new learning or skilled performance on tasks that are known. When information is well known, it can be processed more automatically; when information is new, the processing is more effortful. However, even after a task is learned, individuals can differ in the speed and fluency with which they perform those tasks (Ackerman, 1987). Schneider and Shiffrin (1977) made a distinction between *automatic* and *conceptual* processing. The automatic processes require little attentional resources, whereas the conceptual processes are controlled and require the application of knowledge and strategies. Processing speed appears to be most closely related to the lower-order academic tasks that become increasingly automatic with repeated practice, such as reading words quickly, knowing multiplication facts, or spelling words with accuracy. Two consistent findings have emerged from the research on individuals with SLD: (1) Individuals both with and without SLD exhibit a range of responses on a variety of speeded tasks, and the intercorrelations between different speeded tasks often differ for both groups; and (2) individuals with SLD typically obtain lower scores than typically achieving individuals on a variety of speeded tasks (Ofiesh, Mather, & Russell, 2005).

Thus the issue of extended time has particular relevance for students with SLD. Weis, Dean, and Osborne (2016) have suggested that when making a recommendation for extended time, evaluators should include "(a) history of reading disability, special education, or previous reading accommodations, (b) a current diagnosis of reading disability or disorder, (c) objective evidence of current limitations in reading, and (d) test data suggesting the need for additional time" (p. 486). A further consideration is whether an exam is attempting to measure an individual's knowledge and/or the speed in which they can demonstrate this knowledge. In considering the provision of time accommodations on exams, Kelman and Lester (1997) advised educational authorities to consider whether or not speed is a genuine academic virtue in the particular context. If not, the test should be untimed. In the very few cases where speed is judged to be necessary, no one should be provided with accommodations. As noted by Weis and colleagues (2016), it seems desirable to remove construct-irrelevant barriers for all students, not just those with SLD.

Interventions for Limited Processing Speed

Individuals with limited processing speed often require specific accommodations, particularly when an academic area is compromised as well. These individuals may need extended time, as well as shortened directions and assignments. In addition, copying activities should be limited or eliminated. If slow processing of language is also involved, it may be necessary to increase "wait" time so that an individual has more time to think and respond. In order to suggest appropriate interventions or accommodations, the evaluator must first determine whether the person's processing speed deficit is due to limits in speed, accuracy, or both. For example, some individuals work slowly but accurately, whereas others may work quickly but inaccurately. If performance is slow and accurate, a student may benefit from extra time or shorter assignments. If the student works quickly but makes numerous errors, extra time may not be appropriate or beneficial.

Use of explicit instruction in the relevant academic area is also recommended. Often instructional interventions designed to increase rate and fluency in the academic area of concern are beneficial. For example, repeated practice to build automaticity, speed drills, and computer programs that focus on increasing rate or making decisions quickly may improve the performance of individuals with slow processing speed (Klingberg, 2009; Mahncke, Bronstone, & Merzenich, 2006).

Fluid Reasoning

Fluid reasoning involves the ability to solve novel problems via inductive or deductive reasoning and to transfer or generalize learning. Intelligence tests typically include fluid reasoning tasks, such as matrices, sequences, or analogies. Research has documented the relationship between fluid reasoning and reading comprehension (Cormier et al., 2017; Floyd et al., 2006; McGrew, 1993; Nation et al., 2002), math achievement (Fiorello & Primerano, 2005; Flanagan et al., 2013; Floyd et al., 2003; Fuchs et al., 2006; Geary, 1993, 2007; Hale et al., 2007; Proctor, 2012; Rourke & Conway, 1997), and writing achievement (Cormier et al., 2016; Floyd et al., 2008).

Some individuals with SLD tend to have difficulties in abstracting principles from experiences (Geary, 1993; Swanson, 1987), and some appear to struggle with making generalizations (Ackerman & Dykman, 1995). Unfortunately, little is known about the breadth, depth, and developmental course of children's generalization capabilities (Pressley & Woloshyn, 1995), but a growing body of research indicates that poor inferential reasoning is one cause of reading comprehension problems (Wise & Snyder, 2001). For individuals with math disabilities, research indicates that fluid reasoning is frequently impaired (Geary, 2007; Proctor et al., 2005). These inferential reasoning difficulties then interfere with an individual's ability to "classify an event as belonging to a category" (Bruner, 1971, p. 93), and thereby affect success at mathematical problem solving. A deficit in fluid reasoning may affect the development of other cognitive abilities, especially in the domain of acquired knowledge (Blair, 2006).

Reasons for Differences in Performance

Performance on fluid reasoning tasks may vary for many reasons. One reason for variation is how effectively a student employs strategies. Results from one study indicated that high achievers are more attentive and use more effective strategies (e.g., talking through a task) that help them learn and practice the task at hand, whereas low achievers use less effective strategies for task completion (e.g., guessing, carelessness, attending to inappropriate contextual clues; Anderson, Brubaker, Alleman-Brooks, & Duffy, 1985). Another reason for differences in performance is mental flexibility, or the ability to shift cognitive gears. Individuals who have mental flexibility are able to anticipate what is expected on a task and change the approach when needed, resulting in more successful outcomes (Kronick, 1988). In contrast, individuals with rigid cognitive styles may be unable to use their knowledge except when the context closely resembles the original learning situation (Westman, 1996). Semrud-Clikeman (2012) explored the role of inattention in individuals with ADHD, combined type, and those with primarily the inattentive type. She found that both groups performed significantly more poorly on fluid reasoning tasks than the control group. Performance can also vary, depending on the type of reasoning task. Some tasks require reasoning with language (e.g., analogies), whereas others require nonverbal problem solving and mathematics (e.g., matrices).

Implications for Achievement

Individuals with limited fluid reasoning may require instruction at a reduced level of difficulty—in other words, a modification to instruction rather than an accommodation. These individuals often experience difficulty with higher-level thinking tasks and may struggle with comprehending what they read, solving math problems, or expressing themselves in writing. They may display rigidity when attempting new things and continue to apply a strategy that does not work. Even after learning a skill, they may not be able to apply that skill in a new context.

On the other hand, individuals with high performance on fluid reasoning tasks are likely to succeed in higher-level thinking tasks, such as those involved in reading comprehension, math reasoning, or written expression. They will typically display mental flexibility when approaching problem-solving tasks, shift strategies to accomplish their goals, and demonstrate self-regulation.

Regarding math instruction, reasoning and seeing patterns are at the heart of mathematics. These abilities help the student make connections between new learning and prior knowledge. In some instances, math instruction is too focused on rules and procedures. Effective instruction must help students develop conceptual understanding. Math concepts and connections must be taught explicitly, and students must be actively engaged to understand and acquire mathematical concepts (Richland, Stigler, & Holyoak, 2012).

Interventions for Limited Fluid Reasoning

Students benefit from opportunities to develop their metacognitive and higher-order thinking skills. Such opportunities may include engaging in reflective discussions about lessons, comparing and contrasting concepts, or using thought journals. Teaching students to use self-questioning techniques, identify main ideas and themes, classify and categorize objects, attend to organizational cues, and implement strategies can lead to significant gains in inferential skills. Specific strategy instruction has proven to be effective in improving the performance and achievement of students with SLD (e.g., Deshler, Ellis, & Lenz, 1996; Pressley & Woloshyn, 1995; Swanson, 2001). This type of instruction appears to be more effective for higher-order, conceptual tasks than for lower-order tasks (Deshler et al., 1996), but the strategies must be taught explicitly and practiced (Klauer, Willmes, & Phye, 2002).

Higher-level thinking skills (e.g., analyzing, comparing, evaluating, synthesizing) require the brain to use multiple and complex systems of retrieval and integration (Lowery, 1998). Experiential learning appears to activate the area of the brain responsible for higher-order thinking (Sousa, 1998). Therefore, instruction that combines physical activities with problem-solving tasks can help connect the motor cortex with the frontal lobes, where thinking occurs, and can thus increase memory and learning (Kandel & Squire, 2000). Learning can be demonstrated in multiple ways, such as dramatizations, experiments, visual displays, music, or inventions. Effective instructional principles include strategies to activate prior knowledge, activities that actively engage learners, and provision of explicit instruction.

CONCLUSION

Cognitive assessment is not only relevant, but essential, for the accurate identification of individuals with SLD. As research continues to increase our knowledge of the relationships among cognitive abilities and achievement, more and more evaluators are taking advantage of that knowledge and applying it to their evaluation practices. Although the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013) subsumes all types of SLD under the category of specific learning disorder, distinct cognitive profiles exist that are relevant to the selection of accommodations and interventions (Poletti, Carretta, Bonvicini, & Giorgi-Rossi, 2018).

Interestingly, the research on cognitive–achievement relationships connects comprehensive evaluations to RTI in a way suggesting that they both provide useful data. Advocates of RTI talk about cognitive markers when discussing the need for early screening to identify children at risk for academic difficulties. For example, phonemic awareness, RAN, and vocabulary are often mentioned as predictors of early reading skill; number sense, processing speed, and fluid reasoning are discussed as predictors of math achievement. When an evaluator is establishing a PSW that suggests the presence of an SLD, the focus is on a student who is experiencing academic difficulties. RTI allows for early identification and intervention. Comprehensive evaluations enable us to understand why a student is struggling or not making sufficient progress. If a student fails to respond to intervention, and the results of a comprehensive evaluation indicate that the student has a processing deficit that affects academic performance, both the definitional criteria for SLD and the SLD eligibility criterion of limited response to evidence-based instruction have been addressed, resulting in a balanced model that promotes diagnostic accuracy (Hale et al., 2006).

In discussing the assessment of intellectual functioning, Wasserman (2003) indicated that one of applied psychology's biggest failures of the last century was that intellectual assessments were not systematically linked to effective interventions. Fortunately, progress is being made to correct this. First, research continues to identify and clarify the cognitive correlates and predictors of achievement. This information provides an early warning system that can lead to timely interventions and prevent or ameliorate future learning problems. Second, most modern intelligence tests measure a broad array of abilities that reflect the findings of current research, streamlining the process of gathering diagnostic information. Third, the principles of effective instruction are known, and educators are responsible for implementing evidence-based instruction. Finally, progress is being made in linking evaluation results to evidence-based interventions. As Monroe noted back in 1935, "In all remedial work, the teacher should start first with the child and then find the appropriate method. Fit the method to the child, not the child to the method" (p. 227). Ultimately, all psychologists and educators share the same goal: to provide the best educational experiences and opportunities for each and every student.

REFERENCES

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research, 67,* 461–502.

Aaron, P. G., & Joshi, R. M. (1992). *Reading problems: Consultation and remediation.* New York: Guilford Press.

Abu-Hamour, B. (2009). *The relationships among cognitive ability measures and irregular word, non-word, and word reading.* Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3–27.

Ackerman, P. T., & Dykman, R. A. (1995). Reading-disabled students with and without comorbid arithmetic disability. *Developmentval Neuropsychology, 11,* 351–371.

Ackerman, P. T., Holloway, C. A., Youngdahl, P. L., & Dykman, R. A. (2001). The double-deficit theory of reading disability does not fit all. *Learning Disabilities Research and Practice, 16,* 152–160.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

American Psychiatric Association. (2013). *Desk reference to the diagnostic criteria from DSM-5.* Washington, DC: American Psychiatric Publishing.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anderson, L. M., Brubaker, N. L., Alleman-Brooks, J., & Duffy, G. S. (1985). A qualitative study of seatwork in first-grade classrooms. *Elementary School Journal, 86,* 123–140.

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology, 100,* 48–66.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255–259.

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly, 31*(1), 42–52.

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive developmental research* (pp. 302–338). New York: Springer-Verlag.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44,* 75–106.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130,* 224–237.

Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., & Menon, V. (2013). Neurobiological underpinnings of math and reading learning disabilities. *Journal of Learning Disabilities, 46,* 549–569.

Assel, M. A., Landry, S. H., Swank, P., Smith, K. E., & Steelman, L. M. (2003). Precursors to mathematical skills: Examining the roles of visual–spatial skills, executive processes, and parenting factors. *Applied Developmental Science, 7,* 27–38.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Boston: Allyn & Bacon.

Beck, I., & McKeown, M. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal, 107,* 251–271.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74,* 506–521.

Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology, 86,* 244–255.

Benbow, C. P., & Lubinski, D. (Eds.). (1996). *Intellectual talent: Psychometric and social issues.* Baltimore: Johns Hopkins University Press.

Berninger, V. W. (1990). Multiple orthographic codes: Key to alternative instructional methodologies for developing orthographic phonological connections underlying word identification. *School Psychology Review, 19,* 518–533.

Berninger, V. W. (1996). *Reading and writing acquisition: A developmental neuropsychological perspective.* Boulder, CO: Westview Press.

Berninger, V. W. (2000). Development of language by hand and its connections to language by ear, mouth, and eye. *Topics in Language Disorders, 20,* 65–84.

Berninger, V. W. (2009). Highlights of programmatic, interdisciplinary research on writing. *Learning Disabilities Research and Practice, 24,* 69–80.

Berninger, V. W., & Abbott, R. D. (1994). Redefining learning disabilities: Moving beyond aptitude–achievement discrepancies to failure to respond to validated treatment protocols. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 163–183). Baltimore: Brookes.

Berninger, V. W., Abbott, R. D., Thomson, J., Wagner, R., Swanson, H. L., Wijsman, E., et al. (2006). Modeling developmental phonological core deficits within a working-memory architecture in children and adults with developmental dyslexia. *Scientific Studies in Reading, 10,* 165–198.

Berninger, V. W., & Graham, S. (1998). Language by hand: A synthesis of a decade of research on handwriting. *Handwriting Review, 12,* 11–25.

Berninger, V. W., Raskind, W., Richards, T., Abbott, R., & Stock, P. (2008). A multidisciplinary approach to understanding developmental dyslexia within working-memory architecture: Genotypes, phenotypes, brain and instruction. *Developmental Neuropsychology, 33,* 707–744.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24,* 323–335.

Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly, 26,* 189–200.

Blair, C. (2006). How similar are fluid cognition and general intelligence?: A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29,* 109–160.

Bowers, P. G., Sunseth, K., & Golden, J. (1999). The route between rapid naming and reading progress. *Scientific Studies of Reading, 3,* 31–53.

Bowers, P. G., & Wolf, M. (1993). Theoretical links between naming speed, precise timing mechanisms, and orthographic skill in dyslexia. *Reading and Writing: An Interdisciplinary Journal, 5,* 69–85.

Brown, T. E., Reichel, P. C., & Quinlan, D. M. (2009). Executive function impairments in high IQ adults with ADHD. *Journal of Attention Disorders, 13,* 161–167.

Bruner, J. S. (1971). *The relevance of education.* New York: Norton.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33,* 205–228.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology, 65,* 1–24.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96,* 671–681.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus, 133,* 59–68.

Carlisle, J. F., & Rice, M. S. (2002). *Improving reading comprehension: Research-based principles and practices.* Baltimore: York Press.

Carroll, J. B. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology* (pp. 43–67). Hillsdale, NJ: Erlbaum.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. M., Solity, J., & Shapiro, L. R. (2016). Predicting dyslexia using prereading skills: The role of sensorimotor and cognitive abilities. *Journal of Child Psychology and Psychiatry, 57,* 750–758.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91,* 77–111.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11.

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33,* 300–317.

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1689–1710.

Clay, M. M. (1982). *Observing the young reader.* Auckland, New Zealand: Heinemann.

Cooper, K. L. (2006). *A componential reading comprehension task for children.* Unpublished doctoral dissertation, University of New England.

Corbett, B. A., Carmean, V., & Fein, D. (2009). Assessment of neuropsychological functioning in autism spectrum disorders. In S. Goldstein, J. A. Naglieri, & S. Ozonoff (Eds.), *Assessment of autism spectrum disorders* (pp. 253–289). New York: Guilford Press.

Cormier, D. C., Bulut, O., McGrew, K. S., & Frison, J. (2016). The role of Cattell–Horn–Carroll (CHC) cognitive abilities in predicting writing achievement during the school age years. *Psychology in the Schools, 53,* 787–803.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relationships between the WJ-IV measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment, 35,* 731–754.

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12,* 145–162.

Cruickshank, W. M. (1977). Least-restrictive placement: Administrative wishful thinking. *Journal of Learning Disabilities, 10,* 193–194.

Cunningham, A. E., Stanovich, K. E., & Wilson, M. R. (1990). Cognitive variation in adult college students differing in reading ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 129–159). San Diego, CA: Academic Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25,* 1–18.

De Cruz, H. (2012). How do spatial representations enhance cognitive numerical processing? *Cognitive Processing, 13,* 137–140.

de Jong, C. G. W., Licht, R., Sergeant, J. A., & Oosterlaan, J. (2012). RD, ADHD, and their comorbidity from a dual route perspective. *Child Neuropsychology, 18,* 467–486.

Decker, S. L., Strait, J. E., Roberts, A. M., & Wright, E. K. (2017). Cognitive mediators of reading comprehension in early development. *Contemporary School Psychology.* [Epub ahead of print]

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284,* 970–974.

Denckla, M. B. (1979). Childhood learning disabilities. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 535–573). New York: Oxford University Press.

Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia, 49,* 29–42.

Denckla, M. B., & Rudel, R. (1974). Rapid automatized naming of pictured objects, colors, letters and numbers by normal children. *Cortex, 10,* 186–202.

Deshler, D. D., Ellis, E. S., & Lenz, B. K. (1996). *Teaching adolescents with learning disabilities: Strategies and methods* (2nd ed.). Denver, CO: Love.

Eden, G. F., & Vaidya, C. (2008). ADHD and dyslexia: Neural basis and treatment. *Annals of the New York Academy of Sciences, 1145,* 316–327.

Ehri, L. C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20,* 19–36.

Elksmin, L. K. (2001). Implementing case method of instruction in special education teacher preparation programs. *Teacher Education and Special Education, 24,* 95–107.

Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Educational Effectiveness, 2,* 1–44.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31,* 246–262.

Feifer, S. G. (2011). How SLD manifests in reading. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 21–41). Hoboken, NJ: Wiley.

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to children with reading problems. *Psychology in the Schools, 43,* 835–853.

Fiorello, C. A., & Primerano, D. P. (2005). Research into practice: Cattell–Horn–Carroll cognitive assessment in practice: Eligibility and program development issues. *Psychology in the Schools, 42,* 525–536.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools, 47,* 739–760.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Fletcher, J. M. (1985). Memory for verbal and nonverbal stimuli in learning disabilities subgroups: Analysis by selective reminding. *Journal of Experimental Child Psychology, 40,* 244–259.

Fletcher, J. M., & Foorman, B. R. (1994). Issues in the definition and measurement of learning disabilities: The need for early intervention. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 185–200). Baltimore: Brookes.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based intervention oriented approach. *Journal of School Psychology, 40,* 27–63.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., et al. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research and Practice, 13,* 186–203.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guilford Press.

Fletcher, J. M., Taylor, H. G., Levin, H. S., & Satz, P. (1995). Neuropsychological and intellectual assessment of children. In H. Kaplan & B. Sadock (Eds.), *Comprehensive textbook of psychiatry* (pp. 581–601). Baltimore: Williams & Wilkins.

Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing, 19,* 427–456.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 60,* 155–171.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell–Horn–Carroll (CHC) cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools, 45,* 132–144.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves study. *Journal of Educational Psychology, 88,* 3–17.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29–43.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100,* 30–47.

Furst, A., & Hitch, G. J. (2000). Separate roles for executive and phonological components in mental arithmetic. *Memory and Cognition, 28,* 774–782.

Gardner, D., & Davies, M. (2014). *Determining the technical vocabulary of academic English: A corpus-based analysis.* Paper presented at the annual conference of the American Association for Corpus Linguistics, Flagstaff, AZ.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.* New York: Basic Books.

Gathercole, S. E., & Alloway, T. P. (2008). Working memory and classroom learning. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K–3 classrooms* (pp. 17–40). New York: Routledge.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition. *European Journal of Psychology, 8,* 259–272.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In S. J. Pickering (Ed.), *Working memory and education* (pp. 220–241). Burlington, MA: Elsevier Press.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70,* 177–194.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC: American Psychological Association.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Geary, D. C. (2007). An evolutionary perspective on learning disabilities in mathematics. *Developmental Neuropsychology, 32,* 471–519.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22,* 23–27.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27,* 398–406.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2011). How SLD manifests in mathematics. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 43–64). Hoboken, NJ: Wiley.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9,* 6–10.

Georgiou, G., Parrila, R., & Kirby, J. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading, 2,* 199–220.

Georgiou, G., Parrila, R., Kirby, J., & Stephenson, K. (2008). Rapid naming components and their relationship with phonological awareness, orthographic knowledge, speed of processing, and different reading outcomes. *Scientific Studies of Reading, 12,* 325–350.

German, D. J. (2001). *It's on the tip of my tongue.* Chicago: Word Finding Materials.

Geschwind, N. (1965). Disconnection syndrome in animals and man (Parts I, II). *Brain, 88,* 237–294, 585–644.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hove, UK: Erlbaum.

Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia, 46,* 3–20.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Gould, B. W. (2001). Curricular strategies for written expression. In A. M. Bain, L. L. Bailet, & L. C. Moats (Eds.), *Written language disorders: Theory into practice* (2nd ed., pp. 185–220), Austin, TX: PRO-ED.

Gregg, N., Davis, M., Coleman, C., Wisenbaker, J., & Hoy, C. (2004). *Implications for accommodation decisions at the postsecondary level.* Unpublished manuscript, University of Georgia.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales—Second Edition (neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools, 45,* 838–858.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16,* 31–35.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified?: A response to special issue authors. *Applied Neuropsychology, 14,* 37–51.

Hale, J. B., Kaufman, A. S., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Response to intervention and cognitive assessment methods. *Psychology in the Schools, 43,* 753–770.

Hallahan, D. P., & Cohen, S. B. (2008). Many students with learning disabilities are not receiving *special* education. *Learning disabilities: A Multidisciplinary Journal, 15,* 3–9.

Hammill, D. (2004). What we know about the correlates of reading. *Exceptional Children, 70,* 453–468.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*(1), 103–128.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Brookes.

Hauerwas, L. B., Brown, R., & Scott, A. S. (2013). Specific learning disability and response to intervention: State-level guidance. *Exceptional Children, 80,* 101–120.

Hecht, S., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86,* 277–302.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem-solving. *Journal of Educational Psychology, 91,* 684–689.

Higbee, K. L. (1993). *Your memory, how it works and how you improve it.* New York: Paragon House.

Hirsch, E. D. (2013). Primer on success: Character and knowledge make the difference. *Education Next, 13*(1). Retrieved from *http://educationnext.org/primer-on-success.*

Holmes, V. M., & Castles, A. E. (2001). Unexpectedly poor spelling in university students. *Scientific Studies of Reading, 5,* 319–350.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K.

Werder, & R. W. Woodcock, *WJ-R technical manual* (pp. 197–232). Itasca, IL: Riverside.

Hulme, C. (1981). *Reading retardation and multisensory teaching.* London: Routledge & Kegan Paul.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Exceptional Child Psychology, 96,* 150–166.

Hulme, C., & Snowling, M. (2009). *Developmental cognitive disorders.* Oxford, UK: Blackwell/Wiley.

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading, 9,* 351–365.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12,* 123–129.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1400 (2004).

Jenkins, J., & O'Connor, R. E. (2001, August). *Early identification and intervention for young children.* Paper presented at the U.S. Department of Education LD Summit, Washington, DC.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jordan, N. C., Fuchs, L. S., & Dyson, N. (2015). Early number competencies and mathematical learning: Individual variation, screening, and intervention. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1079–1078). Oxford, UK: Oxford University Press.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21,* 85–97.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490–501.

Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics, 20,* 303–314.

Kandel, E. R., & Squire, L. R. (2000). Neuroscience: Breaking down scientific barriers to the study of brain and mind. *Science, 290,* 1113–1120.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin and Review, 9,* 637–671.

Katusic, S. K., Colligan, R. C., Weaver, A. L., & Barbaresi, W. J. (2009). The forgotten learning disability: Epidemiology of written-language disorder in a population-based birth cohort (1976–1982), Rochester, MN. *Pediatrics, 123,* 1306–1313.

Kavale, K. A., & Flanagan, D. P. (2007). Ability–achievement discrepancy, response to intervention, and assessment of cognitive abilities/processes in specific learning disability identification: Toward a contemporary operational definition. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 130–147). New York: Springer.

Kavale, K. A., & Forness, S. R. (1998). The politics of learning disabilities. *Learning Disability Quarterly, 21,* 245–273.

Kelman, M., & Lester, G. (1997). *Jumping the queue: An inquiry into the legal treatment of students with learning disabilities.* Cambridge, MA: Harvard University Press.

Kibel, M. (1992). Linking language to action. In T. R. Miles & E. Miles (Eds.), *Dyslexia and mathematics* (pp. 42–57). London: Routledge.

Kidd, J. K., Carlson, A. G., Gadzichowski, K. M., Boyer, C. E., Gallington, D. A., & Pasnak, R. (2013). Effects of patterning instruction on the academic achievement of 1st-grade children. *Journal of Research in Childhood Education, 27,* 224–238.

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties.* Hoboken, NJ: Wiley.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Oxford, UK: Blackwell.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology, 27,* 1–25.

Klein, K., & Boals, A. (2001). The relationship of life stress and working memory. *Applied Cognitive Psychology, 15,* 565–579.

Klingberg, T. (2009). *The overflowing brain: Information overload and the limits of working memory* (N. Betteridge, Trans.). New York: Oxford University Press.

Korhonen, T. T. (1991). Neuropsychological stability and prognosis of subgroups of children with learning disabilities. *Journal of Learning Disabilities, 24,* 48–57.

Kovaleski, J. F., Lichtenstein, R., Naglieri, J., Ortiz, S. O., Klotz, M. B., & Rossen, E. (2015). Current perspective in the identification of specific learning disabilities. *Communique, 44*(4), 4–6.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School and Educational Psychology, 4,* 124–136.

Kronick, D. (1988). *New approaches to learning disabilities: Cognitive, metacognitive, and holistic.* Philadelphia: Grune & Stratton.

Kruk, R. S., & Willows, D. M. (2001). Backward pattern masking of familiar and unfamiliar materials in disabled and normal readers. *Cognitive Neuropsychology, 18*(1), 19–37.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14,* 389–433.

Lawrence, J. F., Rolland, R. G., Branum-Martin, L., & Snow, C. E. (2014). Generating vocabulary knowledge for at-risk middle school readers: Contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk, 19,* 76–97.

Leamnson, R. (2000). Learning as biological brain change. *Change, 32*(6), 34–40.

Lehto, J. (1996). Working memory capacity and summarizing skills in ninth graders. *Scandinavian Journal of Psychology, 37*(1), 84–92.

Lesaux, N. K., Kieffer, M. J., Kelly, J., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal, 51,* 1159–1194.

Liberman, I. Y., Rubin, H., Duques, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. In D. B. Gray & J. F. Kavanaugh (Eds.), *Biobehavioral measures of dyslexia* (pp. 163–176). Parkton, MD: York Press.

Light, J. G., & DeFries, J. C. (1995). Comorbidity of reading and mathematics disabilities: Genetic and environmental etiologies. *Journal of Learning Disabilities, 28,* 96–106.

Logie, R. H., Gilhooly, K. J., & Wynn, C. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition, 21,* 11–22.

Lohman, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research, 59,* 333–373.

Lohman, D. F. (1994). Spatial ability. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 1000–1007). New York: Macmillan.

Lowery, L. (1998). How new science curriculums reflect brain research. *Educational Leadership, 56*(3), 26–30.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49,* 344–351.

Lyon, G. R. (1998). Why reading is not natural. *Educational Leadership, 55*(6), 14–18.

Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Brain plasticity and functional losses in the aged: Scientific bases for a novel intervention. *Progress in Brain Research, 157,* 81–109.

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly, 30,* 457–469.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading, 3,* 129–157.

Manis, F. R., Seidenberg, M. S., Stallings, L., Joanisse, M., Bailey, C., Freedman, L., et al. (1999). Development of dyslexic subgroups: A one-year follow up. *Annals of Dyslexia, 49,* 105–134.

Maricle, D. E., & Johnson, W. L. (2016). Instructional implications from the Woodcock–Johnson IV Tests of Cognitive Abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 123–144). San Diego, CA: Academic Press.

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 377–384.

Marzano, R. J. (1992). *A different kind of classroom: Teaching with dimensions of learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mastropieri, M. A. (1988). Using the keyboard (*sic*) method. *Teaching Exceptional Children, 20*(2), 4–8.

Mather, N., & Kaufman, N. (2006). It's about the *what,* the *how well,* and the *why. Psychology in the Schools, 43,* 747–752.

Mather, N., & Wendling, B. J. (2012). *Essentials of dyslexia: Assessment and intervention.* Hoboken, NJ: Wiley.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research and Practice, 16,* 230–239.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties.* New York: Routledge.

McCloskey, G., Whitaker, J., Murphy, R., & Rogers, J. (2012). Intellectual, cognitive, and neuropsychological assessment in three-tier service delivery systems in schools. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 852–890). New York: Guilford Press.

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., et al. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry, 52,* 547–557.

McGrew, K. S. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and reading achievement across the lifespan. *Journal of Psychoeducational Assessment, Monograph Series: WJ R Monograph,* 39–53.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond *g*: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 177–189.

McGrew, K. S., & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment, 13,* 21–38.

McGrew, K. S., & Knopik, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life span. *School Psychology Review, 22,* 687–695.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV.* Itasca, IL: Riverside.

McGrew, K. S., & Wendling, B. J. (2010). CHC cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651–675.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III technical manual.* Itasca, IL: Riverside.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74,* 240–260.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Selective predictive value of rapid automatized naming in poor readers. *Journal of Learning Disabilities, 31,* 106–117.

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Holt.

Miller, S. P., & Mercer, C. L. (1997). Educational aspects of mathematics disabilities. *Journal of Learning Disabilities, 30,* 47–56.

Moats, L. C. (2001). Spelling disability in adolescents and adults. In A. M. Bain, L. L. Bailet, & L. C. Moats (Eds.), *Written language disorders: Theory into practice* (2nd ed., pp. 43–75). Austin, TX: PRO-ED.

Moats, L. C. (2010). *Speech to print: Language essentials for teachers* (2nd ed.). Baltimore: Brookes.

Moats, L. (2015). Many children left behind?: The Common Core and students with reading difficulties. *Perspectives on Language and Literacy, 41*(2), 19–23.

Moats, L. C., & Daken, K. E. (2008). *Basic facts about dyslexia and other reading problems.* Baltimore: International Dyslexia Association.

Moats, L., & Tolman, C. (2009). *Language essentials for teachers of reading and spelling (LETRS): Module 2. The speech sounds of English: Phonetics, phonology, and phoneme awareness.* Boston: Sopris West.

Moll, K., Gobel, S. M., Gooch, D., Landerl, K., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: Processing speed, temporal processing, and working memory. *Journal of Learning Disabilities, 49,* 272–281.

Monroe, M. (1935). Diagnosis of reading disabilities. In G. M. Whipple (Ed.), *The thirty-fourth yearbook of the National Society for the Study of Education: Educational diagnosis* (pp. 201–228). Bloomington, IL: Public School.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90,* 347–373.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665–681.

Muter, V., Hulme, C., Snowling, M. J., & Taylor, S. (1997). Segmentation, not rhyming predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65,* 370–398.

Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 145–172). Hoboken, NJ: Wiley.

Naglieri, J. A., & Otero, T. M. (2012). The Cognitive Assessment System: From theory to practice. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 376–399). New York: Guilford Press.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47,* 91–108.

Nation, K. (2007). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 248–265). Oxford, UK: Blackwell.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47,* 199–211.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology, 72,* 549–560.

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness in normal readers and poor comprehenders. *Applied Psycholinguistics, 21,* 229–241.

National Association for Gifted Children. (2010). *Redefining giftedness for a new century: Shifting the paradigm.* Washington, DC: Author.

National Center for Education Statistics. (2012). *School and staffing survey.* Washington, DC: Author. Retrieved from *https://nces.ed.gov/surveys/sass/tables/sass1112_2016006_t1s.asp.*

National Reading Panel (NRP). (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Niileksela, C. R., Reynolds, M. R., Keith, T. Z., & McGrew, K. S. (2016). A special validity study of the WJ IV: Acting on evidence for specific abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 65–106). San Diego, CA: Academic Press.

Noble, K. G., & McCandliss, B. D. (2005). Reading development and impairment: Behavioral, social, and neurobiological factors. *Journal of Developmental and Behavioral Pediatrics, 25,* 370–378.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427–452.

Oakhill, J., & Cain, K. (2016). Supporting reading comprehension development: From research to practice. *Perspectives on Language and Literacy, 42*(2), 32–39.

Ofiesh, N., Mather, N., & Russell, A. (2005). Using speeded cognitive, reading, and academic measures to determine the need for extended test time among university students with learning disabilities. *Journal of Psychoeducational Assessment, 23,* 35–52.

Ormrod, J. E., & Cochran, K. F. (1988). Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Reading Research and Instruction, 28,* 33–43.

Oslund, E. L., Clemens, N. H., Simmons, D. C., Smith, S. L., & Simmons, L. E. (2016). How vocabulary knowledge of middle-school students from low socioeconomic backgrounds influences comprehension processes and outcomes. *Learning and Individual Differences, 45,* 159–165.

Osmon, D. C., Smertz, J. M., Braun, M. M., & Plambeck, E. (2006). Processing abilities associated with math skills in adult learning disability. *Journal of Clinical and Experimental Neuropsychology, 28,* 84–95.

Pammer, K., Lavis, R., Hansen, P., & Cornelissen, P. L. (2004). Symbol-string sensitivity and children's reading. *Brain and Language, 89,* 601–610.

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology, 33,* 229–250.

Peng, P., Wang, C., & Namkung, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research.* [Epub ahead of print]

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101,* 385–413.

Perfetti, C. A. (1994). Reading. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 923–930). New York: Macmillan.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11,* 357–383.

Perfetti, C. A. (2011). Phonology is critical in reading but a phonological deficit is not the only source of low reading skill. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp. 153–171). New York: Psychology Press.

Perfetti, C. A., Marron, M. A., & Foltz, P. W. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 137–165). Mahwah, NJ: Erlbaum.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18,* 22–37.

Pinel, P., Piazza, D., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41,* 1–20.

Poletti, M., Carretta, E., Bonvicini, L., & Giorgi-Rossi, P. (2018). Cognitive clusters in specific learning disorder. *Journal of Learning Disabilities, 51,* 32–42.

Pressley, M., & Woloshyn, V. (Eds.). (1995). *Cognitive strategy instruction that really improves children's academic performance.* Cambridge, MA: Brookline Books.

Proctor, B. E. (2012). Relationships between Cattell–Horn–Carroll (CHC) cognitive abilities and math achievement within a sample of college students with learning disabilities. *Journal of Learning Disabilities, 45,* 278–287.

Proctor, B. E., Floyd, R. G., & Shaver, R. B. (2005). Cattell–Horn–Carroll broad cognitive ability profiles of low math achievers. *Psychology in the Schools, 42*(1), 1–12.

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology, 116,* 914–929.

Pulido, D., & Hambrick, D. (2008). The virtuous circle: Modeling individual differences in L2 reading and vocabulary development. *Reading in a Foreign Language, 20,* 164–190.

Rack, J. P., Snowling, M., & Olson, R. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 28–53.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91,* 137–157.

Reid, D. K., Hresko, W. P., & Swanson, H. L. (1996). *Cognitive approaches to learning disabilities.* Austin, TX: PRO-ED.

Reynolds, M. C., & Lakin, K. C. (1987). Noncategorical special education for mildly handicapped students: A system for the future. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), *The handbook of special education: Research and practice* (Vol. 1, pp. 331–356). Oxford, UK: Pergamon Press.

Reynolds, M. R., & Turek, J. J. (2012). A dynamic developmental link between verbal comprehension–knowledge (Gc) and reading comprehension: Verbal comprehension knowledge drives positive change in reading comprehension. *Journal of School Psychology, 50,* 841–863.

Richlan, F., Kronbichler, M., & Wimmer, H. (2011). Metaanalyzing brain dysfunctions in dyslexic children and adults. *NeuroImage, 56,* 1735–1742.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist, 47,* 189–203.

Robinson, C. S., Menchetti, B. M., & Torgesen, J. K. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learning Disabilities Research and Practice, 17,* 81–90.

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., & von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia, 47,* 2859–2865.

Rourke, B. P. (1993). Arithmetic disabilities, specific and otherwise: A neuropsychological perspective. *Journal of Learning Disabilities, 26,* 214–226.

Rourke, B. P. (1995). *Syndrome of nonverbal learning disabilities: Neurodevelopmental manifestations.* New York: Guilford Press.

Rourke, B. P., & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning: Perspectives from neurology and neuropsychology. *Journal of Learning Disabilities, 30,* 34–46.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48,* 115–136.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Detection, search, and attention. *Psychological Review, 84,* 1–66.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV cognitive abilities assessment.* Hoboken, NJ: Wiley.

Schrank, F. A., Miller, D. C., Wendling, B. J., & Woodcock, R. W. (2010). *Essentials of WJ III cognitive abilities assessment* (2nd ed.). Hoboken, NJ: Wiley.

Schrank, F. A., Stephens, T. L., & Schultz, E. K. (2016). *The WJ IV core-selective evaluation process for the identification of a specific learning disability* (Woodcock–Johnson IV Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside.

Schultz, E. K., Simpson, C. G., & Lynch, S. (2012). Specific learning disability: What constitutes a pattern of strength and weaknesses? *Learning Disabilities: A Multidisciplinary Journal, 18*(2), 87–97.

Schumm, J. S., Moody, S. W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities, 33,* 477–488.

Semrud-Clikeman, M. (2012). The role of inattention on academics, fluid reasoning, and visual–spatial functioning in two subtypes of ADHD. *Applied Neuropsychology: Child, 1,* 18–29.

Semrud-Clikeman, M., Steingard, R. J., Filipeck, P., Biederman, J., Bekken, K., & Renshaw, P. F. (2000). Using MRI to examine brain–behavior relationships in males with attention deficit disorder with hyperactivity. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 477–484.

Shanahan, M. A., Pennington, B. F., Yerys, B. E., Scott, A., Boada, R., Willcutt, E. G., et al. (2006). Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. *Journal of Abnormal Child Psychology, 34,* 585–602.

Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine, 338,* 307–312.

Shaywitz, S. E. (2003). *Overcoming dyslexia: A new and complete science-based program for overcoming reading problems at any level.* New York: Knopf.

Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology, 59,* 451–475.

Shaywitz, S. E., & Shaywitz, B. A. (2008). Paying attention to reading: The neurobiology of reading and dyslexia. *Development and Psychopathology, 20,* 1329–1349.

Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., et al. (2003). Neural systems for compensation and persistence: Young adult outcome of childhood reading disability. *Biological Psychiatry, 54,* 25–33.

Sinatra, R., Zygouris-Coe, V., & Dasinger, S. (2011). Preventing a vocabulary lag: What lessons are learned from research. *Reading and Writing Quarterly, 28,* 333–357.

Smith, D. D., & Rivera, D. P. (1998). Mathematics. In B. Wong (Ed.), *Learning about learning disabilities* (2nd ed., pp. 346–374). San Diego, CA: Academic Press.

Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81,* 1–23.

Sousa, D. (1998). Brain research can help principals reform secondary schools. *NASSP Bulletin, 82,* 21–28.

Spectrum-K12. (2011). Response to intervention adoption survey. Retrieved February 26, 2018, from *http://sss.usf.edu/Resources/presentations/2010/fasp_summer_inst2010/Resource_SLD/RTI/2010RTIAdoptionSurveyReport.pdf.*

Spencer, M., Quinn, J. M., & Wagner, R. K. (2014). Specific reading comprehension disability: Major problem, myth, or misnomer? *Learning Disabilities Research and Practice, 29,* 3–9.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stanovich, K. E. (1993). The construct validity of discrepancy definitions of reading disability. In G. R. Lyon, D. B. Gray, J. F. Kavanagh, & N. A. Krasnegor (Eds.), *Better understanding learning disabilities: New views from research and their implications for education and public policies* (pp. 273–307). Baltimore: Brookes.

Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of Learning Disabilities, 32,* 350–361.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological–core variable–difference model. *Journal of Educational Psychology, 86,* 24–53.

Stern, W. (1938). *General psychology from the personalistic standpoint.* New York: Macmillan.

Stoeckel, R. E., Colligan, R. C., Barbaresi, W. J., Weaver, A. L., Killian, J. M., & Katusic, S. K. (2013). Early speech–language impairment and risk for written language disorder: A population-based study. *Journal of Developmental and Behavioral Pediatrics, 34,* 38–44.

Strang, R. (1964). *Diagnostic teaching of reading.* New York: McGraw-Hill.

Strawser, S., & Miller, S. P. (2001). Math failure and learning disabilities in the postsecondary student population. *Topics in Language Disorders, 21,* 68–84.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41,* 3–22.

Swanson, E., Vaughn, S., & Wexler, J. (2017). Enhancing adolescents' comprehension of text by building vocabulary knowledge. *Teaching Exceptional Children, 50,* 84–94.

Swanson, H. L. (1987). Information-processing theory and learning disabilities: An overview. *Journal of Learning Disabilities, 20,* 3–7.

Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32,* 504–532.

Swanson, H. L. (2000). Are working memory deficits in readers with learning disabilities hard to change? *Journal of Learning Disabilities, 33,* 551–566.

Swanson, H. L. (2001). Searching for the best model for instructing students with learning disabilities. *Focus on Exceptional Children, 34,* 1–15.

Swanson, H. L. (2014). Does cognitive strategy training on word problems compensate for working memory capacity in children with math difficulties? *Journal of Educational Psychology, 106,* 831–848.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21,* 83–108.

Swanson, H. L., & Hsieh, C. J. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research, 79,* 1362–1390.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research, 76,* 249–274.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79,* 294–321.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73,* 407–440.

Swanson, H. L., & Zheng, X. (2013). Memory difficulties in children and adults with learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 214–238). New York: Guilford Press.

Sweeney, J. E., & Rourke, B. P. (1985). Spelling disability subtypes. In B. P. Rourke (Ed.), *Neuropsychology of learning disabilities: Essentials of subtype analysis* (pp. 133–144). New York: Guilford Press.

Szucs, D. (2016). Subtypes and comorbidity in mathematical learning disabilities: Multidimensional study of verbal and visual memory processes is key to understanding. *Progress in Brain Research, 227,* 277–304.

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2016). Five describing factors of dyslexia. *Journal of Learning Disabilities, 49,* 466–483.

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment, 29,* 446–457.

Tomlinson, C. A. (2006). *An educator's guide to differentiating instruction.* Boston: Houghton Mifflin.

Torgesen, J. K. (1997). The prevention and remediation of reading disabilities: Evaluating what we know from research. *Journal of Academic Language Therapy, 1,* 11–47.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent failure in young children. *American Educator, 22,* 32–39.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7–26.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvin, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593.

Townsend, D., Bear, D., Templeton, S., & Burton, A. (2016). The implications of adolescents' academic word knowledge for achievement and instruction. *Reading Psychology, 37,* 1119–1148.

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing: An Interdisciplinary Journal, 22,* 993–1020.

Travis, L. E. (1935). Intellectual factors. In G. M. Whipple (Ed.), *The thirty-fourth yearbook of the National Society for the Study of Education: Educational diagnosis* (pp. 37–47). Bloomington, IL: Public School.

Uccelli, P., Phillips Galloway, E., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly 50,* 337–356.

Uhry, J. K. (2005). Phonological awareness and reading: Theory, research, and instructional activities. In J. R. Birsh (Ed.), *Multisensory teaching of basic language skills* (2nd ed., pp. 83–111). Baltimore: Brookes.

Unsworth, N. (2016). The many facets of individual differences in working memory capacity. *Psychology of Learning and Motivation, 65,* 1–46.

Urso, A. (2008). *Processing speed as a predictor of poor reading.* Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139,* 352–402.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vellutino, F. R. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press.

Vellutino, F. R. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Exceptional Child Psychology, 59,* 76–123.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ–achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, S. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3–32.

Vukovic, R. K., Wilson, A. M., & Nash, K. K. (2004). Naming speed deficits in adults with reading disabilities: A test of the double-deficit hypothesis. *Journal of Learning Disabilities, 37,* 440–450.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83–103.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101,* 817–835.

Walberg, H. J., & Tsai, S. (1983). Matthew effects in education. *American Educational Research Journal, 20,* 359–373.

Wandell, B. A., Rauschecker, A. M., & Yeatman, J. D. (2012). Learning to see words. *Annual Review of Psychology, 63,* 31–53.

Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading, 16,* 45–62.

Wasserman, J. D. (2003). Assessment of intellectual functioning. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology* (pp. 417–442). New York: Wiley.

Weis, R., Dean, E. L., & Osborne, K. J. (2016). Accommodation decision making for postsecondary students with learning disabilities: Individually tailored or one size fits all? *Journal of Learning Disabilities, 49,* 484–498.

Westman, J. C. (1996). Concepts of dyslexia. In L. R. Putnam (Ed.), *How to become a better reading teacher: Strategies for assessment and intervention* (pp. 65–73). Englewood Cliffs, NJ: Merrill.

Whitaker, S. D. (2001). Supporting beginning special education teachers. *Focus on Exceptional Children, 34,* 1–18.

Whitener, E. M. (1989). A meta-analytic review of the effect on learning of the interaction between prior achievement and instructional support. *Review of Educational Research, 59,* 65–86.

Williams, J., Zolten, A. J., Rickert, V. I., Spence, G. T., & Ashcraft, E. W. (1993). Use of nonverbal tests to screen for writing dysfluency in school-age children. *Perceptual and Motor Skills, 76,* 803–809.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities, 34,* 237–248.

Windfuhr, K. L., & Snowling, M. J. (2001). The relationship between paired associate learning and phonological skills in normally developing readers. *Journal of Experimental Child Psychology, 80,* 160–173.

Wise, B. W., & Snyder, L. (2001, August). *Judgments in identifying and teaching children with language-based reading difficulties.* Paper presented at the U.S. Department of Education LD Summit, Washington, DC.

Wodrich, D. L., Spencer, M. L. S., & Daley, K. B. (2006). Combining use of RTI and psychoeducational testing: What we must assume to do otherwise. *Psychology in the Schools, 43,* 798–806.

Wolf, M. (1999). What time may tell: Towards a new conceptualization of developmental dyslexia. *Annals of Dyslexia, 49,* 3–27.

Wolf, M. (2007). *Proust and the squid: The story and science of the reading brain.* New York: HarperCollins.

Wolf, M. (2010). *RAVE-O: Retrieval through automaticity, vocabulary, elaboration, and orthography.* Longmont, CO: Sopris.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91,* 415–438.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 387–407.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Exceptional Child Psychology, 110,* 481–498.

Zorzi, M., Priftis, K., & Umiltá, C. (2002). Neglect disrupts the mental number line. *Nature, 417,* 138–139.

CHAPTER 29

The Kaufman Test of Educational Achievement, Third Edition, Wechsler Intelligence Scale for Children—Fifth Edition, and Wechsler Intelligence Scale for Children—Fifth Edition Integrated

Integrating Data for a Comprehensive Evaluation of Specific Learning Disability

Jaclyn Danielle Morrison
Jennie Kaufman Singer
Susan Engi Raiford

In 2015, an estimated 2.4 million U.S. students received special education services under the category of specific learning disability (SLD). This group represents about 35% of the national special education population, now more prevalent than any other disability type permissible by the Individuals with Disabilities Education Improvement Act (IDEIA, 2004; Kaufman, Kaufman, Drozdick, & Morrison, 2018; Maki, Burns, & Sullivan, 2017). Recent research has shown that the intellectual profile of children with SLD differs from that of typically developing children. In general, students with SLD demonstrate average overall intelligence (referred to as general cognitive ability, or *g*) but present isolated weaknesses in specific cognitive abilities that directly support academic learning (Toffalini, Giofre, & Cornoldi, 2017).

School psychologists play an essential role in maximizing the learning potential of these students, as emphasized by the National Association of School Psychologists (NASP) in a 2011 position statement, "Expertise in Specific Learning Disabilities is an essential area of specialization for all school psychologists" (p. 4). Reasonably, school psychologists spend most of their time conducting psychoeducational assessments to inform decisions regarding special education eligibility (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002; Filter, Ebsen, & Dibos, 2013).

Despite the steadily increasing prevalence of SLD and the growing body of research investigating the underlying mechanisms, the field of school psychology continues to lack cohesive guidelines on how to best to assess students with suspected SLD. Major challenges stem from the lack of a universally accepted SLD definition, diagnostic methods that vary across disciplines and regional boundaries, overlapping nomenclature that creates confusion (e.g., What is the difference between a cognitive ability and a cognitive process?), and consistent debates among experts and researchers who advocate disparate approaches for test interpretation and linking data to intervention. In light of advancements in both theoretical and applied psychometrics, it is necessary for school psychologists to reexamine their own professional practice in the identification and intervention of students with SLD.

The main purpose of this chapter is to synthesize current research on SLD diagnosis and interventions, geared specifically for school psychologists working in the U.S. school system. A large focus is placed on the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014a, 2014b), a valuable tool for assessing a student's reading, writing, mathematics, and oral language skills as part of a comprehensive SLD evaluation. We begin with a review of the KTEA-3, including the underlying test theory and structure, psychometric properties, and the various quantitative, norm-referenced scores available. Next, we highlight the KTEA-3 interpretative options that go beyond the standard profile analysis to investigate a student's intra-individual processing strengths and weaknesses. These qualitative methods include task and demands analyses and the unique KTEA-3 Error Analysis system. We then present a comprehensive summary of recent research using the KTEA-3 with specific clinical groups. Finally, we conclude with a discussion of issues specifically related to the assessment of SLD by school psychologists within the educational paradigm. An appendix to the chapter provides an interpretive case study that demonstrates the synthetization of achievement and cognitive data to inform special education eligibility decisions and intervention design.

KTEA-3 THEORY AND STRUCTURE

The KTEA-3 (Kaufman & Kaufman, 2014a, 2014b) is an individually administered diagnostic test of academic achievement for children and youth from ages 4 through 25, or from prekindergarten (PK) through grade 12. The KTEA-3 offers a wealth of information that can be used for multiple purposes in educational, clinical, and research settings alike. It is an effective tool for identifying academic strengths and weaknesses; making eligibility, placement, and diagnostic decisions; planning individualized instruction; monitoring academic progress; measuring a student's response to intervention; and conducting research. Although there are other popular standardized measures available, the KTEA-3 is the only achievement test that offers a systematic and norm-referenced system for error analysis (Flanagan, Mascolo, & Alfonso, 2017). These procedures allow the clinician to look beyond standardized scores and identify specific skill and processing weaknesses to target for intervention.

The KTEA-3 offers two versions: a Brief Form (Kaufman & Kaufman, 2015) and a Comprehensive Form (Kaufman & Kaufman, 2014a). The Brief Form is used to screen for weaknesses in three core academic areas (reading, writing, and math), while the Comprehensive Form assesses a fourth academic area (oral language) and additional skills in reading-related and cross-domain areas. The Comprehensive form offers two versions (A and B), which makes it useful for progress monitoring with a reduced potential for practice effects. The six subtests on the Brief Form were taken directly from the Comprehensive Form B, which allows practitioners to extend a screening assessment without the need to re-administer subtests. The Comprehensive Form (hereinafter referred to simply as the KTEA-3) is the focus of discussion in this chapter, but administration and scoring rules for mirrored subtests are identical.

The KTEA-3 represents a substantial revision of the Kaufman Test of Educational Achievement—Second Edition (KTEA-II; Kaufman & Kaufman, 2004b), updated with valuable modifications that enhance the clinical utility of the test. Changes from the previous version include updated norms (new PK norms), a decrease in the lower age limit from 4 years, 6 months (4:6) to 4 years, 0 months (4:0), the addition of four new subtests, improved content coverage of existing subtests and items, simplified administration procedures, and enhanced artwork. The KTEA-3 is based on a clinical model of development that advocates for both interpersonal and intrapersonal analysis.

The content of the KTEA-3 was developed with strong guidance from curriculum experts and practicing professionals. Items were designed to resemble popular textbooks and to include skills that align with state standards when possible. The breadth and depth of content coverage enhance interpretation and provide critical information for diagnostic and instructional decision making. The KTEA-3 measures all areas of SLD identified in IDEIA 2004 (e.g., Oral Expression, Listening Comprehension, Basic Reading Skills, Reading Comprehension, Reading Fluency Skills, Written Expression, Mathematics Calculation, and Mathematics Problem Solving), as well as the areas of impairment provided by the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013). Administration, scoring, and interpretive procedures offer optimal flexibility for designing an individualized battery that allows examiners to administer a single subtest or any combination of subtests specific to the referral concerns.

Like the two previous versions, the KTEA-3 includes a comprehensive criterion-referenced error analysis system that has been carefully developed over time. Curriculum experts in all academic areas helped to define the specific skills that are measured by each subtest and the types of errors that students are likely to make. The test developers also conducted a literature review on instructional theory and practice, consulted with practicing clinicians, and analyzed actual errors made by students during the data collection phases. This optional interpretive process allows an examiner to compare a student's error patterns with those of a normative group to determine specific skill deficits.

COMPOSITES AND SUBTESTS

The KTEA-3 includes 19 subtests that can be grouped into three core composites, 10 supplemental composites, and a global Academic Skills Battery (ASB) composite. The three core composites include Reading, Math, and Written Language. The 10 supplemental composites are organized into three subgroups: Reading-Related composites, which include Sound–Symbol, Decoding, Reading Fluency, and Reading Understanding; Oral composites, which include Oral Language and Oral Fluency; and Cross-Domain composites, which include Comprehension, Expression, Orthographic Processing, and Academic Fluency. Not all subtests and composites are available for all ages and grades. Most notably, different subtests contribute to the ASB for students in PK, kindergarten (K), and those in grades 1–12+. Figure 29.1 illustrates the composites and subtests that contribute to the ASB by age. The descriptions below provide specific details about the composites (task requirements, age ranges, etc.), followed by subtest specific explanations.

Core Composites

- *Academic Skills Battery (ASB).* This composite provides a measure of overall academic achievement in Reading, Math, and Written Language. Unlike the global score offered on the KTEA-II (the Comprehensive Achievement Composite, or

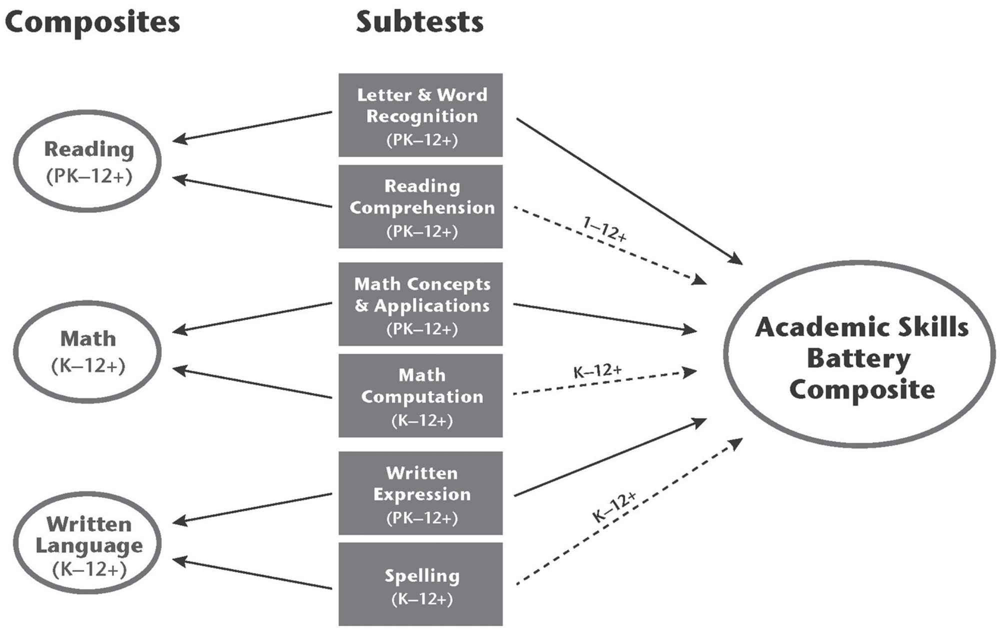


FIGURE 29.1. Core composite structure of the KTEA-3. Figure found in the *Manual for the Kaufman Test of Educational Achievement, Third Edition (KTEA™-3)*. Copyright © 2014 by NCS Pearson, Inc. Reproduced with permission. All rights reserved. "KTEA" is a trademark, in the US and/or other countries, of Pearson Education, Inc. or its affiliates(s).

CAC), the ASB does not include Oral Language skills. Results of joint confirmatory analyses conducted with the KTEA-II and the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004a) revealed that the Oral Language subtests were excellent measures of crystallized intelligence, but not directly related to any core academic skill; in other words, oral language is better conceptualized as a foundational skill that is essential for building academic skills, rather than as an academic skill in and of itself (S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). The ASB covers all three core academic areas for all ages, but different subtests contribute to the score for PK, K, and grades 1–12+.

- *Reading (grades PK–12+; ages 4–25).* This composite provides a measure of overall reading ability, including basic reading skills and reading comprehension. Weak scores may indicate deficits in basic reading skills, language comprehension, or both. Subtests: Letter and Word Recognition; Reading Comprehension.
- *Math (grades K–12+; ages 5–25).* This composite provides an overall measure of math achievement, including math computation and problem-solving skills. Subtests: Math Computation; Math Concepts and Applications.
- *Written Language (grades K–12+; ages 5–25).* This composite provides a measure of achievement in expressive written language. Subtests: Written Expression; Spelling.

Supplemental Composites

Reading-Related

- *Sound–Symbol (grades 1–12+; ages 6–25).* This composite measures phonological processing and decoding skills. Weak scores may indicate deficits in phonological, orthographic, and/or morphological awareness, as well as the phonological loop of working memory. A phonological core deficit is a common underlying cause of dyslexia. Subtests: Phonological Processing; Nonsense Word Decoding.
- *Decoding (grades 1–12+; ages 6–25).* This composite measures broad basic reading skills, including recognizing regular, irregular, and nonsense words. Weak scores may indicate deficits in phonological, orthographic, and or morphological awareness; long-term storage and retrieval; receptive vocabulary; and the phonological loop of working memory. Subtests: Letter and Word Recognition; Nonsense Word Decoding.
- *Reading Fluency (grades 3–12+; ages 8–25).* This composite provides a measure of reading automaticity across a range of speeded conditions, including real and nonsense words, words in isolation and in context, and both oral and silent responses. Subtests: Silent Reading Fluency; Word Recognition Fluency; Decoding Fluency.
- *Reading Understanding (grades 1–12+; ages 6–25).* This composite measures comprehension of written narrative containing literal or inferential information, and the ability to identify or infer the meaning of words. Low scores may indicate weakness in vocabulary knowledge, the ability to comprehend written passages, or both. Subtests: Reading Comprehension; Reading Vocabulary.

Oral

- *Oral Language (grades PK–12+; ages 4–25).* This composite measures the ability to comprehend literal and inferential information from oral narratives, to describe a picture orally, and to name words fluently in a sematic category. Low scores may indicate difficulties with oral expression in the areas of fluency, pragmatics, or grammar, or with comprehending formal speech. Subtests: Associational Fluency; Listening Comprehension; Oral Expression.
- *Oral Fluency (grades PK–12+; ages 4–25).* This composite measures verbal fluency and automaticity with tasks that require naming pictured objects or words in a semantic category, as quickly as possible. Also known as *rapid automatic naming (RAN)*, this ability is well supported in research as a predictor of reading and spelling difficulties. Low scores may indicate weak word retrieval difficulty or deficits in long-term memory. Subtests: Associational Fluency; Object Naming Facility.

Cross-Domain

- *Comprehension (grades PK–12+; ages 4–25).* This composite provides a measure of receptive language skills in both written and oral modalities. Low scores may indicate a receptive language deficit and/or weakness in verbal working memory, inattention, or distractibility. Subtests: Reading Comprehension; Listening Comprehension.
- *Expression (grades PK–12+; ages 4–25).* This composite measures expressive language skills in both written and oral modalities. Low scores

may indicate an expressive language deficit and/or weakness in grammar, verbal working memory, and self-monitoring. Subtests: Written Expression; Oral Expression.

- *Orthographic Processing (grades 1–12+; ages 6–25).* This composite measures the ability to form, store, recognize, and retrieve orthographic representations. Low scores may indicate poor orthographic memory; difficulty learning phoneme–grapheme correspondences; or weakness in phonological processing, auditory verbal working memory, and/or cognitive efficiency. Subtests: Spelling; Letter Naming Facility; Word Recognition Fluency.

- *Academic Fluency (grades 3–12+; ages 8–25).* This composite assesses the ability to perform rudimentary academic tasks quickly and accurately, while maintaining focus under time pressure. This information helps to determine the impact of automaticity on academic performance. Low scores may indicate weaknesses in automaticity and processing speed. Additionally, pairwise comparisons can help determine whether slow performance is a function of a skill deficit versus an emotional factor such as anxiety or low motivation. Subtests: Writing Fluency; Math Fluency; Decoding Fluency.

Subtest Descriptions

Basic Reading

- *Letter and Word Recognition (grades PK–12+; ages 4–25).* The student identifies letters and pronounces words of gradually increasing difficulty. The initial items assess knowledge of letter names and sounds, followed by easy word items including high-frequency, phonetically regular words that can be read by using principles of phonological decoding. As the difficulty level increases, the proportion of irregular words increases; this ensures that the subtest measures word recognition (reading vocabulary) in addition to decoding skills.

- *Nonsense Word Decoding (grades 1–12+; ages 6–25).* The student reads invented words of increasing difficulty. This task primarily measures phonological decoding.

Reading Understanding

- *Reading Comprehension (grades PK–12+; ages 4–25).* On this untimed measure of silent reading comprehension, the item types vary depending on the difficulty level. The easiest items require matching a symbol or word with its corresponding picture. Subsequent items ask the student to read a simple instruction and respond by performing an action. On later items, the student reads passages of increasing difficulty and answers literal or inferential questions about them. The most difficult items require rearranging five sentences into a coherent paragraph and answering related questions.

- *Reading Vocabulary (grades 1–12+; ages 6–25).* Newly designed for the KTEA-3, this subtest measures contextual reading vocabulary by requiring the student to determine the meaning of words that are embedded in the context of a sentence. On early items, the student is required to point to one of three words that have the same meaning as a picture or target word. For remaining items, the student reads a sentence (silently or aloud) and then points to the word in the sentence that has the same meaning as a target word.

Reading Fluency

- *Word Recognition Fluency (grades 1–12+; ages 6–25).* The student reads a list of real words aloud as quickly as possible for two 15-second trials.

- *Decoding Fluency (grades 3–12+; ages 8–25).* The student reads a list of invented words aloud as quickly as possible for two 15-second trials.

- *Silent Reading Fluency (grades 1–12+; ages 6–25).* New to the KTEA-3, this subtest was designed to measure both reading fluency and literal comprehension. The student silently reads simple sentences and indicates whether each is true or false by marking “yes” or “no” in the response booklet, completing as many items as possible within a 2-minute time limit.

Language Processing

- *Phonological Processing (grades PK–12+; ages 4–25).* The student responds orally to items that require manipulation of sounds. The subtest is separated into five sections of phonemic awareness tasks: Blending, Rhyming, Sound Matching, Deleting Sounds, and Segmenting.

- *Object Naming Facility (grades PK–12+; ages 4–25).* The student names pictured objects as quickly as possible for two short trials.

- *Letter Naming Facility (grades K-12+; ages 5-25):* The student names upper- and lowercase letters as quickly as possible for two short trials.

Math

- *Math Concepts and Applications (grades PK–12+; ages 4–25).* The student is required to respond orally to items that require the application of various mathematical principles to real-life situations. Skills measured include number concepts, operation concepts, time/money, measurement, geometry, fractions/decimals, data investigation, and higher math concepts.
- *Math Computation (grades K–12+; ages 5–25).* The student writes answers to as many math calculation problems as possible. Skills measured include simple counting and number identification; addition, subtraction, multiplication, and division operations; fractions and decimals; square roots and exponents; and algebra.
- *Math Fluency (grades 1–12+; ages 6–25).* The student writes answers to as many addition, subtraction, multiplication, and division problems as possible within a 60-second time limit.

Written Language

- *Written Expression (grades PK–12+; ages 4–25).* Students in grades PK and K are required to trace, copy, and write letters, words, and sentences from dictation. Older students are required to complete writing tasks in the context of a grade-appropriate story format with items that include writing sentences from dictation, adding punctuation and capitalization, filling in missing words, completing sentences, combining sentences, writing compound and complex sentences, and retelling a story in essay format.
- *Spelling (grades K–12+; ages 5–25).* On early items, the student writes single letters that represent sounds. On subsequent items, the student writes regular and irregular words from dictation at increasing difficulty levels.
- *Writing Fluency (grades 2–12+; ages 7–25).* In the response booklet, the student writes one sentence for each picture presented and completes as many items as possible within 5 minutes.

Oral Language

- *Listening Comprehension (grades PK–12+; ages 4–25).* The student listens to either a sentence read by the examiner or a recorded passage played from the audio files on the flash drive, and then responds orally to comprehension questions (literal and inferential) asked by the examiner.
- *Oral Expression (grades PK–12+; ages 4–25).* For each item, the student orally responds with a complete sentence that describes a picture presented in the stimulus book. Harder items require the student to use one or two target words in each response, while the most difficult items require the student to begin each response with a target word or phrase.
- *Associational Fluency (grades PK–12+; ages 4–25).* Within a 60-second time limit, the student says as many words as possible that belong to a given semantic category.

PSYCHOMETRIC PROPERTIES

Standardization Sample

Standardization of the KTEA-3 was conducted with an age-norm sample of 2,050 examinees (ages 4:0–25:11) and a grade-norm sample of 2,600 students (grades PK–12). All age levels included a range of 120–160 examinees, apart from age 4 (sample of 100) and ages 19–20 and 21–25 (each with a sample of 75). Each grade level was represented by 150–200 students, and the cumulative sample was split into semesters, with approximately half tested in the fall and the other half in the spring. The normative sample was closely matched to the 2012 American Community Survey of the U.S. Census Bureau along the variables of age/grade, sex, parental education level, geographic region, and special group designation. Approximately 8–10% of the norm samples at each grade or age group included students with one of more of the following educational classifications: SLD in reading and/or writing, SLD in math, language disorder, attention-deficit/hyperactivity disorder (ADHD), mild intellectual disability, and academic giftedness. To account for the KTEA-3 parallel forms, approximately half of the norm sample was administered Form A, and the remaining half was administered Form B. Data were collected by 490 examiners in 48 states between August 2012 and July 2013.

Reliability

For both Forms A and B, the internal-consistency reliability of the KTEA-3 Comprehensive Form is strong. For all grade levels and age ranges, reliabilities are in the .90s for the ASB composite (.97–.99), as well as the Reading (.92–.97), Math (.95–.98), Written Language (.91–.97), and Reading-Related

(.92–.99) composites. The Oral Language composite has a mean reliability of .86, while the Cross-Domain composites have a mix of reliabilities in the .80s and .90s.

Alternate-form reliabilities were calculated by administering both versions of the KTEA-3 (Forms A and B) to a sample of 306 children. The average time elapsing between administrations was 7.5 days. The values for the alternate-form reliabilities are comparable to the high internal-consistency reliability values. The ASB composite has very high consistency values across time and forms (mid-.90s). The Reading, Math, and Written Language composites have alternate-form reliabilities in the high .80s to mid-.90s. The Decoding and Reading Fluency composites also have alternate-form reliabilities in the high .80s to mid-.90s. These strong values suggest that the KTEA-3 alternate forms will be useful in reducing practice effects.

The Written Expression and Oral Expression subtests require subjective judgment in scoring because examinees' responses can vary significantly. As such, these subtests are most susceptible to differences in scoring among examiners. The interrater reliability was calculated as the percentage of agreement between trained scorers for all double-scored cases in the norm sample, weighted by the number of protocols scored. Agreement rates were calculated for both forms combined. There were 10 scorers for Oral Expression and 13 for Written Expression, each scoring approximately 300 protocols. Interrater agreement was 90% for Oral Expression and 95% for Written Expression, which supports the reliability of the KTEA-3 scoring criteria. Although other KTEA-3 subtests also require some degree of selective judgment in scoring (i.e., Reading Comprehension, Listening Comprehension, Writing Fluency, and Associational Fluency), their respective agreement rates were consistently high (≥98%).

Validity

There is strong evidence supporting the interpretation of KTEA-3 scores for their intended purpose. Revisions of the KTEA-3's content were based on careful consideration of examiner feedback, school curriculum guidelines, and expert and panel reviews, as well as on special-group studies conducted with clinical populations. The KTEA has a history of strong theoretical and empirical evidence of its validity based on response processes. Further evidence collected during the development of the KTEA-3 included consultation with experts in the field, extensive literature reviews, and empirical and qualitative evaluations. Results continue to indicate that the KTEA-3 items are measuring the constructs as expected.

The internal structure of the KTEA-3 also shows evidence of strong construct validity. Correlations among the Reading, Math, Written Language, Sound Symbol, Decoding, and Reading Fluency composites are mostly within the .70s and .80s across groups. The Oral Language and Oral Fluency composites correlate with the other composites at lower levels (.40s to .50s). These lower correlations are expected, given that language processes are considered to constitute a foundation for building academic skills, rather than an academic skill area in and of themselves.

Additional confirmatory factor analysis was conducted to investigate the relationship between the KTEA-3 composites and subtests in a more systematic way than simply investigating intercorrelations. Confirmatory factor analysis was applied to the six subtests that make up the ASB composite, plus the additional three Oral Language subtests. Results support the validity of composite structure through a four-factor model with constructs related to the domains of reading, mathematics, written language, and oral language. The final model had good fit statistics, and all subtests had high loadings on all factors.

Parkin (2018) recently investigated the factor structure of the KTEA-3 reading, writing, and oral language measures to test for invariance across grade levels. Results of his analyses suggest that the factors analyzed demonstrated the same amount of variance across grade levels and are associated with each other to the same degree across groups. Parkin's findings support the use of the KTEA-3 across grade levels for interpretation of both its core academic composites and supplemental cross-domain composites related to comprehension, expression, and decoding. This research also suggested that subtests could contain systematic variance for multiple abilities. Performance on Reading Comprehension, for example, may reflect a general reading ability and a comprehension ability. Spelling might be a summative score reflecting general writing skill and decoding (Parkin, 2018).

The KTEA-3 demonstrates evidence of concurrent validity when compared to other similar measures of academic achievement. Correlations were calculated with the KTEA-3 Comprehensive Form and the KTEA-3 Brief Form, as well as the

Wechsler Individual Achievement Test—Third Edition (WIAT-III; Pearson, 2009), the Woodcock–Johnson III Tests of Achievement (WJ III ACH; Woodcock, McGrew, & Mather, 2001), and the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). Like the KTEA-II composites, the KTEA-3 composites show correlations with the WIAT-III and WJ III ACH in the mid- to high .80s; correlations between most of the total achievement scores range from the high .80s to .95. The CELF-4 Formulated Sentences was used to evaluate the validity of the new KTEA-3 Oral Expression subtest. They show a correlation of .64, which supports the validity of the new KTEA-3 subtest.

Correlation studies were also conducted with tests of cognitive ability, including the KABC-II (Kaufman & Kaufman, 2004a), the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007), and the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014). Correlations between the KTEA-3 core composites and the global scales of the KABC-II, DAS-II, and WISC-V range from .49 to .82. KTEA-3 composites generally correlate highest with measures of crystallized, fluid, and learning ability.

ADMINISTRATION AND SCORING

Subtest Selection

A key asset of the KTEA-3 is the flexibility for practitioners to individually design assessments that are specific to each student. If only one domain of academic functioning is of concern, an examiner may choose to administer a single subtest or any combination of subtests in the relevant academic domain. If multiple domains need to be measured, then all age-appropriate subtests can be administered to obtain the desired composite score(s). The KTEA-3 manual (Kaufman & Kaufman, 2014b, p. 21) provides suggestions for designing an initial battery based on the referral concerns, but the examiner is encouraged to take a fluid assessment approach by modifying the battery as data are collected that confirms or refutes initial theories and hypotheses.

Administration Considerations

The administration of subtests on standardized tests typically follows a strict sequential format that is predetermined by the test publishers. Another aspect of the KTEA-3's flexible design is that the order of subtest administration is mostly left to the examiner's discretion. There are two sets of subtests that must follow a sequence (Letter and Word Recognition must be given before Word Recognition Fluency, and Nonsense Word Decoding must be given before Design Fluency), but other subtest administration guidelines offered in the KTEA-3 stimulus book and record form are mostly suggestions. Administration times can vary, depending on the subtests selected for administration, the examiner's testing style, and examinee-related factors (such as personality, mood, and rapport with the examiner). In general, each of the core academic composites takes 10–35 minutes to administer; obtaining the ASB composite varies from approximately 15 minutes for the youngest children to 85 minutes for students in grade 3+.

As in other standardized tests of achievement, the KTEA-3 items are ordered by difficulty, with the easiest items being administered first. Grade-based starting points are listed on the record form and on the first page of a subtest's directions in the stimulus book. The KTEA-3 subtests include fewer sample, teaching, or practice items than are typically seen in many standardized cognitive tests. Only eight subtests direct the examiner to record time during administration: Six subtests enforce time limits (Writing Fluency, Reading Fluency, Math Fluency, Word Recognition Fluency, Decoding Fluency, and Associational Fluency), while two require completion time (Object Naming Facility and Letter Naming Facility). On certain subtests, the examiner may be encouraged or required to record oral responses, depending on the depth of information needed for scoring. If an examiner chooses to apply error analysis procedures, verbal responses may need to be recorded verbatim. For a subtest-by-subtest review of administration, see Breaux and Lichtenberger (2016).

Scoring Methods

Q-Global™

The KTEA-3 offers convenient digital scoring through Q-global™, Pearson's web-based scoring platform. Benefits of digital scoring include efficient organization of student information and test data, instantaneous calculation of scores, and assorted options for generating reports that can be downloaded and customized. Additionally, Q-global offers features that are not available through hand scoring, but provide information that is exceptionally valuable within the context of a comprehensive SLD assessment. These features include:

- A pattern of strengths and weaknesses (PSW) discrepancy analysis.
- Intervention statements and teaching objectives, based on patterns of errors in the student's performance and skill deficits aligned with the Common Core State Standards.
- Recommended interventions for the home environment, which include fun and engaging educational activities that family members can use to strengthen a child's basic academic skills at home.

Hand Scoring

Examiners may also score the KTEA-3 by hand. All forms and normative data required for hand scoring are preloaded onto a flash drive that is included as part of the test kit. Detailed instructions for hand scoring are provided in the KTEA-3 manual (Kaufman & Kaufman, 2014b).

Types of Scores

The KTEA-3 yields several types of derived scores that practitioners can use to inform their clinical judgment. Analyzing different types of scores helps an examiner to understand a student's test performance in a systematic and meaningful way. Standard scores, percentile ranks, grade and age equivalents, and growth scale values (GSV) are all available. Age-based norms are available for all ages (4:0–25:11), and grade-based norms are available for all grade levels (PK–12). Grade-based norms are separated by trimester: fall (for students assessed in August through November), winter (for students assessed in December through February), and spring (for students assessed in March through July). Age and grade equivalents are available to describe the norms of the test and the average performance of students across ages and grades. The GSV scale is an equal-interval scale of academic skills that can be used to measure academic progress throughout the school years. Whereas standard scores allow for comparison of an examinee's performance relative to his or her peers, GSV scores offer a measurement of the examinee's rate of progress compared to the typical growth rate in the normative sample. The test authors urge examiners to use caution when interpreting GSV scores to describe an examinee's achievement or skill level, as they can often be misleading.

INTERPRETATION

One of the primary goals of the KTEA-3 is to identify students' strong and weak areas of academic functioning from both *ipsative* (person-based) and *normative* (age-based or grade-based) perspectives (Breaux & Lichtenberger, 2016). Because the KTEA-3 allows examiners to customize the battery for each unique assessment situation, the approaches to interpretation must also allow for flexibility. The KTEA-3 manual (Kaufman & Kaufman, 2014b) offers a systematic method for interpretation of both the comprehensive and selective test batteries. However, examiners should be careful not to overestimate the significance or clinical utility of any score in isolation. KTEA-3 results can be quite valuable for generating hypotheses about why a student is struggling to learn, but these hypotheses should always be investigated or substantiated with multiple pieces of supportive data.

Most approaches to standardized test interpretation are based on the concept of *drilling down*, which begins with interpreting a student's global level of functioning and then subsequently interpreting the more specific, underlying skills and abilities that contribute to overall functioning (Flanagan et al., 2017). The KTEA-3 interpretive approach for the comprehensive battery models this hierarchical process, beginning with interpretation of the global ASB composite, followed by evaluation of composite and subtest scores, and concluding with a step-by step procedure for identifying a student's PSW. Additional guidance is provided below, but any person intending to administer the KTEA-3 or interpret the resulting data should refer to the test manual.

Standard Profile Interpretation

- *Step 1.* The first step is reporting the ASB standard score, along with its confidence interval and percentile rank. The ASB battery is composed of three subtests for students in PK, five subtests for kindergarteners, and six subtests for examinees in grades 1–12+. The ASB was designed to provide a reliable normative overview of the examinee's academic achievement in the core domains of Reading, Written Language, and Math. The ASB represents a midpoint for determining an examinee's relatively strong and weak areas of academic achievement.

- *Step 2.* The second interpretive step involves calculating the standard scores and confidence intervals of all composites and subtests administered. Depending on the normative sample chosen, these scores can be evaluated to identify the examinee's strengths and weaknesses, compared to those of same-age or same-grade peers.
- *Step 3.* The third interpretive step takes an ipsative approach by analyzing the examinee's performance on individual composites relative to his or her overall academic achievement. This is done by comparing any composite scores that were derived to the overall ASB score. Any significant difference in which the composite score is greater than the ASB score represents a personal strength, and any significant difference in which the composite score is less than the ASB score represents a personal weakness.
- *Step 4.* If any significant personal strengths or weaknesses are identified, the test authors also recommend reporting how unusual the difference is within the normative sample. Generally, if the difference occurs in 10% or less of the population, it is considered infrequent and of particular clinical interest. Step 4 then involves applying the same ipsative approach at the subtest level.
- *Step 5.* The last step in the formal interpretive process is making pairwise comparisons between specific academic or reading-related skills, which may provide useful information for diagnostic and instructional planning purposes. The authors recommend two specific subtest comparisons that should be analyzed as part of any standard battery: the Oral Expression subtest with the Written Expression subtest, and the Reading Comprehension subtest with the Listening Comprehension subtest.

Error Analysis

"The key to effective interpretation of test performance after administration is careful observation of test performance during administration. Integration of what was observed during administration with what is scored after administration enables the clinician to characterize more accurately the specific cognitive strengths and weaknesses of the child" (McCloskey, Hartz, & Slonim, 2016). Systematically examining the errors made by students with disabilities on tests of cognitive and academic performance can enhance diagnostic procedures, improve clinical judgment, and identify more specific and curriculum-relevant skills to target with educational interventions (Avitia, DeBiase, et al., 2016; Avitia, Pagirsky, et al., 2016). Following publication of the original KTEA, error analysis procedures were developed as a method for identifying patterns of errors that could be indicative of specific skills deficit. The original system accurately identified the number of errors in each category; however, it did not provide a rationale for understanding the inaccurate thought processing that led to the error (Breaux, 2017; McCloskey, 2017).

Traditionally, criterion-related tests have determined a student's skill level with specific cut-off scores, largely based on the judgments of curriculum experts. However, this type of subjective numerical value provides little information about the statistical significance of the student's level of mastery. Instead of using such arbitrary cut-off scores, the KTEA-3 compares a student's total errors within predetermined categories, with the average number of errors made by the reference group (normative comparison) to indicate whether the student's performance is above average, average, or below average.

The KTEA-3 uses two types of error classification methods. *Item-level error analysis* is available on five subtests and enables automatic classification of errors based on the item score (Phonological Processing, Written Expression, Math Concepts and Applications, Listening Comprehension, and Reading Comprehension). In this approach, each item is classified according to the process, concept, or skill assessed and these classifications constitute the error categories. Four additional subtests offer *within-item error analysis*, which requires the examiner to manually classify errors based on a qualitative analysis of the student's response (Letter and Word Recognition, Nonsense Word Decoding, Spelling, and Oral Expression). One subtest (Math Computation) includes both item-level and within-item error analyses. In a significant improvement from previous versions, the KTEA-3 error analysis incorporates aspects and skills that can be directly mapped to the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

Language-Related Analysis

Language is not a unitary construct. Students with and without learning disabilities (regardless of subtype) exhibit different profiles of strengths and weaknesses across all language areas. There is strong evidence to suggest that four distinct language systems exist: language by eye (reading), language by hand (written), language by ear (listening), and language by mouth (speaking). Oral language skills typically begin to develop sooner than written language; however, the four language systems (listening, speaking, reading, and writing) then develop concurrently as integrated strands that influence one another (Berninger & Abbott, 2010; Carretti, Motta, & Re, 2016). Over the past 40 years, a sizable amount of research publications have focused on contrasting reading and writing disabilities and investigating ways to differentially diagnose SLD subtypes based on the levels of language (subword, word, syntax) and language systems (by ear, mouth, eye, hand) that are impaired, the developmental trajectory, and their relationships to instructional needs (Nielsen et al., 2017). Researchers have long promoted the value of error analysis procedures for gaining insight into how individuals learn to read, how literacy skills develop, and the strategies upon which students rely (Liu et al., 2017).

Letter & Word Recognition, Nonsense Word Decoding, and Spelling

The Letter & Word Recognition (LWR), Nonsense Word Decoding (NWD), and Spelling subtests assess an examinee's ability to connect speech sounds to letter patterns. LWR includes many words with both predictable and unpredictable letter patterns. NWD is designed to measure a student's ability to apply decoding and structural analysis skills to typically occurring letter patterns. The Spelling error analysis gives information about the student's ability to relate speech sounds to letter patterns.

The error analysis system for these subtests is made up of categories corresponding to letters and letter combinations that have a predictable relationship to their sound. The error categories are constructed to align with patterns and rules that are generally taught in school. The error analysis depends on dividing the words or nonsense words into parts based on orthographically predictable patterns. For example, the word *point* would be broken down into its parts: *p* (single consonant), *oi* (vowel team/diphthong), and *nt* (consonant blend). One-syllable words and words without affixes are divided into consonant and vowel parts, while multisyllabic words with roots and affixes are divided into morphemic word parts. Error Categories are similar across the three subtests and include:

Consonants: Single/Double Consonant, Initial Blend, Medial/Final Blend, and Consonant/Digraph

Vowels: Wrong Vowel, Short Vowel, Long Vowel, Vowel Team/Diphthong, and *R*-controlled Vowel

Other: Silent Letter, Prefix/Word Beginning, Suffix/Inflection, Hard/Soft *C, G, S*, Unpredictable Pattern, Initial/Final Sound, Syllable Insertion/Omission, Misordered Sounds, Nonphonetic, and Whole Word Errors

Phonological Awareness

Phonological processing is the ability to manipulate phonemes, the smallest units of sound heard in spoken language. Deficits in phonological processing skills—which include rhyming, matching, blending, segmenting, and manipulating sounds—are among the most prominent cognitive weaknesses exhibited by students with a specific learning disability in reading (SLD-R; Fletcher, Lyon, Fuchs, & Barnes, 2007). Phonological processing has long been acknowledged as a strong predictor of emerging literacy skills and an effective way for practitioners to identify poor readers. Children with poor phonological awareness may struggle with the following: identifying isolated sounds in words; manipulating sounds; perceiving a word as a fluid sequence of sounds; or isolating the beginning, middle, or ending sounds in a word.

On the KTEA-3 Phonological Processing subtest, error categories include blending, rhyming, sound matching, deleting sounds, and segmenting. Choi and colleagues (2017) compared errors made on the KTEA-3 Phonological Processing subtest across two error factors. Basic Phonological Awareness (BPA) reflects basic sound awareness, or the ability to identify and distinguish sounds in words, and includes the error categories of rhyming, blending, and phoneme matching skills; Advanced Phonological Processing (APP) reflects phonological skills that allow an individual to hold phonological information within working memory and decompose or manipulate it, which include deleting and segmenting. Choi and colleagues' research revealed some interesting results. First, both BPA and APP skills were significantly related to all subreading skills (including processing speed–related tasks); Spelling and

Written Expression (but not Writing Fluency); and Listening Comprehension, Oral Expression, and Associational Fluency (except Object Naming Facility and Letter Naming Facility). Furthermore, APP predicts students' reading, writing, and oral language skills better than BPA across ages. Second, segmenting errors are the most difficult and distinct on the KTEA-3. Third, statistically meaningful correlations were found between phonological processing and comprehension. Error analysis of phonological processing performance is especially important for poor readers because they may achieve a high score on easier BPA tasks (e.g., blending) but a poor score on the more advanced APP tasks (e.g., deleting), or vice versa. Students with oral language and literacy acquisition difficulties need more differentiated phonological processing tests because the overall cut-off scores do not distinguish the levels of difficulty or target specific skills for intervention. Although phonological processing is well established as a predictor of reading at younger ages, Choi and colleagues found that it is also predictive of reading performance in older children (Mather & Wendling, 2016). This emphasizes the importance of including comprehensive measures of phonological processing for all students referred for reading concerns, regardless of age.

Reading Comprehension and Listening Comprehension

Listening and reading comprehension are two essential receptive communication skills that serve as significant predictors of factors related to language development. Linguistically speaking, the most crucial factors involved in listening comprehension are phonological processing and knowledge of syntax/semantics. For reading comprehension, the most important linguistic factors include decoding, word recognition, syntax, semantics, and discourse (Hatcher et al., 2017). The error categories derived from these two KTEA-3 comprehension subtests are categorized as literal, inferential, narrative, and expository. Literal comprehension requires recognition or recall of information that is explicitly stated in a text. The KTEA-3 literal items require the student to find an answer that is stated verbatim or similarly paraphrased directly within the passage. Inferential comprehension requires the generation of novel ideas beyond those explicitly stated in the text. The KTEA-3 inferential items require the student either to deduce the central thought of the passage, make inferences about the content or the author's purpose, and recognize tone or mood.

The narrative category identifies errors that are made on both literal and inferential items about narrative passages, which include fictional information presented in a literary structure. The expository category identifies errors that are made on both literal and inferential items about expository passages, which include information presented in a nonliterary format. The distinction between these receptive skills is particularly important in grade school curricula, as the CCSS are specific to the literal and inferential comprehension of narrative and expository texts. Understanding the relationship between a student's errors in language comprehension of passages across oral and written modalities may assist in differentiating a specific reading weakness from a more general comprehension problem.

Written Expression and Oral Expression

Written and oral expression are primarily concerned with the production of language output and jointly measure a student's ability to communicate with words. Oral language establishes the basis for written language; therefore, weaknesses in oral expression are likely to emerge comparably in written expression. Despite these similarities, they also involve distinctive skills that develop independent of one other. Written expression exclusively involves letter formation, word formation, and text formation whereas oral expression encompasses vocabulary, syntactic and semantic knowledge, memory, comprehension, and storytelling (Hatcher et al., 2017).

Error analyses for Written Expression and Oral Expression subtests are somewhat parallel, developed to break down the speaking and writing processes to determine where a student exhibits weakness. Error categories identify the pragmatic aspects of the task (producing comprehensible and functionally effective writing or speech that adheres to the task demands), the structural aspects (ability to construct well-structured sentences), and the word-level morphological and grammatical aspects (using the correct forms of words). Written Expression includes a fourth category, which addresses the mechanical aspects of writing (capitalization and punctuation). Comparing a student's errors across skillsets can help the examiner to differentiate a specific writing weakness from a global language problem and identify specific skills on which to focus intervention.

Math-Related Analysis

Math Concepts and Applications

Math concepts are the ideas and relationships on which the system of mathematics is founded, and from which all applications are derived. For the Math Concepts and Applications subtest, items are classified into error categories based on the concept or application required to successfully solve the problem. In earlier grades, error categories include classifying objects into sets and identifying quantitative concepts (*more than, less than,* etc.). Primary grades include skills such as place value and regrouping, while errors for secondary students include linear functions, quadratic equations, and hypothetical problems such as reading graphs.

Math Computation

For the Math Computation subtest, the error analysis system addresses both item-level and within-item errors. Item-level errors are automatically categorized according to the math domains in which the student is weak, average, or strong (e.g., addition, subtraction, multiplication, division, fractions, decimals, and exponents or root). Examiners can also manually categorize within-item errors based on the specific processes that have not been mastered (wrong operation, fact or computation error, regrouping for both addition and subtraction, converting the common denominator, etc.).

Qualitative Observations

Qualitative observations are intended to provide information based on the student's testing behavior that can aid the examiner in developing, confirming, or refuting hypotheses about factors affecting test performance. According to Kranzler and Floyd (2013), Qualitative indicators allow the examiner to monitor and record potential construct-irrelevant influences on item-level performance by completing a brief checklist of notable behaviors after each subtest. Behaviors that can disrupt performance may include the failure to self-monitor or sustain attention, impulsive responding, perseverating despite feedback, task refusal, reluctant to commit to a response when uncertain, and worrying about time limits. Behaviors that may enhance performance might include closing one's eyes to concentrate, requesting repetitions, perseverance, trial-and-error, hyperfocus, verbalizing related knowledge or strategies for recall, and working quickly but carefully (Kranzler & Floyd, 2013).

On the KTEA-3, qualitative observations are available for 15 subtests and provide indicators of potential cognitive processing weaknesses organized by IDEIA domain. Cross-validating information suggested by the qualitative observations with other sources of assessment data (e.g., standard scores, error analyses, and measures of cognitive processing) can significantly enhance clinical judgment in the test interpretation process.

Recent Research with the KTEA-3

By systematically reviewing the performance of clinical groups on tests of intelligence and achievement, practitioners can observe specific cognitive profiles and error patterns that may emerge consistently for students with specific diagnoses or skill deficits. Breaux and Lichtenberger (2016) presented substantial research that focused on identifying cognitive and achievement profiles across different clinical groups. More recently, Alan Kaufman and colleagues published a group of studies using data from the KTEA-3 standardization sample to analyze the specific kinds of errors that students make in reading, writing, math, and oral language (Breaux, Bray, Root, & Kaufman, 2017). Results from this research are discussed below that may inform test selection, enhance clinical judgment, and improve tailored intervention design (Breaux et al., 2017).

Cognitive Processing and Academic Achievement

Difficulties in reading, writing, and math may be directly related to common cognitive processes and the shared neurological mechanisms that are needed for the development of these skills (Berninger & Richards, 2010). Numerous studies focus on identifying relationships between cognitive and academic achievement profiles. In one study, researchers analyzed score patterns of children with reading disabilities on the KTEA-3 and KABC-II (which were conormed). When a sample of 141 students ages 6–18 with SLD-R were compared to a matched, nonclinical reference group, both subtest and composite scores were significantly lower for children with SLD-R than for the reference group on both the KTEA-3 and KABC-II (Kaufman & Kaufman, 2004a, 2004b). The largest difference between the nonclinical group and the SLD-R group was found on the KABC-II Learning/Glr scale (with 1 *SD* difference between mean scores). Kaufman, Lichtenberger, Fletcher-Janzen, and Kaufman (2005) posit that the Learning/Glr scale is most demanding because it requires that all cognitive processes work together. Hence, it makes sense for students with SLD-R to achieve lower scores than a matched group on this subtest.

A similar pattern was observed in children with writing disabilities (SLD-W). In a study using the KABC-II, a sample of 122 students ages 6–18 with SLD-W were found to have significantly worse scores on all KABC-II scales than the nonclinical reference group (which was a sample matched on gender, race, and parent education). The lowest index score for the SLD-W group was also found on the Learning/Glr scale. Like the previous study, the Learning scale may have yielded the lowest for this group due to the large cognitive demand of both written expression and Glr tasks (Kaufman et al., 2005).

Breaux and Lichtenberger (2016) compared a sample of 96 students ages 6–8 with math disabilities (SLD-M) to a matched control group on both the KABC-II and the KTEA-3. As in the sample of students with SLD-R, scores on all scales were significantly lower for the group with SLD-M than for the matched control group. The largest standard score difference between the group with SLD-M and the control group (about 16 points) was on the Planning/Gf scale. Some studies have found that children with disabilities in math can be helped by implementing a plan of remediation related to planning and fluid reasoning (Rourke, 1989; Teeter & Semrud-Clikeman, 1998).

Specific Learning Disability

Reading disorders are often comorbid with disorders of written expression. Another study examined performance on the KTEA-3 by a sample of 67 students in grades 1–12 with SLD-R and/or SLD-W (Kaufman & Kaufman, 2014b). Results indicated that about 40% of the students were identified as having a specific learning disability in reading only, 5% (three students) had a disorder in written expression only, and a little more than half of the subjects had a disorder in both reading and writing. The KTEA-3 scores of the SLD R/W group were compared to those of a matched control group. The SLD R/W group scored significantly lower than the comparison group on all KTEA-3 subtests and composite scores. Results further support the comorbidity of reading and writing disorders.

Avitia, Pagirsky, and colleagues (2016) sought to compare error types made by children diagnosed with SLD R/W, children diagnosed with language impairment (LI), and two demographically matched control groups. Results suggest that children with SLD R/W and LI exhibit distinct patterns of academic errors, providing evidence that the two groups have distinct academic profiles. The children with SLD R/W made significantly more errors in word reading and decoding than the matched control group, particularly on error factors concerning contextual vowel pronunciation and letter-sound knowledge. The children with LI differed from their control group on intermediate sound knowledge. Both groups performed significantly worse than matched controls on tasks of phonological processing, reading comprehension, and written expression. Both groups also performed significantly lower than matched controls on tasks of listening comprehension and oral expression, although the authors reported that this was unexpected. The children with SLD R/W demonstrated significantly more errors compared with their control group on the error factors of math calculation and miscellaneous math concepts, whereas no differences were found between the children with LI and their matched control group on any of the math error factors. Results suggest that a student with SLD R/W may require more individualized interventions to address vowel pronunciations that require context and basic letter-sound knowledge, whereas students with LI may require stronger support in learning consonant digraphs and blends.

Koriakin and Kaufman (2017) sought to identify error patterns associated with specific reading disability profiles by comparing four groups of students with differentially diagnosed SLD-R-basic reading difficulties (BRD), reading fluency difficulties (RFD), reading comprehension difficulties (RCD), and typical readers with no formal diagnosis. No significant differences were found on any error factors between the RFD and RCD groups. However, compared to typical readers, the RCD group demonstrated significant difficulties on both error factors while the RFD group did not. This suggests that the RCD group was more affected by language-related difficulties than any other group, providing evidence for a significant relationship between language and reading comprehension skills. This also reinforces the importance of measuring language comprehension as part of any evaluation for dyslexia or SLD-R. On the Oral Expression error factors (which include task-oriented mechanics of speaking and general oral expression), the RFD group made significantly more errors than typical readers, while the RCD group did not. This indicates that children with reading fluency difficulties may experience difficulties with oral expression, but not reading comprehension. The RFD group also demonstrated a specific pattern of weakness in spelling problems compared to typical readers, which may be a function of phonetic deficits, which often co-occur with poor fluency (Koriakin & Kaufman, 2017).

Students identified with specific learning disabilities in math (SLD-M) also present with unique profiles. In one study using the KTEA-3 (Kaufman & Kaufman, 2014a), 51 students in grades 1–12 with SLD-M were administered the KTEA-3 Comprehensive Form. Compared to a matched control group, students with SLD-M scored significantly lower on all KTEA-3 subtests and composites. Approximately 15% of the SLD-M group had a comorbid reading disability, which Breaux and Lichtenberger (2016) suggest may have contributed to the deficits this group displayed on reading and writing tasks. The SLD-M group demonstrated their lowest score on the overall Math composite.

In another study examining SLD-M, the WIAT-III was administered to a sample of 90 students in grades 2–12, ages 7–19 (Breaux & Lichtenberg, 2016). About 9% of the group with SLD-M also had a diagnosis of SLD-R, which contributed to significant weaknesses in reading and writing subtests consistent with other research. When the group with SLD-M was compared to a matched control group, students with SLD-M earned the lowest mean scores on the math subtests.

A recent study was designed to evaluate the types of errors made by students with disabilities in reading and/or writing (SLD-R/W) compared to errors made by students with a SLD in math (SLD-M). In all academic areas assessed (reading, writing, language, and math), Avita, DeBiase, and colleagues (2016) found more similarities than differences between the two groups. The SLD-M group performed lower within some error categories that were not related to their area of disability when compared with the SLD-R/W group. These results support the clinical utility of error analysis procedures for these two clinical populations, as those with learning disabilities in reading and writing may require some assistance in math, and vice versa.

Attention-Deficit/Hyperactivity Disorder

ADHD is estimated to affect about 5% of children in the United States (American Psychiatric Association, 2013). Approximately one out of three children diagnosed with ADHD also meet criteria for SLD, and these rates may be higher among children who exhibit written language difficulties (Barkley, 2015). However, the relationship between ADHD and SLD remains complicated and the directionality is not quite clear. Symptoms of ADHD affect a child's ability to work productively in the classroom and perform well on academic material due to their inattention, restlessness, or impulsive behavior (Barkley, 2015). Barkley (2015) has also noted that children with ADHD are more likely to require academic tutoring, repeat a grade, or be placed in special education classes.

For school psychologists, it is important to understand the specific skill deficits exhibited by children with ADHD so that appropriate interventions can be put in place. On standardized tests of academic achievement, the ADHD population tends to earn standard scores that are 10–30 points lower than their peers in the areas of reading, spelling, math, and reading comprehension (Breaux & Lichtenberger, 2016). One study using the WIAT-III found that the lowest scores earned by the ADHD clinical group were on the Written Expression, Spelling, and Written Language subtests. This group's highest scores were on Listening Comprehension, Oral Expression, and the Oral Language Composite. Of interest, there were large gaps between typical students and those dually diagnosed with both ADHD and SLD. The lowest scores for the group with both ADHD and SLD were on Written Language, which requires sustained effort, planning, self-monitoring, and organizing ability (Breaux & Lichtenberger, 2016).

Another study measured performance on the WIAT-II by children with ADHD, combined type (ADHD-C) and ADHD, inattentive type (ADHD-I), and compared their achievement scores with those of students who were either nonreferred children or children who had been referred for testing for other reasons than ADHD (McConaughey, Ivanova, Antshel, & Eiraldi, 2009). When the two groups with ADHD were compared to the nonreferred comparison group, children in these two groups scored significantly lower on all three WIAT-II composites (Reading, Mathematics, and Written Language). The group with ADHD-C group did not significantly differ from the group with ADHD-I. The authors concluded that the

WIAT-II effectively separated those clients with ADHD (either type) from those who did not have the disorder, but that the tool may not be useful in distinguishing between the subtypes of ADHD. However, it should be noted that behavioral observations during assessment were found to be an effective way to distinguish these two groups.

In another study, the KTEA-3 was administered to a sample of 91 students in grades K–12 with ADHD only (no LI or SLD concerns) and their scores were compared to a matched sample of students with no diagnosed disability. Although many of the KTEA-3 composite scores were not significantly lower for the group with ADHD, the Written Language, ASB, Oral Language, and Oral Fluency composites were significantly lower for this group than for the control group (Breaux & Lichtenberger, 2016). However, these composite scores for the group with ADHD were still within the average range when compared to scores earned by other SLD sample groups, whose composite scores tended to be more than 1 *SD* lower than those of the matched controls.

The Comprehensive Assessment of SLD in the Schools

The purview of school psychological practice is ultimately dictated by professional codes of ethics, federal and state regulations, and school district policies. Approaches to special education assessments must be aligned with legal and ethical standards set forth by the following regulations:

General Education Law

- No Child Left Behind (NCLB; 2012)
- Every Student Succeeds Act (ESSA; 2015)
- Americans with Disabilities Act (ADA; 2008)
- Family Educational Rights and Privacy Act (FERPA; 1974)

Special Education Law

- IDEIA (2004)
- State special education regulations
- District policies

Professional Codes of Ethics

- American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (2016)
- National Association of School Psychologists' Principles for Professional Ethics (2010)

Testing Standards

According to the NASP (2011) position statement on SLD, the primary purposes of a comprehensive evaluation are to determine whether a child has an SLD and to make recommendations regarding educational placement and instructional interventions. NASP supports the following best practices: Evaluations must include a variety of assessments and other evaluation methods, must not be discriminatory on a racial or cultural basis, and must be administered in the language and form most likely to yield accurate information; measures must be used only for purposes for which they demonstrate reliability and validity; they must be administered by trained and knowledgeable personnel and in accordance with instructions provided by the test producer; and they must encompass all areas of suspected disability. After analyzing the codes of ethics, professional standards, and federal laws that address psychological assessment, five broad ethical-legal concerns resulted. School psychologists must strive to ensure that psychoeducational evaluations are multifaceted, comprehensive, fair, valid, and useful (Jacob, Decker, & Lugg, 2016).

The Role of Intelligence in SLD Assessment

Cognitive abilities and academic achievement exist as highly related, but distinct constructs. A growing body of research supports an empirical relationship among specific academic skills, cognitive abilities, and neuropsychological processes (Decker, Hale, & Flanagan, 2013; Dehn, 2013a; Miller, 2013). This is especially salient among the school-age population (Breaux et al., 2017). However, experts consistently disagree over whether it is necessary to include measures of intelligence as part of a learning disability evaluation. This debate can be misleading and problematic for school psychologists, as federal regulations require that any student who is referred for a special education evaluation must be assessed in all areas of suspected disability, including health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities (34 CFR § 300.304[c][4]). Additionally, federal definitions of SLD emphasize that a student must demonstrate average intellectual ability overall (*g*), but present specific weaknesses in certain cognitive processes that directly support academic learning (Toffalini et al., 2017). To practice in accordance with these guidelines, school psychologists are strongly encouraged to include measures of intelligence as a standard component of SLD assessments.

General Intelligence and the Intelligence Quotient (IQ)

Misunderstanding and potential misuse of intelligence tests frequently occur. This happens when cognitive ability scores are treated as measures of innate capacity when intelligence does not truly exist as a measurable trait. General intelligence is not housed somewhere in the brain, much like a person's "athletic ability" is not actually located within the body or how a car's maximum speed does not exist somewhere *in* the car (Kievit et al., 2012). Intelligence essentially reflects differences between people, so it can't really be "in" a person; rather, it represents a hypothesized statistical concept that scientists have inferred based on individual differences in how large groups of individuals perform on cognitive and achievement tasks (S. B. Kaufman et al., 2012).

Furthermore, the concepts of general Intelligence (*g*) and IQ are not synonymous. Accurate interpretation of test scores requires that school psychologists understand the difference. A student's general intelligence, or *g*, is essentially a latent trait, whereas their overall cognitive ability (traditional IQ) and academic achievement are observable and measurable outcomes (Horton & Reynolds, 2015). Kievit and colleagues (2012) compare cognitive abilities to physical properties, such as "height," whereas *g* is more akin to a construct, such as "physical fitness," that is dependent on a range of physical properties working together (e.g., lung capacity, metabolism, etc.).

Theories of Intelligence

Scientific theories allow us to make predictions about phenomena that has yet to be observed. Ultimately, intelligence tests have been used to define the theory of intelligence that the test is intended to measure (Goldstein & Cunningham, 2009). School psychologists should use approaches to link assessment to intervention that are supported by contemporary clinical practice, theory, and research.

Cognitive Abilities versus Psychological Processes

Within intelligence theories, the terms *process* and *ability* are often used interchangeably which is inconsistent with widely held definitions of cognition (Dehn, 2006). Cognitive abilities (Gf, Gc, Gwm, etc.) are essentially behaviors that can be observed as people complete similar tasks that are grouped together because they require common neurobiological proficiency for successful performance. Cognitive abilities are measurable factors that have been hypothesized to make predictions about the underlying brain functions that facilitate thinking. Psychological processes refer to the actual neurobiological mechanisms that are responsible for cognitive(conscious) and emotional (unconscious) thinking.

Psychometric Theories of Cognitive Ability: CHC

Cattell–Horn–Carroll (CHC) theory is considered the most empirically valid psychometric model for understanding human intelligence and the most widely used framework for developing and interpreting cognitive assessments (Cormier, McGrew, Bulut, & Funamoto, 2017; Keith & Reynolds, 2010; Schneider & McGrew, 2012). (For a comprehensive description of the current CHC taxonomy, see Schneider & McGrew, Chapter 3, this volume.) The KTEA-3 provides coverage of 20 narrow CHC abilities and eight broad abilities: fluid reasoning (Gf), comprehension–knowledge (Gc), long-term storage and retrieval (Glr), processing speed (Gs), auditory processing (Ga), reading and writing (Grw), quantitative knowledge (Gq), and psychomotor speed (Gps) (Kaufman & Kaufman, 2014b, p. 6).

The broad and narrow abilities defined by in CHC theory are essentially labels that refer to the many unseen steps that are required for completing a task. Cognitive abilities reflect the sum, or outcome, of underlying psychological processes. However, the CHC psychometric taxonomy of abilities cannot possibly account for all the factors that affect performance during an assessment. Ultimately, it is not important to determine how high or low an ability is; rather, it is what a student does with the ability that counts. The way in which an examinee completes a task *matters*. As such, CHC theory best serves school psychologists as a starting point when trying to figure out what a student's test score actually means.

Neuropsychological Theories of Cognitive Processing

When an individual is given any task to complete—educational, adaptive, or otherwise—an array of cognitive or neuropsychological processes are required to meet the demands successfully. Consider the mental tasks that might be required for a student who is asked to solve a simple oral arithmetic problem:

- Focus attention on the problem as it is presented (attention).
- Plan a strategy for problem solving; choose a strategy to use; initiate application of the strategy (executive functions).
- Transcribe the text into a mathematical equation (fluid reasoning).
- Remember arithmetic facts required to solve the problem (long-term retrieval).
- Check the solution for accuracy (executive functions).
- Express the response in the required modality (oral language).
- Hold information in mind long enough to solve the problem (working memory).

A task that seemingly lasts only a few seconds involves countless numbers of brain-based processes to solve. A deficit in any one of these processes might make the problem difficult or even impossible for the student to complete. Furthermore, students can arrive at answers and solutions to problems—either correct or incorrect—by employing a variety of different strategies. Variations in input, processing, and/or output demands can greatly affect performance on tasks involving identical or similar content. The only way to truly gauge what a subtest is measuring is to consider how the child performs.

Approximately 50 years ago, cognitive psychologists proposed a theory of mental processing and learning, known as *information processing theory*, that "described the flow of information through sensation, perception, cognition, memory, and expression" (Dehn, 2013a, p. 120; Dehn, 2013b). Since then, research in neuropsychology has greatly enhanced our understanding of how the brain functions during learning and memory. Based on the principles of neurodevelopment, neuropsychological theories examine the specific neurological processes that can be localized in the brain within the context of its typical developmental trajectory.

A plethora of research supports the fact that neuropsychological processes play a critical role in a person's ability to learn academic skills (Dehn, 2013a). Practitioners who employ these methods are primarily concerned with stimulus inputs that must be processed for a person to complete a task, as well as the ways in which information is apprehended, encoded, stored, organized, retrieved, and mentally manipulated (Jensen, 2006). Neuropsychological approaches to interpretation deemphasize the importance of understanding a student's performance relative to peers in favor of interpreting the student's intrapersonal pattern of processing strengths and weaknesses. Many neuropsychological approaches to interpretation have been developed that focus on cognitive processing, information processing, and task/demands analysis, including:

Lurian- and PASS-based approaches
Edith Kaplan's process approach
Milton Dehn's (2013b) processing assessment method
Hale and Fiorello's cognitive hypothesis testing (CHT)
Dan Miller's (2013) school neuropsychological approach

These approaches are discussed at length within other chapters of this text. School psychologists should use theoretical models of intelligence and related assessment approaches as a framework for designing a comprehensive SLD assessment. In general, school psychologists should never be wedded to a single instrument and the unique array of subtests they include in the comprehensive assessment of a student should be systematic. The following validated methodologies have been developed to aid in selecting the best assessment tools based on the school psychologist's theoretical orientation(s) and the student's referral concerns:

Dawn Flanagan's XBA (based on CHC psychometric theory) (Flanagan, McGrew, & Ortiz, 2000)
Jack Naglieri's (2015) PASS model (based on Luria's neuropsychological processing theory)
Sally Shaywitz's (2003) approach to diagnosing dyslexia
George McCloskey's interpretative system via process analysis (McCloskey, Whitaker, Murphy, & Rogers, 2012)

Flanagan and colleagues (2017; see also Chapter 22, this volume) have proposed an operational definition of SLD that is particularly useful for school psychologists because is reflects the most current federal regulations, theory, and research most salient to our field, including (1) the nature of SLD, (2) the methods of evaluating various elements and concepts inherent in SLD definitions, and (3) criteria for establishing SLD as a specific condition that cannot be explained by undifferentiated low achievement or below-average overall cognitive functioning. Primarily grounded in the CHC theory of intelligence, this model is referred to as the dual discrepancy/consistency (DD/C) operational definition of SLD. The DD/C encourages a continuum of data-gathering methods, beginning with curriculum-based measurements and progress monitoring and culminating in a comprehensive evaluation for students who do not respond adequately to high-quality instruction and intervention within general education. The DD/C model provides a framework for organizing data from multiple sources to evaluate whether an individual's PSW is consistent with the SLD construct.

Synthesizing Data from the KTEA-3, WISC-V, and WISC-V Integrated

A chapter appendix provides an interpretive case study that integrates assessment of academic achievement, cognitive abilities, and neuropsychological processes in a comprehensive assessment for intervention for a child referred for difficulties in mathematics. The KTEA-3 was specifically designed to be paired with measures of cognitive functioning to identify a student's unique profile of strengths and weaknesses. These relationships are important for both diagnosis and intervention planning. The two other major batteries used in this case study are the WISC-V (Wechsler, 2014) and the WISC-V Integrated (Wechsler & Kaplan, 2015). (For comprehensive reviews of the WISC-V and the WISC-V Integrated, see Wahlstrom, Raiford, Breaux, Zhu, & Weiss, Chapter 9, and Raiford, Chapter 11, this volume, respectively.)

APPENDIX 29.1

Illustrative Case Report

IDENTIFYING AND DEMOGRAPHIC INFORMATION

Name: Joseph M.

Date of birth: 3/10/2006

Age: 12 years, 2 months

Race/ethnicity: European American

Parents: Mr. and Mrs. M.

School: C. Middle School

Grade: 7

Date of Testing: 5/16/2017

Date of Report: 5/21/2017

Examiner's name: Susan Engi Raiford, PhD

REFERRAL QUESTIONS AND ANSWERS

Mr. and Mrs. M. requested an assessment of their son, Joseph, because he has received low grades in math in the past and because C. Middle School's admissions testing indicated that his math skills were below expectations for his grade. Testing is being done to document the specific nature of these difficulties, so that accommodations can be made and an intervention plan developed before he enters middle school. Joseph's parents wish to learn the answers to the following questions:

- Does Joseph have a specific learning disability in mathematics? If yes, what skills are impaired?
- What are Joseph's cognitive strengths and needs?
- What accommodations are needed?
- What other recommendations can be used in developing a plan to meet Joseph's educational needs?

ANSWERS TO REFERRAL QUESTIONS

Does Joseph Have a Specific Learning Disability in Mathematics?

Yes. I have diagnosed Joseph with specific learning disorder with impairment in mathematics: number sense, memorization of arithmetic facts, and accurate calculation, moderate; and fluent calculation, severe.

What Are Joseph's Cognitive Strengths and Needs?

Joseph's abilities to access, apply, and express knowledge he has gained about words and their meanings, and to reason with verbal material, are personal strengths relative to his other cognitive abilities. Joseph's visual–spatial processing—that is, his ability to evaluate visual details and to understand relationships of visual parts in space, and to use that information to assemble a geometric design that matches a model (either pictured or real) —is a personal weakness relative to his other cognitive abilities. He also has a personal weakness in naming facility (efficiency), which involves recognizing and recalling overlearned information (e.g., letters and numbers or quantities) as quickly as possible.

What Accommodations Are Needed?

Joseph's diagnosis of specific learning disorder in mathematics qualifies as a disability as defined by the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), which governs free public school education. Accommodations for this condition, if available within the private school environment, are appropriate. Refer to the accommodations listed in the "Recommendations" section of this report. Some of the most common accommodations would involve giving additional time to complete work or assessments related to math, allowing use of a calculator, providing content mastery support, and decreasing reliance on visual materials when teaching math. The accommodation plan should be discussed and agreed upon in a follow-up meeting with parents, educators, and Joseph himself.

What Intervention Recommendations Can Be Used to Develop a Plan to Meet His Educational Needs?

Joseph may benefit from math fact drills, math problem attack strategies, and summer or after-school math programs.

EVALUATION METHODS AND PROCEDURES

- History and background review
- Parent interview (Note: Joseph is coming from a home-schooling situation)
- Child interview
- Behavioral observations
- Review of admissions testing results
- Psychological testing

Psychological tests administered were as follows (see "Complete Test Data" at the end of this report for full results):

- Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V)
- Wechsler Intelligence Scale for Children—Fifth Edition Integrated (WISC-V Integrated)
- Kaufman Test of Educational Achievement, Third Edition (KTEA-3)

HISTORY OF PRESENTING PROBLEM

Joseph has good academic success in subjects other than mathematics. Since he began memorizing basic math facts as a first grader, however, he has struggled to learn math. He needs a great deal of support to keep up. He tends to learn math skills and move on, but at a review later he cannot reproduce the use of the skills he has learned. He can learn a skill in the moment and at this point has little trouble memorizing math facts.

Later in elementary school, math reasoning and word problems were difficult for Joseph; drawing pictures of problems to solve them was helpful. If he didn't draw pictures, such problems were very difficult or impossible for him. Mental arithmetic is very challenging for Joseph, as are changes to the problem or requirements of a task.

FAMILY HISTORY

Mr. and Mrs. M., Joseph's biological parents, have been married for 15 years. Joseph has one brother who is 3 years older and does not have academic difficulties. Joseph's mother struggled with math throughout school. She is beginning work as a preschool teacher after several years as a stay-at-home mother. Joseph's father presently works as an information technology professional in the private sector.

MEDICAL/DEVELOPMENTAL HISTORY

Mrs. M.'s pregnancy with Joseph was normal. Joseph was born after a full-term pregnancy with normal delivery. He had no health problems after his birth. He walked at 16 months (at the higher end of normal limits). He met his language developmental milestones on time and was toilet-trained on time.

ACADEMIC HISTORY AND STATUS

At age 3, Joseph began to recognize numbers, letters, colors, and shapes. He began reading well at age 5. Mrs. M. stated that his lowest grade last year was in math (78), and that he reads and writes at above grade level.

A review of Joseph's achievement test scores and his home-schooling records confirmed that his reading and writing both were at the seventh-grade level. He has a good vocabulary, and his spelling and grammar skills are average. Joseph states that he wants to be a police officer when he grows up.

PATIENT'S STRENGTHS, COPING MECHANISMS, AND AVAILABLE SUPPORT SYSTEMS

Joseph enjoys soccer and baseball and is on teams for both throughout the year. He is involved in his church's youth group. He also enjoys building model planes with his father.

MENTAL STATUS AND BEHAVIORAL OBSERVATIONS

Joseph presented dressed casually, with good grooming and hygiene. He appeared slightly older than his stated age. His rate of speech was normal. He was appropriately behaved and well mannered. He answered quickly when he seemed confident about an answer, and became quieter when he appeared less sure. His eye contact was good. His affect was congruent with stated mood, which was "OK."

Joseph was aware that the evaluation was to be used to help him in middle school and to understand better how he learns. He seemed clear about the purpose of testing and was self-disclosing with me, opening up about his problems with math. Rapport was easily established and maintained. Joseph put forth good effort on all tasks and persevered on challenging tasks. I believe that the test findings provide good estimates of his true abilities. Two breaks were offered and given, to ensure that Joseph could give his best effort and not become fatigued.

TEST RESULTS AND INTERPRETATION

Cognitive Functioning

Intellectual Ability

Joseph's performance on cognitive ability measures suggests that his overall intellectual ability is in the average range compared with that of other young people his age. In large studies, out of all the scores on cognitive tests, overall intellectual ability is the best score at predicting children's school success. However, it doesn't always tell us everything we need to know about a child's individual strengths and needs. An individual's cognitive picture is usually better and more deeply understood when narrower areas of ability are also considered.

Language

Speech

Joseph responded readily to items that required expressive responses. He elaborated sufficiently.

Verbal Comprehension

Joseph's ability to access, apply, and express knowledge he has gained about words and their meanings, and to reason with verbal material, is in the high average range and is a personal strength (relative to his other cognitive abilities). Because his verbal skills and vocabulary are relatively strong compared with his other cognitive abilities, teachers may expect him to perform math-related tasks more easily than he does in reality.

Visual–Spatial Processing

As noted above, visual–spatial processing is the ability to evaluate visual details and to understand relationships of visual parts in space, and to use such information to assemble a geometric design that matches a model (either pictured or real). Sometimes the expression of this ability is tested through manipulation of real objects.

Joseph's visual–spatial processing ability is in the low average range compared with that of other children his age, and is a personal weakness (relative to his other abilities). Such a weakness is common in children who have problems learning math, because learning math requires the ability to see objects and lines in one's head and use one's imagination to understand them and how they are constructed or related.

Joseph's difficulty in this area is similar on a task that involves trial-and-error problem solving and physical (motor) manipulation of components of a construction, and on a task that involves mental imagery. He performed slightly better on a task that required him to recognize correct solutions among multiple choices than to construct them himself.

Interestingly, assembling model planes is one of Joseph's favorite hobbies. This activity should be encouraged, to help him enjoy and feel confident about activities in the visual-spatial realm.

Reasoning and Problem Solving

Reasoning is the ability to detect and apply the underlying rules or relationships that define how objects or ideas are understood as a group. Joseph's visual reasoning ability is average compared with that of other children his age. He had greater difficulty on problems that required multiple operational steps or division than on those that merely required simple addition, subtraction, and multiplication. However, he was slow to complete the problems; even on very easy items, he often hesitated more than is typical and worked up to the time limit before responding. He also requested that items be repeated more frequently than is typical of other children. He was still working at the time limit for several problems he did not receive credit for.

Giving more time and allowing Joseph to use pencil and paper on the math problems did not assist him in better expressing quantitative reasoning, relative to his own prior performance or to that of his same-age peers. He continued to request a greater number of repetitions than typical of other children until he reached the phase in the task for which he received the pencil and paper. On a task for which the problem-solving load is reduced to requiring only simple responses to written math problems, he responded correctly to *fewer* items. He meticulously worked the math problems. Some of the items he missed were simple calculation errors at the last step, whereas his work was off track from the first step on others.

Learning and Memory

Learning and memory were assessed to examine their role in, and implications for, Joseph's future academic success. The two are closely related to each other and are very important to school achievement. Joseph's performance on a broad measure of his long-term storage and retrieval, which involves tasks that require accurate and fluent retrieval of overlearned associations as well as accurate retrieval of new associations, is in the low average range.

Working Memory

Working memory is the ability to take in, keep, and manipulate information in one's awareness to get some type of output that can be expressed. Joseph's working memory ability is in the average range relative to that of other children his age. Because he showed low visual–spatial processing skills, his visual working memory skills (i.e., visual stimuli and spatial locations) were also assessed, and they are also in the average range. However, his performance on a task that involved remembering and tracing spatial locations was significantly lower than one that involved remembering meaningful visual material in sequence. Joseph's working memory performance was consistent across items rather than variable, which suggests that his attention was consistent during the tasks.

Associative Memory

Joseph's visual–verbal associative memory, or the ability to form new associations between symbols and meanings, is in the average range relative to that of other young people his age. His performance suggests uniform ability when he is recalling associations he learned within the past few seconds or minutes, or when he is trying to recall them half an hour later. He performs better when he is permitted to select the association from among options read aloud, rather than recall them from memory without a cue half an hour later.

Rapid Automatic Naming

Rapid automatic naming involves recognizing and recalling overlearned information, like letters and numbers or quantities, as efficiently as possible. Overall, Joseph's rapid automatic naming is in the very low range. His ability to recognize, recall, and recite letters and numbers quickly is in the low average range relative to that of other children his age. His ability to rapidly recognize and name quantities is less efficient, and is in the very low range compared with that of others his age.

Cognitive Speed

Processing Speed

Processing speed is the ability to make speedy and accurate judgments about visual information and act on those judgments. Slow processing speed can lower academic performance, because information that cannot be processed quickly tends to be lost. Joseph's processing speed is in the low average range relative to that of other children his age, and is a personal weakness relative to his other cognitive abilities. He scored lower on a task that required him to scan symbols within a group to see if they match two other symbols or to indicate the symbol is not present, than he did on a subtest that required him to quickly associate and write codes that were paired in a key with numbers. He also scored somewhat lower on a task that required him to scan a page of objects for objects from a single meaningful category (i.e., animals).

Rapid Automatic Naming

Rapid automatic naming is also sometimes thought of as an aspect of cognitive speed; it has been discussed above under "Learning and Memory."

Achievement Skills

Math Skills

Math Computation

Math computation includes skills with the most basic building blocks of math (like addition and subtraction), all the way through more complicated skills like using fractions, decimals, algebra, roots and exponents, signed numbers, and so on. Joseph's skills in this area are in the low average range relative to those of other young people his age. He had more difficulty than is typical on problems involving multiplication, multistep problems, and word problems.

Math Fluency

Math fluency involves quickly answering simple math problems using basic operations like addition, subtraction, multiplication, and division that should be mastered by the end of elementary school. Joseph's math fluency is in the very low range relative to that of others his age.

Math Problem Solving

Math reasoning involves the application of reasoning and mathematical concepts to solve meaningful problems. Joseph's math reasoning skills are in the average range relative to those of other young people his age. He had more difficulty than is typical on items involving simple calculations and basic math operations.

SUMMARY AND DIAGNOSTIC IMPRESSIONS

Summary

Joseph qualifies for a diagnosis of specific learning disorder in math (see "Diagnostic Impressions" below for details). He is able to achieve at his present level with great support and effort, as often seen with specific learning disorders. He will need to receive assistance from a math specialist as he makes the transition to more difficult coursework in middle school.

There are various approaches to diagnosing specific learning disabilities. A traditional methodology, known as *ability–achievement discrepancy*, involves examining the difference between an indicator of cognitive ability (e.g., WISC-V Full Scale IQ) and a measure of achievement (e.g., KTEA-3 Math composite).

A newer method involves examining the child's cognitive ability and achievement results to determine whether there is a processing strength (e.g., WISC-V Verbal Comprehension Index) that is discrepant with from both an area of achievement (e.g., KTEA-3 Math composite) and from a processing weakness (e.g., WISC-V Naming Speed Index) that is empirically or theoretically associated with a specific learning disability in that skill. The Verbal Comprehension Index has served as the processing strength in Joseph's case, and it is not theoretically associated with math disabilities. It is discrepant from the KTEA-3 Math composite and also from the Naming Speed Index, which is commonly impaired in children with math disability. In particular, poor performance on the subitizing task (Naming Speed Quantity) is commonly found in children with mathematics problems in the published literature.

Regardless of the model used, the test results support the hypothesis that Joseph has a specific learning disability in mathematics.

Diagnostic Impressions

F81.2: Specific learning disorder with impairment in mathematics: number sense, memorization of arithmetic facts, and accurate calculation, moderate; and fluent calculation, severe

RECOMMENDATIONS

Several possible accommodations may be effective:

- Allow Joseph additional time (50% more) to complete math in-class work, quizzes, and tests, as well as standardized math achievement tests.
- Allow use of a calculator unless a test is explicitly designed to test math computation.
- Content mastery support is recommended for math.
- Support from a math specialist outside school hours and over the summer is recommended.
- Due to Joseph's low visual–spatial ability, use visual materials related to math sparingly, and supplement with verbal/auditory materials.
- Encourage Joseph to make his own flashcards to practice addition, subtraction, multiplication, and division facts and become more fluent. Various math quiz apps are also available; these will enable Joseph to practice math facts and complete problems more easily.
- While Joseph's math reasoning is presently at the low end of the average range, his math fluency and math computation skills are low. He could benefit from a streamlined and systematic approach to solving math problems: Teach him to analyze math problems, select a strategy, and then monitor his problem solving to ensure it is completed correctly. For each of the steps below, have Joseph say the purpose, ask what needs to be done, and check to ensure it has been done correctly. At home, ensure that he knows the calculations necessary to solve each problem and correct errors as needed. The steps for Joseph to use are as follows:
  1. Read the problem carefully, and clear up anything that is uncertain, such as words or math terms you don't understand.
  2. Restate the problem in your own words.
  3. Draw or create a visual to represent the problem.
  4. Decide how to best solve the problem, and create a plan to do it.
  5. Estimate the answer to the problem, using shortcuts such as rounding.
  6. Use the plan to calculate the answer.
  7. Check the calculation of each step, and compare the obtained answer to the estimated answer from step 6 to make sure the answer is reasonably accurate.

Thank you for the opportunity to work with Joseph. Please let me know if I can be of further assistance.

COMPLETE TEST DATA

Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) and WISC-V Integrated

Subtest Score Summary

Index	Subtest name	Total raw score	Scaled score
Verbal Comprehension	**Similarities**	37	15
	Vocabulary	30	10
Visual Spatial	**Block Design**	22	7
	Block Design Multiple Choice[a]	17	9
	Visual Puzzles	13	7
Fluid Reasoning	**Matrix Reasoning**	18	8
	Figure Weights	23	10
	(Arithmetic)	19	8
	Arithmetic Process Approach Part A[a]	19	9
	Arithmetic Process Approach Part B[a]	19	9
	Written Arithmetic[a]	18	9
Working Memory	**Digit Span**	26	10
	Picture Span	33	11
	Spatial Span[a]	15	8

Index	Subtest name	Total raw score	Scaled score
Verbal Comprehension	**Coding**	37	15
	Symbol Search	23	6
	Cancellation	56	7
Complementary subtests			
Naming Speed	Naming Speed Literacy	52	83
	Naming Speed Quantity	32	77
Symbol Translation	Immediate Symbol Translation	65	91
	Delayed Symbol Translation	47	91
	Recognition Symbol Translation	29	99

Note. Subtests used to derive the Full Scale IQ are given in boldface. Secondary subtests are given in parentheses. Results on most subtests are reported as scaled scores. Average scaled score is 10. About 68% of scores fall between 7 and 13. About 95% fall between 4 and 16. Complementary subtest scores are reported as standard scores. Average standard score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

[a]WISC-V Integrated subtest.

Process Observations

	Raw score	Base rate
Arithmetic item repetitions	10	0.5
Arithmetic Part A item repetitions	3	1.3

Composite Score Summary

Composite	Sum of scaled/ standard scores	Composite score	Percentile rank	95% confidence interval	Qualitative description
Verbal Comprehension Index (VCI)	**25**	**113**	**81**	**104–120**	**High average**
Visual Spatial Index (VSI)	**14**	**84**	**14**	**78–93**	**Low Average**
Fluid Reasoning Index (FRI)	**18**	**94**	**34**	**87–102**	**Average**
Quantitative Reasoning Index (QRI)	18	94	34	88–101	Average
Working Memory Index (WMI)	**21**	**103**	**58**	**95–110**	**Average**
Visual Working Memory Index (VWMI)	19	97	42	89–106	Average
Symbol Translation Index (STI)	281	92	30	86–99	Average
Storage and Retrieval Index (SRI)	170	80	9	74–88	Low Average
Processing Speed Index (PSI)	**15**	**86**	**18**	**79–97**	**Low Average**
Naming Speed Index (NSI)	160	78	7	72–89	Very low
Full Scale IQ (FSIQ)	***69***	***99***	***47***	***93–105***	***Average***

Note. Boldface, primary index score. Italics, global composite score. Average composite score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Form A

Composite/Subtest Score Summary

Composite/subtest	Subtest raw score	Sum of subtest standard scores	Standard score	90% confidence interval	Percentile rank	Descriptive category
Math Composite	—	171	84	80–88	14	Below average
Math Concepts and Applications	58	—	90	85–95	25	Average
Math Computation	39	—	81	75–87	10	Below average
Math Fluency	17	—	78	71–85	7	Very low

Note. All KTEA-3 scores are reported as standard scores. Average standard score is 100. About 68% of scores fall between 85 and 114. About 95% fall between 70 and 129.

Error Analysis

	Math Concepts and Applications (last item administered: 70)				Math Computation (last item administered: 47)			
Error category	Items attempted	Average # of errors	Student's # of errors	Skill status	Items attempted	Average # of errors	Student's # of errors	Skill status
Number concepts	22	0–1	1	A				
Addition	3	0	0	A	14	0	1	W
Subtraction	3	0	0	A	12	0	4	W
Multiplication	3	0	2	W	7	0–1	1	A
Division	3	0–1	1	A	4	0–1	2	W
Tables and graphs	3	0	0	A				
Time and money	9	0–1	2	W				
Geometry	2	0–1	1	A				
Measurement	7	0–2	2	A				
Fractions	4	0	0	A	1	0	0	A
Decimal					—	—	—	—
Decimals and percents	2	0–1	1	A				
Data investigation	2	0–1	1	A				
Multistep problems	4	0–2	4	W				
Word problems	11	0–1	3	W				
Exponent or root					1	0–1	1	A
Algebra	7	0–1	1	A	—	—	—	—
Wrong operation					38	0–1	0	A
Fact or computation					38	0–2	0	A
Regrouping: Addition					2	0	0	A
Regrouping: Subtraction					4	0	0	A
Subtract smaller from larger					4	0	0	A
Add or subtract numerator and denominator					1	0	0	A

	Math Concepts and Applications (last item administered: 70)				Math Computation (last item administered: 47)			
Error category	Items attempted	Average # of errors	Student's # of errors	Skill status	Items attempted	Average # of errors	Student's # of errors	Skill status
Equivalent fraction/ common denominator					—	—	—	—
Multiply/divide fraction					—	—	—	—
Mixed number					—	—	—	—
Incorrect sign					—	—	—	—
Uncodable					38	0	0	A

Note. A, average; W, weakness.

Ability–Achievement Discrepancy

KTEA-3 subtests/composite	Predicted KTEA-3 score	Actual KTEA-3 score	Difference	Critical value (.05)	Significant difference?	Base rate
Math Concepts and Applications	109	90	19	8	Yes	≤5%
Math Computation	107	81	26	8	Yes	≤2%
Math Fluency	—	78	—	—	—	—
Math Composite	108	84	24	8	Yes	≤2%

Pattern of Strengths and Weaknesses Analysis

- Area of processing strength: WISC-V Verbal Comprehension Index (113)
- Area of processing weakness: WISC-V Naming Speed Index (78)
- Area of achievement weakness: KTEA-3 Math (84)

Comparison	Relative strength score	Relative weakness score	Difference	Critical value (.05)	Significant difference?	Supports SLD hypothesis?
Processing strength/ achievement weakness	113	84	29	10	Yes	Yes
Processing strength/ processing weakness	113	78	35	11	Yes	Yes

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Psychological Association. (2016). Revision of ethical standard 3.04 of the "Ethical Principles of Psychologists and Code of Conduct" (2002, as amended 2010). *American Psychologist, 71,* 900.

Americans with Disabilities Act Amendments of 2008. The Americans with Disabilities Act of 1990 is codified at U.S.C. §§ 12101 et seq.

Avitia, M., DeBiase, E., Pagirsky, M., Root, M. M., Howell, M., Pan, X., et al. (2016). Achievement error differences of students with reading versus math disorders. *Journal of Psychoeducational Assessment, 35*(1–2), 111–123.

Avitia, M., Pagirsky, M., Courville T., DeBiase, E., Tawnya Knupp, T., & Ottone-Cross, K. (2016). Differences in errors between students with language and reading disabilities. *Journal of Psychoeducational Assessment, 35*(1–2), 149–154.

Barkley, R. A. (2015). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed.). New York: Guilford Press.

Berninger, V. W., & Abbot, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102,* 635–651.

Berninger, V., & Richards, T. (2010). Inter-relationships among behavioral markers, genes, brain, and treatment in dyslexia and dysgraphia. *Future Neurology, 5*(4), 597–617.

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39,* 327–335

Breaux, K. C. (2017). See the error, see the individual—and see the intervention. Retrieved from *www.socialsciencespace.com/2017/01/see-error-see-individual-see-intervention.*

Breaux, K. C., Bray, M. A., Root, M. M., & Kaufman, A. S. (2017). Introduction to special issue and to KTEA-3 error analysis. *Journal of Psychological Assessment, 35*(1–2), 4–6.

Breaux, K. C., & Lichtenberger, E. O. (2016). *Essentials of KTEA-3 and WIAT-III assessment.* Hoboken, NJ: Wiley.

Carretti, B., Motta, E., & Re, A. M. (2016). Oral and written expression in children with reading comprehension difficulties. *Journal of Learning Disabilities, 49*(1), 65–76.

Choi, D., Hatcher, R. C., Dulong-Langley, S., Liu, X., Bray, M. A., Courville, T., et al. (2017). What do phonological processing errors tell about students' skills in reading, writing, and oral language? *Journal of Psychoeducational Assessment, 35*(1–2), 24–46.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment, 35*(8), 731–754.

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools, 50*(3), 300–313.

Dehn, M. J. (2006). *Essentials of processing assessment.* Hoboken, NJ: Wiley.

Dehn, M. J. (2013a). Enhancing SLD diagnoses through the identification of psychological processing deficits. *Australian Educational and Developmental Psychologist, 30*(2), 119–139.

Dehn, M. J. (2013b). *Essentials of processing assessment* (2nd ed.). Hoboken, NJ: Wiley.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Every Student Succeeds Act of 2015 20 U. S. C. § 1118 (2015).

Family Educational Rights and Privacy Act of 1974, 20 U.S.C.A. § 1232g. Regulations appear at 34 C.F.R. Part 99

Filter, K. J., Ebsen, S., & Dibos, R. (2013). School psychology crossroads in America: Discrepancies between actual and preferred discrete practices and barriers to preferred practice. *International Journal of Special Education, 28*(1), 88–100.

Flanagan, D. P., Mascolo, J. T., & Alfonso, V. C. (2017). Utility of KTEA-3 error analysis for the diagnosis of specific learning disabilities. *Journal of Psychoeducational Assessment, 35*(1–2), 226–241.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guilford Press.

Goldstein, S., & Cunningham, S. (2009). Current issues in the assessment of intelligence, specific learning disability, and attention deficit hyperactivity disorder. In J. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 11–26). Hoboken, NJ: Wiley.

Hatcher, R. C., Breaux, K. C., Liu, X., Bray, M. A., Ottone-Cross, K. L., Courville, T., et al. (2017). Analysis of children's errors in comprehension and expression. *Journal of Psychoeducational Assessment, 35*(1–2), 57–73.

Horton, A. M., & Reynolds, C. R. (2015). Common and variable aspects of intelligence. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 367–380). New York: Springer.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1400 (2004).

Jacob, S., Decker, D. M., & Lugg, E. T. (2016). *Ethics and law for school psychologists* (7th ed.). Hoboken, NJ: Wiley.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Childre—Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement—Second Edition (KTEA-II), Comprehensive Form.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014a). *Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Comprehensive Form.* Bloomington, MN: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2014b). *Manual for the Kaufman Test of Educational Achievement, Third Edition (KTEA-3).* Bloomington, MN: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2015). *Kaufman Test of Educational Achievement, Third Edition (KTEA-3), Brief Form.* Bloomington, MN: Pearson.

Kaufman, A. S., Kaufman, N. L., Drozdick, L. W., & Morrison, J. (2018). *Kaufman Assessment Battery for Children, second edition, normative update, manual supplement.* Bloomington, MN: NCS Pearson.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment.* Hoboken, NJ: Wiley.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. (2012). Are cognitive *g* and academic *g* one and the same *g*?: An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123–138.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650.

Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., et al. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience, 3*(2), 89–97.

Koriakin, T. A., & Kaufman, A. S. (2017). Investigating patterns of errors for specific comprehension and fluency difficulties. *Journal of Psychoeducational Assessment, 35*(1–2), 138–148.

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents.* New York: Guilford Press.

Liu, X., Marchis, L., DeBiase, E., Breaux, K. C., Courville, T., Pan, X., et al. (2017). Do cognitive patterns of strengths and weaknesses differentially predict the errors on reading, writing, and spelling? *Journal of Psychoeducational Assessment, 35*(1–2), 186–205.

Maki, K. E., Burns, M. K., & Sullivan, A. (2017). Learning disability identification consistency: The impact of methodology and student evaluation data. *School Psychology Quarterly, 32*(2), 254–267.

Mather, N., & Wendling, B. J. (2017). Implications of error analysis studies for academic interventions. *Journal of Psychological Assessment, 35*(1–2), 215–225.

McCloskey, G. (2017). Error analysis: Past, present, and future. *Journal of Psychoeducational Assessment, 35*(1–2), 242–250.

McCloskey, G., Hartz, E., & Slonim, J. (2016). Interpreting the WISC-V using George McCloskey's neuropsychologically oriented process approach to psychoeducational evaluations. In A. S. Kaufman, S. Raiford, & D. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 493–496). Hoboken, NJ: Wiley.

McCloskey, G., Whitaker, J., Murphy, R., & Rogers, J. (2012). Intellectual, cognitive, and neurological assessment in three-tier service delivery systems in schools. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 852–882). New York: Guilford Press.

McConaughy, S. H., Ivanova, M. Y., Antshel, K., & Eiraldi, R. B. (2009). Standardized observational assessment of attention deficit hyperactivity disorder combined and predominantly inattentive subtypes: I. Test session observations. *School Psychology Review, 38*(1), 45–66.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Naglieri, J. A. (1999). *Essentials of CAS assessment.* New York: Wiley.

Naglieri, J. A. (2015). Hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D Princiotta, & J. A. Naglieri (Eds.). *Handbook of intelligence: evolutionary theory, historical perspective, and current concepts* (pp. 295–316). New York: Springer.

National Association of School Psychologists (NASP). (2011). *Position statement: Identification of students with specific learning disabilities.* Bethesda, MD: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *The Common Core Standards for English language arts and literacy in history/social studies, science, and technical subjects.* Washington, DC: Authors.

Nielsen, K., Andria-Habermann, K., Richards, T., Abbot, R., Mickail, T., & Berninger, V. (2017). Emotional and behavioral correlates of persisting specific learning disabilities in written language during middle childhood and adolescence. *Journal of Psychological Assessment.* [Epub ahead of print]

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2002).

Parkin, J. R. (2018). Do achievement factors change across development?: An investigation with the KTEA-3. *Journal of Psychoeducational Assessment.* [Epub ahead of print]

Pearson. (2009). *Examiner's manual: Wechsler Individual Achievement Test—Third Edition.* San Antonio, TX: Author.

Rourke, B. P. (1989). *Nonverbal learning disabilities: The syndrome and the model.* New York: Guilford Press.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level.* New York: Vintage Books.

Teeter, P. A., & Semrud-Clikeman, M. (1998). *Child clinical neuropsychology: Assessment and interventions for neuropsychiatric and neurodevelopmental disorders of childhood.* Boston: Allyn & Bacon.

Toffalini, E., Giofre, D., & Cornoldi, C. (2017). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder. *Clinical Psychology Science, 5*(2), 402–409.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D., & Kaplan, E. (2015). *Wechsler Intelligence Scale for Children—Fifth Edition Integrated.* Bloomington, MN: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement.* Itasca, IL: Riverside.

PART VI

Contemporary and Emerging Issues in Intellectual, Cognitive, and Neuropsychological Assessment

CHAPTER 30

Using the Joint Test Standards to Evaluate the Validity Evidence for Intelligence Tests

Alyssa Montgomery
Erica Torres
Jamie Eiseman

Anyone involved with assessment has at least two main legal and ethical responsibilities: (1) to administer and interpret tests correctly, and (2) to use tests that are adequate for their intended purpose. Since 1966, the *Standards for Educational and Psychological Testing* volume, or one of its revisions, has been considered the "gold standard" in guidance on testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Collaboratively prepared by AERA, APA, and the National Council on Measurement in Education (NCME), the most recent edition of the *Standards* (2014) is a product of the Joint Committee for the Revision of the Standards for Educational and Psychological Testing. The purpose of the *Standards* is to "promote sound testing practices and to provide a basis for evaluating the quality of these practices" (AERA et al., 2014, p. xx).

For the 2014 *Standards,* the Joint Committee responsible for overseeing the revisions was tasked with examining five specific areas: "consider the accountability issues for use of tests in educational policy; broaden the concept of accessibility of tests for all examinees; represent more comprehensively the role of tests in the workplace; broaden the role of technology in testing, and provide for a better organizational structure for communicating the standards" (AERA et al., 2014, p. 3). To meet the goals of considering accountability, broadening accessibility, comprehensively representing tests in the workplace, broadening technology in testing, and restructuring to better communicate the standards, several changes were made in the 2014 edition. One important change involved the issue of fairness. In 1999, fairness was woven into multiple chapters of the *Standards.* In 2014, by contrast, one chapter was dedicated specifically to fairness. The authors of the 2014 *Standards* believed that writing a separate chapter about fairness would allow for a broader application of it, rather than applications only for subgroups of test takers (e.g., individuals with disabilities, individuals from culturally and linguistically diverse backgrounds). Although subgroups are referenced frequently, the authors did not want readers to think that these are the only groups to which fairness considerations apply. Despite the wide range of considerations presented throughout the 2014 *Standards,* the present chapter, like the chapters by Braden and Niebling (2005, 2012) in previous editions of *Contemporary Intellectual Assessment,* focuses on issues related to validity—of which fairness is one.

According to the *Standards,* "Validation evidence allows for a summary judgment of the intended interpretation that is well supported and defensible" (AERA et al., 2014, p. 22). The guiding principle regarding the standards for validity is that tests provide "clear articulation of each intended test interpretation . . . and appropriate validity evidence in support of each intended interpretation" (p. 23). As such, we have integrated and analyzed the evidence available to support the validity of several tests, based on the *Standards*' guiding principles and individual themes. We have used the original conceptual framework of validity set forth by Braden and Niebling (2005, 2012) and applied that framework when determining evidence of validity based on the *Standards*.

Evaluating an intelligence test's validity involves a summary judgment of the collective evidence. Validity "cannot be determined by using a checklist" (AERA et al., 2014, p. 6). Therefore, this chapter outlines examples of evidence that can be considered in assessing validity, and provides clear and explicit examples of a test's validity found in the technical manuals of current measures of intelligence and cognitive abilities. This chapter is not intended to serve as an exhaustive and comprehensive list of evidence for each test, but rather as a guide to help clinicians evaluate a test's validity. Other issues to be considered by the reader, as outlined by the *Standards,* include: (1) When a test publisher claims to have made efforts to adhere to the *Standards,* the publisher should offer supporting evidence; (2) the reader should understand that the *Standards* do not set forth particular methods of statistical reporting, so a "generally accepted equivalent" (p. 6) can be implied and used as support; and (3) while the *Standards* make notes regarding relevant legal requirements, legal consultation should always be sought when appropriate.

Ensuring the validity or accuracy of any assessment is "the most fundamental consideration in developing tests and evaluating tests" (AERA et al., 2014, p. 11). Messick (1989) defines validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p. 5). Validity can be thought of in terms of the extent to which a test measures what it intends to measure. Ensuring and evaluating test validity are the responsibilities of test developers and test users, respectively. For test developers, creating tests that accurately reflect the constructs intended to be measured is critical to a test's success. However, this chapter is meant primarily for test users, and is intended to serve as a reference for how to interpret the validity of the intelligence tests and the scores obtained from them. Users of these instruments maintain a responsibility to be able to say, with confidence, that the results of a particular test can be considered accurate and that the interpretations based on those results are relevant.

To provide examples of how to interpret the validity of intelligence tests, we have selected the following tests, all of which are described at length in separate chapters in this book: the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014a: see Wahlstrom, Raiford, Breaux, Zhu, & Weiss, Chapter 9, this volume); the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014a; see Naglieri & Otero, Chapter 15, this volume); the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014; see Schrank & Wendling, Chapter 14, this volume); and the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007a; see Elliott, Salerno, Dumont, & Willis, Chapter 13, this volume). To identify evidence regarding each test's validity, information was gathered from their respective technical manuals and, in some cases, their administrative and scoring manuals. The following areas were assessed for each test.

TEST CONTENT

One source of evidence that practitioners should consider when evaluating the validity of a psychological test is its content validity. Critical validity evidence can be taken from "an analysis of the relationship between the content of a test and the construct it is intended to measure" (AERA et al., 2014, p. 14). *Test content* is defined as "the themes, wording, and format of the items, tasks, or questions on a test, as well as the guidelines for procedures regarding administration and scoring" (AERA et al., 2014, p. 14). Content includes the development of the instrument, including the specific steps that were designed and used in the process of creating it (Sullivan, 2011). With regard to the validity of test content, the items should cover the constructs that the test authors claim it measures (Huck, 2012). According to Goodwin and Leech (2003), test content validity is based on "logical analyses and experts' evaluations of the content of the measure, including items, tasks, formats, wording, and processes required of examinees" (p. 183). In other words, this type of validity is used to assess whether the content of a test represents a specific content domain accurately.

When evaluating an intelligence test's content validity, users should consider the test's "sufficiency, clarity, relevancy, and the match between the items and tasks" (Goodwin & Leech, 2003, p. 183). The test should provide the user with a clear definition of the broad and narrow constructs that the test intends to measure, as highlighted in Standard 1.1 (AERA et al., 2014). For example, the DAS-II clearly defines its overall composite score, General Conceptual Ability (GCA), as "the general ability of an individual to perform complex mental processing that involves conceptualization and the transformation of information" (Elliott, 2007b, p. 17). To further support its content validity, the DAS-II cluster scores and subtests are also clearly defined for the reader throughout Chapter 2 of the test's introductory and technical handbook (Elliott, 2007b).

In addition to clear definitions of constructs, it is important to consider the theoretical basis of the test structure and the relationship of the individual items to the specific domains (Sireci & Faulkner-Bond, 2014). This information helps the user better understand the foundation upon which the interpretations are made. These elements are also discussed in Standard 1.2, which states, "The rationale should indicate what propositions are necessary to investigate the intended interpretation. The summary should combine logical analysis with empirical evidence to provide support for the test rationale" (AERA et al., 2014, p. 23). Again, we use the DAS-II as an example: Elliott (2007b) explains that the test was designed to follow Cattell–Horn–Carroll (CHC) theory (Schneider & McGrew, 2012; see Schneider & McGrew, Chapter 3, this volume, for a complete explanation of current CHC theory). Seven broad factors (i.e., comprehension–knowledge [Gc], fluid reasoning [Gf], visual processing [Gv], short-term memory [Gsm] or short-term working memory [Gwm], long-term storage and retrieval [Glr], cognitive processing speed [Gs], and auditory processing [Ga]) that "appear to be the most robust and replicable" (Elliott, 2007b, p. 13) are incorporated into the test, and their relationships to the individual subtests are explained.

Finally, the appropriateness of the test development process should be evaluated, as explained in Standard 1.11: Information related to the "procedures followed in specifying and generating test content should be described and justified" (AERA et al., 2014, p. 26). As an example, the WISC-V technical manual (Wechsler, 2014b, pp. 37–41) outlines the development stages of this fifth edition of the test, including the use of an advisory panel, expert research, and feedback from professionals to guide its development.

RESPONSE PROCESS

Response process, according to the *Standards,* refers to the process utilized by a test taker before providing his or her response. The consideration of response process as a component of validity was first introduced into the *Standards* in the 1985 revision (Padilla & Benítez, 2014). Standard 1.12 of the 2014 revision of the *Standards* states:

> If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided. (AERA et al., 2014, p. 26)

Analyzing an individual's response process allows the examiner to ascertain "the fit between the construct and the detailed nature of performance or response actually engaged in by test takers" (AERA et al., 2014, p. 15). Did the examinee approach the items in a way that makes sense, given the nature of what the item is purported to measure? In the 2014 *Standards,* the example is given that if an examinee is working on a segment about mathematical reasoning, analyzing the response process will highlight "whether test takers are, in fact, reasoning about the material given instead of following a standard algorithm" (p. 15).

Questioning the examinee is one form of demonstrating evidence of response processes. Other methods of demonstrating evidence of this type of validity may include drafts throughout the writing process, response time, and eye movement (AERA et al., 2014). Methods such as analyzing multiple drafts of the same writing assignment, documenting how long it takes for an individual to respond, or tracking eye movements can be broken into two categories, according to Padilla and Benítez (2014): "those that directly access the psychological processes or cognitive operations (think aloud, focus group, and interviews), compared to those which provide indirect indicators which in turn require additional inference (eye tracking and response times)" (p. 139).

One example of how a test might employ this response process comes from the CAS2 Planning subtests (Naglieri et al., 2014a). On these subtests, guidelines for observing and quantifying strategies used by the test taker are provided. A Strategy Assessment Checklist is printed in the examiner's record form, offering brief descriptions of various approaches often employed. Additionally, there is a section on the checklist to document observed/reported strategies not included on the checklist. On the CAS2, questioning of the examinee takes place after a Planning subtest is completed, and examples of such queries are "Tell me how you did these," or "How did you find what you were looking for?" (Naglieri, Das, & Goldstein, 2014b, p. 7). These queries are designed to explore how the examinee arrived at his or her answers and to shed light on the meaning of obtained scores.

In addition to learning about how a test taker approaches various tasks, analyzing response processes helps identify what abilities, apart from those assumed to be needed based on what the items are intended to measure, may contribute to an examinee's performance. Response process studies from different subgroups can help itemize such abilities, and provide valuable information regarding variability in score interpretation (AERA et al., 2014).

INTERNAL STRUCTURE

Another type of validity evidence to be considered is *internal structure*. Rios and Wells (2014) outline three types of internal structure that can be assessed: *dimensionality, measurement invariance,* and *reliability*. This type of evidence is necessary to determine "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (AERA et al., 2014, p.16). Analyzing the internal structure of a test gives test developers and test users a better understanding of whether the various test components (test items, composites, etc.) measure what was intended. Therefore, gathering evidence of the internal structure of a test assists in making well-grounded inferences based on test results.

Dimensionality is a component of a test's internal structure that relates to whether the relationships among subtest items support suggested inferences. Factor analysis is commonly used to illustrate the relationship among subtest components. Standard 1.13 states that when a test reportedly is unidimensional, it is "supported by a multivariate statistical analysis, such as a factor analysis, showing that the score variability attributable to one major dimension was much greater than the score variability attributable to any other identified dimension, or showing that a single factor adequately accounts for the covariation among test items" (AERA et al., 2014, p. 27). Conversely, when tests are multidimensional and more complex, the relationships between the scores "should be shown to be consistent with the construct(s) being assessed" (p. 27). Multidimensional tests that yield composite scores should provide the user with additional information related to the rationale for and interpretation of these subscores. As stated in Standard 1.14, evidence should be provided that supports the subscores' relationship to the construct and demonstrates that they can be validly and reliably interpreted.

Measurement invariance, or equivalence, is another type of internal structure that should be evaluated carefully. Standard 3.6 states: "Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups" (AERA et al., 2014, p. 65). Because current intellectual assessments are administered to individuals from many different racial, cultural, geographic, or educational backgrounds, it is important that the tests are fair for all. Analyzing potential bias within a test allows clinicians to determine whether results are a fair representation of a client's skills, or whether his/her group membership influenced the score. One way to assess the fairness of a test across different groups is to evaluate the differential item functioning (DIF) (AERA et al., 2014, p. 51). Although DIF provides information regarding whether the scores of particular groups are significantly different, it does not, by itself, determine whether a test is biased (see Ortiz, Piazza, Ochoa, & Dynda, Chapter 25, this volume, for a discussion).

The final type of internal structure to be assessed is a test's reliability. The *Standards* volume includes a separate chapter on reliability because it is recognized as an independent characteristic of scores that must be considered. Before a practitioner can make statements about a test's validity, its reliability must be evaluated because it has implications for validity (AERA et al., 2014, p. 34). Simply put, if a test is not reliable, it cannot be valid for any purpose.

RELATIONS TO OTHER VARIABLES

When test users are examining the evidence of validity, it is important to consider whether the test predicts an expected outcome, as well as whether the test results relate to the results of other tests thought to measure a similar construct (AERA et al., 2014). In other words, would similar results be found, regardless of which intelligence test was administered? Test results can be compared to some "gold standard" of measurement, to determine whether the scores from one test correlate as expected with scores from another. When the relationships between tests are explored, three specific types of evidence are considered: (1) convergent and discriminant evidence, (2) test–criterion relationships, and (3) validity generalization.

Convergent evidence demonstrates relationships between measures that are expected to assess similar constructs (e.g., intelligence tests compared to intelligence tests), while *discriminant* evidence presents the lack of relationships between measures that assess dissimilar constructs (e.g., intelligence tests compared to behavior rating scales). An example of convergent evidence on the CAS2 is found by examining two subtests in the Attention composite scale, Expressive Attention and Number Detection. Both are intended to measure the construct of attention, but responses for the Expressive Attention subtest are oral, while the responses for the Number Detection subtest are written. In theory, the subtests should correlate highly, since they are both measuring attention. Discriminant evidence, for example, might be found by comparing the Planned Codes and Word Series subtests of the CAS2. The Planned Codes subtest is part of the Planning scale, which measures a test taker's ability to create a plan, follow it, reflect on whether the plan worked, and alter the plan accordingly, while the Word Series subtest is part of the Successive scale, which assesses an examinee's capacity to make sense of or understand information presented in a strict order. Although these two subtests should correlate (they are both assumed to be measuring some aspect of intelligence), they should have lower correlations to each other (they are assumed to be measuring different aspects of intelligence) than to subtests assessing the same construct of intelligence. Reviewing a correlation matrix and finding that the correlations between the two subtests are dissimilar would be evidence of discriminant validity.

For example, Naglieri and colleagues (2014b) compared the CAS2 and the CAS2: Brief. They found that the two assessments were "highly correlated (Full Scale r = .78 for Core and .80 for Extended Batteries)" (p. 88), and asserted that each scale of the PASS correlated highest with the same scale on the other assessment (e.g., the Planning composite on the CAS2 had the highest correlation with the Planning subtest on the CAS2: Brief).

In the 2014 *Standards,* Standard 1.16 states that when other variables are being considered with test results, "the rationale for selecting the additional variables should be provided" (AERA et al., 2014, p. 19). On the CAS2, for example, Naglieri, Das, and Goldstein (2014c) itemized disorders (e.g., attention-deficit/hyperactivity disorder [ADHD], anxiety disorder, autism spectrum disorder [ASD]) as other variables considered. Although these authors did not provide a clear rationale for why the specific variables were included, one can assume that this was done because these are among the disorders most commonly encountered by school practitioners. When examining the group with ASD, Naglieri and colleagues stated that those with an ASD diagnosis have been shown to have deficits in shifting their attention. An attention deficit was supported by the CAS and CAS2 results, as well as by results from independent researchers. Specifically, the group with ASD performed most poorly on the CAS2 Attention scale, which supports the validity of the Attention scale.

Figures 30.1 through 30.4 provide examples of validity evidence that were collected from the four intelligence tests referenced in this chapter: the WISC-V, WJ IV COG, DAS-II, and CAS2, respectively. Although these figures do not exhaust all evidence for each test, they provide examples of the types of information necessary to examine when evaluating a test's validity. In addition, Table 30.1 provides a few examples of how these four intelligence tests compare across various types of validity evidence.

IMPLICATIONS

Using the principles set forth by the 2014 *Standards,* we have reviewed the most recent editions of four contemporary intelligence tests in this chapter, and have discussed the evidence that the authors of these tests provided for them. Test developers have a responsibility to provide the information necessary for a practitioner to assess a test's validity; however, it is imperative that test users understand how to interpret and evaluate

Cross-cutting standards (Standards 1.1, 1.2, 1.3, 1.4, 1.8, 1.10, 1.23)

- Score interpretation and intended use are defined. (pp. 14–21)
- Research on the theories and models are addressed. (pp. 22–28)
- Factor analysis is discussed. (pp. 77–84)
- The Full Scale and the primary, ancillary, and complementary indexes are described. (pp. 20–21)
- Clear subtest descriptions are provided. (pp. 7–14)
- Neurodevelopment research, specific cognitive abilities, and working memory models support the rationale for test interpretation. (pp. 23–28)
- Guidelines for interpretation are provided. (pp. 149–186)
- Users are cautioned to use the results as part of a decision-making process, not as a sole factor. (pp. 147, 157, 186)
- Test developers explicitly indicate the norms upon which the test was used. (pp. 38, 39)
- Norming sample is clearly defined and relates to the target population. (pp. 42, 43)
- The standardization and development of the normative sample are clearly explained. (p. 31)

Test content (Standards 1.9, 1.11)

- Subtests and indexes are described in detail. (pp. 7–15)
- The updated structure and subtest modifications are explained, and the purpose of the changes is supported. (p. 6)
- An advisory panel was assembled in order to provide feedback regarding the clinical utility and appropriateness of the items. (p. 38)

Response processes (Standards 1.7, 1.12, 1.25)

- Test developers provide evidence to support the intelligence models and research that contributed to the foundation of the test. (pp. 22, 70)
- Response results were examined to ensure that the test was measuring its intended purposes. (p. 70)
- Specifically, responses were assessed to investigate the possibility of unintentionally providing a correct response, and then changes were made. (p. 70)

Internal structure (Standards 1.13, 1.14, 1.15)

- Intercorrelations for subtest, process, and composite scores are provided for all ages. (pp. 72–77)
- Evidence is provided for interpretation of composite scores, as well as subscores and score differences. (pp. 71–73)
- Confirmatory factor analyses provide evidence for the five-factor model chosen by test developers. (pp. 77–78)
- Evidence is provided to support the interpretation of four levels: the Full Scale, primary index, ancillary index, and complementary index levels. (pp. 20, 77, 78, 157–159, 165, 166)

Relations to other variables (Standards 1.16, 1.17, 1.18, 1.19, 1.20, 1.21, 1.22)

- The WISC-V developers examined two types of external variables: scores on other instruments designed to measure the same or similar constructs, and results for special groups.
- Evidence about correlations is provided with regard to other intelligence tests, achievement tests, adaptive scales, and rating scales.
- Special groups that were examined included children identified as intellectually gifted, children with mild or moderate intellectual disability, children with borderline intellectual functioning, children with specific learning disorders, children with ADHD, children with disruptive behavior, children with traumatic brain injury, children who were English-language learners, and children with autism spectrum disorder. Limitations of these studies are explained.
- Data suggest that the current edition correlates highly with the prior edition. (pp. 85–147)

(continued)

FIGURE 30.1. Examples of WISC-V validity evidence. All page numbers refer to the WISC-V technical and interpretive manual (Wechsler, 2014b).

Fairness, bias, and consequences of testing (Standards 1.5, 1.6, 1.7, 1.10, 1.17, 1.18, 1.19, 1.22, 1.24, 1.25)

- WISC-V scales have been evaluated using score differences between groups; race, ethnicity, sex, and socioeconomic status were specifically examined. (p. 147)
- Subtests and items were assessed for potential bias by expert examiners. Items that were deemed to be problematic were identified. (p. 32)

FIGURE 30.1. *(continued)*

Cross-cutting standards (Standards 1.1, 1.2, 1.3, 1.4, 1.8, 1.10, 1.23)

- Descriptions of tests and clusters are clearly and comprehensively explained and supported by their relation to the CHC theory. (Mather & Wendling, 2014, pp. 13–25)
- CHC abilities are defined. Implications for demonstrated strengths/weaknesses of skills are provided. (Mather & Wendling, 2014, pp. 7, 21–24, 84–96, 101)
- Evaluating individual strengths and weaknesses is encouraged, and procedures for doing so are provided. (Mather & Wendling, 2014, pp. 7, 92)
- Several types of scores can be interpreted, depending on the purpose of the assessment. (Mather & Wendling, 2014, pp. 77–84)
- Four levels of interpretation are encouraged: qualitative, level of development, proficiency, and relative standing in a group. (Mather & Wendling, 2014, pp. 75–77)
- When modifications are necessary to accommodate the test taker, the administrator is provided with recommendations, cautions, and implications for interpretation. (Mather & Wendling, 2014, pp. 41–53)
- The normative sample includes over 7,416 people ages 2–90. A detailed description of the sample (including age, sex, race, ethnicity, geographic location, and level of parent education) assists users in determining the test's appropriateness. (McGrew, LaForte, & Schrank, 2014, p. 1)

Test content (Standards 1.9, 1.11)

- Multidimensional scaling (MDS) was used to assess expert judgments in measuring content validity. (McGrew et al., 2014, pp. 129–134)
- All broad and narrow CHC factors are clearly described and explained. (McGrew et al., 2014, pp. 120–125)
- **Response processes (Standards 1.7, 1.12, 1.25)**
- Theoretical support and evidence are provided for the use of CHC theory. (McGrew et al., 2014, pp. 1–9)
- Broad and narrow abilities within the CHC model are identified and explained for use in interpretation. (Mather & Wendling, 2014, pp. 21–25, 84–90)

Internal structure (Standards 1.13, 1.14, 1.15)

- Cluster analysis is provided. (McGrew et al., 2014, pp. 143–149)
- Test and cluster correlations are provided for all age groups. (McGrew et al., 2014, pp. 143–149)
- To ensure that all items in each test measured the same narrow ability or trait, stringent fit criteria based on the Rasch model were employed during the process of item pool development and test construction. (McGrew et al., 2014, pp. 44–46)
- Broad- and narrow-ability constructs are justified by the CHC theory. (McGrew et al., 2014, pp. 120–125)
- Guidelines for interpreting performance are provided, which include four levels of test information to consider. Again, these levels are known as qualitative, level of development, proficiency, and relative standing in a group. (Mather & Wendling, 2014, pp. 75–77)
- Scores can be interpreted by grade equivalents, age equivalents, relative proficiency indexes, cognitive–academic language proficiency levels, percentile ranks, and standard scores. (Mather & Wendling, 2014, pp. 77–84)

(continued)

FIGURE 30.2. Examples of WJ IV COG validity evidence.

Relations to other variables (Standards 1.16, 1.17, 1.18, 1.19, 1.20, 1.21, 1.22)

- In the technical manual, results of 15 studies are charted to illustrate the relationships among other cognitive measures, oral language measures, and achievement measures. (McGrew et al., 2014, pp. 220–223)
- A three-stage model of structural validity is employed. (McGrew et al., 2014, pp. 149–167)
- Nine clinical groups are clearly identified, and group studies were completed to identify differences across groups. These groups included individuals with giftedness; intellectual disability/mental retardation; learning disabilities in reading, writing, or math; language delay; ADHD; head injury; and ASD. (McGrew et al., 2014, pp. 221–222)
- Data suggest that the current edition correlates highly with the prior edition. (McGrew et al., 2014, p. 2)

Fairness, bias, and consequences of testing (Standards 1.5, 1.6, 1.7, 1.10, 1.17, 1.18, 1.19, 1.22, 1.24, 1.25)

- Outside experts were consulted to ensure construct representation and limit confounding variables. (McGrew et al., 2014, pp. 43–44)
- Bias and sensitivity issues were evaluated for women, individuals with certain disabilities, and cultural or linguistic minorities. (McGrew et al., 2014, pp. 53–56)
- Differential item functioning was evaluated by using the Rasch iterative-logit method. Items were flagged if the difference was considered significant, and were removed from the test as appropriate. (McGrew et al., 2014, pp. 53–56)

FIGURE 30.2. *(continued)*

Cross-cutting standards (Standards 1.1, 1.2, 1.3, 1.4, 1.8, 1.10, 1.23)

- The normative sample includes 3,480 children between the ages of 2 and 17. Information on sex, race, ethnicity, level of parental education, and geographic location is included. The sample is broken down by group and compared to the U.S. population. Inclusion and exclusion criteria for the participants are included to further define the normative sample. (Chapter 6)
- The DAS-II follows a hierarchical structure and is influenced by various theories and perspectives, including CHC theory and neurodevelopmental research. (pp. 8–16)
- Subtests are related to associated broad and narrow CHC abilities to aid in interpretation. (pp. 9–11)
- A process of analysis is suggested, involving stages of comparison (e.g., core comparisons, diagnostic comparisons, ability–achievement comparisons). (Chapter 5)
- Examiners are encouraged to utilize multiple sources of information when interpreting the results of the evaluation. (p. 36)

Test content (Standards 1.9, 1.11)

- Clusters are described and related to CHC abilities, and interpretive considerations are provided. (Chapter 4)
- For each subtest, specific information is provided, including the subtest's purpose, contributing factors, interpretive considerations, and (for some subtests) alternate methods of administration. (Chapter 4)

Response processes (Standards 1.7, 1.12, 1.25)

- Specific clusters are supported by factor analysis. (pp. 153–158)
- An alternative stop point rule was incorporated, due to the probability of a child's providing an accurate response after a certain amount of incorrect responses. (p. 4)
- Rasch scaling was used to determine scoring. (pp. 115–118)

(continued)

FIGURE 30.3. Examples of DAS-II validity evidence. All chapter and page numbers refer to the DAS-II introductory and technical handbook (Elliott, 2007b).

Internal structure (Standards 1.13, 1.14, 1.15)

- Intercorrelations of subtests and composites are illustrated and comprehensively explained. (pp. 242–259)
- Subtest and composite correlation coefficients are provided for three age groups (2:6–3:5, 3:6–6:11, 7:0–17:11). (pp. 150–158)
- Confirmatory factor analysis is provided for core and diagnostic subtests. (pp. 158–162)
- CHC theory is used to support the rationale for test interpretation and implications. (pp. 8–14)
- Age equivalents are provided for norm-referenced subtests; limitations are outlined. (p. 86)

Relations to other variables (Standards 1.16, 1.17, 1.18, 1.19, 1.20, 1.21, 1.22)

- Correlations between other cognitive and academic achievement assessments are illustrated and discussed. (pp. 163–184)
- Special-group studies were conducted to identify differences between groups. The groups included individuals with intellectual giftedness, intellectual disability, reading disorders, reading and writing disorders, math disorders, ADHD, ADHD and learning disorders, expressive language disorder, mixed receptive–expressive language disorder, limited English proficiency, developmental risk, and deafness/hearing impairment. Groups are compared to matched controls. (pp. 184–218)
- Data suggest that the current edition correlates highly with the prior edition. (pp. 164–165)

Fairness, bias, and consequences of testing (Standards 1.5, 1.6, 1.7, 1.8, 1.10, 1.17, 1.18, 1.19, 1.22, 1.24, 1.25)

- A review panel was established and included members from a variety of cultural backgrounds. The panel evaluated the assessment to ensure fairness of test items. The technical handbook includes detailed description of the process involved in evaluating item bias and associated modifications of the findings. (p. 101)
- Fairness of prediction was assessed to determine whether the test can predict achievement similarly for different groups. (pp. 222–224)
- Examples of what factors should be considered before considering the scores valid and interpreting the results are presented. (pp. 21–24)

FIGURE 30.3. *(continued)*

Cross-cutting standards (Standards 1.1, 1.2, 1.3, 1.4, 1.8, 1.10, 1.23)

- Norming sample is clearly defined and relates to the target population of school-age children in the U.S. population in 2011. (Naglieri et al., 2014c, p. 42)
- Test authors explain the assessment's roots in planning, attention, simultaneous, successive (PASS) theory and in Luria's work on brain functioning. (Naglieri et al., 2014c, p. 1)
- Subtests and PASS scales are clearly defined. (Naglieri et al., 2014c, pp. 7, 10)
- The CAS2 is described as being able to determine factors such as individual strengths and weaknesses, processing compared to that of same-age peers, and PASS scores as they relate to achievement abilities. (Naglieri et al., 2014c, p. 12)
- Flexibility is discussed in regard to the administration of instructions to bilingual, deaf, and hard-of-hearing students. (Naglieri et al., 2014b, p. 6)

Test content (Standards 1.9, 1.11)

- Subtests and composites are described in detail. (Naglieri et al., 2014c, pp. 6–12)
- Test improvements and their purposes are explained and connected to PASS theory. (Naglieri et al., 2014c, pp. ix–x)
- PASS theory is explained and related to the measurement of PASS abilities. (Naglieri et al., 2014c, pp. 1–6)

(continued)

FIGURE 30.4. Examples of CAS2 validity evidence.

Response processes (Standards 1.7, 1.12, 1.25)

- The CAS2 is based on PASS theory and consists of four scales: Planning, Attention, Simultaneous, and Successive. (Naglieri et al., 2014b, p. 1)
- Test developers provide evidence to support this model. (Naglieri et al., 2014c, p. 4)
- Questioning about strategies utilized on various subtests gives insight into how test takers determine answers, and how the strategies relate to the standard score achieved. (Naglieri et al., 2014c, p. 99)

Internal structure (Standards 1.13, 1.14, 1.15)

- Confirmatory factor analysis (CFA) was used; "each subtest is permitted to load only on the factor that it represents." (Naglieri et al., 2014a)
- CFA models were utilized regarding the PASS theory in four age intervals. (Naglieri et al., 2014c, p. 111)
- Based on the criteria used by the authors, factor loadings range from moderate to very large. (Naglieri et al., 2014c, p. 112)
- All subtest correlations are significant. (Naglieri et al., 2014c, p. 116)
- When subtests were correlated with their designated PASS scale, subtests correlated highest with their specific PASS scale and lowest with the scales in which they are not utilized. (Naglieri et al., 2014c, p. 116)

Relations to other variables (Standards 1.16, 1.17, 1.18, 1.19, 1.20, 1.21, 1.22)

- Evidence is provided about the correlations between the CAS2 and other intelligence tests and achievement tests. (Naglieri et al., 2014c, p. 84)
- Special groups are explained (e.g., the bilingual and hard-of-hearing groups). (Naglieri et al., 2014b, p. 6)

Fairness, bias, and consequences of testing (Standards 1.5, 1.6, 1.7, 1.10, 1.17, 1.18, 1.19, 1.22, 1.24, 1.25)

- Scores were evaluated based on different groups: gender, region, ethnicity, Hispanic status, exceptionality status, household income and parental education (Naglieri et al., 2014c, p. 43)
- The test authors analyzed the potential for gender, ethnicity, and racial bias in test items. (Naglieri et al., 2014c, p. 82)

FIGURE 30.4. *(continued)*

validity claims. As such, this chapter is intended to be used as a reference for test users to guide them in evaluating whether any given test is valid for its intended use. The authors of the *Standards* state: "It is commonly observed that the validation process never ends, as there is always additional information that can be gathered to more fully understand a test and the inferences that can be drawn from it" (AERA et al., 2014, p. 12). The authors therefore encourage test users to reference multiple sources of validity evidence, and to develop opinions about validity that are "supported and defensible" (p. 13).

In addition to gathering evidence regarding test validity, we have presented strengths and limitations of the validity evidence presented in the test manuals of the four intelligence tests reviewed. Most manuals include a separate section on validity, which is often broken down into validity types. For example, Naglieri and colleagues (2014c) allocated 48 pages in the technical manual of the CAS2 to content description validity, criterion predication validity, and construct identification validity. Organizing validity evidence by specific types of validity assists users in accessing the information needed to make well-founded decisions about whether to use the test for a given purpose.

A strength of the intelligence tests reviewed is that their manuals contain comprehensive and explicit data for their respective standardization samples. This level of detail allows test users to determine whether the test is appropriate for the population of individuals they intend to evaluate. Including members of a particular group (e.g., individuals with specific learning disabilities) in a test's norming sample does not, of course, automatically make the test fair for that group, and omission of members of a group does not automatically invalidate the test for that group. However, the manuals for the tests referenced herein provide data to allow test users to judge the fairness of the test for specific groups. It should be noted that not

TABLE 30.1. A Comparison of Data Related to Validity Evidence across Tests

Norming sample and intended uses clearly explained	Evidence of internal structure	Correlations with clinical samples
	WISC-V	
N = 2,200 20 age groups from ages 6:0 to 16:11 are created and include 200 children in each group	Verbal Comprehension Index and Similarities = .92 Processing Speed Index and Similarities = .29	Mean General Ability Index for those with traumatic brain injury = 85.4 Mean GAI for matched control group = 103.9
	WJ IV COG	
N = 7,416 Groups are based on individual ages from 2 to 19, and grouped by tens starting at 20. Number of participants varies based on age. Examples (age/number of participants): 3 = 203 11 = 329 20–29 = 759	Oral Vocabulary and other Gc tests = .62–.70 Oral Vocabulary and Gs tests = .26–.32	Mean Spelling score for those with learning disability in reading = 74.2 Mean Spelling score for those with learning disability in math = 85.9
	CAS2	
N = 1, 342 Ages: 5.0–17.11 11 age groups are reported, and number of participants varies per group 686 males 656 females	Ages 5–7: Matrices and Verbal–Spatial Relations = .40 Figure Memory and Expressive Attention = .19 Ages 8–18: Planned Connections and Planned Number Matching = .46 Number Detection and Sentence Questions = .18	Mean Full Scale IQ for Core and Extended Batteries: Giftedness/talent: Core Battery = 111.8; Extended Battery = 112.9 Speech–language impairment: Core Battery = 90.9; Extended Battery = 91.2 Learning disability: Core Battery = 87.1; Extended Battery = 86.4 ADHD: Core Battery = 93.5; Extended Battery = 92.5 Emotional disturbance: Core Battery = 91.9; Extended Battery = 90.8 Anxiety disorder: Core Battery = 91.2; Extended Battery = 90.4 ASD: Core Battery = 83.8; Extended Battery = 82.1
	DAS-II	
N = 3,480 Children divided into 18 age groups from ages 2:6 to 17:11, with 200 children per group	Early Years: Naming Vocabulary and Verbal ability = .91 Naming Vocabulary and Nonverbal Reasoning ability = .56	Mean General Conceptual Ability (GCA) for children with ADHD = 100.2 Mean GCA for matched control group = 105.9

all data necessary for reviewing validity information for the WISC-V are contained in the manuals that come with the test. A special supplement (Wechsler, 2014c) must be downloaded from the test company's (Pearson's) website.

Test publishers and authors over the years have generally adhered to the *Standards*. In the 2012 edition of this book, Braden and Niebling noted, "We appreciate that test developers cannot be held accountable for all possible consequences of test scores" (p. 755). While this remains true overall, it appears that current test manuals provide sufficient information to allow test users to make sound decisions about test use and interpretation.

ACKNOWLEDGMENTS

It is with sincere gratitude that we acknowledge the guidance, dedication, and instructional support provided by Ronald Dumont and John O. Willis. Additionally, we extend our respectful appreciation to Jeffrey P. Braden and Bradley C. Niebling for their validity standards framework, and for their chapters in the previous editions of this book. We would also like to acknowledge and thank Dawn P. Flanagan for offering this publication opportunity.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

Braden, J. P., & Niebling, B. C. (2005). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed., pp. 615–630). New York: Guilford Press.

Braden, J. P., & Niebling, B. C. (2012). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 739–757). New York: Guilford Press.

Elliott, C. D. (2007a). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Elliott, C. D. (2007b). *Differential Ability Scales—Second Edition: Introductory and technical handbook.* San Antonio, TX: Harcourt Assessment.

Goodwin, L. D., & Leech, N. L. (2003). Assessment in action: The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36,* 181–191.

Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Boston: Pearson.

Mather, N. & Wendling, B. J. (2014). *Examiner's manual: Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock– Johnson IV.* Rolling Meadows, IL: Riverside.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014a). *Cognitive Assessment System—Second Edition.* Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014b). *Cognitive Assessment System—Second Edition: Administration and scoring manual.* Austin, TX: PRO-ED.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014c). *Cognitive Assessment System—Second Edition: Interpretive and technical manual.* Austin, TX: PRO-ED.

Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema, 26*(1), 136–144.

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema, 26*(1), 108–116.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 99–144). New York, NY; Guilford Press.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*(1), 100–107.

Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education, 3*(2), 119–120.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children—Fifth Edition: Technical and interpretive manual.* Bloomington, MN: Pearson.

Wechsler, D. (2014c). *Wechsler Intelligence Scale for Children—Fifth Edition: Technical and interpretive manual supplement: Special group validity studies with other measures and additional tables.* Bloomington, MN: Pearson.

CHAPTER 31

Using Confirmatory Factor Analysis to Aid in Understanding the Constructs Measured by Intelligence Tests

Timothy Z. Keith
Matthew R. Reynolds

Factor analysis is inexorably linked with the development of intelligence theory and intelligence tests. Early intelligence theories and factor-analytic methods were developed in tandem, and the connection continues to this day. Carroll's (1993) three-stratum theory of intelligence was developed in part through the use of *exploratory factor analysis* (EFA).

For much of its history, the term *factor analysis* meant what is now called EFA. In its simplest form, EFA involves making a series of decisions about the method of factor extraction to use, the number of factors to retain, the method of rotation to use, and the criterion for meaningfulness of factor loadings. For researchers who do not wish to make these decisions, most computer programs will default to a common method if none is specified (not always a wise choice). The output from the analysis consists of factor loadings of each variable on each factor and, if an oblique rotation was used, correlations among the factors. The researcher then assigns names to the factors based on the loadings of the variables on the factors, along with relevant theory and previous research.

In the hands of an expert, EFA can be much more complex and sophisticated than the simple approach just described. A variety of extraction methods can be used, depending on the questions of interest; complex decision rules and expert judgment can be used to determine how many factors should be extracted; and a variety of graphical and mathematical methods can be used to rotate the extracted factors to simple structure. For example, Carroll (1993, Ch. 3) outlined an EFA approach that was an elegant combination of consistency and judgment; see also McDonald (1999, p. 187). Whether simple or complex, EFA involves judgment on the part of the researcher: judgment concerning the decisions required, and judgment concerning the meaning of the extracted factors. It is this aspect of EFA that can be disconcerting to those wanting yes–no answers to questions, but it is also this requirement for theory, thought, and judgement that makes the approach so alluring and so powerful. Although researchers sometimes choose EFA to allow the data to "speak for themselves," it is always ultimately the researcher who does the "speaking," given the judgment needed.

In contrast, and in its simplest form, *confirmatory factor analysis (CFA)* requires the researcher to decide, in advance, the nature of the factor structure underlying the data. He or she must specify the number of factors and the variables that load on each factor. So, for example, the researcher may specify that variable 1 loads on factor 1, but not on factor 2. The researcher may specify that factors are correlated, or may specify that they are uncorrelated. The results of the analysis provide *fit statistics,* which provide feedback as to the adequacy of the specified factor structure, or the degree to which the *model* (factor structure) *fits* the data (reproduces the covariances among the data). In CFA, in other words, the researcher tests the adequacy of a particular factor structure by restricting the factor solution (thus the method is sometimes called *restricted factor analysis*) and seeing whether that restricted solution is consistent with the data. In contrast, in EFA, the researcher examines and imparts meaning to the best factor structure (given the decision rules used).

CFA is often described as a more theory-driven approach than EFA. This assertion probably involves some overstatement. It is possible, for example, to use EFA in a theory-driven, hypothesis-testing manner (cf. Thorndike, 1990; Thurstone, 1947), just as it is possible to use CFA in an exploratory, theory-absent manner. Indeed, Jöreskog and Sörbom (1993, p. 115) argued that a combination exploratory–confirmatory "model-generating" approach is likely the most common approach to structural equation modeling (SEM) and CFA. Nevertheless, because CFA requires the specification of a model—and thus knowledge about the probable structure of the characteristic being measured—some sort of theory (formal or informal, strong or weak) is required. EFA can easily be conducted in the absence of theory, although theoretically driven analyses are almost invariably more complete and informative than atheoretical ones. EFA can be a valuable tool for *developing* theory, whereas CFA may be better suited for *testing* existing theory.

This chapter demonstrates the use of CFA to understand the constructs measured by modern intelligence tests. It begins with a "simple" CFA model, and gradually moves to CFA methods that provide a more complete evaluation of the theories underlying tests (e.g., hierarchical analysis, the comparison of alternative models, testing hypotheses about tests and constructs, and multisample analyses). Our emphasis is on the use of CFA to *test hypotheses* about *tests* and *theories.*

There are several computer programs available that conduct CFA. At this writing, it is possible to obtain manuals for all of these programs as .pdf files online. These programs are designed to conduct latent-variable SEM; SEM includes and subsumes CFA (the *measurement model,* in the jargon of SEM), and thus SEM programs also conduct CFA. The oldest program is LISREL (LInear Structural RELations; Jöreskog & Sörbom, 1996); extensive information about the program is available at *www.ssicentral.com.* Other common programs include EQS (Bentler, 1995; *www.mvsoft.com*) and Mplus (Muthén & Muthén, 1998–2010; *www.statmodel.com*), perhaps the most powerful and flexible SEM program. The analyses presented in this chapter were conducted using Amos (Analysis of MOment Structures; Arbuckle, 2014; *www.amosdevelopment.com* or *www-03.ibm.com/software/products/en/spss-amos*), likely the easiest-to-use such program. Amos uses a drawing program to provide pictorial input and output of models. The figures in this chapter were drawn with Amos, and the graphical output shown was produced by the Amos program. Finally, the free statistical software R has several SEM packages: Lavaan, OpenMx, and SEM (*www.r-project.org*).

AN INTRODUCTORY EXAMPLE

The Kaufman Adult and Adolescent Intelligence Test (KAIT) is an older measure of cognitive abilities for adolescents and adults ages 11 through adulthood (Kaufman & Kaufman, 1993). According to the manual, the KAIT was designed to measure intelligence according to the Cattell–Horn theory of intelligence, although only two of the Cattell–Horn factors are included in the KAIT model (*fluid intelligence,* also known as *novel reasoning,* and *crystallized intelligence,* also known as *comprehension–knowledge*). In addition, the KAIT was designed to measure *delayed recall,* or a person's memory for material learned earlier in the test. Thus the KAIT includes 10 subtests designed to measure three abilities: fluid intelligence (Gf), crystallized intelligence (Gc), and delayed recall.

Figure 31.1 shows this theory underlying the KAIT in figural form. The 10 subtests are shown in rectangles, and the abilities they are designed to measure are enclosed in ovals. Directed arrows or paths point from the abilities to the subtests, in recognition of the implicit assumption that the abilities residing within a person are what cause him or her to score a certain way on a subtest. So, for example, examinees' levels of fluid intelligence are the primary determinant of their scores on the Mystery Codes subtest.

Figure 31.1 is also the beginning of a CFA model; a more complete CFA model is shown in Figure 31.2.[1] In the jargon of CFA, the variables enclosed in rectangles (the subtests) are the *measured,* or *manifest,* or *observed* variables, whereas

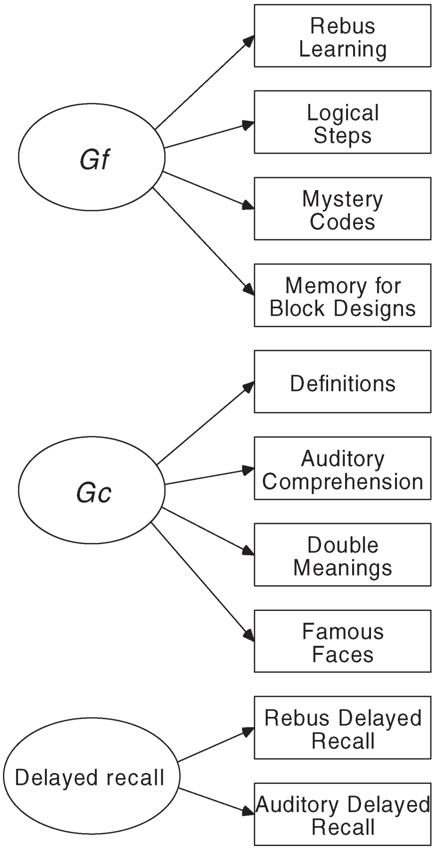


FIGURE 31.1. Theoretical structure of the KAIT.

the variables enclosed in ovals (i.e., the KAIT abilities) are *unmeasured* or *unobserved* variables, also known as *latent* variables, or *factors*. The paths from latent to measured variables represent the factor loadings. Notice that not all possible paths are drawn. Theoretically, the lack of a path from, say, Gf to Double Meanings means that the test authors believe Double Meanings does not measure fluid intelligence, or that variation in fluid intelligence does not produce variation in scores on the Double Meanings subtest. Rather, scores on Double Meanings are a reflection of crystallized intelligence, as indicated by the path drawn from Gc to Double Meanings. At a practical level, the path from Gc to Double Meanings means that the factor loading will be estimated in the analysis, and the lack of the path from Gf to Double Meanings means that the factor loading of Double Meanings on G*f* will be fixed to 0.

Figure 31.2 includes information beyond that included in Figure 31.1. The curved, two-headed arrows between factors represent covariances (or, in the standardized output, correlations). Although the KAIT manual does not say so explicitly, it is reasonable to expect that the abilities measured by the KAIT are not independent of (uncorrelated with) one another. Modern intelligence theories recognize this relation among factors (e.g., Carroll, 1993, Chs. 2–3), and CFAs of intelligence tests should specify correlated factors (unless the theory underlying the tests maintains that the factors are independent). The figure also includes small ovals, labeled u1 through u10, with paths drawn to each of the subtests. The factors are not the only cause of a person's scores on the subtest; each subtest is also partially the result of other influences that are unique to each subtest, generally called *unique* or *specific* variances. In addition, each subtest is also affected by errors of measurement (*error* variance). These unique and error variances, combined, are represented by u1 through u10; they are enclosed in ovals because they are unmeasured. These unique and error variances are hereafter termed *unique variances*, although they are referred to by a variety of names here and in the literature, including *errors* and *residuals*.

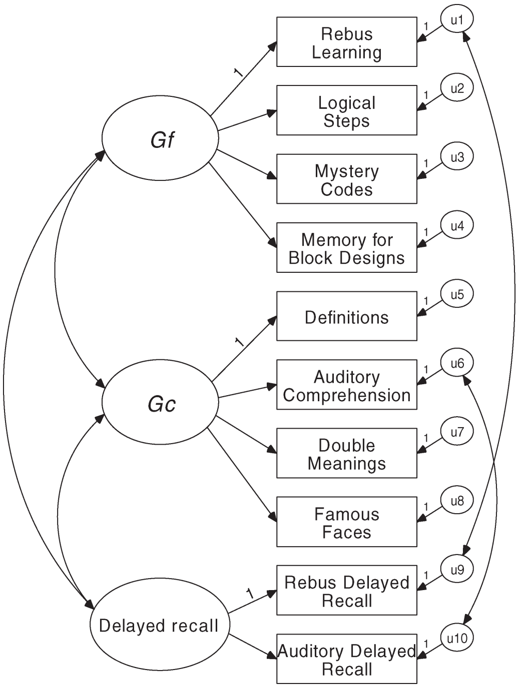


FIGURE 31.2. A CFA model of the structure of the KAIT.

Several of the paths in Figure 31.2 have the value 1 beside them. The measured variables in the model have a defined scale, and that scale is whatever scale was used for each subtest (e.g., a scaled score from 1 to 19). But none of the latent variables (neither the factors nor the unique variances) has a predetermined scale. The 1's beside the paths serve the purpose of setting the scales of the latent variables by setting the path to 1.0. The path of 1.0 from Gf to Rebus Learning, for example, sets the scale of the Gf factor to be the same as the scale for Rebus Learning. Thus each factor includes one path, or loading, of 1.0, and the path from each of the unique variances to the measured variables is set to 1.0 (to set the scale of the unique variances to be the same as the corresponding measured variable). The use of 1.0 is arbitrary—any value could be used—and once all of the parameters of the model that were not fixed to a value are estimated (called the *unstandardized* solution), all values are restandardized (the *standardized* solution).

Finally, the model in Figure 31.2 includes a less common characteristic: correlations among the unique variances, as represented by the curved lines connecting two pairs of the unique variances (e.g., u1 and u9). The Rebus Delayed Recall test on the KAIT requires examinees to remember rebuses learned earlier in the Rebus Learning subtest. Since Rebus Delayed Recall builds on Rebus Learning, it seems likely that the unique variances affecting Rebus Delayed Recall will be related to those affecting Rebus Learning. Similarly, the *error* variances affecting Rebus Delayed Recall may well be related to those affecting Rebus Learning. These possibilities may be built into the CFA model by specifying that the unique variances of the Rebus Delayed Recall and Rebus Learning subtests are allowed to correlate. There are other ways of doing so as well, including the specification of minor factors.

The KAIT standardization data were used to estimate the model. The KAIT manual includes correlation matrices of subtests for the KAIT at each age level, along with an average correlation matrix for the entire sample (Kaufman & Kaufman, 1993, p. 136). The correlation matrix and standard deviations for the entire standardization sample were used as input for the Amos computer program (Arbuckle, 2014) that converts them to covariances for the analysis. The average sample size ($N = 143$) for the different age groups was used as the sample size.

The results of the analysis are shown in Figure 31.3. The fit indices for the model are shown in the lower left of the figure. The model is *overidentified,* meaning that we could have estimated many more parameters in the model than we actually did. For example, the path from Gf to Definitions was fixed to 0, as were many other possible paths (note that the lack of a path is equivalent to fixing the path to 0). In an overidentified model, the number of parameters estimated is less than the number of variances and covariances among the measured variables. As a result, the model has positive degrees of freedom. *Degrees of freedom (df)* are an index of the degree of overidentification of a model; they are not, as in most other statistical analyses, related to the sample size. For the current model, *df* = 30 means that we could have estimated 30 additional parameters (e.g., paths, correlations).

It is also possible to conduct the analysis in reverse: to estimate the correlation or covariance matrix from the solved model. That is, we could have used the factor solution shown in Figure 31.3 to estimate what the covariances among the subtests should be, given the model. But because the model is overidentified, the estimates of the predicted covariances will not be identical to the covariance matrix used to estimate the model in

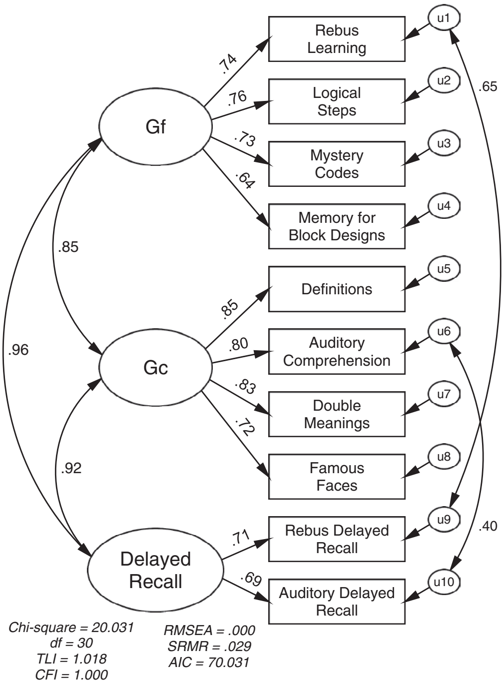


FIGURE 31.3. Results of a CFA of the structure of the KAIT. The results are for the standardized solution.

the first place (see Keith, 2015, Ch. 13, or Kline, 2016, Ch. 7, for a discussion of overidentification). The fit indices shown in the figure are all measures of the degree to which the matrix *implied* by the model differs from the matrix *actually* used to estimate the model.

Five fit indices are shown in the figure, although there are dozens to choose from. Chi-squared (χ^2) is the most commonly reported fit statistic. It has the advantage of allowing a statistical test of the fit of the model; it can be used with the *df* to determine the probability that the model is, in some sense, "correct." Thus a large χ^2 in comparison to the *df* (and a small probability—e.g., $p < .05$) suggests that the actual and implied covariance matrices are statistically significantly different, that the model provides a poor fit to the data, or that the model could not have produced the data. The model, therefore, is not a good representation of the "true" factor structure. In contrast, a small χ^2 in comparison to the *df* is desirable; if it is statistically nonsignificant ($p > .05$), it suggests that the model *could* have produced the data, and that the model does provide a reasonable explanation of the data.

Although χ^2 fits well within the tradition of significance testing in psychology (somewhat confusingly, however, the researcher wants it to be nonsignificant), it also has well-known problems (as does the tradition of either–or significance testing itself; cf. Cohen, 1994; Little, 2013). In particular, it is directly related to sample size, so that with large samples, virtually all χ^2's will be significant, even when the model is only trivially incorrect (see Tanaka, 1993, among others, for further discussion). With small samples, even inadequate models may have a good fit, as judged by χ^2. For this and other reasons, fit indices have been developed; the ones listed for this analysis were chosen because they highlight different dimensions of fit, and have shown promise in simulation studies (Fan, Thompson, & Wang, 1999; Hu & Bentler, 1998, 1999). Methodologists generally recommend using a combination of criteria.

One criticism of χ^2 is that it is a measure of the exact fit of the model to the data, whereas, at best, models are designed to *approximate* reality. The *root mean square error of approximation (RMSEA)* is, in contrast, a measure of approximate fit. Smaller values suggest a better fit, with values of .06 or smaller suggesting a good fit (Hu & Bentler, 1999), and those of approximately .08 suggesting an adequate fit (Browne & Cudeck, 1993). The RMSEA is sometimes reported along with its 90% confidence interval. The *standardized root mean square residual (SRMR)* may be one of the more intuitively appealing measures of fit. Recall that fit indices are derived from the similarity or dissimilarity between the actual covariance matrix used to estimate the model and the matrix implied by the model. The *root mean square residual (RMR)* represents the average of these differences, and the SRMR is the standardized average of these differences. Because a standardized covariance is a correlation, the SRMR, therefore, represents the average difference in the actual *correlations* among the measured variables and those implied by the model. Values of .08 or less suggest little difference in the two matrices (Hu & Bentler, 1999).

The *Tucker–Lewis index (TLI)* and the *comparative fit index (CFI)* both compare the fit of the model with that of a *null* model, one in which the measured variables are assumed to be unrelated (the null model is also referred to as the *independence* model; there can be other types of null models). The TLI appears to be relatively unaffected by sample size; the CFI is designed to estimate the fit in the population. For both, values of .95 or greater suggest a good fit of the model to the data (Hu & Bentler, 1999), with values above .90 suggesting an adequate fit. We sometimes report one of these indices (because they often are fairly similar in magnitude), sometimes both.

Briefly, all of the fit indices suggest that the KAIT model provides an excellent fit to the standardization data. χ^2 is small and statistically nonsignificant ($p = .668$); the TLI and CFI are large; and both the RMSEA and the SRMR are quite small. Thus the "theory" underlying the KAIT appears to fit the KAIT standardization data; the test appears to measure what the authors designed it to measure; and the structure of the KAIT appears valid. The next step, then, is to interpret the substantive results.[2]

The paths from latent to measured variables show the factor loadings. They are all large, and examination of their standard errors and *z* values (shown in the detailed printout, but not included here or in the figure) shows that they are all statistically significant. Likewise, the factor correlations are large and statistically significant, ranging from .96 for the correlation between the latent delayed-recall and Gf factors to .85 between Gf and Gc. Finally, the correlations among unique variances suggest that there is a substantial correlation between the variance of the Rebus Learning test that is not accounted for by the Gf factor and the variance of the Rebus Delayed Recall subtest that is not accounted for by the delayed-recall factor (*r*

= .65). Similarly, the unique variances of Auditory Comprehension and Auditory Delayed Recall are substantially correlated.

These findings generally support the validity of the KAIT. One curious finding, however, is the magnitude of the correlation between the delayed-recall and the Gf and Gc factors (.96 and .92, respectively), which are higher than the correlation between the two presumably more intellectual factors, Gf and Gc (.85). Such a finding could be investigated by comparing this model with plausible alternative models, as shown below.

An Alternative Method of Specifying Factor Models

Before we illustrate the comparison of competing alternative models, it is worth illustrating an alternative method of scaling the factors. The previous model set the scale of the latent factors by fixing a single factor loading from each factor to 1.0 (what Kline, 2016, refers to as *unit loading identification [ULI]*). An alternative is to fix the *factor variances* to 1.0 (*unit variance identification [UVI]*) and then estimate all factor loadings; a KAIT model using this method is shown in Figure 31.4. This procedure has advantages over the method of fixing a factor loading. When the factor variances are set to 1, the factor covariances in the unstandardized solution are in fact correlations. Because constraints to models are made in the unstandardized metric, this method makes it possible to set factor correlations to 1, or 0, or some other value to see what this constraint does to the fit of the model. This ability to constrain and test the magnitude of correlations is used later in this chapter. The UVI method also has its limitations. With hierarchical models (to be discussed below), only the highest factor level can use this alternative method, and so a ULI approach is common. The ULI method is also more common in SEM.

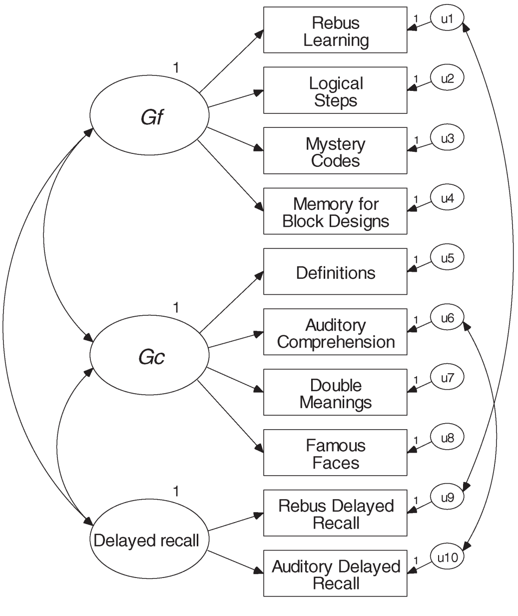


FIGURE 31.4. An alternative method of specifying factor models. Factor variances rather than factor loadings are set to 1 to set the scale of the factors.

The results of the analysis of the model shown in Figure 31.4 are identical to those shown in Figure 31.3 and are not repeated here. It is worth noting, however, that this method can occasionally produce slightly different results than the ULI method does (Millsap, 2001).

COMPARING ALTERNATIVE MODELS

The previous section has presented the basics of CFA, but has also presented a fairly sterile approach: the simple rejection of or support for a single model. Such an approach is problematic. The simple fact that a model does or does not fit the data provides only partial support for the validity of a test or theory. A model may fit the data well, but there may be other, competing models that fit the data as well or better. Or a given model may provide an inadequate fit to the data, but may be the best model among the alternatives. Formal or informal theory may suggest several possible models to explain the test scores we observe; if all fit the data well, how is one to decide which is the *best* model? What is needed is the ability to compare competing models, or to compare a target model with one or several alternative models. "At best, a given model represents a tentative explanation of the data. The confidence with which one accepts such an explanation depends, in part, on whether other, rival explanations have been tested and found wanting" (Loehlin & Beaujean, 2017, p. 63).

Figure 31.5 shows an alternative model of the constructs measured by the KAIT. The reasoning behind this model is that if the two nondelayed versions of these tasks (Rebus Learning and Auditory Comprehension) measure Gf and Gc abilities, then shouldn't the delayed versions also measure these underlying abilities? Since, for example, Rebus Delayed Recall asks examinees to recall material first learned in the Rebus Learning test,

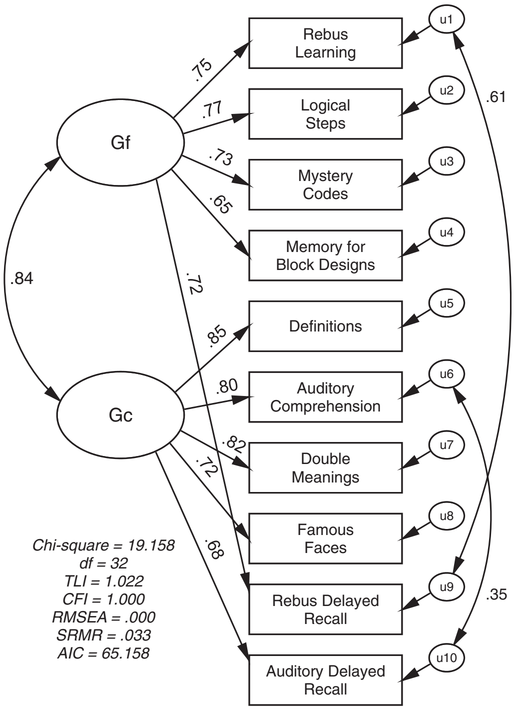


FIGURE 31.5. An alternative model of the structure of the KAIT. Rebus Delayed Recall and Auditory Delayed Recall are assumed to measure Gf and Gc, respectively, instead of delayed-recall ability.

isn't Rebus Delayed Recall really a measure of Gf? Similarly, is Auditory Delayed Recall a measure of Gc instead of delayed recall? This supposition may also help explain why the correlations between the delayed-recall and the Gf and Gc factors are so high: They measure overlapping constructs.

The fit indices shown in Figure 31.5 suggest that it also shows a good fit to the data. The SRMR, for example, suggests an average difference between the actual correlations and those implied by the model of .033, versus .029 for the initial model. The TLI barely changes with the model in Figure 31.5. Can we conclude that one model is better than the other—and thus come to a conclusions about the likely constructs measured by the KAIT?

Fortunately, fit statistics can be used to *compare* competing models as well. If two models are *nested*—that is, if one model can be derived from another by placing additional constraints on the model—then the χ^2 for the two models can be compared. The χ^2 for the less constrained model (the model with smaller *df* and smaller χ^2) can be subtracted from the χ^2 for the more constrained model. The χ^2 difference ($\Delta\chi^2$) can then be compared to the change in *df* (Δdf) to determine whether the constraints added to the model result in a statistically significant increase in χ^2. If the $\Delta\chi^2$ is not statistically significant—if one model is not significantly better than another—then, as scientists, we generally prefer the more parsimonious (more constrained, higher-*df*) model. If the $\Delta\chi^2$ is statistically significant, then the less parsimonious model is preferred. Parsimony is reflected in CFA models by *df*; the larger the *df,* the more constraints in the model, and therefore the more parsimonious the model. Again, nested models are those in which it is possible to get from one model to the other by constraining one or more parameters to some value (or, alternatively, by freeing existing parameter constraints). This is a powerful use of CFA, and allows the testing of specific hypotheses about the constructs being measured by a test. In our opinion, this is one of the biggest advantages of CFA over EFA: the ability to test hypotheses derived from such questions. It is often possible to do so statistically (via $\Delta\chi^2$) using carefully planned, nested models.

The *Akaike information criterion* (*AIC*; Akaike, 1987) is another useful measure for comparing competing models, and does not require that the models be nested. The AIC is not interpreted in isolation; rather, the AICs for two or more competing models are compared, with the smaller AIC suggesting the better model. Other, related fit indices include the *Bayes information criterion* (BIC; Schwarz, 1978), and the sample-size-adjusted BIC (aBIC; Sclove, 1987). The AIC and aBIC have performed well in intelligence test CFA simulation research (Keith, Caemmerer, & Reynolds, 2016). Finally, a highly constrained "baseline" model may be used instead of a null model in the calculation of fit indices such as TLI and CFI. For example, a one-factor *g* model may provide a useful baseline model in CFAs of intelligence tests. Of course, with these alternative baseline models, the resulting values will generally not approach the .95 cutoff recommended for these fit statistics when null models are used.

Testing Hypotheses by Comparing Alternative Models for the KAIT

Figure 31.3 displays the results of an initial CFA of the KAIT, and Figure 31.5 shows an alternative. Unfortunately, these models are not nested: To get from Figure 31.3 to Figure 31.5, we have freed two parameters (the paths from Gf to Rebus Delayed

Recall and Gc to Auditory Delayed Recall), but we have also fixed the paths from these two Delayed Recall tests to 0 (and deleted the now-empty delayed-recall factor), and the two correlations of the delayed-recall factor with Gf and Gc have also been constrained to 0. Nested models require the addition or the relaxing of constraints, but not both. It is possible to compare these two models via the AIC, however, with the model with the smaller AIC being the preferred one. As shown in the figures, and in Table 31.1 (see the model labeled "4. Gf-Gc (Figure 31.5)"), the initial model has an AIC of 70.03, versus 65.16 for the Figure 31.5 model. Our research thus suggests that the delayed-recall tests should be considered as further measures of Gf and Gc, rather than as a measure of a separate delayed-recall factor.

The use of nested-model comparisons is a powerful methodology, however, so it is useful to see how we *could* get from Figure 31.3 to Figure 31.5 through a series of nested models. The model shown in Figure 31.6 evaluates the hypothesis that the two delayed-recall subtests also measure Gf and Gc abilities *in addition to* delayed recall by freeing two constraints. In addition to allowing Rebus Delayed Recall to load on the delayed-recall factor, it is allowed to load on the Gf factor; Auditory Delayed Recall loads on both the delayed-recall and the Gc factors. One more change in this model (compared to the one shown in Figure 31.3) is required for the model to be properly identified: Some sort of additional constraint is required for the two Delayed Recall subtests. To allow estimation, the factor loadings for Rebus Delayed Recall and Auditory Delayed Recall on delayed recall need to be constrained to be equal.[3] Because going from Figure 31.3 to Figure 31.6 involves first adding a constraint (two paths equal) and then freeing other constraints (freeing two paths), it is conducted in two steps. The model fit results are shown in Table 31.1.

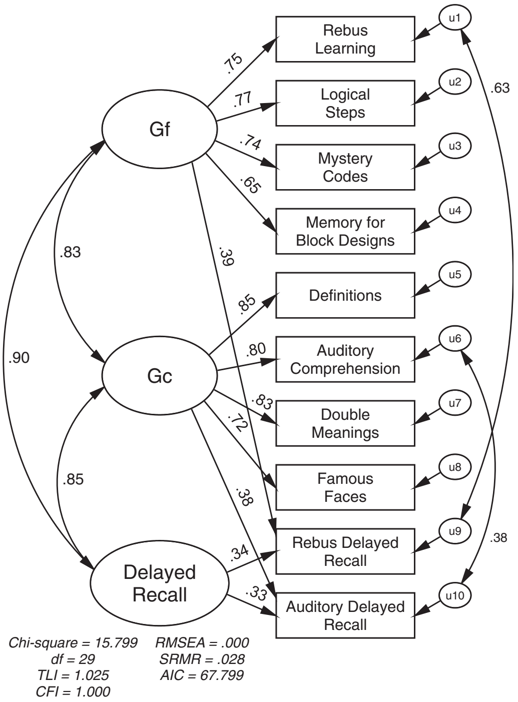


FIGURE 31.6. Another alternative model of the structure of the KAIT. The two Delayed Recall subtests (Rebus and Auditory) are assumed to measure Gf and Gc and a separate delayed-recall ability.

As shown in Table 31.1, constraining the two delayed-recall paths from the model in Figure 31.3 to be equal (this type of constraint is easy to do in all SEM programs) results in 1 extra *df* and an increase in χ^2 of only 0.01 ("2. Equal delayed-recall loadings"). This model is a more constrained version of the model in Figure 31.3, and is therefore nested with that model. As shown in the table, this change in $\Delta\chi^2$ is not statistically significant. If $\Delta\chi^2$ is not significant, this finding is generally

TABLE 31.1. Comparison of Various CFA Models of the KAIT

Model (Figure no.)	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	TLI	CFI	SRMR	AIC
1. Actual structure (Figure 31.3)	20.031	30				1.018	1.00	.029	70.031
2. Equal delayed-recall loadings	20.032	31	0.001	1	.97	1.019	1.00	.029	68.032
3. Delayed recall and Gf-Gc (Figure 31.6)	15.799	29	4.233	2	.12	1.025	1.00	.028	67.799
4. Gf-Gc (Figure 31.5)	19.158	32	3.359	3	.34	1.022	1.00	.033	65.158
5. Subtests affect delayed recall (Figure 31.7)	19.158	32				1.022	1.00	.033	65.158

Note. All comparisons are with the preceding model.

taken as support for the more parsimonious model, which is the model with more *df* (the table does not include RMSEA values, which do not vary from model to model). Model 2 has nearly equivalent χ^2 and a larger *df* than model 1, and thus is preferred. In other words, specifying these two paths (factor loadings) as equal is reasonable, and it is reasonable to use this model to compare to other nested models.

Figure 31.6 (model 3 in Table 31.1) shows the results for the next step: A model in the previous constraint is retained (the delayed-recall paths are constrained to be equal), *and* paths are allowed from Gf to Rebus Delayed Recall and from Gc to Auditory Delayed Recall. This model and the previous one are nested. The results of this analysis are also shown in Figure 31.6 and Table 31.1. Like the previous models, this variation provides an excellent fit to the averaged standardization data. But does the model provide any better fit than the structure as intended by the test authors? The χ^2 for the model is 15.779 for Figure 31.6 (model 3) versus 20.032 for the previous model (i.e., model 2, with equal loadings for the delayed tests on the delayed-recall factor). As shown in Table 31.1, the change in χ^2 is 4.233, which, given the change in *df* ($\Delta df = 2$), is not statistically significant ($p > .05$). Thus specifying that Rebus Delayed Recall and Auditory Delayed Recall measure Gf and Gc in addition to delayed recall improves the fit of the model, but not to a statistically significant degree. It would be perfectly reasonable to stop model comparisons at this point and conclude that model 2 (with the addition of equal loadings on the delayed-recall factor) is a reasonable explanation of the KAIT structure.

It is worth comparing the model in Figure 31.5 (model 4 in Table 31.1) with the one in Figure 31.6 (model 3), however, and this is also done in the table. These models are nested because it is possible to get from Figure 31.6 to Figure 31.5 by deleting the two paths from delayed recall and by deleting the two correlations of Gf and Gc with delayed recall. This comparison results in an increase in *df* of 3, as opposed to 4, because the two delayed-recall paths are constrained to be equal, and thus they are equivalent to 1 *df* in the model. This model resulted in a $\Delta\chi^2$ of 3.359, which is not statistically significant. Once again, with the $\Delta\chi^2$ not statistically significant, we would prefer the more constrained (higher-*df*) model—the model that considers the delayed-recall tests as further reflections of Gf and Gc. So now we have a dilemma. On the one hand, the Figure 31.5 model is "better" (equivalent fit based on χ^2, more parsimonious) than the Figure 31.6 model. But on the other hand, the Figure 31.6 model is not better than the original Figure 31.3 model. And on yet a third hand, the Figure 31.5 model fits better than the Figure 31.3 model, according to the AIC. Although nested-model comparisons provide a powerful method for testing hypotheses about the constructs measured by intelligence tests, the results are not always clear-cut![4]

The substantive interpretation of the model in Figure 31.5 is straightforward. Rebus Delayed Recall and Auditory Delayed Recall provide measures of Gf and Gc abilities, respectively, that are almost as strong as the tests intended to measure those abilities. The factor loadings (paths) are all reasonable, as is the factor loading between Gf and Gc (.84).

One final model for the KAIT is shown in Figure 31.7 and is used to illustrate two points. In the models shown so far, the presumed overlap between the delayed-recall tests and the nondelayed versions of the tests (i.e., Rebus Learning and Auditory Comprehension) is modeled using correlated error terms. This is a common method of dealing with such overlap, and means that the two tests are thought to measure something in common beyond the factors that are shown as affecting them (e.g., Gf, Gc, delayed recall). This is not the only way of dealing with the overlap. It is not simply the case that performance on the delayed-recall tests *share* unique and error variance with the tests from which they were derived. Rather, performance on the delayed-recall tests should depend in part on how well the material was learned when first presented. Thus, in addition to being affected by delayed recall, Gf, and Gc, these two tests should also be *affected* by their original versions. Figure 31.7 shows a model that embodies this reasoning; it builds on the Figure 31.5 model, but has Rebus Learning affect Rebus Delayed Recall and has Auditory Comprehension affect Auditory Delayed Recall. As shown in Figure 31.7 and Table 31.1, this model fits the data quite well. Note the strong effect of Rebus Learning on Rebus Delayed Recall, and the moderate effect of Auditory Comprehension on Auditory Delayed Recall.

The first points we believe this example illustrates nicely are the power and the flexibility of CFA (and, more generally, SEM), and their ability to model conceptions of the way the world works. Researchers who can think through the way they believe that intelligence works on various outcomes (including the subtests used to measure it)

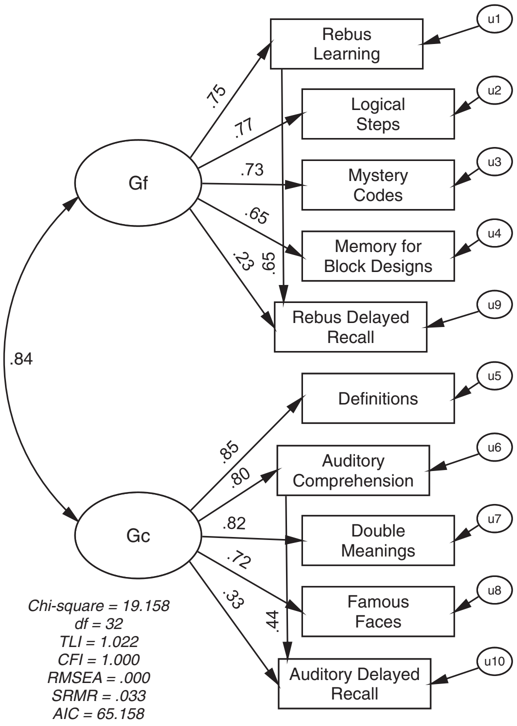


FIGURE 31.7. Yet another alternative model of the structure of the KAIT. The Rebus Learning and Auditory Comprehension tests are assumed to *affect* performance on the subsequent Rebus Delayed Recall and Auditory Delayed Recall tests.

can probably model that notion. This flexibility is further illustrated by the fact that we have used the KAIT to illustrate the basics of CFA in previous editions of this chapter, but have analyzed different models each time!

The second point this model illustrates is that sometimes it is not possible to differentiate plausible alternative models based on fit. As already noted, this model in Figure 31.7 fits the data well. But note that this model has an identical fit to the previous model in Figure 31.5 (see Table 31.1). These equivalent models have the same fit, and thus cannot be compared. It is always worth considering alternative models to one's preferred model; Keith (2015) and Kline (2016), among others, have demonstrated rules for generating equivalent, alternative models (see also Lee & Hershberger, 1990). Before turning to the next topic, we simply note that these models do not constitute an exhaustive examination of the structure of the KAIT; there are other plausible models from other perspectives (see, e.g., Cole & Randall, 2003; Flanagan & McGrew, 1998).

HIERARCHICAL CFA

Many modern theories of intelligence recognize a factor that is more general and broader than the specific abilities tested in first-order CFA. This *general factor* is often considered to subsume, affect, or partially cause the more narrow abilities, and is often symbolized as *g*. For example, Carroll's (1993) three-stratum theory of intelligence includes *g* as the most general, highest-order factor. Although conceptually similar to Spearman's *g*, most modern theories assume that *g* is a higher-order factor (cf. Burt, 1949; Vernon, 1950). Most modern intelligence tests also tacitly recognize such a general, overall factor by summing subtests or subscales into an overall score. Although this general score goes by a variety of names—Full Scale IQ, General Cognitive Ability, or General Intellectual Ability—it generally represents an overall, general, summative ability.

If *g* is recognized in formal theory (as a latent variable that partially explains the positive correlations among all tests of mental ability) and through the informal theory of the scoring of intelligence tests, it is also valuable to test such a construct through CFA. One approach has been to specify a single-factor model, such as the one shown in Figure 31.8 (a *g*-model version of the Wechsler Intelligence Scale for Children—Fifth Edition [WISC-V; Wechsler, 2014]). This model suggests that scores on the individual subtests are a product of a general factor and of unique and error variances. Such an approach mirrors, to some degree, the common practice of isolating a *g* factor in EFA by examining the unrotated first component in principal-components analysis. It is also an unsatisfying solution for several reasons. First, *g* is generally recognized as a *hierarchical* factor; a more realistic structure for the tests from Figure 31.8 is shown in Figure 31.9, in which the subtests are explained (in part) by first-order factors, and the first order factors are, in turn, explained (in part) by a second-order *g* factor (a higher-order hierarchical model). Second, if Figure 31.9 represents the *true* structure of the abilities measured by the 16 WISC-V subtests, then Figure 31.8 represents an inadequate test of the second-order *g* factor. The presence of first-order factors means that the verbal tests (Gc) measure something in common other than general intelligence; the working memory

(Gsm) tests measure something in common other than *g;* and so on. These first-order factors may be identified by characteristics shared among groups of subtests, whereas *g* is not. To reiterate, Figure 31.8 represents an inadequate test of a higher-order *g* factor. It may be, however, a useful comparison model for other models. If we were interested in whether a test *only* measures general intelligence, versus general intelligence plus other cognitive abilities, for example, the fit of the models shown in Figures 31.8 and 31.9 should likely be compared.

Figure 31.10 shows an alternative conception for a hierarchical general factor, often referred to as a *bifactor* model (or, alternatively, as the *nested-factors* model or *direct hierarchical* model). Like a higher-order model, a bifactor model includes both general intelligence and more narrow abilities, but all are first-order factors. Although not strictly hierarchical in organization, the bifactor model is commonly referred to as a hierarchical model, and has become more popular in recent years.

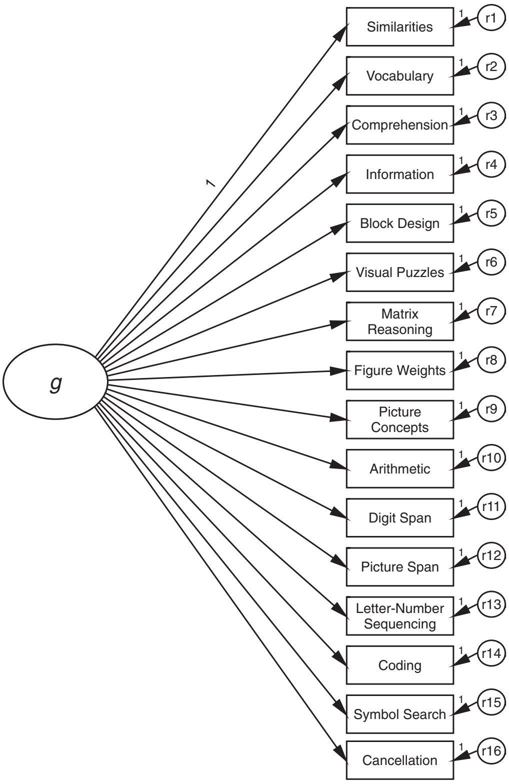


FIGURE 31.8. A general intelligence, or *g*, model of the structure of cognitive abilities. The model assumes that the Wechsler tests are reflections of general intelligence only, rather than more specific, shared abilities, such as verbal or spatial abilities.

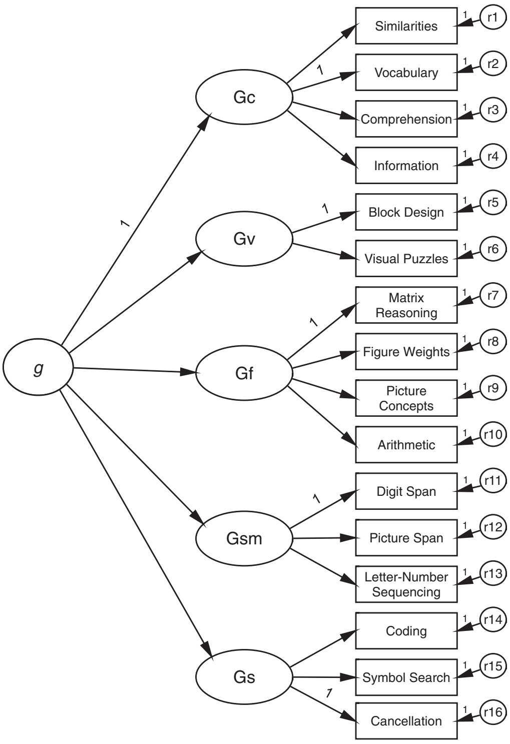


FIGURE 31.9. A higher-order model of the structure of the WISC-V. The model assumes that each test is a reflection of narrow, shared abilities (e.g., Gc or verbal ability), and that these narrow abilities are in turn partially a product of *g*. Note that the model is incomplete as a CFA model; the unique variances associated with the first-order factors are not shown.

Higher-Order Model

Figure 31.11 shows a CFA model to test the higher-order structure of the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007). In scoring the School-Age version of the DAS-II, two "core" subtests are each added together to form Verbal, Nonverbal Reasoning, and Spatial scores, respectively, and all six of those scores are added together to form a General Cognitive Ability score. Thus the core subtests are designed to measure both a *general* ability and more specific abilities (Gc, Gf, Gv). The informal theory underlying the test is clearly hierarchical in nature. There are three memory tests included as well, two of which are added together to form a Working Memory com-

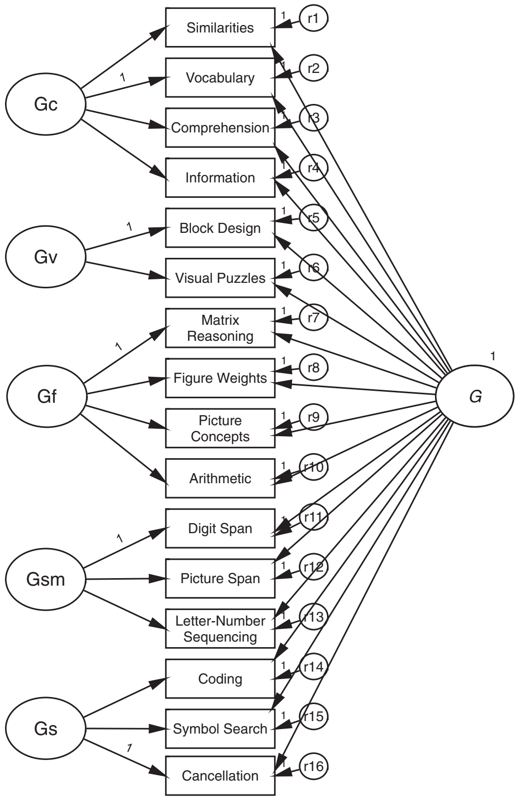


FIGURE 31.10. A bifactor model of the structure of the WISC-V. The model assumes that each test is affected by general intelligence and a broad ability. Both general intelligence and the broad abilities are first-order factors, however, and are unrelated to one another beyond this shared influence.

posite. Here we have used all three as measures of Gsm, which is also affected by a second-order *g* factor. To simplify presentation, these analyses do not include the subtests assessing long-term retrieval or processing speed.

One interesting feature of the DAS-II is that although the structure of the test changes in going from the Early Years to the School-Age version, all tests were given to children ages 5–8 in the standardization sample; these age levels, with multiple indicators of each construct, are the ones used here. The structure shown (including the correlated errors and cross-loading) is based on findings presented in Keith, Low, Reynolds, Patel, and Ridley (2010; see this article for more complete analyses of the structure of the DAS-II, including cross-age tests of measurement invariance).

The figure is quite similar to the higher-order model from Figure 31.9, except for the presence of small ovals (f1 through f4) pointing to the first-order factors. These represent the *unique* variances of each of the first-order factors. Just as the first-order factors are not the only causes of the scores on the subtests, *g* is not the only cause of a person's Verbal or Spatial scores; there is also something unique about Verbal ability as opposed to Spatial ability. This *unique factor* variance, with the test-specific and general factor variance removed, is recognized in the model through the presence of the latent variables f1 through f4.[5] As in earlier models, each latent variable (including *g* and the unique factor variances) has its scales set by fixing one loading to 1.0.

As noted above, the model has been estimated on the basis of the standardization data for ages 5–8. The results of the higher-order CFA of the DAS-II are shown in Figure 31.12. Although χ^2 is statistically significant, the other fit statistics suggest that the model fits the data well, and thus support the scoring structure of the DAS-II. The first-order factor loadings suggest that the core subtests generally measure their corresponding factors well, although the Verbal Comprehension subtest appears to require both Verbal (or Gc) and Nonverbal Reasoning (Gf) ability.

The second-order factor loadings are perhaps even more interesting. In scoring the DAS-II, subtests from the Verbal, Nonverbal Reasoning, and Spatial domains are combined to create an overall score. The figure shows that the three corresponding latent factors are indeed strongly affected by *g*, thus supporting the hierarchical structure of the scale. Also of interest, the Nonverbal Reasoning factor has a very high loading, .92, on the second-order *g* factor. Although the factor is statistically distinguishable from *g* (based on $\Delta\chi^2$), the loading is quite high in practical terms. This finding supports the claim that the Nonverbal Reasoning scale of the DAS-II should be considered a measure of Gf (Elliott, 2007; Keith et al., 2010), given evidence that Gf is generally quite similar to, and very often indistinguishable from, *g* (e.g., Carroll, 1993, Ch. 15; Gustafsson, 1984). Interestingly, if the correlation between the Gf and Gv disturbances were removed, the *g*-Gf loading would be .98. As a result, the unique variance of Gf (the variance of the disturbance f2) would be nonsignificantly different from 0. The finding of indistinguishable (or nearly indistinguishable) *g* and Gf factors is not the result of having relatively few measures of Gf from a single battery; cross-battery CFA of four different tests, including seven measures of Gf, showed a

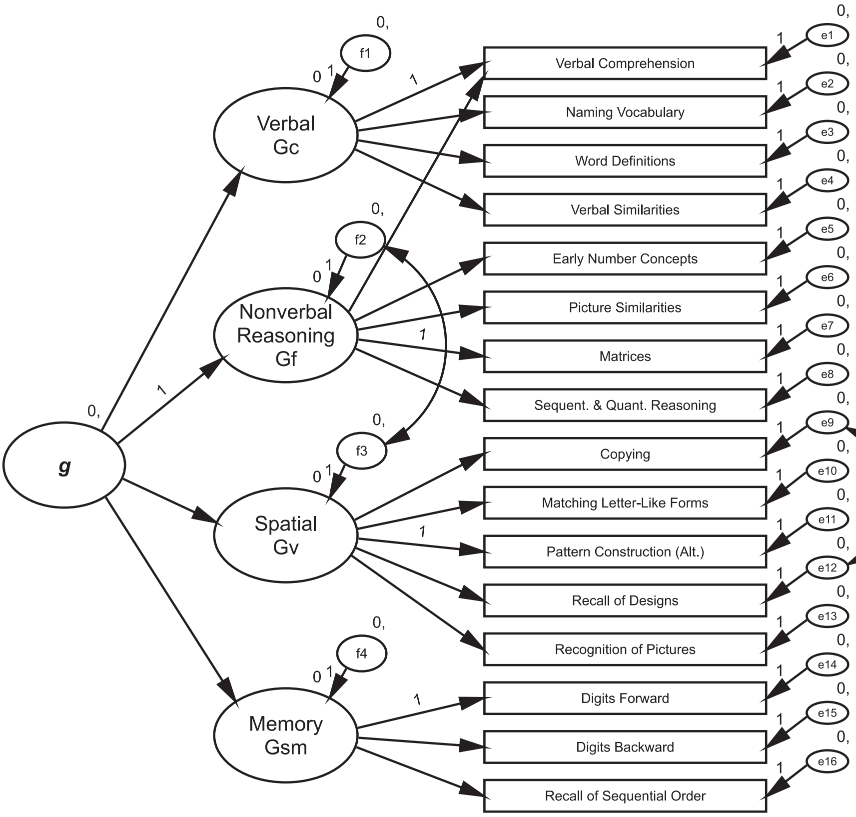


FIGURE 31.11. A hierarchical, higher-order model of the structure of the DAS-II. Digits Forward, Recall of Digits—Forward; Digits Backward, Recall of Digits—Backward.

loading of .98 for a Gf factor on a higher-order *g* factor (Reynolds, Keith, Flanagan, & Alfonso, 2013). G and Gf are hard—often impossible—to distinguish. The high loading by Gsm for the DAS-II is less common, and may suggest that this factor is more cognitively complex (e.g., working memory) than our label of Gsm would suggest.

Researchers and clinicians are often interested in the loading of the subtests on *g*, or the relative strength or weakness of tests as measures of *g*. It is not necessary to resort to *g*-only models (Figure 31.8) or bifactor models (discussed below) to get such estimates, however. In higher-order models, the loadings of the subtests on higher-order factors are calculated as the total indirect (and total) effects of the hierarchical factors on the subtests, through the intermediate factors. So, for example, in Figure 31.12, the loading of Sequential and Quantitative Reasoning on *g* is .736 (.92 × .80). SEM (and CFA) programs will easily calculate these indirect effects. The *g* loadings of the DAS-II are shown in Table 31.2 (arranged from largest to smallest, based on the higher-order loadings). Note that these *g* loadings are model-specific; the estimates shown in Table 31.2 apply to the model shown in Figure 31.12. If the model is correct, then the *g* loadings will be accurate.

Bifactor Model

As noted above, an alternative method of testing hierarchical models is through what is called a *bifactor* model. In a bifactor model, all subtests are loaded directly on both a G factor (see below regarding G versus *g*) and on narrow factors, with the factors generally orthogonal or uncorrelated (cf. Carroll, 1995; Gignac, 2008; Gustafsson & Balke, 1993; Keith, 2015; Reynolds & Keith, 2013).

A bifactor version of the DAS-II model is shown in Figure 31.13. (The ovals representing the unique variances of the subtests are not shown in the figure to help simplify it, but were included in the

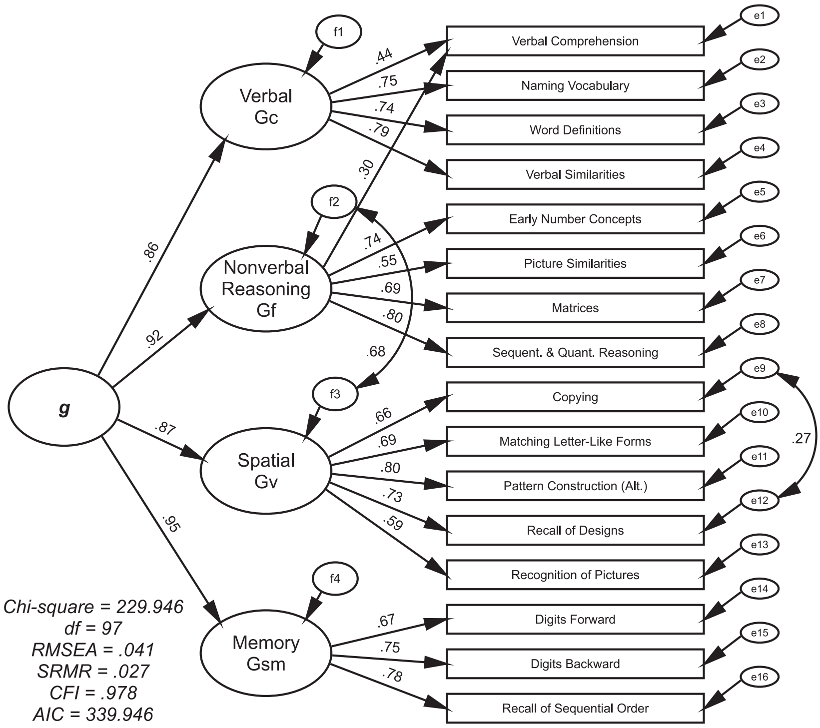


FIGURE 31.12. Standardized solution for a higher-order model of the structure of the DAS-II.

TABLE 31.2. Loadings of DAS Subtests on a Higher-Order *g* Factor and a Bifactor *G* Factor

Subtest	Higher-order *g*	Bifactor G
Recall of Sequential Order	.74	.73
Sequent. and Quant. Reasoning	.74	.75
Recall of Digits—Backward	.72	.72
Pattern Construction (Alt.)	.70	.72
Verbal Similarities	.68	.69
Early Number Concepts	.68	.69
Verbal Comprehension	.65	.63
Naming Vocabulary	.65	.64
Matrices	.64	.65
Word Definitions	.64	.62
Recall of Designs	.64	.62
Recall of Digits—Forward	.63	.62
Matching Letter-Like Forms	.60	.61
Copying	.58	.57
Recognition of Pictures	.51	.50
Picture Similarities	.51	.51

analysis.) This initial version of the model did not estimate well, and required modification. (The residual variance of the Sequential and Quantitative Reasoning test was negative, among other issues.) We have constrained the loadings of the subtests on the Gf factor to be equal (except Verbal Comprehension). The results of this model are shown in Figure 31.14. This model also fits the data well, although slightly worse than the higher-order model (with the changes, the models are not nested, so this comparison is based on the AIC).[6]

We first focus on the findings of this model, and then discuss how this model relates to the previous higher-order model. First, as shown in Table 31.2, the loadings of the subtests on the (first-order) G factor are similar, but not identical, to those shown as indirect effects for the higher-order model (we refer to such first-order general factors as G, rather than g, to make the distinction between first- and second-order general factors explicit). In contrast, the loadings of the subtests on the Gc, Gv, and Gsm factors are much smaller than in the previous model (see Table 31.3). The practical and conceptual differences in the two models lead to these differences.

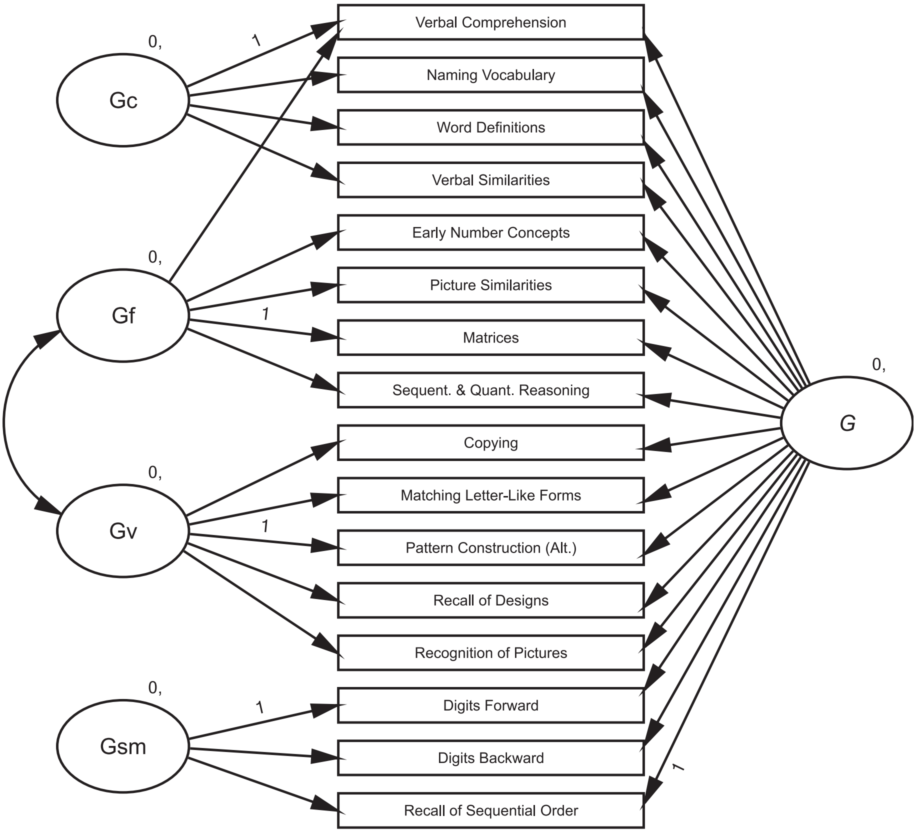


FIGURE 31.13. A bifactor model of the structure of the DAS-II.

Bifactor versus Higher-Order Models

How do the models differ, and in what ways are they similar? Some researchers treat bifactor models as equivalent to higher-order models, but clearly they are not. Although both models suggest that the subtests are affected both by general intelligence and by one or more abilities that are less broad, the nature of that influence differs. The higher-order model assumes that *g* influences individual tests through the broad abilities, or multiple intelligences, and that *g* indeed influences, or is related to, the broad abilities (or the broad abilities are to some extent also reflections of *g*). The bifactor model seemingly makes no assumptions about the relation between *g* and the broad (first-order) factors, instead only specifying that the subtests measure both G and broad abilities. Further examination shows this initial impression is not correct, however. This model in fact asserts that G is *unrelated* to the broad abilities, and (usually) the broad abilities are unrelated to each other. Relatedly, the bifactor model assumes only direct effects for G on the measured variables, whereas the higher-order model asserts indirect effects.

The higher-order model is more restricted than is the bifactor model. Because *g* only affects the subtests via the broad abilities in the higher-order model, the indirect effects of *g* on the subtests are restricted by the effect of *g* on the broad abilities. The bifactor model has no such restrictions. Because of these restrictions, the higher-order model is a more constrained model than is the bifactor model. In fact, it is possible to go from one model to another in several ways (Yung, Thissen, & McLeod, 1999). Most simply, both models may be considered subsets of a model in which *g* influences all first-order factors and subtests directly (a model that would be underidentified without additional constraints). With such a model, if the effects of *g* on the first-order factors are deleted, the bifactor model ensues; or, if the direct effects of *g* on the subtests are deleted, the higher-order model results.

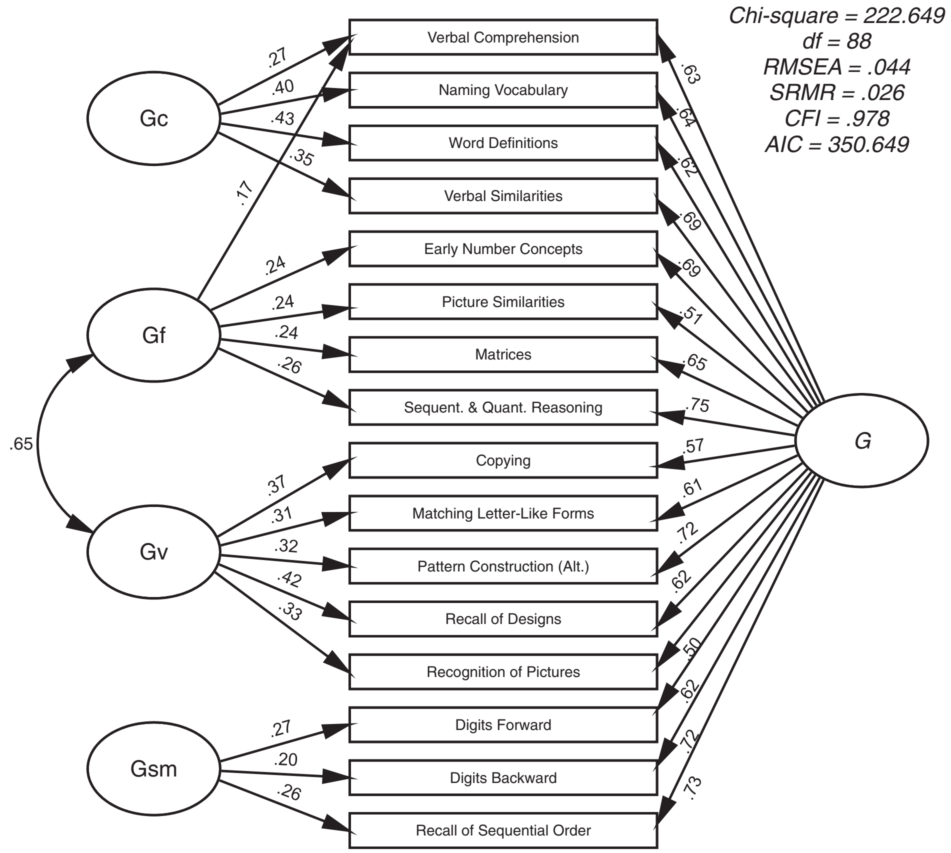


FIGURE 31.14. Standardized solution for a DAS-II bifactor model.

It is also possible to get equivalent-fitting bifactor and higher-order models by adding paths from *g* directly to the subtests in the higher-order model. The number of additional paths is equal to the difference in *df* for the two models. And the difference in *df* for a strict higher-order and a strict bifactor model equals the number of subtests minus the number of factors. To do so, one adds *g*-to-subtest paths for every subtest, minus 1, per factor. So, for example, if factor 1 has four measured variables, we would add three additional *g*-to-subtest paths in addition to the existing factor 1-to-subtest paths (and so on, for the other factors) (Mansolf & Reise, 2017).

One seeming advantage of the bifactor approach is that when all factors are included at the same level, the paths may be considered *partialed* effects. That is, the loadings of the Gc, Gf, and Gsm factors are the effects of these factors on the subtests, *with* G *statistically removed,* or partialed out. As a result, the findings from such models are often similar to a Schmid–Leiman transformation in a higher-order EFA, in which first-order and higher-order loadings are orthogonalized to aid in interpretation (Schmid & Leiman, 1957). This would seem to be an important (practical) advantage for the bifactor approach. But it is also very simple to obtain such orthogonalized results in a higher-order model as the indirect effect of the factor disturbances on the subtests (f1, f2, etc., in Figure 31.12). (To obtain the paths from disturbances to first-order factor, UVI can be used for scaling the disturbances. This has been done in Figure 31.12.) Because the factor disturbances are the factors with the effects of *g* removed, these indirect effects (i.e., the path from the disturbance to the first-order factors times the subtest loading on the factor) are then the loadings of the subtests on the factors, with *g* removed (see Reynolds & Keith, 2013 or 2017, for illustrations; the 2013 chapter also compares these models in more detail). The difference, as shown in Table 31.3, is

TABLE 31.3. Loadings of DAS Subtests on Broad-Ability Factors in a Higher-Order and a Bifactor Model

	Higher-order		Bifactor
Factor/subtest	Direct	Unique	Direct
Verbal (Gc)			
Verbal Comprehension	.44	.22	.27
Naming Vocabulary	.75	.38	.40
Word Definitions	.74	.37	.43
Verbal Similarities	.79	.40	.35
Nonverbal Reasoning (Gf)			
Verbal Comprehension	.30	.11	.17
Early Number Concepts	.74	.28	.24
Picture Similarities	.55	.21	.24
Matrices	.69	.26	.24
Seq. and Quant. Reasoning	.80	.30	.26
Spatial (Gv)			
Copying	.66	.33	.37
Matching Letter-Like Forms	.69	.34	.31
Pattern Construction	.80	.39	.32
Recall of Designs	.73	.36	.42
Recognition of Pictures	.59	.29	.33
Memory (Gsm)			
Recall of Digits—Forward	.67	.21	.27
Recall of Digits—Backward	.75	.23	.20
Recall of Sequential Order	.78	.24	.26

Note. The "Direct" column in the higher-order model shows the uncorrected loadings. The "Unique" column shows the direct values, with *g* removed from the broad abilities. The "Direct" bifactor values are the values with G removed from the subtests.

that the higher-order approach allows the estimation of the effects of both the actual broad abilities on the subtests (the simple effects) *and* the effects of the broad abilities with *g* partialed out (the partialed effects). A bifactor approach only produces the partialed effects; that is the only interpretation allowed.[7] The partialed effects for the two solutions are often quite similar (see, e.g., Reynolds & Keith, 2017).

Another potential advantage of the bifactor approach is that it appears agnostic as to the relation between the *g* factor and the other factors. Said differently, this approach may be somewhat more exploratory than a higher-order approach. But this advantage does not come without a cost. All CFA models (and, indeed, all EFA models) make assumptions about the nature of intelligence. The models shown in Figures 31.11 and 31.13 make very different assumptions about the nature of intelligence. The higher-order model says that *g* is a superordinate ability—an ability that is further removed from the subtests, indicating a higher-level of abstraction. In a strict higher-order model, there are no direct effects from *g* to the subtests, which seems consistent with the inability to define *g* based on surface characteristics of specific tests (see Jensen, 1998, Ch. 4). *g* is more abstract than the broad abilities,[8] and it is also more general because it influences every subtest, albeit indirectly. The second-order loadings in the higher-order model are also useful, in that they provide validity information about the broad abilities and their relation to *g*. Carroll (1993), for example, arranged the stratum II factors (broad abilities) by their loadings on *g*, with Gf being the most closely related. Many studies have shown this same relation with higher-order intelligence models.

The bifactor approach, in contrast, says that *g* is no more important than other abilities; that is, it does not affect those abilities, but affects test scores directly. The bifactor method is a more top-down approach, with G taken into account concurrently with the broad abilities, whereas the higher-order approach is more of a bottom-up approach (Gustafsson, 2002). In sum, models imply theories, and researchers should know what theory their model is implying! Given the intertwined nature of measurement and theory, it is important to base the choice of a model on theoretical grounds as well as well as practical ones.

A practical disadvantage of the bifactor approach is that the model is sometimes difficult to estimate. Factors must include more than two indicators, or empirical underidentification will result. To avoid this, additional indicators, or meaningful factor intercorrelations (cf. Keith, Reynolds, Patel, & Ridley, 2008), are required; or one must make additional constraints, which may result in a worse fit and which require justification. A common method for dealing this problem is to constrain the loadings of the two indicators of a factor to be equal to each other (here we have constrained four loadings to be equal). This method is useful in getting the models to converge, but should be explained in the research results. Even with three or more tests per factor, bifactor models can be touchy. In our experience, it sometimes helps to specify start values for many parameters.

The bifactor model often fits better than the higher-order model. As already noted, the higher-order model is a more constrained version of the bifactor model. Because the effects of *g* on a broad ability (e.g., Gc) are the same for each subtest within that broad ability in the higher-order model, the ratio of *g* loadings to unique loadings within each broad ability is proportional. The bifactor model does not have this underlying proportionality condition. This condition is often referred to as a *proportionality constraint,* and research shows that inducing such proportions in simulated data results in equivalent-fitting models for the two methods (Gignac, 2016). As noted by Mansolf and Reise (2017), this is not a CFA model constraint in the common sense of the term. It is in fact the result of a unique set of underlying tetrad constraints for the higher-order model.

The often better-fitting nature of the bifactor model may seem a reason to prefer it, but research has shown that the bifactor model often fits better even when it is incorrect (Murray & Johnson, 2013). Unmodeled complexity (e.g., small, unmodeled correlated errors or cross-loadings) may increase the tetrad violations and make a bifactor model fit better than a higher-order model even when a higher-order model is the true model (Mansolf & Reise, 2017). The bifactor model is prone to overfitting and may fit even in the presence of unlikely or invalid data (Reise, Kim, Mansolf, & Widaman, 2016). These findings make sense if we consider that in a bifactor model the correlations among measured variables are assumed to reflect a general factor, even if they do not result from an underlying latent variable that is the "cause" of those correlations. "A bifactor model may provide a superior fit, not because it more accurately reflects the substantive phenomenon, but because fit indices are biased in its favor when models are misspecified (i.e., pure models are applied to data that contain unmodeled complexities)" (Mansolf & Reise, 2017, p. 121). At any rate, it appears that the choice between a bifactor and a higher-order model should not be based on model fit.

Our position is that one's model choice should reflect one's guiding theory. Theoretically, the model analyzed should, at a minimum, reflect the theory underlying the test, or an alternative theoretical specification. The higher-order model would seem to have a clear advantage in this regard, given that most theories of intelligence that include a general factor suggest that it is a higher-order, hierarchical factor (Jensen, 1998). As already noted, it may also be considered useful for theory testing, since effects of *g* on the first-order factors should have predictable effects across studies. Tests are not always scored this way, of course (e.g., subtests, rather than scales, are often added together to form IQ composites), but most test manuals reference a higher-order type structure if theoretical justification for the test structure is presented. In contrast, the bifactor model is not consistent with most modern theoretical orientations with which we are familiar. See, however, Beaujean (2015) for a spirited defense of the bifactor model as reflecting Carroll's three-stratum theory. Indeed, even the Schmid–Leiman transformation was originally developed to transform, for interpretive purposes, results of *higher-order* EFAs. For these reasons, we generally prefer a higher-order model. But we also believe that all models are incorrect to some extent, and that all models to some extent have their own set of practical uses. Thus we routinely look at both types of models, and interpret them in combination rather than in competition. The bifactor model may be especially useful for isolating measured variables that do not belong on a broad ability (Keith et al., 2016). See Gustafsson and Åberg-Bengtsson (2010) and Reynolds and Keith (2013) for further discussion.

TESTING HYPOTHESES ABOUT THE SIMILARITY OF FACTORS

CFA is also a useful method for testing hypotheses about how factors or tests should be interpreted. Such research often involves testing the similarity of factors within a test or between two tests. CFA can be used to test whether two tests designed to measure the same factors (e.g., the G*f* factors from the Woodcock–Johnson IV [WJ IV] and the KAIT) do, in fact, measure statistically indistinguishable factors. By the same token, the method can be used to test whether two tests designed to measure *different* abilities do in fact measure distinguishable factors (e.g., Keith, Kranzler, & Flanagan, 2001). A brief within-test example follows.

Are Gf and Gv Separable for the WISC-V?

The WISC-V (Wechsler, 2014) is designed to measure five distinct yet correlated cognitive abilities consistent with the constructs Gc, Gf, Gv, Gsm, and Gs (crystallized ability, fluid reasoning, visual–spatial ability, working memory, and processing speed, respectively) from Cattell–Horn–Carroll (CHC) theory. One major difference between the WISC-V and the previous version, the WISC-IV, is the separation of the nonverbal tests into a Fluid Reasoning Index (Gf) and a Visual Spatial Index (Gv); the WISC-IV had such tests combined into a single Perceptual Reasoning Index. Recent research has questioned this separation in the WISC-V, however, suggesting instead that the Fluid Reasoning Index and Visual Spatial Index measure one set of underlying abilities, not two (e.g., Canivez, Watkins, & Dombrowski, 2017).

Figure 31.15 shows a first-order version of the WISC-V's structure, with the 16 subtests and the five factors (with CHC-type labels) they measure, according to the test publisher. Note that the Arithmetic test is thought to measure both fluid reasoning and working memory, and thus has a cross-loading on both Gf and Gsm. There are several additions to the figure that need explanation. First, note that UVI (factor variances set to 1) has been used to define the factors. This helps in the hypothesis testing described below. Note also the values of 0 above the factors and the residuals. Means and intercepts have been analyzed for these models; latent means are here (and commonly) fixed to 0. Finally, note the labels for some of the correlations (eq1, eq2, etc.). These are used to constrain these values sequentially.

Figure 31.16 shows the standardized solution for the model, with various fit indices. The model generally fits well, and all factor loadings are substantial. Because the UVI method has been used to set the scale of the latent factors, the unstandardized solution would show the same values for the factor covariances as shown here for the factor correlations (with the factors standardized, the

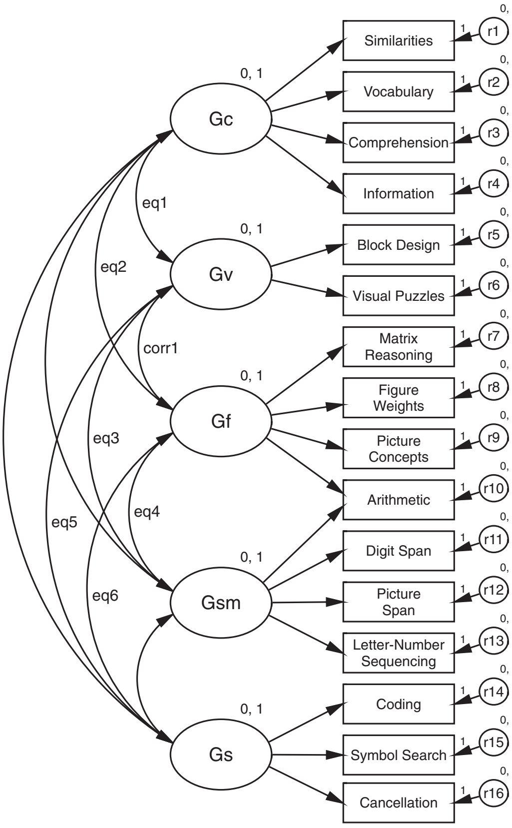


FIGURE 31.15. A first-order CFA model of the WISC-V. This model is designed to test the similarity or dissimilarity of the Gf and Gv factors.

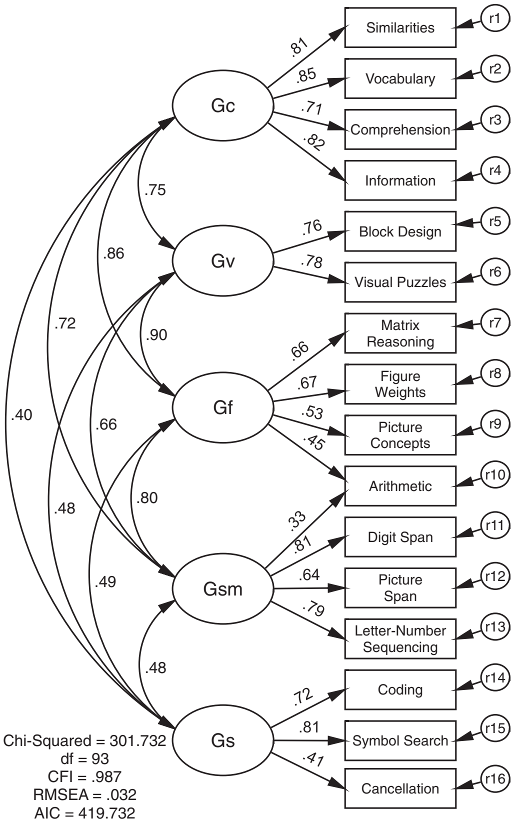


FIGURE 31.16. Standardized solution for the WISC-V CFA model. The correlation of the Gf and Gv factors is higher than the other factor correlations (.90).

covariances are standardized also; standardized covariances are correlations). Note the very high correlation between Gf and Gv (.90). This value certainly suggests considerable overlap between these two variables. But are they statistically distinguishable; are they separable?

For the next step in this analysis, we have constrained the correlation between Gf and Gv to 1 (corr1 = 1). As shown in Table 31.4, the fit of this model is considerably worse than that of the original five-factor model ($\Delta\chi^2 = 43.71$, $df = 1$, $p < .01$). This finding would seem to suggest that the two factors are indeed different. But if Gf and Gv were the *same* factor, then they would also have the same correlation with other variables, including other factors. Thus a more complete test of the identity of the two factors would also set correlation eq1 equal to eq2, correlation eq3 equal to eq4, and correlation eq5 equal to eq6. These additional constraints further degrade the fit of the model, as shown in Table 31.4 ("3. Correlations same" model).

A more direct way of testing the equivalence of these two factors would be to simply combine them into a single perceptual reasoning factor, similar to the Perceptual Reasoning Index from the WISC-IV. It may be unclear, however, whether this model is to be nested with the initial model. As shown in Figure 31.17, this method produces exactly the same fit and *df* as the "correlations same" model does, demonstrating that this model can indeed be considered the same as model 3, and thus nested with model 1 (see Table 31.4).

One final model is worth examining. We have so far rejected the notion that Gf and Gv on the WISC-V overlap to such an extent that they should be combined into a single factor. But is the Gf-Gv correlation of larger-than-normal magnitude for cognitive factors? For model 5, the correlation between Gf and Gv has been set to the level of the median correlation between other pairs of factors (.66). As shown in Table 31.4, this model also leads to a degradation in model fit. The correlation between these two factors is not 1.0, but it is also statistically significantly larger than the average (median) correlation among the other cognitive factors.

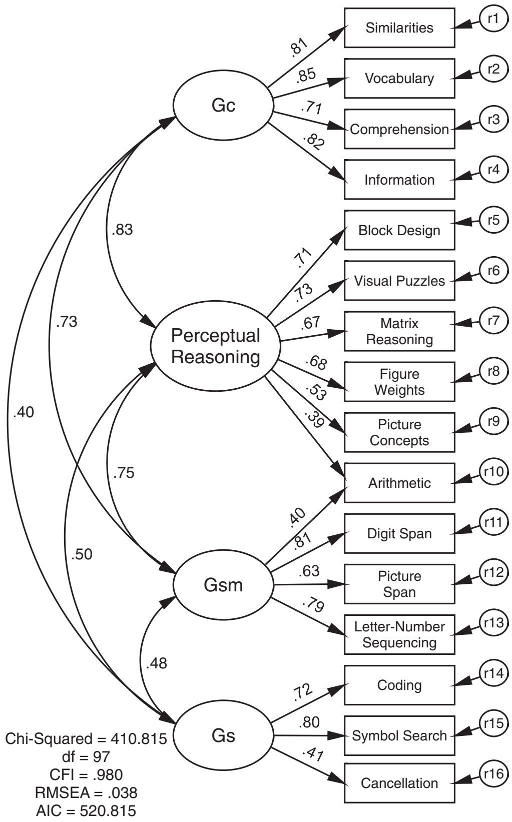


FIGURE 31.17. A model in which the Gf and Gv factors are combined. This model fits worse than the models with separate factors.

TABLE 31.4. Comparison of Competing Hypotheses about the WISC-V Structure (Gf and Gv Factors)

Model (Figure no.)	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	RMSEA	AIC
1. Five factors (Figure 31.16)	301.73	93				.99	.032	419.73
2. Gf = Gv	345.44	94	43.71	1	<.01	.98	.035	461.44
3. Correlations same	410.81	97	109.08	4	<.01	.98	.038	520.81
4. Four factors (Figure 31.17)	410.81	97	109.08	4	<.01	.98	.038	520.81
5. Median correlation	381.32	94	79.59	1	<.01	.98	.037	497.32

Note. Each model is compared to the first model.

It is worth noting that the combined Gf-Gv factor shown here differs from that proposed by Canivez and colleagues (2017), in that we show Arithmetic as loading on both the Gf and the Gsm factors, whereas Canivez and colleagues have included this subtest only on Gsm. The conclusions (that the factors are indeed separable) would be the same with that starting model, however. Our starting model is consistent with the first-order model supported in Reynolds and Keith (2017).

For a further illustration of the power of CFA to test competing theory- and research-driven hypotheses about the constructs measured by tests, see Keith and colleagues (2001). In that study, the authors tested whether the Cognitive Assessment System (CAS) measured CHC abilities, or whether it measured distinct processes as outlined in the Planning, Attention, Simultaneous, and Successive theory of mental processing. In addition to testing the similarity of factors, the authors tested a series of joint hierarchical and integrated models from both theoretical orientations, and whether the higher-order *g* factor assessed by the two instruments was the same underlying factor (cf. Floyd, Reynolds, Farmer, & Kranzler, 2013; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004).

TESTING THE SIMILARITY OF FACTOR STRUCTURE ACROSS GROUPS: INVARIANCE

An important subset of questions about the constructs measured by intelligence tests involves questions about whether the test measures the same constructs across groups. Does a multiple-age battery, for example, measure the same set of constructs for 8-year-olds as it does for 18-year-olds? Does a new intelligence test measure G*f* and G*c* abilities for European American students, but merely test-taking skills for students from other ethnic backgrounds? Does a verbal comprehension factor have the same meaning for boys as it does for girls? These questions ask, in essence, whether the factor structure and corresponding factor loadings of the tests vary across groups. These are questions of factorial invariance.

The method of *multigroup CFA (MG-CFA)* provides an excellent method for answering such questions. In MG-CFA, any of the parameters that are estimated or fixed in a model can be specified as being invariant across two or more groups. For example, suppose we wish to determine whether a set of tests measures the same constructs for boys and girls. We could specify that the factor loadings of that series of tests on their associated factors are the same for boys and girls. Additionally, we could also specify that the factor covariances, the factor variances, and the unique and error variances are identical across groups.

In one of the earliest examples of using CFA to test for invariance (Jöreskog, 1971), MG-CFA was used to test for invariance of covariance structures. Meredith (1993) extended and formalized the testing of factorial invariance under the more general scope of measurement invariance. Most of the models tested thus far in this chapter have focused on structuring covariances (solving for CFA models using variances and covariances), and have not involved structuring the means. Meredith demonstrated the use of MG-CFA with mean structures, often referred to as *multigroup mean and covariance structure analysis (MG-MACS),* to test various aspects of measurement invariance. This model-based framework compares various models specifying different degrees of invariance in a structured approach using MG-CFA. The following is an example of this approach.

This example uses test score data from 128 girls and 122 boys who were part of the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004) norming sample to test for measurement invariance across the sexes. The example tests whether the KABC-II measures the CHC abilities in the same way for 4-year-old boys as it does for 4-year-old girls. For children who are 4 years of age, the KABC-II provides four index scores designed to reflect four CHC abilities: Gsm, Glr (long-term retrieval), Gv, and Gc. These composite scores are generated from scores obtained from two or three subtests. The factor model is shown in Figure 31.18. This factor model is consistent with the scoring structure of the KABC-II. Note that the KABC-II also includes a global composite score, but in this example we focus on testing for measurement invariance in the broad-ability composites only and not the global composite (for higher-order models for the KABC-II, see Potvin, Keith, Caemmerer, & Trundt, 2015; Reynolds, Keith, Fine, Fisher, & Low, 2007; Reynolds, Keith, Ridley, & Patel, 2008).

As with the WISC-V models, the model shown is slightly different from previous models. Note the values of 0 next to all the latent variables ("0," in Figure 31.18). Again, these are the means of the latent variables, which are initially fixed to 0.

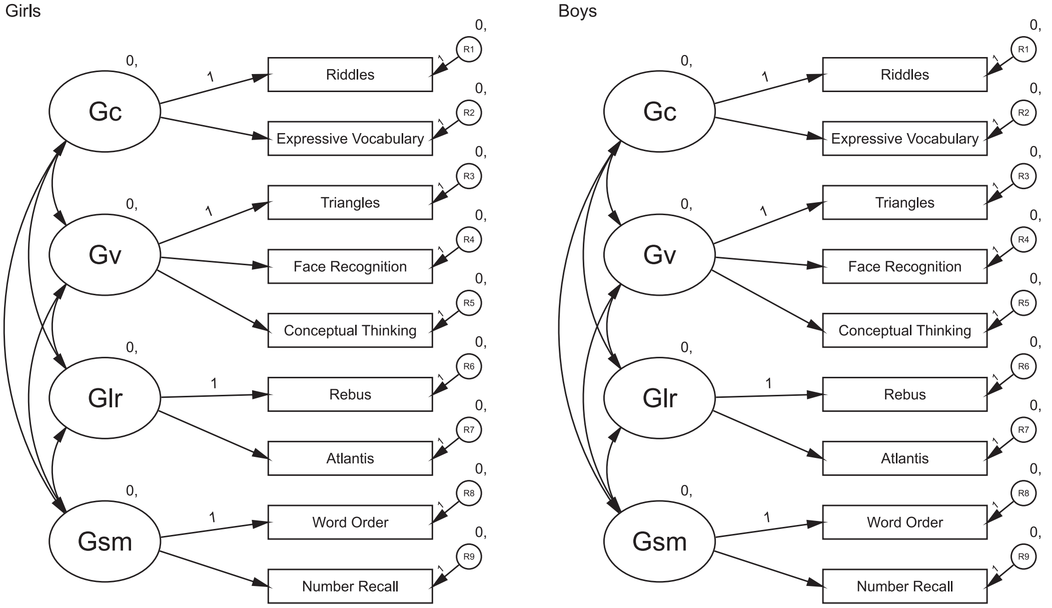


FIGURE 31.18. MG-CFA setup for testing configural invariance across sexes for the KABC-II for 4-year-olds. The same configuration of factors is specified for girls and boys.

This model also estimates the intercepts for the subtests, although these are not shown in the figure. The intercepts, when estimated, may be interpreted in a way similar to those in a regression model, with the factor being the independent variable and the subtest the dependent variable. Thus the intercept for the Riddles subtest, for example, is the expected Riddles score for someone with a value of 0 on the latent Gc factor. In previous models, we have not been concerned with means or intercepts; however, it is necessary to consider, and explicitly model, means and intercepts when testing for measurement invariance. All SEM programs can explicitly model means and intercepts (i.e., mean structures).

Steps in Invariance Testing

Testing for measurement invariance via MG-CFA typically involves the application of increasingly restrictive sets of equality constraints on parameters across groups. Different researchers use different terms to describe the different facets of invariance; we generally use the hierarchy discussed by Meredith (1993). In the first step, the same factor structure is imposed on each group. This initial step is often referred to as *configural invariance* (cf. Thurstone's [1947] *configurational invariance*). This type of invariance requires that the CFA model is configured similarly across groups. This step may be performed with multigroup analysis; alternatively, models can be estimated separately for each group. After that step, steps for testing for *weak factorial invariance* (also known as *metric* or *factor-loading invariance*), *strong factorial* (*scalar* or *intercept*) *invariance*, and *strict factorial* (*item/subtest residual*) *invariance* are followed. Greater detail about the meaning of invariance at each step is explained as the example is worked through. This particular order is not strictly required. For example, some might first test for invariance of the covariance structure (factor loadings and residuals) and then introduce the means (intercepts). There are some important considerations, however. For example, factor loadings should be invariant if the intercepts associated with those loadings are to be tested for equality.

Configural Invariance

Configural invariance simply entails establishing the same configuration of factors across the groups. In this example, the model shown in Figure 31.18 has been estimated for both boys and girls by using MG-CFA (the model for girls is on the left, the model for boys on the right). The factors and pattern of free and fixed loadings within groups are the same across the groups. One loading per factor has been fixed to 1 to scale the factors; this is an important consideration because the factor variances (and covariances) should vary freely across groups. There are alternative ways to scale the factors, such as the *fixed-factor method* and *effect coding.* These will result in the same model fit statistics, but may provide some interpretive advantages. These variations are beyond the scope of the current chapter (see Little, 2013).

The fit of this MG-CFA model is excellent (Table 31.5). The parameter estimates within each group are reasonable. The fit within each group is excellent as well. The factor model appears to be a good fit to the data for boys and girls. Note that if these models were analyzed separately, the sum of the χ^2 and *df* for boys (χ^2 = 17.5; *df* = 21) and girls (χ^2 = 23.3; *df* = 21) would be equal to the χ^2 and *df* from the MG-CFA. Configural invariance is typically thought of as a baseline model, one that future models may be tested against. If modifications are needed to the factor model, this step is the best place to make those modifications.

Weak Factorial Invariance

The configuration of the factor structure has been established in each group in the configural invariance model. The next step is to restrict corresponding factor loadings to be equal across groups, and is referred to as a test of weak factorial (or metric or factor loading) invariance. Each factor has one loading already fixed to 1, and those remain.[9] The factor loadings estimated freely within each group in the configural model are now constrained to be equal across groups. These constraints are made to the unstandardized loadings. So for example, the (unstandardized) loading of Expressive Vocabulary on Gc is constrained to be equal across the sexes. Such a constraint, if supported in this example, means that a one-unit increase in Gc should result in the same unit increase in Expressive Vocabulary performance for boys as for girls. In all, five constraints have been added: one loading on each Gc, Glr, and Gsm factor, and two loadings on the Gv factor. Figure 31.19 shows the models for boys and girls, with these constraints. In Amos, these are accomplished by labeling the corresponding parameters with the same names across the two groups (a for the factor loading of Expressive Vocabulary, etc). Other programs use other methods, but such cross-group constraints are possible in all SEM programs. These constraints resulted in a change of 5 *df.* Because this model is nested within the configural model, $\Delta\chi^2$ may be used to test for degradation in model fit. The inclusion of these five additional constraints does not result in degradation of model fit ($\Delta\chi^2$ = 3.18 [5], *p* = .67). Our experience suggests that the $\Delta\chi^2$ criterion may be overly sensitive for *measurement* invariance tests in larger samples (especially for tests beyond weak factorial invariance), but it is useful for specific aspects of *structural* invariance, which we discuss later. It works well here, in part because of the small (for CFA) sample sizes.

Simulation research has supported change in CFI (ΔCFI) as a useful fit index for comparison of models in invariance testing, with the decreases in CFI of less than .01 providing support for invariance (Cheung & Rensvold, 2002), and ΔCFI is often our preference for comparing models in invariance testing at the measurement level, especially if the goal of the research is not related to the studying the scale itself (χ^2 and *df* should, however, always be reported for all models). Here, however, almost all CFI values are 1.0, so we have used the $\Delta\chi^2$ criterion instead. Weak factorial invariance is tenable. This finding means that the relations between the factors and the subtests are the same across groups, or that the factors are the "same" (in some sense of the word) for both groups. Group comparisons of *factor* variances and covariances would now be considered acceptable. And

TABLE 31.5. Tests of Factorial Invariance for the KABC-II Data for 4-Year-Olds

Invariance model	$\chi2$	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	ΔCFI
Configural	40.82	42				1.000	.000
Weak (metric)	44.36	47	3.54	5	.62	1.000	.000
Strong (intercept)	49.54	52	5.18	5	.39	1.000	.000
Strict (residual)	62.96	61	13.42	9	.14	.997	–.003

Note. Each model is compared to the previous model.

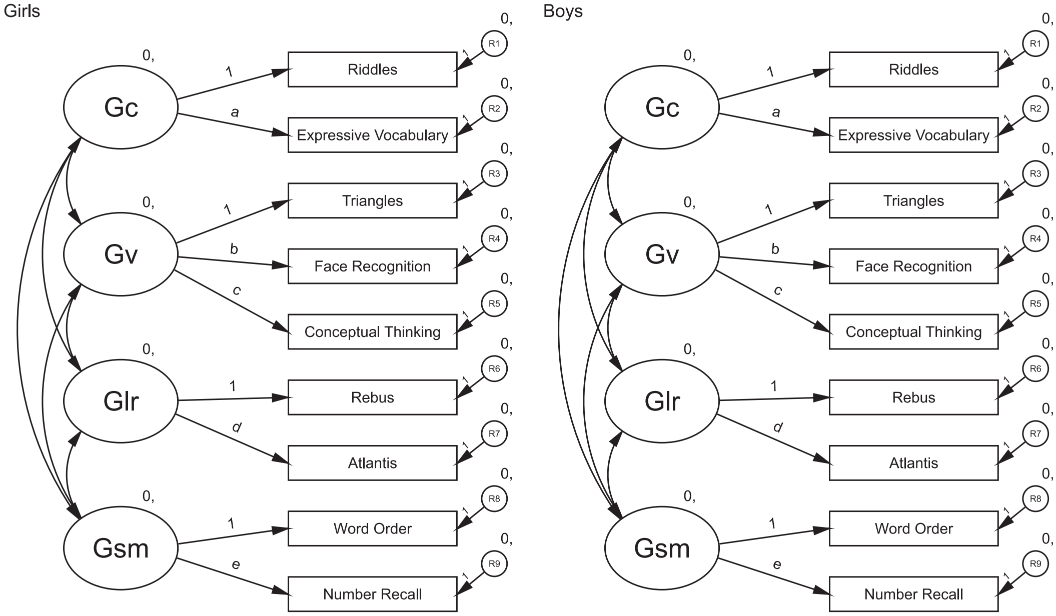


FIGURE 31.19. Testing for weak factorial (metric) invariance. Factor loadings are constrained to be equal for girls and boys.

in our experience, weak factorial invariance is in fact commonly supported, especially in research using intelligence subtests, which often do not involve many indicators per factor and have relatively strong loadings (i.e., reliable factor indicators). On the rare occasions when it is not supported, there are often very large numbers of groups. Even though weak factorial invariance is tenable, this is not yet enough evidence to say that boys and girls with the same latent levels of the CHC factors would indeed have the same observed scores.

Strong Factorial Invariance

Strong factorial (or intercept or scalar) invariance needs to be established before group comparisons may be made of the common factor means (or composite score means). Strong invariance requires that covariances *and* means are analyzed and modeled. In the weak invariance model, the corresponding unstandardized factor loadings have been constrained to be equal across groups. As in previous models, the means have not been modeled. Or more exactly, the factor means have been fixed to 0, and subtest intercepts have been allowed to vary across groups. A test of strong factorial invariance requires that the subtest intercepts (means) also be modeled. Specifically, factor mean differences are allowed across groups, while all corresponding subtest intercepts are constrained to be equal across groups. This specification imposes the restriction that all subtest mean differences *are the results of difference in means on the common factors*. Unstandardized results are shown in Figure 31.20. Nine equality constraints, one per subtest, have been added to the corresponding subtest intercepts (e.g., the Expressive Vocabulary intercepts have been constrained to be equal across boys and girls). In previous models, the latent factor means have been fixed to be 0 (this is the default). To allow differences in the common factor means across groups, the latent means of one group have been freed, while the other "reference" group means remain fixed to 0. Here we denote girls as the reference group and keep their latent means fixed to 0. Boys are the comparison group, and their latent factor means are freed and estimated (note that the 0's are replaced by other values for the boys' factors in Figure 31.20). Technically speaking, the latent means are not estimated; rather, the boys' latent mean estimates represent the difference from the means of the girls because the constraints on corresponding intercepts essentially force mean differences "up" through the factor means. Freeing these four factor means results in 4 fewer degrees of freedom. Therefore, the Δdf

for this model compared to the weak factorial invariance model should equal 5 (i.e., 9 subtest intercept constraints minus 4 latent means freed). It is important to check the *df* changes in the models; as reviewers for journals, we have often noticed *df* values that are not consistent with the intended invariance test. For example, a common mistake is that researchers keep latent factor means set to 0 for both groups after they constrain the intercepts. It is important both conceptually and statistically that the factor means be allowed to differ because strong factorial invariance is a test that the factor mean differences can fully account for differences in the observed means. It is not simply a test of the difference in observed means across groups.

Given that girls are the reference group and their latent means are fixed to 0, positive latent mean values for boys indicate that boys have higher latent means, and negative values indicate that girls have higher latent means. Before those latent means may be interpreted meaningfully, however, strong factorial invariance must be tenable. That is, those differences in latent means should account for the differences across the sexes in the observed test scores. If the strong factorial constraints have been added and the model fit has been degraded, then the researchers should investigate the reasons for such misfit, and the latent mean differences at this point would be misleading. Similarly, the composite scores from the test may be misleading across the sexes. It is important not to stop at this point, but rather to investigate the model to determine why invariance is not tenable. The answers may suggest that it is of very little practical importance, or they may provide new insights (e.g., one group differs on a subtest-specific mean rather than the common factor mean).

To return to our example: Because the weak factorial invariance model is tenable, all prior constraints are maintained; nine additional intercepts constraints are applied; and the boys' Gc, Gv, Gsm, and Glr factor means are freed. The addition of these constraints does not result in a statistically significant degradation in model fit ($\Delta\chi^2 = 5.18$ [5], $p = .36$). Strong factorial invariance is tenable. The unstandardized results are shown in

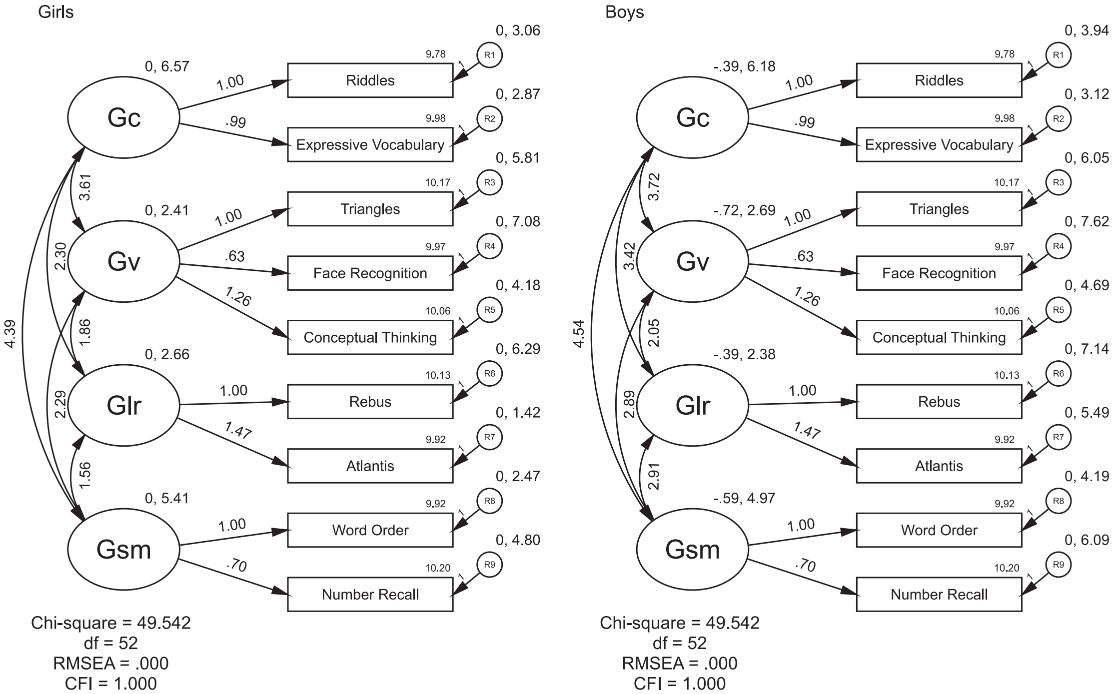


FIGURE 31.20. Unstandardized results for a test of strong factorial (intercept) invariance. Note the numbers above each measured variable. These are the intercepts, which are constrained to be equal across the sexes in this step. The numbers above the latent variables are the means and variances. The latent means for the broad abilities are allowed to vary for boys, and the numbers shown are the latent mean differences for boys as compared to girls.

Figure 31.20. Differences in subtest means across the sexes are explained by differences in the latent factor means between boys and girls. Stated differently, boys and girls with the same level of Gc (or Gv, etc.) would earn the same score on the subtests. It would thus be acceptable for researchers to compare the factor means or the observed composite means for boys with girls. It may also be said that KABC-II CHC scores are unbiased with respect to sex.

Strict Factorial Invariance

The last step in testing for measurement invariance is referred to as strict factorial or residual invariance. In tests of strict factorial invariance, equality constraints are imposed on the variances of the residuals (error + specific variance; R1 through R9 in the figures). Strict factorial invariance allows for cross-group comparisons of both observed means and variances. There is disagreement about the necessity of strict factorial invariance (cf. Little, 1997; Lubke & Dolan, 2003). Some of it boils down to the purpose of the research. Strictly speaking, however, strict factorial invariance is consistent with measurement invariance (Meredith & Teresi, 2006).

The application of these nine additional constraints on the residuals does not result in a statistically significant degradation of model fit (Table 31.5). All sex differences in observed (subtest) score means and variances are due to differences in the common factor means and variances/covariances. The CHC composites from the KABC-II are completely unbiased with respect to sex (for 4-year-olds). Observed scores on the KABC-II depend on the examinees' latent abilities, not on sex.

In this example, measurement invariance has been established with relative ease. In many situations, however, the findings may be more complicated, and partial measurement invariance may be investigated. Such a topic is beyond the scope of this chapter (see Byrne, Shavelson, & Muthén, 1989).

Substantive Questions

We have thus far focused on tests of measurement invariance, which may subsume many interesting research questions. Researchers are often interested in testing for construct bias for a test across two or more groups (e.g., Trundt, Keith, Caemmerer, & Smith, 2017). Such questions may be modeled by using tests of factorial invariance.

Other interesting research questions may also be answered. For example, we have mentioned that differences in latent means are estimated and can be compared because strong factorial invariance is tenable (and even better, strict factorial invariance is). Girls' latent means for each factor are 0, while the boys' are estimated. Negative latent mean values thus indicate that boys' levels of abilities are lower. Each latent mean is negative (Gc = –.39; Gv = –.72, Glr = –.39, Gsm = –.59), although only Gv is statistically significant at the $p < .05$ level (see Figure 31.20). To be clear, a mean difference on this latent factor does *not* indicate that the test is biased; rather, it indicates that there is a "true" mean difference between boys and girls in these latent constructs. That mean difference in the latent factor accounts for the differences in the observed scores, so at age 4, boys on average score slightly lower on Gv (it may also be of interest to include a *g*-factor and test for differences using higher-order models) (see Keith et al., 2008; Reynolds et al., 2008). Latent mean differences can also be tested with model-fitting procedures by constraining the factor mean to be equal for boys and girls (by fixing the boys' factor means back to 0) and checking for degradation in model fit.

Other questions may also be of interest. Are the factor variances equal across groups? Or are the latent factors equally differentiated across groups? To answer those questions, researchers may want to test for equality of the variances and covariances/correlations (see Little, 1997). It is once again important to note that these are substantive questions, and measurement invariance must be established before these comparisons are made. These questions may follow tests of measurement invariance, but *during* testing for measurement invariance, it is imperative to allow for group differences in factor variances and covariances, and (when the means are modeled) differences in the factor means. Furthermore, the use of the χ^2 difference test is more commonly used to test specific *substantive* research questions (Little, 2013).

Intercepts and Means

As already noted, means are often not modeled in factor analysis, so it is worth describing strong factorial (intercept) invariance in a bit more detail. For example, if there is a difference in intercepts, this means that for equal levels of the latent trait, a constant score is added to the mean of one group; however, this constant is specific and is not related to the common factor.

Say, for example, that researchers are interested in comparing boys and girls on a composite score from a test battery designed to measure visual–spatial ability (Gv). The composite score includes four subtests that are supposed to measure four aspects of visual–spatial ability (e.g., closure speed, visual memory, spatial scanning, and spatial visualization). Measurement invariance is investigated. Weak factorial invariance is supported. Strong factorial (intercept) invariance, however, is not. Specifically, the intercept associated with spatial visualization differs across groups, such that boys score higher on this test even when boys and girls have the same levels of Gv. When the constraint on the intercept for this subtest is removed, partial intercept invariance is tenable. This means that if a Gv composite score is calculated that includes spatial visualization, boys score higher even when boys and girls have equal levels of Gv. Without this test in the composite, however, boys and girls with equal levels of true Gv obtain the same score on the composite.

If the factor model is considered in terms of a linear equation, then the expected score for a person equals the intercept plus the product of the factor loading times the "true latent ability." Imagine a boy and girl with exactly the same latent Gv mean (e.g., 10). Weak factorial invariance is established, so that the factor loading (.5) is equivalent across groups. Strong factorial invariance is supported for all of the tests except for spatial visualization. The intercept values for spatial visualization are 15 for boys and 10 for girls. The expected score for boys is therefore $15 + (.5 \times 10) = 20$. The expected score for girls is $10 + (.5 \times 10) = 15$. The expected score for boys is higher, even though the latent Gv ability is identical. Note that in contrast, if the intercept are invariant, the expected scores are the same for boys and girls (see Gregorich, 2006, for more examples).

Finding intercept differences (or other differences) may be discouraging for a researcher. But from a substantive point of view, and from that of research and understanding what intelligence tests measure, the finding may be quite informative. In this hypothetical Gv example, *partial* strong factorial invariance is demonstrated, and it is found that there are no sex differences in the broad Gv factor, but only specific subtest mean differences. In our experience, such intercept differences are often the results of a subtest's measuring a specific narrow ability, and one on which there are differences. Such knowledge is interesting substantively, and may lead to a better understanding of sex differences and a better understanding of the measurement of Gv. What should be obvious, however, is that the application of a model-based framework allows for more specificity and sophistication in investigating such measurement and substantive questions.

TESTING FACTOR–COMPOSITE OVERLAP

Some of the methods discussed previously are useful for determining whether two tests measure the same underlying construct. A related question is this: How well does a test *composite* or cluster measure a construct? There are different ways to go about answering this question, and it is important to think about what construct the composite is intended to measure.

Cross-battery analysis, or joint factor analysis of more than one test battery administered to the same group of examinees, may be useful for this purpose. For example, Keith and colleagues (2001) found that a higher-order *g* factor from the WJ III correlated perfectly with a higher-order *g* factor from the CAS. In another analysis, they then correlated the higher-order *g* factor from the WJ III with the Full Scale score from the CAS and found that it was lower (.79). A similar analysis was conducted by Farmer, Floyd, Reynolds, and Kranzler (2014) with other intelligence batteries, including versions of the WISC, DAS, and WJ. Research had already demonstrated that the *g* factors from these various test batteries were either perfectly or nearly perfectly correlated (Floyd et al., 2013). Farmer and colleagues showed that range-corrected correlations of the *g* factor from one battery (e.g., the WJ) with a global composite from another battery (e.g., the DAS) ranged from .88 to .95. Thus there is considerable variability (.79 to .95) in the degree that overall or global composite scores (such as Full Scale IQs) correlate with latent *g* factors (cf. Reynolds, Hajovsky, Pace, & Niileksela, 2016). Taken together, the findings suggest that all these test batteries measure the same *g* factor, but that the composites from some batteries do a better job of measuring *g* than others do. The finding may be particularly important in deciding which test to use for a possible diagnosis of intellectual disability if intellectual disability is thought to be primarily the result of a deficit in *g*.

The correlations of global composites with *g* have been described as the *g-loading* of the composites (Jensen, 1998). Reports of factor analysis typically report subtest (or item) loadings on factors, but it is the composites that are interpreted as indexes of the factors or constructs. Knowledge of the loadings of the composites on the factors would thus be very useful. Conceptually, obtaining this estimate would be akin to also including, for example, the global composite in a factor analysis of subtests and interpreting the composite factor loading on *g*, but this solution is not practical for other reasons. Applying cross-battery CFA to estimate the correlations of a global composite with a *g*-factor from another battery has been one way to estimate the *g*-loading of that global composite. It may also be thought of as a way to estimate the validity of a global composite as an index of *g*.

Having data from multiple batteries administered to the same examinees is a luxury, however. How might such estimates be obtained if data from multiple batteries are not available? Spearman (1927) originally provided a formula for calculating the correlation of a composite with a factor (see Jensen, 1998). If that correlation is squared, the estimate reflects the variance in a composite explained by the factor. Other methods to calculate the variance explained in composites by factors are used today and are commonly calculated from estimates obtained via factor analysis. A popular coefficient, if the test is unidimensional, is referred to as *coefficient omega*, or the proportion of variance explained in the global composite by the general factor (McDonald, 1999). Of course, intelligence tests measure multiple factors, not a single factor, and that is why hierarchical models are employed. Fortunately, however, such calculations are easily applied to multidimensional hierarchical models to estimate the overlap of the multiple factors in composites. In this context, the estimates are sometimes referred to as *omega hierarchical* (Zinbarg, Yovel, Revelle, & McDonald, 2006). For example, using estimates from a factor model, an omega hierarchical coefficient for the *g* factor can be calculated as the square of the sum of subtest *g* loadings divided by the total variance in the subtest scores included in the composite (Gustafsson, 2002). This coefficient is an estimate of the *g* saturation in a global composite, and the square root of that estimate is essentially the *g*-loading of that global composite.

Omega hierarchical coefficients and related *g* loadings associated with global composites in the WISC-IV, DAS-II, and KABC-II were estimated in this way in a study using the standardization samples from those tests (Reynolds, Floyd, & Niileksela, 2013). The square roots of those estimates (i.e., the *g*-loadings) ranged from .88 to .93, and were very similar to those found in Farmer and colleagues' (2014) cross-battery approach. Although conceptually very similar (given that all tests appear to be measuring the same *g*), the cross-battery coefficients are more likely to be described as *validity indices*, whereas the within-test coefficients are likely to be called *composite reliability estimates*. Researchers almost always refer to all of these estimates as composite reliability coefficients. Others (e.g., Gustafsson & Åberg-Bengtsson, 2010; McDonald, 1999) however, have described the omega hierarchical estimates as validity estimates (cf. Bollen, 1989), representing indices of how well certain latent variables are measured in the composites. We believe that this validity-based interpretation makes a great deal of sense. We also tend to prefer the unsquared versions (e.g., construct–composite correlations) because these indices tend to be more easily translatable into the real world—unlike variances.

Omega-type estimates may be used for all kinds of composites, including the composites associated with the broad abilities. These estimates can be calculated both with and without *g* included in the factor (although the broad-ability *observed* scores do not have *g* removed from them).

Rather than demonstrating the calculation of omega hierarchical by using the formula described above, we demonstrate how to obtain these estimates by using model-implied estimates from a CFA that includes a phantom variable. We think that this method may provide some conceptual clarity. Earlier, we have mentioned that a conceptually clear but unrealistic way to obtain the *g*-loading of a global composite is to include that composite in a factor analysis of the subtests. Here we demonstrate how this method is approached via the use of phantom composite variables in CFA (see Fan, 2003; Gignac, 2007). The data are again from the DAS-II, and take the form of a covariance matrix, averaged across the age ranges 7 and older. We use a different mix of DAS-II tests than in previous examples, in order to provide greater comparability to Reynolds, Floyd, and Niileksela (2013).

As the first step, we have developed a model that explains the data (see Figure 31.21). Shown in the figure is a higher-order model, but we also report omega hierarchical estimates obtained from a bifactor model for comparison. The higher-order model provides an adequate description of the covariances among all of the DAS-II subtests.

Our initial interest is in the *g* loading of General Conceptual Ability (GCA in the figure) from

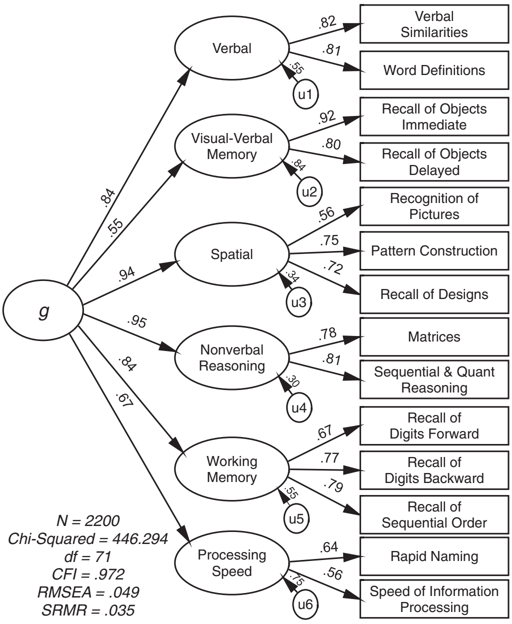


FIGURE 31.21. DAS-II higher-order model. This model includes additional tests and factors, compared to earlier models.

the DAS-II. This is the global composite, and it includes six subtests, with each subtest contributing equally to it. Therefore, in our model we include what is referred to as a "phantom variable" to represent the GCA. As shown in the model setup in Figure 31.22, there are arrows pointing directly to this phantom composite variable from the six subtests included in it, and all of the arrows are weighted by 1. This weighting simply means that this phantom variable is an equally weighted composite of these six subtests. The model is reestimated with this phantom variable, and we request the implied moments (covariances and correlations) in the output to obtain the correlation of the GCA composite with the *g* factor. Note that the correlation is not included in the actual model shown in the diagram; it is only produced in the implied moments output. Note also that model fit, *df*, and all other aspects of the model do not change at all. The implied correlation, or *g*-loading, of the GCA is .91 (see Table 31.6). The loading squared, or *g* saturation (or omega hierarchical), is .83 (Gignac, 2007, has referred to this as *phantom omega*). The estimate with a bifactor model using the phantom composite (not shown) is the same (*g*-loading = .91). These estimates are the same as the estimate from Reynolds, Floyd, and Niileksela (2013), who used the omega hierarchical calculations with output from a bifactor model.

Several other aspects of this model are worthy of comment. Note that we have included *all* of the available subtests in our CFA model (i.e., 14), to obtain what would be considered the best *g* estimates, but have only used the estimates from subtests included in the global composite (e.g., 6 from the DAS-II) to obtain the model-implied correlation (see also Reynolds, Floyd, & Niileksela, 2013; Reynolds & Keith, 2017). Including all of the subtests provides a more rigorous test of the CFA model (and the CFA model is important, since all of the estimates are obtained from the model) and likely results in the best possible estimate of a *g* factor. This view is consistent with the view that these estimates are closer to indices of validity. Note also that the estimates should reflect and be associated with the composite that is actually used. For example, a 14-subtest composite including every subtest included in the DAS-II is not available, so estimates associated with that composite are not that interesting. To summarize, we prefer to model all of the data to obtain the best estimates for the factors, but to use composites that reflect the actual composites on the test of interest. We follow this method for the rest of the analysis.

In addition to the *g*-loading, the relations of the other composites with their respective factors may be of interest. The DAS-II also includes Verbal, Nonverbal Reasoning, Spatial, Working Memory, and Processing Speed composites. The Working Memory composite does not include the Recall of Digits—Forward score, so it is not used in the calculation of the factor saturation in that composite, although it is included in the factor. All of the estimates are shown in Table 31.6. The first column shows the estimates using the higher-order model. Located in the column to the right of the higher-order model estimates (the "Unique" column) are the estimates where *g* has been removed from the first-order factors before each factor is correlated with its associated composite (here *g* has been partialed from the broad ability, but has not been partialed from the composite). In the model-implied correlation output, the correlations of the unit-weighted composite with the unique variance are interpreted (Unit variance identification should be used to scale the disturbances). For example, the model-implied u1 correlation with the Verbal composite (a unit-weighted composite of Verbal Similarities and Word Definitions) is interpreted. Recall that u1 represents Verbal ability

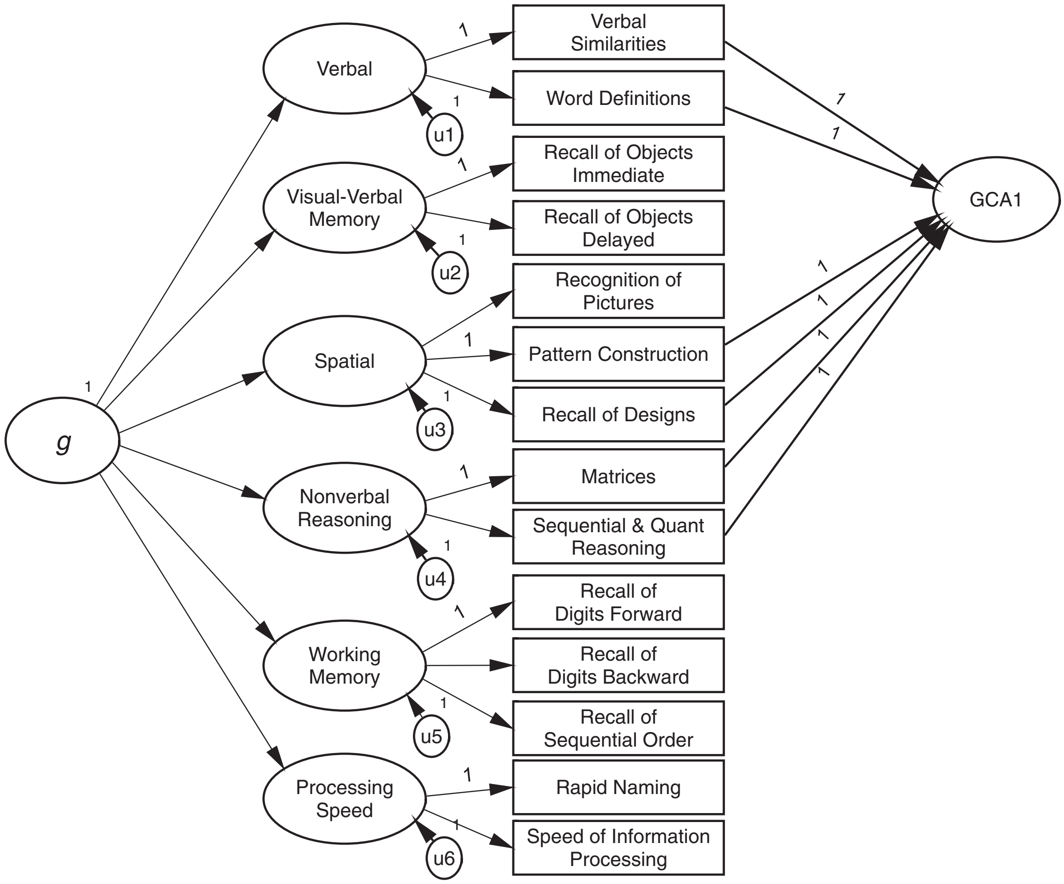


FIGURE 31.22. Setup to determine the model-implied correlation between the DAS-II GCA composite and *g*. Omega hierarchical (the correlation squared) estimates the GCA variance explained by *g*. The unique variances associated with the subtests have been removed for clarity, but are included in the analysis.

from which *g* has been partialed (or Verbal ability with *g* removed, if such a thing exists). Finally, also included are the bifactor model estimates. The unique estimates from the higher-order model are very similar to those from the bifactor model because, of course, *g* has been partialed from the subtests in the bifactor model. These estimates, sometimes referred to as *omega subscale,* of course will be lower than the estimates without *g* partialed.

One potential use of the estimation of the construct composite correlations is to see how the use of different subtests might affect the relation of the construct with that composite (see McDonald, 1999). For example, the DAS-II GCA may be considered an efficient measure of *g* because it only includes six subtests—and from three broad abilities that are traditionally associated strongly with *g*. Additional subtests in the composite, however, would seem to increase the *g* measurement in the composite because all of the subtests load positively on *g*. Would it be worth it to have a composite with additional subtests? As shown in the bottom portion of Table 31.6, including all 14 DAS-II subtests in a composite (e.g., a hypothetical unit-weighted composite of all 14 subtests) does not result in a substantial improvement in *g* measurement. Other potential composites are also tested—for example, one that includes a working memory and processing speed test, since most global composites from intelligence tests include measures of these constructs. Again, there is no improvement. Finally, we test a composite that is similar to the new seven-subtest composite on the WISC-V. It includes two Verbal subtests, two Nonverbal Reasoning subtests, and one subtest from each of the Spatial (Pattern Construction), Working Memory (Re-

TABLE 31.6. Model-Implied Correlations of DAS-II Composites with Their Corresponding Factors

	Correlation			Variance (correlation squared)		
Composite	Higher-order	Unique[a]	Bifactor	Higher-order	Unique[a]	Bifactor
GCA	.91		.91	.83		.83
Verbal	.89	.49	.50	.79	.24	.25
Spatial	.84	.29	.32	.71	.08	.10
Nonverbal Reasoning	.88	.27	.27	.77	.07	.07
Working Memory	.87	.48	.42	.76	.23	.18
Processing Speed	.73	.54	.54	.53	.29	.29
Possible alternative composites not on the DAS-2[b]						
All 14 DAS-2 subtests	.92		.92	.85		.85
GCA + DB/SIP	.92		.92	.85		.85
GCA WISC-V-like	.91		.91	.83		.83

Note. GCA, General Cognitive Ability; DB, Recall of Digits—Backward; SIP, Speed of Information Processing.

[a] The *g* factor has been partialed in the higher-order model.

[b] All of these unit-weighted composites are correlated with the *g* factor.

call of Digits—Backward), and Processing Speed (Speed of Information Processing) broad abilities. Again, a composite of these variables shows about the same relation with *g* as the six-subtest GCA does. It may be concluded that the six-subtest composite from the DAS-II is indeed an efficient measure of *g*. The additional subtests used here do not provide a noticeable improvement. Of course, we do not recommend simply adding and deleting tests to try to maximize the measurement of *g*. Theory is essential. For example, some believe that it is important to sample from a variety of broad abilities for *g* measurement. In that case, some may view the measurement of *g* by the DAS-II as lacking breadth. Others may think that it is more important to measure *g* from subtests and constructs that are more closely associated with *g*.

REFERENCE VARIABLES: COMBINING SAMPLES

A major difficulty in conducting CFA-guided intelligence test research is the time commitment involved for both researchers and participants. An individually administered test of intelligence will generally take an hour and a half to administer, and substantial sample sizes are needed (150 or more are useful, although for more detail on power analysis and sample size, see Brown, 2015; Keith, 2015; Kline, 2016). Research using tests from multiple batteries allows for powerful tests of competing theories (for a further discussion, see Keith & Reynolds, 2010). Data collection for such research also requires a considerable increase in time commitment. It also likely reduces participation rates; how many participants (or their parents) are willing to spend 3–4 hours taking tests to aid in a research project? Even if they were willing, examinee fatigue would probably become an issue, so data collection for one examinee would be likely to require multiple sessions. On the other hand, imagine how useful the inclusion of additional batteries or parts of batteries would be, allowing for more comprehensive tests of specific questions and even underlying theories. Needless to say, despite the appeal, such studies are quite rare.

One possible solution is the use of a *reference variable* approach (McArdle, 1994) as a way of combining data from different samples, if those samples share a subset of subtests given to all participants. McArdle (1994) outlined the approach as a method of reducing data collection requirements by instituting *planned missingness*. That is, a large number of measures could be administered to participants, but no one would take all measures. Instead, each participant might take one, two, or a few tests per factor measured, or subsets of participants would take different subsets of measures. The methodology builds on several topics already discussed, especially multisample analysis and invariance testing, and it requires expanding our thinking about what "latent" variables are. Cheung and Chan (2005) advocated the method as a way of combining correlation matrices for meta-analysis. Keith and colleagues (2010) illustrated the use of the method to conduct cross-age invariance testing when the structure of a test changes across age groups. Reynolds, Keith, and colleagues (2013) used the method to combine data from multiple batteries to test the validity of CHC theory and of those tests as measures of CHC constructs. Here, we likewise illustrate the method as a way of combining overlapping batteries from multiple samples, but on a smaller scale.

Stone (1992) conducted a cross-battery CFA of the original DAS (Elliott, 1990) and the second edition of the WISC, the Wechsler Intelligence Scale for Children—Revised (WISC-R; Wechsler, 1974). His primary purpose was to compare models similar in structure to those guiding the WISC-R to those guiding the DAS (the DAS-type structure was more strongly supported). The data presented by Stone, however, can also be used to test specific hypotheses about the abilities measured by either measure. As a huge help to those who might want to investigate further, he reported the correlation matrix among the tests in these two batteries, thus allowing us (or anyone else) to further analyze those data. We have used those data in previous versions of this chapter to illustrate cross-battery CFA and its utility in answering questions about the constructs measured by tests (viz., whether the Block Design subtest of the WISC-R measures visual processing [Gv] or fluid intelligence [Gf]).

As noted earlier, one enduring question concerning the Wechsler scales is what exactly is being measured by the Arithmetic test. In previous versions of the WISC, Arithmetic was a component of the Verbal scale, but it was often combined with Digit Span and Coding into a (probably misnamed) Freedom from Distractibility factor in EFA. In the WISC-IV, the WISC-V, and the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV), there is evidence that Arithmetic measures Gf (Benson, Hulac, & Kranzler, 2010; Keith, Fine, Reynolds, Taub, & Kranzler, 2006; Reynolds & Keith, 2017; Weiss, Keith, Zhu, & Chen, 2013), but that it may also measure short-term memory (Gsm) and crystallized intelligence. A cross-battery CFA of the WISC-III with the WJ III Tests of Cognitive Abilities (WJ III COG) suggested that Arithmetic is best considered a measure of Gq, or mathematics achievement (Phelps, McGrew, Knopik, & Ford, 2005). Thus, although the Stone (1992) data do not focus on the most recent versions of the WISC and DAS, they may still be useful for helping us understand the construct measured by previous (and likely current) versions of the Arithmetic test.

The DAS includes two good measures of Gf, one of which (Sequential and Quantitative Reasoning) appears to measure quantitative reasoning (RQ). Thus a joint CFA of the WISC-R and the DAS could determine whether the Arithmetic test "fits" on the same factor with Sequential and Quantitative Reasoning. But what about alternative competing hypotheses? The original DAS also included three short measures of achievement (Basic Number Skills, Spelling, and Word Reading), so with a cross-battery CFA of the WISC, the DAS, and the DAS achievement scale, it should be possible to test whether the WISC Arithmetic test measures Gf, achievement, neither, or both. Because the two measures also include other measures of Gsm (each includes a digit span measure), it should also be possible to test whether Arithmetic measures short-term memory skills.

The second sample in this research is a portion of the original DAS standardization sample, ages 8–15 (the age range of the Stone DAS/WISC-R data), using the DAS cognitive and achievement tests. These data are accessible as age-related matrices in the DAS technical handbook (Elliott, 1990). Two samples are thus used in this example. The first sample includes 115 children ages 8–15 who were administered both the DAS and the WISC-R. The second sample includes children ages 8–15 in the standardization sample who were administered the DAS cognitive battery and the three DAS achievement tests. For both samples, the correlation matrix of tests and standard deviations have been used as input for the analysis.[10] The DAS serves as a "link" between DAS achievement and WISC-R tests.

Figure 31.23 shows the model developed and tested for the DAS/WISC-R data, and Figure 31.24 shows the model developed and tested for the DAS/achievement data (the Recall of Digits residual has been constrained, based on the results of the DAS/WISC data). As shown in the figures, both models fit the data well, and we are likely to accept them as reasonable (all factors have been correlated, although these are not shown in the figures in order to simplify them). Figure 31.25 shows the setup for the reference variable analysis, with the DAS/WISC-R sample on the top and the DAS/achievement sample on the bottom. This is a basic multi-

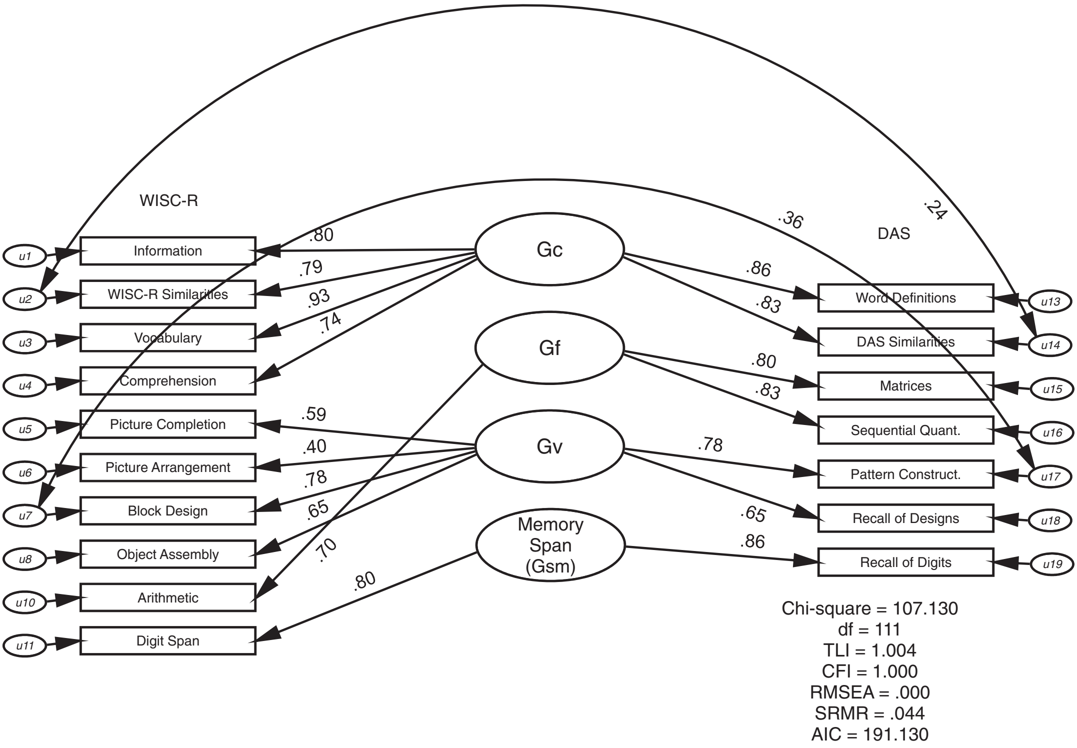


FIGURE 31.23. Cross-battery CFA of the DAS and the WISC-R.

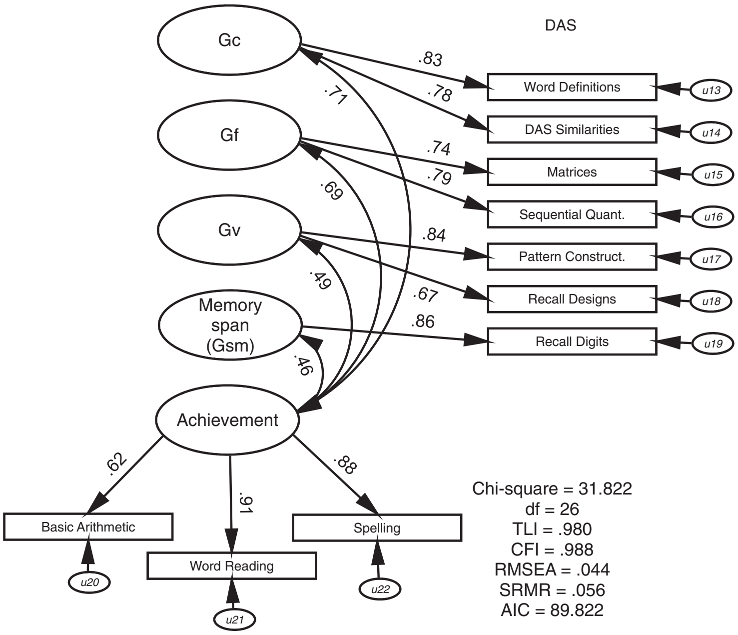


FIGURE 31.24. CFA of the DAS and the DAS achievement tests.

sample analysis, but with different models across samples. What is unusual in these are that some variables that are normally measured variables appear as latent variables in one sample (e.g., the subtests of the WISC-R in the lower portion of Figure 31.25). This may seem strange, but consider some alternative ways of thinking about latent variables: They are unmeasured or even imaginary variables, and an imaginary variable is also a *missing variable*. The latent variables Gc, or u13, do not appear anywhere in either dataset; they are missing. Likewise, in the DAS/WISC-R sample, there is no Basic Arithmetic variable, and thus it is missing, or latent (see the upper portion of Figure 31.25).

Both models will be underidentified (and therefore impossible to estimate) if analyzed separately, but they can be analyzed in tandem, with certain parameter constraints. These are also shown in Figure 31.25. Thus the path from achievement to Basic Arithmetic in the DAS/WISC-R sample is constrained to the value estimated in the DAS/achievement sample, and the path from Gc to Information in the DAS/achievement sample is constrained to the value estimated in the DAS/WISC-R sample, and so on. In fact, every parameter that has the same label across the two figures has been constrained to be equal across the groups. These include some of the factor loadings, factor variances, variances of residuals, and covariances among factors and residuals. The constrained parameters are those related to measured variables that do not exist in one sample versus the other.

These constraints are *required* for estimation, but are they reasonable? What if the DAS/WISC-R sample were very different from the DAS/achievement sample? In that case, assumptions concerning the equivalence of these parameters would not be justified. It is possible, however, to get an idea of the reasonableness of these assumptions by testing for invariance for the parts of the model that exist in *both* groups. The DAS cognitive battery was administered to both samples (these are the *reference variables*), and thus it is possible to go through the invariance testing steps for these portions of the models. Table 31.7 shows the results of such testing. The configural model is the model shown in Figure 31.25, with no cross-sample constraints beyond those required for identification and estimation. For the metric invariance model, the paths from the factors to the DAS cognitive tests have been constrained to be equal across groups. As shown in the table, these constraints do not result in a statistically significant decrement in model fit, and thus we are likely to accept these loadings as equal across groups. The table shows that the results of each level of invariance testing are plausible.[11] If the common structure is equivalent across groups, it makes sense to assume that the unmeasured/missing portions of that structure are equivalent also. As a result, we can now proceed confidently to use these combined data to test hypotheses of interest. If we were using raw data (and if intercept invariance were also tested and established), it would now also be reasonable to combine the two datasets and let the SEM program deal with the missing data. (Given that most SEM programs, by default, use maximum-likelihood methods for dealing with missing data, the results should be the same as in the reference variable approach, with structures estimated via maximum-likelihood estimation.)

The model shown in Figure 31.25, and those discussed so far in Table 31.7, have assumed that the WISC-R Arithmetic test is a measure of quantitative reasoning, and thus Gf, fluid intelligence. This categorization is consistent with the current version of the WISC, the WISC-V. With the Test Arithmetic 1 model, a cross-loading was also allowed for Arithmetic on the Achievement factor. As shown in the table, this model relaxation results in an improvement in χ^2, but that improvement is not statistically significant. We are likely to accept the stricter model, the one specifying that Arithmetic is a measure of RQ/Gf, over the one specifying that Arithmetic measures both Gf and achievement. Interestingly, in the model in which both loadings are allowed, neither is statistically significant (probably due to the smallish sample size), but they are of similar magnitude (.41 for Gf, .39 for Achievement). If we stopped our model testing here, we would be likely to conclude that Arithmetic is better considered a measure of RQ and Gf than of achievement.

The Test Arithmetic 2 model specifies that Arithmetic loads only on the Achievement factor. This model can also be compared to the Test Arithmetic 1 model. The difference is not statistically significant, suggesting that a model with Arithmetic loading only on the Achievement factor is better than a model allowing it to load on both Achievement and Gf. The model comparisons are not definitive in this case (e.g., AIC of 256.271 vs. 256.198), so it is still not clear whether Arithmetic is better considered a measure of Gf or of achievement. Larger sample sizes, additional measures of RQ, and additional measures of arithmetic achievement would help in making this determination.

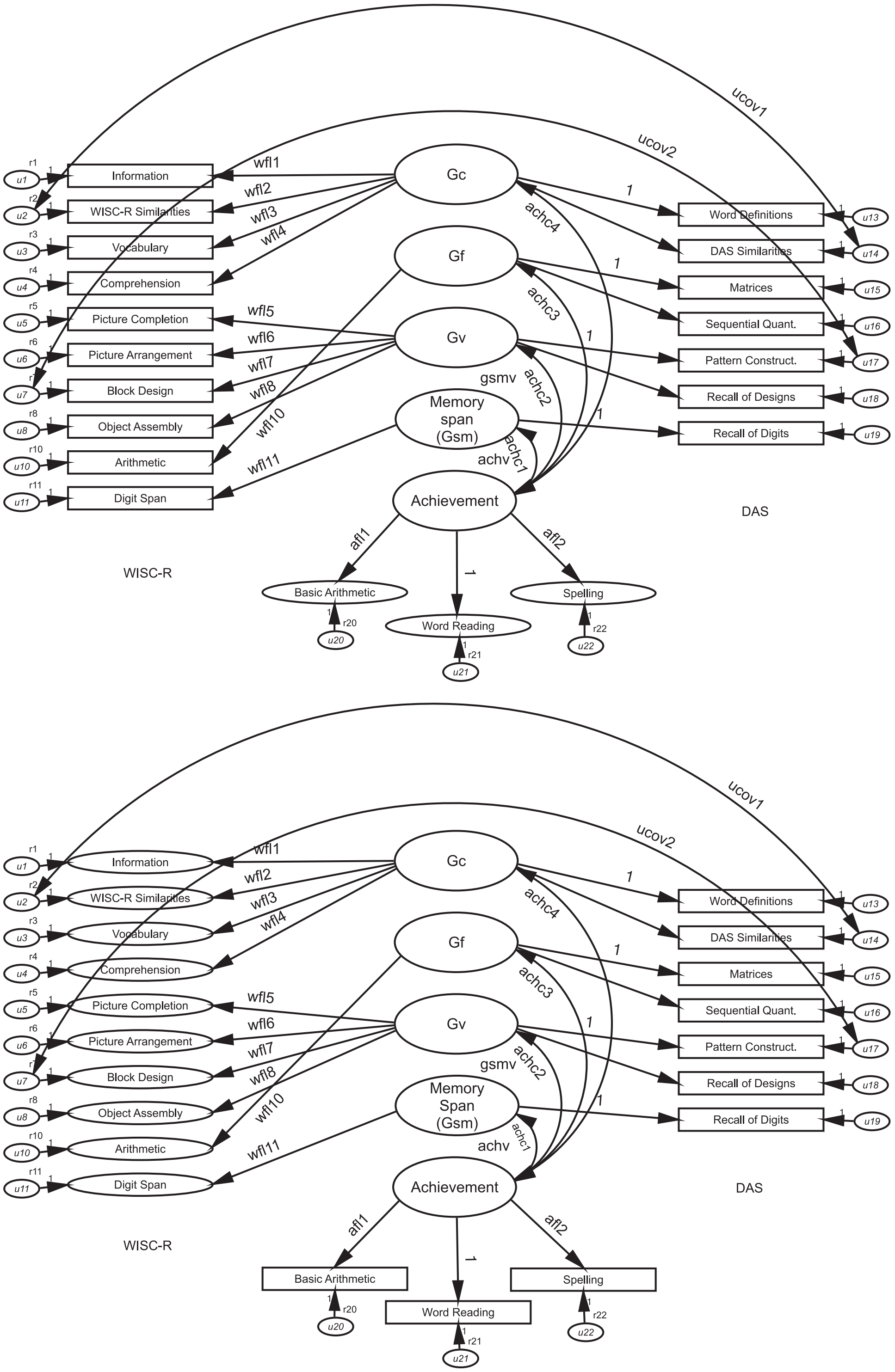


FIGURE 31.25. Reference variable setup for a multigroup, cross-battery analysis of the DAS/WISC-R and the DAS/DAS achievement tests. Variables that are measured in one sample and latent (missing) in the other are shown in ovals.

TABLE 31.7. Comparisons of Reference Variable Models: Invariance and Hypothesis Testing

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	RMSEA	SRMR	AIC
Configural	138.952	137				.008	.044	280.952
Metric	139.029	140	0.077[a]	3	.994	.000	.044	275.029
Subtest residual	143.471	147	4.442[a]	7	.728	.000	.045	265.471
Factor variances	143.675	150	0.204[a]	3	.977	.000	.046	259.675
Factor covariances	152.271	156	8.596[a]	6	.198	.000	.054	256.271
Test Arithmetic 1	150.431	155	1.840[a]	1	.175	.000	.052	256.431
Test Arithmetic 2	152.198	156	1.767[a]	1	.184	.000	.052	256.198
Test Gsm 1	151.351	155	0.920[b]	1	.337	.000	.053	257.351
Test Gsm 2	178.475	156	28.044[a]	1	<.001	.025	.074	282.475

[a] Compared to the previous model.

[b] Compared to the factor covariances model.

The WISC-IV included the Arithmetic test on a Working Memory Index, suggesting that it is a measure of the narrow ability working memory and the broad ability short-term memory (Gsm). Table 31.7 also shows two models testing this possibility. The results of these comparisons are more definitive. Allowing Arithmetic to cross-load on a Gsm factor does not improve model fit, and with a model allowing such cross-loading, the path from Gf to Arithmetic is statistically significant, whereas the path from Gsm to Arithmetic is not. In addition, a model allowing Arithmetic to load only on the Gsm factor fits statistically significantly worse than a model allowing cross-loadings does, and worse (based on the AIC) than a model allowing it to load only on a Gf factor does. The models suggest that Arithmetic (at least the version of Arithmetic on the WISC-R) likely measures Gf more than it measures Gsm. Of course, additional measures of Gsm and additional measures of working memory abilities on the Gsm factor would improve these comparisons as well. The issue is not settled by any means; current research with the WISC-V suggests that Arithmetic measures primarily general intelligence, with a small loading on a memory factor (Reynolds & Keith, 2017).

Figure 31.26 shows another way to specify this model (this is the configural invariance model). With this specification, variables that are measured in one sample but missing in the other simply are not included in that second sample. Note that the latent variable referencing achievement does appear in both, with its variance and covariances constrained to be equal across groups. Note that the fit statistics for this version of the model match those for the original configural invariance specification shown in Table 31.5.

Although these results and speculation are interesting, the main purpose of this example has been to illustrate the reference variable approach. Clearly, this is a useful approach for increasing the number of tests and broadening the factor representation in CFA. It can be useful for combining extant datasets, as we have done here, but a more useful approach would be to use this to plan data collection so that the breadth of measurement is increased without increasing the time commitment per participant. (For example, a useful plan would be that there would be one indicator variable per factor administered to all participants.) And the approach is not limited to two samples. Reynolds, Keith, and colleagues (2013), for example, combined data from four samples in a comprehensive test of CHC theory. This research also tested for measurement invariance in many of the major tests used (e.g., the WISC-IV, the WJ III, and the reference test, the KABC-II) against their standardization data, to ensure that the smaller cross-battery samples were representative of the underlying factor structures.

TESTING THEORIES OF INTELLIGENCE

Most of the examples used in this chapter are examples of CFA used to understand the constructs measured by specific tests. The method is equally applicable, however, for asking and answering questions about *theories* of intelligence. Of course, some analyses serve both functions.

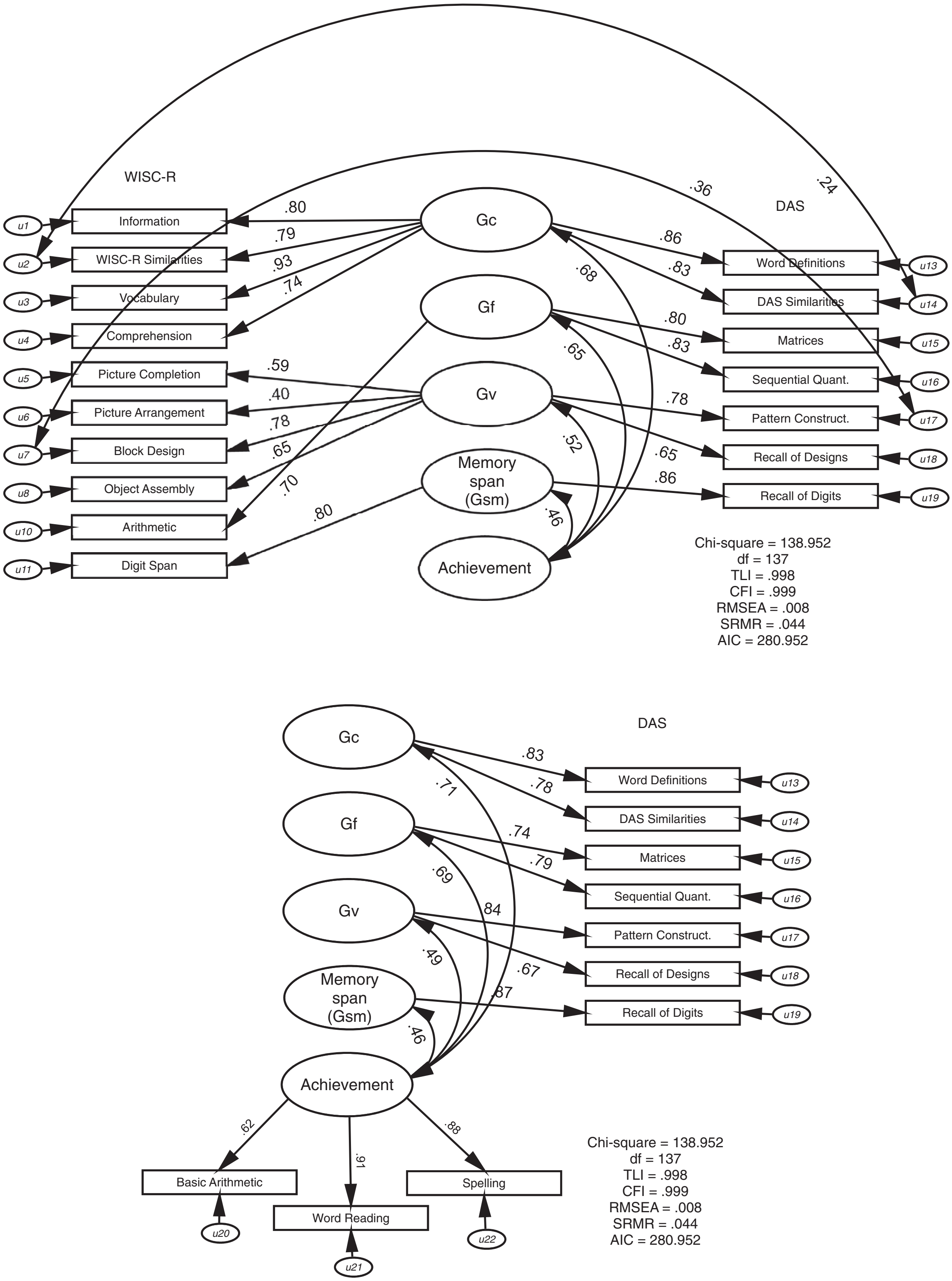


FIGURE 31.26. An alternative model specification for the reference variable approach.

CHC and Three-Stratum Theory

CHC theory is based, in part, on Carroll's three-stratum theory of cognitive abilities. Both theories posit a higher-order model of intelligence with narrow-order abilities at the bottom and the most general ability, *g*, at the apex. The theory is described in detail in Carroll (1993) and in Schneider and McGrew (2012; see also Schneider & McGrew, Chapter 3, this volume).

Carroll speculated that there might well be *intermediate* factors between his second-stratum (e.g., Gf, Gc, Gv) and third-stratum abilities, but left the task of describing this intermediate structure up to other researchers. Bickley, Keith, and Wolfle (1995) addressed the possibility of intermediate factors and tested one such model, and Keith (1997) explored several such possible models, but neither pursued the matter in depth. One difference between three-stratum theory and Gf-Gc theory (the other component of CHC theory) is the nature of quantitative reasoning and quantitative knowledge. Gf-Gc theory has traditionally treated quantitative skills as a separate achievement-related construct, Gq, whereas Carroll focused on quantitative reasoning (RQ) and found it to be a part of fluid or novel reasoning, Gf. To demonstrate CFA's applicability to testing theory, one model with intermediate factors is explored here, and several models exploring the nature of quantitative reasoning are tested.

The WJ III COG is based on CHC theory, and research suggests that it provides valid measures of CHC constructs (McGrew & Woodcock, 2001; Taub & McGrew, 2004), so it is a good tool for testing basic questions about CHC theory. A basic CHC model is shown in Figure 31.27 (cf. Floyd, Keith, Taub, & McGrew, 2007). The data used are from the matrix of correlations and standard deviations for children ages 9–13 from the WJ III standardization data (McGrew & Woodcock, 2001). Not all WJ III COG tests are used; the model includes three good measures of each factor. The sample sizes for tests in this matrix vary, so an overall sample size of 1,000 is used in these analyses (the expectation minimization algorithm was used to deal with incomplete data in the calculation of the matrices).

Figure 31.27 also shows the results of the analysis of this initial model for 9- to 13-year-olds in the WJ III standardization sample. As shown in the figure, the initial model provides a good fit to the WJ III data; the SRMR, RMSEA, TLI, and CFI all suggest an excellent fit to the data. The RMSEA information in the figure is a little different from that presented previously; the figure shows the point value of the RMSEA (.043) surrounded by the 90% confidence interval of the RMSEA (.039–.047).

The model shown in Figure 31.28 presents one possible set of intermediate factors between the second and third strata from CHC/three-stratum theory (in the model, these intermediate factors are second-order factors, and *g* is a third-order factor). Woodcock (1993) proposed a *cognitive performance model (CPM)* of abilities as a method of explaining how abilities work in concert to affect a person's overall functioning, and this model was refined and to some degree built into the scoring of the WJ III and the WJ IV. The CPM includes three intermediate factors between the broad abilities (Gf, Gc, etc) and *g*: *verbal ability, thinking ability,* and *cognitive efficiency.* Two of these intermediate factors are included in the model (as "Think" and "Effic"); verbal ability and Gc are the same, so there is no need to build a verbal ability intermediate factor into the model (an intermediate factor could be built into the model, but the fit would be the same as that for the more simple model shown).

As shown in the figure and in Table 31.8, this categorization of second-stratum abilities into thinking abilities and cognitive efficiency leads to an improvement in the fit of the model. In particular, the intermediate CPM factors produce a statistically significant decrease in χ^2, thus suggesting the division of some of the second-stratum abilities into thinking and cognitive efficiency as a worthwhile addition to the three-stratum theory.

One interesting aspect of this model is the essential equivalence of *g* and thinking ability. (The path from *g* to thinking ability is actually 1.02, but is not statistically significantly different from 1.) We have already discussed how such an equivalence is a common occurrence with *g* and Gf. A common fix would be to constrain the disturbance of the thinking ability factor to 0, which would fix the standardized loading to 1.0. Another, equivalent method for dealing with the issue would be to recognize that this means that one of the factors—*g* or thinking ability—is redundant. Shown in Figure 31.29 is a model in which the *g* and thinking ability factors are combined; this model is equivalent to one in which the thinking ability disturbance is set to zero. We have labeled the highest-order factor "Think" in the figure, but "g"

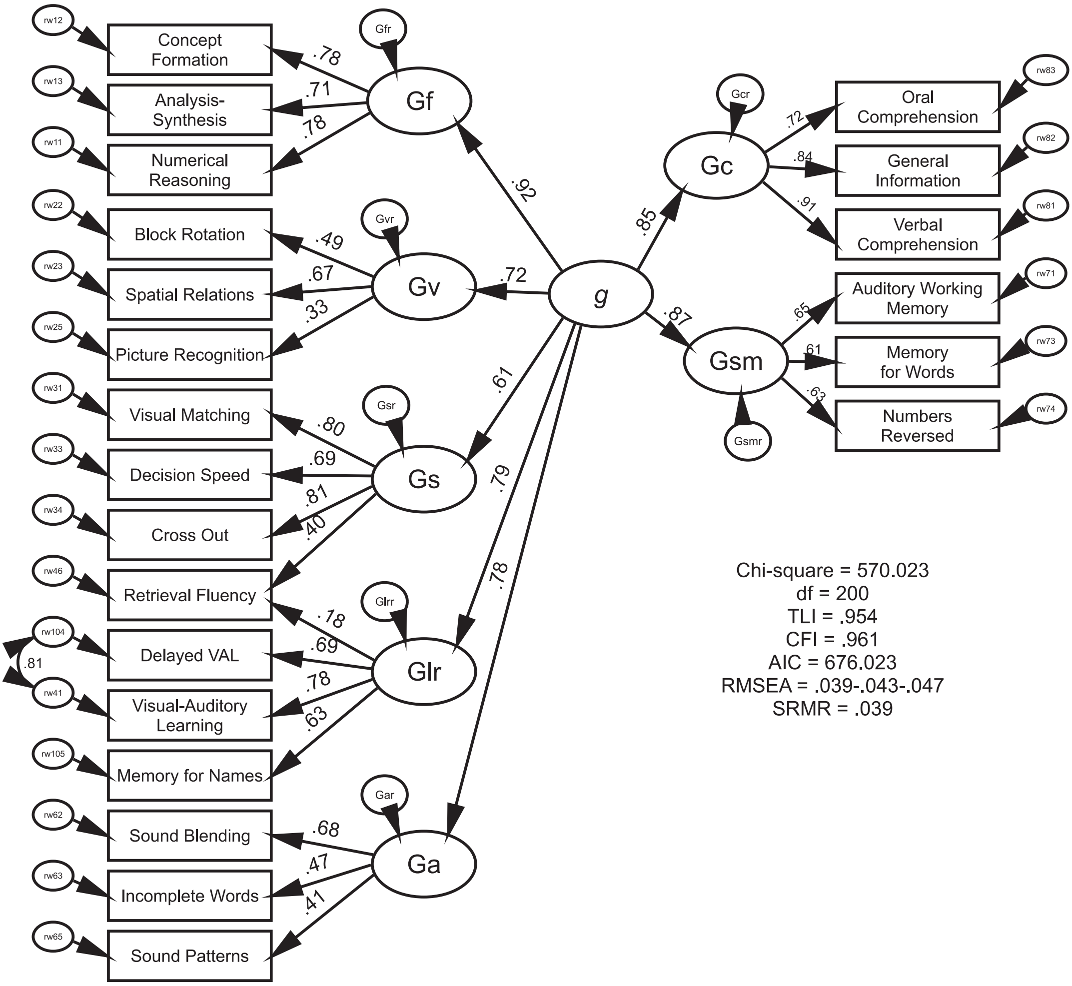
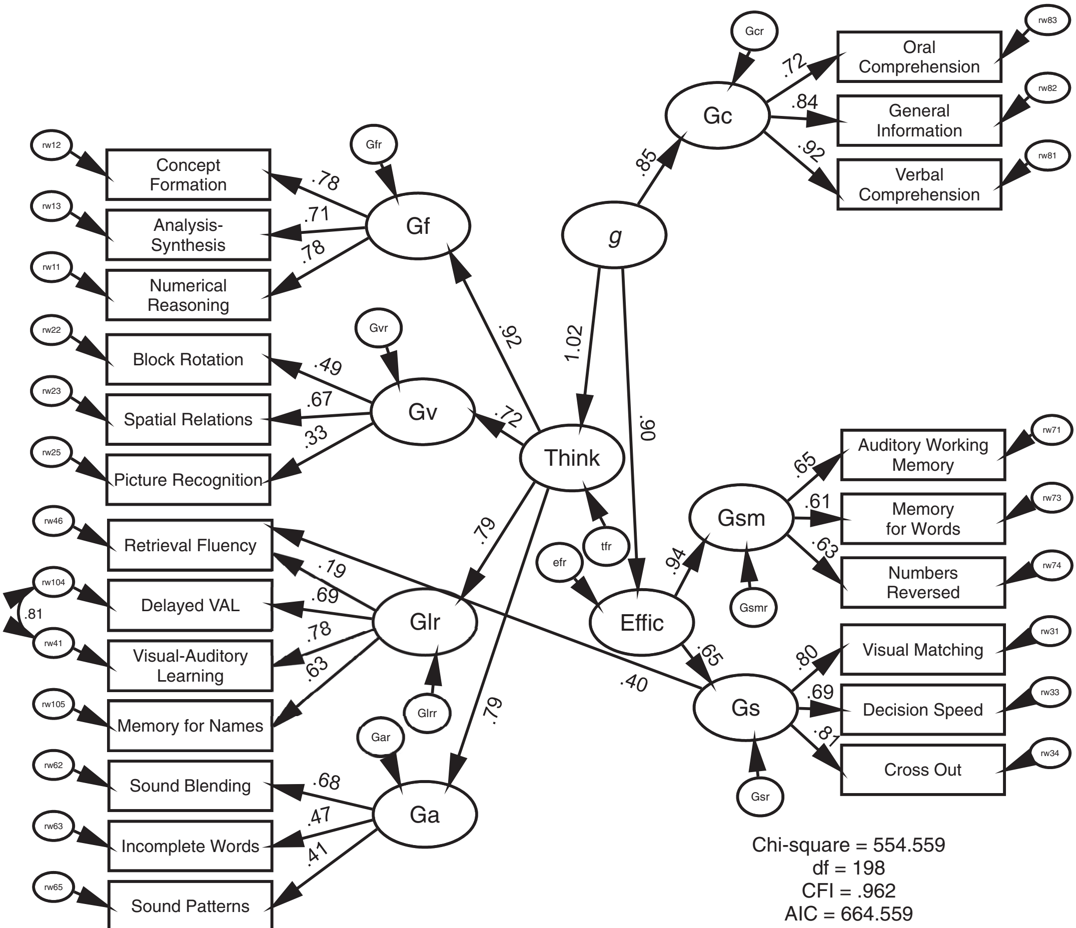


FIGURE 31.27. A three-stratum-theory-derived model of the WJ III COG. The model fits the data well.

would be an equally valid label. As shown in Table 31.8, this model does not result in a statistically significant increase in χ^2, and thus is preferable as a more parsimonious version of the CPM. This modified model supports the combination of the Gs and Gsm factors into a hierarchical cognitive efficiency factor, but suggests that *g* and thinking ability are statistically indistinguishable.

One final variation of this model is mentioned briefly. One could argue that the path from cognitive efficiency to *g* should be reversed, so that efficiency affects *g* rather than the reverse. This modification could be based on the assumption that processing speed and short-term memory (and cognitive efficiency) are fundamental mental skills that influence one's level of general intelligence—essentially, gatekeepers of general cognitive ability (cf. Schneider & McGrew, 2012). Unfortunately, without further modification, this model is statistically equivalent to the modified CPM, and the two cannot be distinguished on the basis of fit statistics. The rules for generating equivalent and nonequivalent models (e.g., Keith, 2015, Ch. 13) could be used to develop some nonequivalent versions of these two alternative models, thus allowing a test of whether *g* should be considered an influence on cognitive efficiency, or an effect.

The final two models (Figures 31.30 and Figure 31.31) in this chapter test Carroll's contention that quantitative reasoning (RQ) is a part of Gf



Chi-square = 554.559
df = 198
CFI = .962
AIC = 664.559
RMSEA = .038-.042-.047
SRMR = .038

FIGURE 31.28. A test of possible intermediate factors between Carroll's stratum II and stratum III abilities. The model is based on Woodcock's cognitive performance model (CPM).

TABLE 31.8. Comparison of the Fit of Models Testing Different Intermediate-Level Factors in the Three-Stratum Theory, and the Relation of Gf and RQ

Model description	χ^2	*df*	$\Delta\chi^2$ (*df*)	Δ*df*	*p*	AIC
1. Initial model: No intermediate factors	570.023	200				676.023
2. Cognitive performance model (CPM)	554.559	198	15.464 (2)[a]	2	<.001	664.559
3. CPM 2	556.827	199	2.268 (1)[a]	1	.132	664.827
4. Initial Gf-RQ model: Separate factors	1914.426	198				2068.426
5. Gf subsumes Gf (narrow) and RQ	1873.621	197				2029.621

[a] Compared to the previous model.

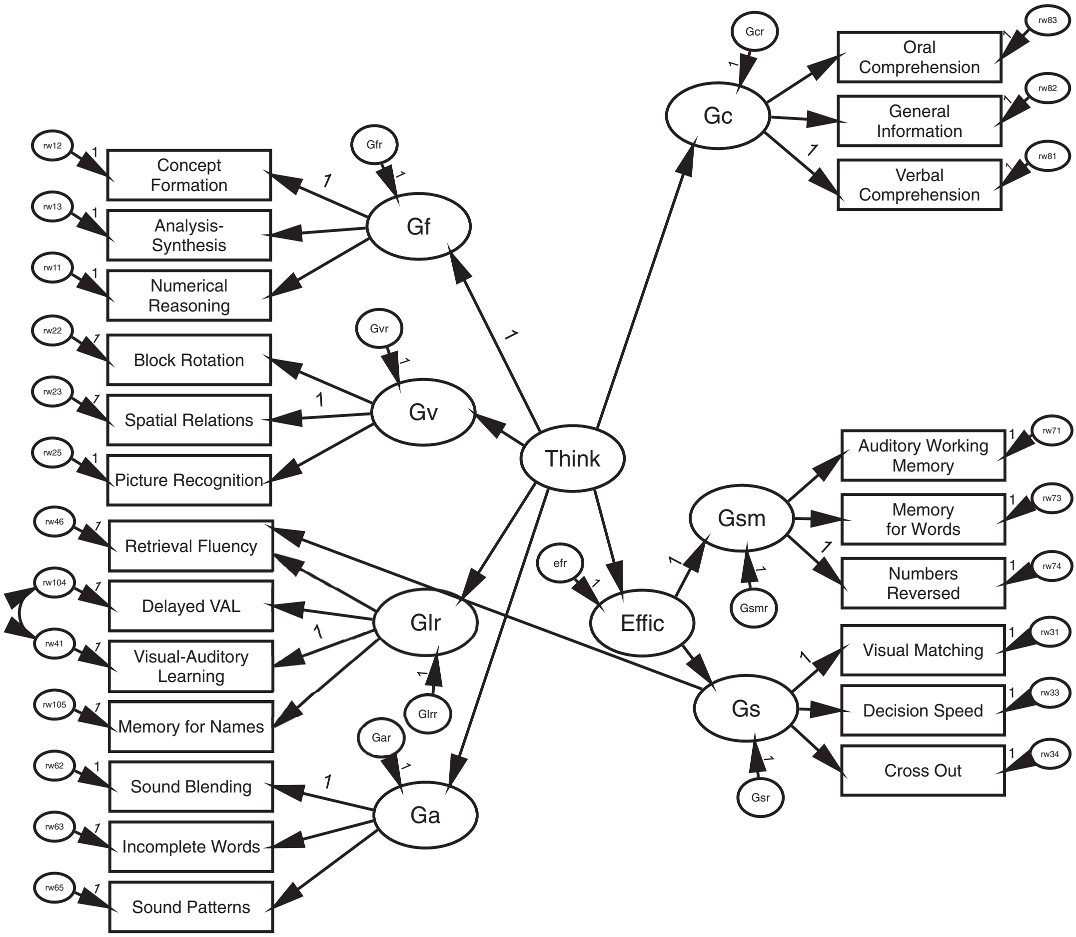


FIGURE 31.29. A simplified version of the CPM in Figure 31.28.

rather than a separate second-stratum factor. The models do not address the existence or nature of a Gq (quantitative knowledge) factor. These analyses have been conducted on over 5,000 participants from the WJ III standardization sample. The advantage of these data is that they include two clear measures of quantitative reasoning: number series and number matrices (these two tests were combined into a Numerical Reasoning test when the WJ III was first released). The sample is described in more detail by Keith and colleagues (2008).

The initial quantitative reasoning model, set up with separate Gf and RQ factors (in the upper part of the figure), is shown in Figure 31.30; the fit statistics for comparing models are shown in the figure and in the lower part of Table 31.8. As shown in the figure, the initial quantitative reasoning model provides a good fit to the data using common criteria. There are missing cases in the data, and Amos does not produce SRMR when there are missing data (some other programs, such as Mplus, do); thus SRMR is not reported.

There are several possible ways to test whether the RQ factor should be subsumed under a broader Gf factor. One common method—the most agnostic approach—would be to specify correlated errors for the disturbances of the Gf and RQ factors. Such a model suggests that these two factors measure something in common besides general intelligence. Figure 31.31 shows a model in which the narrow Gf (symbolized as “Gf narrow”) and

RQ factors are subsumed under a broader Gf factor. This model is in fact statistically equivalent to the correlated disturbance model, but better symbolizes the hypothesis of interest. A third method would be to delete the path from *g* to RQ and include one from Gf to RQ (this model is neither shown nor tested here). As shown in Table 31.8, the loading of the original Gf and RQ factors onto a broader Gf factor results in a considerable improvement in model fit over the initial quantitative reasoning model. Said differently, the broad Gf factor may be considered an intermediate factor between the narrow Gf and RQ factors and *g*. This preliminary investigation supports the contention that quantitative reasoning is a part of fluid/novel reasoning (Gf).

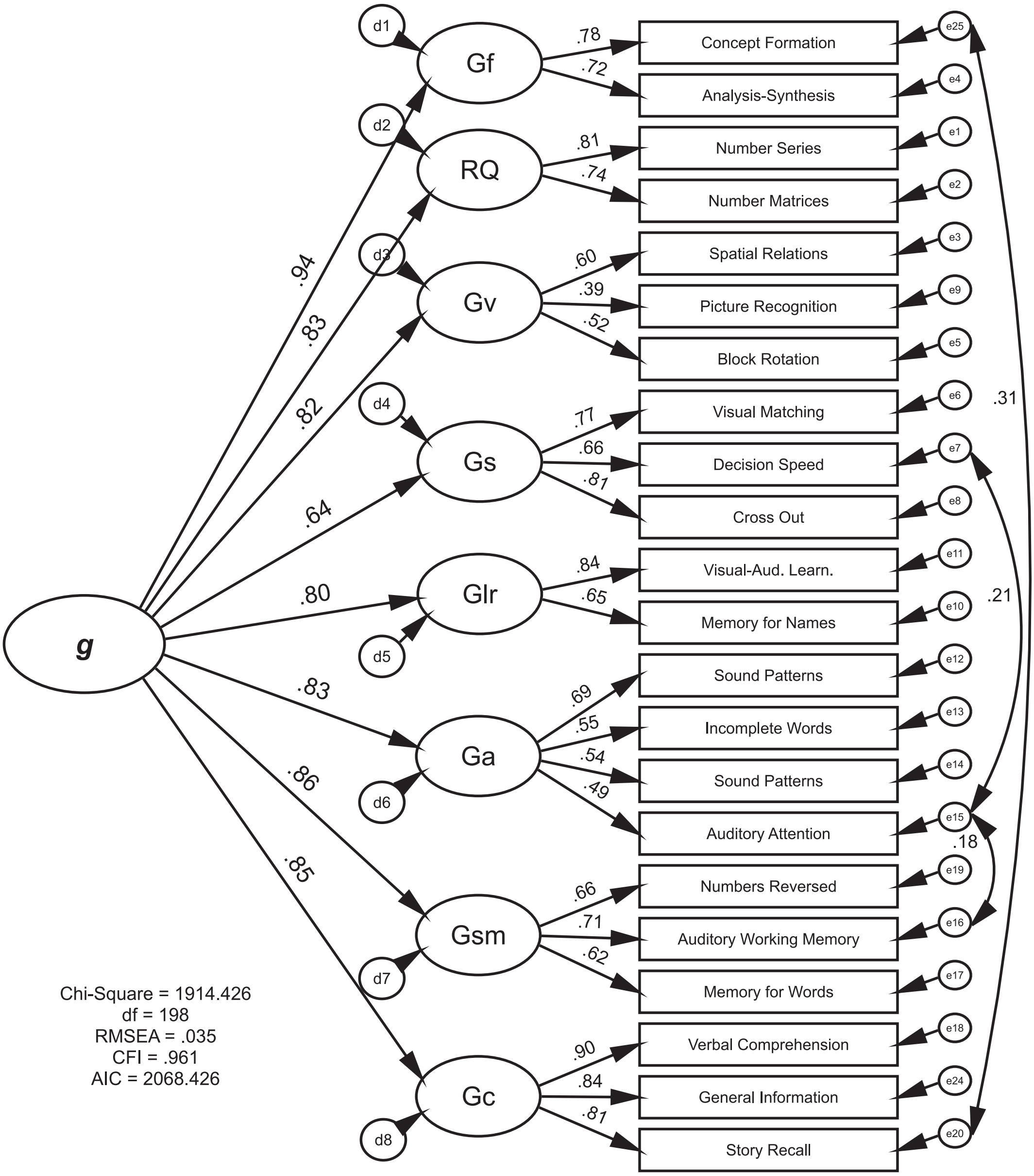


FIGURE 31.30. A model designed to probe the nature of the relation between Gf and RQ.

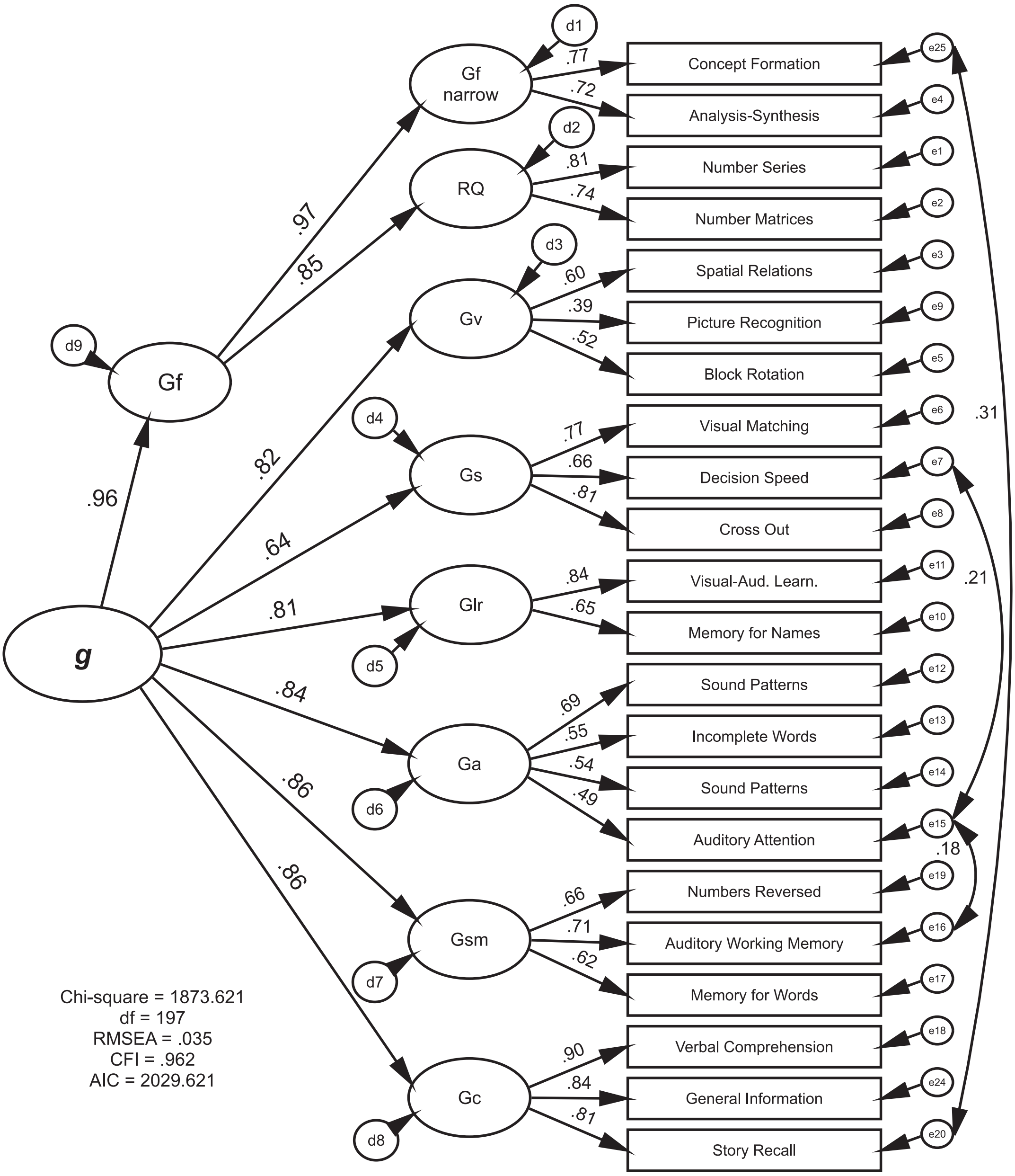


FIGURE 31.31. In this model, Gf subsumes RQ and a narrow Gf factor. The model shows an improvement in fit over the previous model.

SUMMARY

This chapter has provided an overview of and introduction to the method of CFA, with particular attention to the use of the method as an aid in understanding the constructs measured by modern tests of intelligence. The chapter has covered “simple” CFA—in other words, first-order CFA, a method that is fairly common in the factor-analytic/intelligence literature.

We believe, however, that additional uses of CFA are needed for a real understanding of intelligence constructs. Thus we encourage the comparison of *meaningful* alternative explanations to a researcher's pet theory through the testing and comparison of alternative factor models. This practice, too, is becoming more common, although alternative models are not always meaningful. We also encourage the use of carefully ordered nested models to test specific hypotheses concerning intelligence, intelligence constructs, and intelligence tests.

Many modern theories of intelligence are hierarchical in nature, with the most prominent example being the three-stratum theory, a "metatheory" developed and tested by John Carroll (and incorporated into CHC theory). Most tests of intelligence tacitly recognize a hierarchical nature of intelligence as well (Keith & Reynolds, 2010). We strongly believe that our research should therefore test these hierarchical notions of intelligence if we are to fully understand the constructs we are measuring. Put another way, a test of a first-order version of a hierarchical theory/test is not a complete test of that theory. Furthermore, higher-order hierarchical analysis provides a more thorough understanding of the first-order abilities (Carroll, 1993, Ch. 3). This chapter has demonstrated several variations of hierarchical CFA, using second- and even third-order factors.

We also encourage researchers to think about intelligence by using different types of models, even if they cannot be distinguished on the basis of fit. For example, in all of the models presented here, we have conceptualized *g* as affecting broad abilities or subtests. It may be that *g* does not give rise to the correlations among subtests or factors, but alternatively, it simply may arise from those correlations (Kovacs & Conway, 2016). If that is the case, the arrows should flow into *g* and not from *g*, and the nature of that variable would be likely to take on a whole different meaning.

This chapter has demonstrated several other important uses of CFA: to compare the constructs measured across different tests, and to compare the constructs measured by one test across different groups. Many of the most vexing problems in the intelligence field revolve around these issues, and CFA is an excellent method for answering these important questions. Cross-battery CFA can be a powerful method for answering questions about the nature of constructs measured by specific tests, and for understanding the nature of intelligence. Use of a reference variable approach can enable larger and more comprehensive tests of such theories. Multisample CFA provides an organized, effective method for testing for the equivalence of structures across groups, and for testing for construct bias across groups. Finally, CFA provides a powerful method for testing theory, and especially for testing competing theories of intelligence.

We have not covered or tried to cover all possible uses of CFA; those uses are limited primarily by the imagination of the researcher. In addition, CFA is a subset of a more general approach—SEM—and that broader approach is also useful for understanding the nature of the constructs measured by tests of intelligence. To mention only two examples, SEM provides an excellent method for testing the stability over time of intelligence *constructs,* independent of the method of measurement; the dynamic relations among intelligence constructs across time; and the presence of *predictive* bias in intelligence measures (cf. Borsboom, Romeijn, & Wicherts, 2008; Ferrer & McArdle, 2004; McArdle, 1994). Nevertheless, we hope that this chapter has provided enough of an overview to stimulate thought and further study, and to fire the imaginations of future CFA researchers. For those interested in additional study, there are numerous resources available (e.g., Brown, 2015; Keith, 2015; Kline, 2016; Loehlin & Beaujean, 2017; Reynolds & Keith, 2013).

NOTES

1. The factor model tested is one that matches the scoring structure of the KAIT, a good starting point for an informal theory of what a test measures. Other models (e.g., Flanagan & McGrew, 1998) are certainly plausible, but are not evaluated here.

2. Note that one reason the χ^2 is so small is that the sample size used is 143, the average sample size for each age level. If a value 10 times larger were used (1,430), the χ^2 would be almost 10 times larger and statistically significant, thus suggesting a poor fit. Furthermore, although a good fit may be indicated for the *overall* model, it does not mean that there are not some *local* areas of misfit related to specific parameters.

3. Without this constraint, this portion of the model would have been underidentified, meaning that we would not have enough information to solve for all of the parameters in the model. Although these kinds of equality constraints are common ways of dealing with underidentification, researchers should always report when these constraints are made.

4. Readers may want to go over this section several times. It is confusing in part because we first compare different theoretical models (via the figures), but then

change the order of model presentation in the table. The table is ordered as a series of hypotheses that can be tested via nested models by first adding a constraint (model 2), then relaxing two constraints (model 3), then adding three constraints (model 4). It doesn't help that we would come to different answers using ??2 versus the AIC. According to the AIC, model 4 (Figure 31.5) is the best-fitting of these models. This difference is also illustrative, however.

5. In SEM, these are known as *disturbances* and represent all other causes not included in the model. Thus, for the DAS-II example, f1 represents all causes of Verbal ability/Gc other than *g*.

6. The variance of the Gsm factor is nonsignificant for this model, but becomes statistically significant when these three loadings are also constrained to be equal. When UVI is used to scale the factors to check the statistical significance of the constrained loadings, all are statistically significant. If our purpose were an in-depth investigation of the bifactor structure of the DAS-II, we would investigate these anomalies in considerably more depth. For teaching purposes, and for comparison with a higher-order model, the model shown in Figure 31.14 serves well, however.

7. Another difference is that with the higher-order model, the effects of *g* are removed from the first-order factors; with the bifactor model, G is partialed at the subtest level. There are several different ways to obtain the partialed effects with a higher-order model. First, as noted, these can be calculated as the indirect effect of the factor disturbance on the subtest. Another method is to use the squared multiple correlations associated with the first-order factors, easily obtainable from any SEM/CFA program. As noted, one seeming advantage of the bifactor approach is that it produces estimates that look like a Schmid–Leiman transformation. In fact the approach outlined here with the higher-order model is much more similar to the Schmid–Leiman EFA technique.

8. Anyone who has had to explain IQ test scores to parents will probably note that it is much easier to talk about broad-ability scores than it is about the global composite because those tests are grouped together according to explainable, surface characteristics of tests. This is a reflection of the abstractness of *g* versus the broad abilities.

9. See Widaman and Reise (1997) for alternative specifications.

10. The DAS standardization sample included approximately 1,600 children, but the sample size has been set to 115 in the multisample analyses so as not to overwhelm the findings for the other group. The matrix reported in Stone did not include standard deviations, which are needed to test for later levels of invariance. For purposes of illustration, we have set these to the average *SD* for each test (10 for the DAS and 3 for the WISC-R). Information presented in the DAS manual (the source for these data) suggests that these values are likely to be reasonable.

11. The variance steps may be plausible, in part, because we have guessed at values for the subtest variances in the DAS/WISC-R sample. If our purpose were to rigorously test hypotheses instead of illustrate the method, this would be an important limitation. In this example, we are interested in the covariance structure and not the mean structure; thus we have not tested for intercept invariance or the equality of latent means. If questions of latent means were important, it would also be important to include these steps in the invariance testing.

REFERENCES

Akaike, H. A. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Arbuckle, J. L. (2014). *IBM SPSS Amos 23 user's guide.* Crawfordville, FL: Amos Development.

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3,* 121–136.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22,* 121–130.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20,* 309–328.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borsboom, D., Romeijn, J. W., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible. *Psychological Methods, 13*(2), 75–98.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Burt, C. L. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology, 19,* 100–111, 176–199.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for equivalence of factor covariance and means structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children—Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29*(4), 458–472.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1995). Review of the book *Assessment of cognitive processes: The PASS theory of intelligence. Journal of Psychoeducational Assessment, 13,* 397–409.

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage process. *Psychological Methods, 10,* 40–64.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Cole, J. C., & Randall, M. K. (2003). Comparing cognitive ability models of Spearman, Horn and Cattell, and Carroll. *Journal of Psychoeducational Assessment, 21,* 160–179.

Elliott, C. D. (1990). *Differential Ability Scales: Introductory and technical manual.* San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition: Introductory and technical manual.* San Antonio, TX: Harcourt Assessment.

Fan, X. (2003). Using commonly available software for bootstrapping in both substantive and measurement analyses. *Educational and Psychological Measurement, 63*(1), 24–50.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6,* 56–83.

Farmer, R. L., Floyd, R. G., Reynolds, M. R., & Kranzler, J. H. (2014). IQs are very strong but imperfect indicators of psychometric *g*: Results from joint confirmatory factor analysis. *Psychology in the Schools, 51,* 801–813.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40,* 935–951.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from modern Gf-Gc theory: Joint confirmatory factor analysis of the WJ-R and Kaufman Adolescent and Adult Intelligence Test (KAIT). *Journal of School Psychology, 36,* 151–182.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22,* 200–233.

Floyd, R. G., Reynolds, M. R., Farmer, R. L., & Kranzler, J. H. (2013). Are the general factors from different child and adolescent intelligence tests the same?: Results from a five-sample, six-test analysis. *School Psychology Review, 42,* 383–402.

Gignac, G. E. (2007). Working memory and fluid intelligence are both identical to *g*?!: Reanalyses and critical evaluation. *Psychology Science, 49,* 187–207.

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor. *Psychological Science Quarterly, 50* (1), 21–43.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55,* 57–68.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?: Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44,* S78–S94.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179–203.

Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. In H. L. Braun, D. G. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73–95). Mahwah, NJ: Erlbaum.

Gustafsson, J.-E., & Åberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. Embretston (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 97–121). Washington, DC: American Psychological Association.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence, 32,* 95–107.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Erlbaum.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide.* Lincolnwood, IL: Scientific Software.

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual: Kaufman Adolescent and Adult Intelligence Test.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition: Technical manual.* Circle Pines, MN: American Guidance Service.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft,

& P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373–402). New York: Guilford Press.

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling.* New York: Routledge.

Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence, 54,* 37–54.

Keith, T. Z., Fine, J. G., Reynolds, M. R., Taub, G. E., & Kranzler, J. H. (2006). Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 35,* 108–127.

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment (CAS) measure?: Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability (3rd Edition). *School Psychology Review, 30,* 89–119.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales—II: Consistency across ages 4 to 17. *Psychology in the Schools, 47,* 676–697.

Keith, T. Z., & Reynolds, M. R. (2010). CHC and cognitive abilities: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock Johnson III Tests of Cognitive Abilities. *Intelligence, 36,* 502–525.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27,* 151–177.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research, 25,* 313–334.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theorectical issues. *Multivariate Behavioral Research, 32,* 53–76.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford Press.

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural analysis* (5th ed.). New York: Routledge.

Lubke, G. H., & Dolan, C. V. (2003). Can unequal residual variances across groups mask differences in residual means in the common factor model? *Structural Equation Modeling, 10,* 175–192.

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence, 61,* 120–129.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29,* 409–454.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock–Johnson III.* Itasca, IL: Riverside.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11), S69–S77.

Millsap, R. E. (2001). When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling, 8,* 1–17.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41,* 407–422.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Authors.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (*g*), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. *School Psychology Quarterly, 20,* 66–88.

Potvin, D. C. H., Keith, T. Z., Caemmerer, J. M., & Trundt, K. M. (2015). Confirmatory factor structure of the Kaufman Assessment Battery for Children—Second Edition with preschool children: Too young for differentiation? *Journal of Psychoeducational Assessment, 33,* 522–533.

Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses?: Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research, 51,* 818–838.

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric *g* indexed by global composites?: Evidence from three popular intelligence tests. *Psychological Assessment, 25,* 1314–1321.

Reynolds, M. R., Hajovsky, D. B., Pace, J. R., & Niileksela, C. R. (2016). What does the Shipley-2 measure for children and adolescents?: Integrated and conjoint confirmatory factor analysis with the WISC-IV. *Assessment, 23,* 23–41.

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Soklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 48–83). New York: Oxford University Press.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence, 62,* 31–47.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. F., & Low, J. A. (2007). Confirmatory factor analysis of the Kaufman Assessment Battery for Children—Second Edition: Consistency with Cattell–Horn–Carroll theory. *School Psychology Quarterly, 22,* 511–539.

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51,* 535–555.

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence, 36,* 236–260.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464.

Sclove, S. L. (1987). Applications of some model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Stone, B. J. (1992). Joint confirmatory factor analyses of the DAS and WISC-R. *Journal of School Psychology, 30,* 185–195.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. S. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of CHC theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19,* 72–87.

Thorndike, R. M. (1990). Would the real factors of the Stanford–Binet Fourth Edition please come forward? *Journal of Psychoeducational Assessment, 8,* 412–435.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of the mind.* Chicago: University of Chicago Press.

Trundt, K. M., Keith, T. Z., Caemmerer, J. M., & Smith, L. V. (2017). Testing for construct bias in the Differential Ability Scales, Second Edition (DAS-II): A comparison among African American, Asian, Hispanic, and Caucasian children. *Journal of Psychoeducational Assessment.* [Epub ahead of print]

Vernon, P. E. (1950). *The structure of human abilities.* New York: Wiley.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31,* 114–131.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Woodcock, R. W. (1993). An information processing view of Gf-Gc theory. *Journal of Psychoeducational Assessment, Monograph Series: Woodcock–Johnson Psycho-Educational Battery—Revised,* 80–102.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64,* 113–128.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement, 30,* 121–144.

CHAPTER 32

Functional Cattell–Horn–Carroll Nomenclature for Practical Applications

Richard W. Woodcock
Denise E. Maricle
Daniel C. Miller
Ryan J. McGill

The purpose of this chapter is to review the evolution of Cattell–Horn–Carroll (CHC) theory, in an effort to examine the state of the state of the art in this dominant theory defining the structure of cognitive abilities. The nomenclature associated with CHC theory has become more complex as scholars have elaborated upon the theory, introducing their additions and revisions to the taxonomy. With the publication of the Woodcock–Johnson IV (WJ IV), the preeminent instrument operationalizing CHC theory, change was seen in the conceptualization of CHC theory. In particular, the shift seemed to be toward a more "scientific" and less "functional" or clinician/consumer-friendly CHC nomenclature. This scientific taxonomy is exemplified by what is known as the *CHC periodic table of human abilities* (McGrew, LaForte, & Schrank, 2014; Schneider & McGrew, 2012). The application of CHC theory to measures of cognitive abilities, and the wide acceptance of the theory in research and clinical practice, have given rise to a specialized vocabulary (i.e., nomenclature or taxonomy) that can be difficult for consumers of the information, such as parents, clients, or educators, to understand. Scientific taxonomies are needed to provide precision and to guide scientific development, but more functional or user-friendly nomenclature would benefit consumers. Given recent changes in CHC theory, it seems appropriate to summarize the history of the theory, as well as to discuss current proposals for the continued evolution of the theory being presented by various factions within CHC. The chapter concludes with a description and discussion of a more functional and clinician/consumer-friendly CHC nomenclature.

HISTORICAL PRECEDENTS TO CHC THEORY

Spearman's Two-Factor Theory

Inspired by the statistical developments of his cousin Karl Pearson, British statistician Charles Spearman proposed the first unified theory of cognitive abilities with the 1904 publication of his paper "'General Intelligence,' Objectively Determined and Measured." Although Spearman did not label his theory, it is generally referred to in the contemporary literature as *g* theory. Spearman (1904) administered a battery of assessments to a sample of 60 school-age children in a small village in England. The tests measured subjects such as the classics, French, English, math, vocal pitch, and sound discrimination. He determined that all of the measures tended to correlate—a phenomenon later referred to as *positive manifold* (Thurstone, 1947). Spearman arranged all of the coefficients between the tests into a matrix that he then analyzed, using a primitive form of factor analysis called the method of *tetrad differences*. In his analysis, he found that 62.9% of the total variance between all of the tests was accounted for by a single factor, which he identified as a general factor or *g*. The remaining 37.1% of variance was attributed to specific factors unique to the individual tests themselves, which Spearman labeled *s*.

Spearman postulated that all tests of cognitive ability are composed of some form of *g* variance—an observation he referred to as the "indifference of the indicator" (Spearman, 1927, p. 197). However, he appeared to remain ambivalent about the exact nature of *s*, whose influence he stated was largely negated by the combination of individual test scores into larger composites.

Despite criticism, he resisted the notion of including additional common factors in his model because he felt it would open the door for the inclusion of an infinite number of hypothesized subordinate factors. By the end of his career, however, he acknowledged that cognitive ability may be better represented by a second-order *g* factor, with an underdetermined number of first-order common factors representing more discrete cognitive skills (Spearman & Jones, 1959).

The Rise of Multiple-Factor Theories

One of Spearman's critics was Edward L. Thorndike, a psychologist at Columbia University. Thorndike developed and facilitated the administration of a test to 63 primary and secondary school students that purported to measure several psychoeducational abilities, such as sensory discrimination, quantitative reasoning, and vocabulary development. After reviewing correlational data, he and his colleagues concluded that "there is nothing whatsoever common to all mental functions, or even half of them" (Thorndike, Lay, & Dean, 1909, p. 368). Thorndike rejected the notion of a general intelligence factor in favor of a model that emphasized multiple faculties of the mind.

Although Spearman spent the latter part of his career defending his theory from researchers like Thorndike, advances in research methodology and statistical techniques allowed for the discovery of group abilities in cognitive assessment data. Spearman eventually acknowledged these findings when specific cognitive tasks were found to load on group factors subordinate to *g*. These discoveries helped pave the way for the development of more empirically derived theories of mental ability.

L. L. Thurstone (1938) developed a model of mental ability derived from a statistical technique that allowed factors to be extracted from an extant dataset—a method known as *factor analysis*. Using data from a battery of 56 mental tests administered to 240 college students, he extracted seven factors that he described as visual–spatial, perception of visual detail, numerical, verbal logic, verbal words, memory, and induction. Thurstone called these factors *primary mental abilities*, and this term soon became associated with his model of intelligence. Thurstone eventually reconfigured his model to account for eight primary abilities. The influence of Thurstone's work on the modern-day understanding of the structure of human cognitive abilities cannot be underestimated, as many of the group factors that he identified served as the foundation for subsequent models of intelligence.

Thurstone's initial reluctance to acknowledge a general factor may have been an artifact of the methods he utilized to identify his group factors. Thurstone used rotation techniques in his factor analyses that left various broad abilities orthogonal (not correlated) to each other. With this method, it was almost impossible for a general factor to be derived because little common variance in the factors could be extracted. Although Thurstone later accepted the existence of a general factor, and subsequent research using oblique rotations indicated that his broad abilities were correlated (Jensen, 1998), he stated that the use of a single score to estimate overall mental ability was inadequate for clinical decision making; he encouraged the synthesis of an individual's profile of scores across several measures of cognitive functioning to determine individual cognitive strengths and weaknesses.

The Emergence of Gf-Gc Theory and the Fluid–Crystallized Model of Intelligence

A dichotomous model of intellectual ability, the *fluid–crystallized model* or Gf-Gc theory, was proposed by Raymond Cattell in the early 1940s. In a commentary discussing issues unresolved in the measurement of adult intelligence, Cattell (1943) postulated that cognitive ability was best represented by two general factors that he identified as fluid intelligence (Gf) and crystallized intelligence

(Gc). Cattell described fluid intelligence as a general facility in reasoning, wherein prior knowledge cannot be used to solve problems, and crystallized intelligence as the storage, retrieval, and use of prior knowledge. Two decades later, Cattell (1963) conducted the first experimental analysis of the theory by administering a series of nine cognitive tasks to a sample of 277 school-age individuals and then subjecting the results to a factor analysis. His results indicated that each of the tasks primarily loaded on one of the two factors.

Cattell chose not to include a general factor in his model, despite the fact that he acknowledged that Gf and Gc were highly correlated and that a third-order factor solution was tenable. Rather, Cattell posited that *g* operated largely through Gf, and he proposed *investment theory* as a vehicle for describing the interaction between Gf and Gc. According to Cattell (1987), fluid ability serves as a limiting factor in how much information individuals can acquire from the environment. Therefore, learning is a function of the interaction between inherited levels of fluid ability and interpersonal metacognitive factors (motivation, drive, personality) that regulate how much that fluid ability is invested by the individual within the environment. The product of that investment is later expressed in the form of developed crystallized ability. Cattell proposed that this interaction helped explain why Gf and Gc were so highly correlated.

The first replication of Gf-Gc theory was conducted by John Horn in his doctoral dissertation supervised by Cattell at the University of Illinois. Horn (1965) administered 31 cognitive and personality tasks to a sample of 297 adults. He extracted several second-order factors from the data, which he identified as fluid intelligence (Gf), crystallized intelligence (Gc), general visualization (Gv), general speediness (Gs), facility (a forerunner of long-term storage and retrieval), carefulness (general cognitive accuracy), premsia (PRM, literacy and artistic ability), and positive self-image (PSI). Horn then extracted two general factors that he did not identify further. The first general factor was composed of Gf, Gv, Gs, and facility. The second general factor was composed of Gc, Gf, and PSI. From their first joint publication (Horn & Cattell, 1966) through the late 1990s, Horn and Cattell collaborated in a systematic program of research aimed at validating and adding additional second-order factors to the Gf-Gc model. Specifically, Horn (1986) laid out an expansion of Gf-Gc theory designating eight broad abilities that were later modified and used to guide the organization of the WJ-R Tests of Cognitive Ability (Woodcock & Johnson, 1989). By the early 1990s, the Gf-Gc model had expanded to include nine broad second-order abilities: fluid intelligence (Gf), crystallized intelligence (Gc), short-term acquisition (Gsm), visual intelligence (Gv), auditory intelligence (Ga), long-term storage and retrieval (Glr), cognitive processing speed (Gs), correct decision speed (CDS), and quantitative knowledge (Gq). At about this time, Woodcock (1990) proposed the additional inclusion of a reading and writing ability factor (Grw).

One of the more consequential discoveries from this Gf-Gc research program has been the demonstration of differential declines in various broad abilities over the course of the human lifespan. In general, it has been demonstrated that Gc tends to increase throughout adulthood, with small declines emerging at around age 70 (Ackerman, 1996). Conversely, Gf skills have been shown to peak in early adulthood (i.e., ages 25–30) and then to decline throughout the rest of the lifespan (Verhaeghen & Salthouse, 1997). Using regression growth models, Noll and Horn (1998) estimated that the loss of Gf ability in adulthood was equivalent to 0.5 to 1.0 IQ units per decade. McGrew and Woodcock (2001) later argued that in spite of strong Gf-Gc correlations, such developmental validity evidence demonstrated that Gf and Gc were in fact orthogonal, unrelated abilities.

The Emergence of Hierarchical Models of Intelligence and Carroll's Three-Stratum Model

Philip Vernon (1950) is credited with articulating the first hierarchical model of cognitive abilities. He posited that a higher-order *g* factor presides over two lower-order factors, which he identified as verbal ability and spatial ability. In his model, the lower-order factors were composed of dozens of narrow abilities, such as psychomotor coordination, attention, fluency, reasoning, and reaction time. Vernon stated that his model was most likely underidentified and went on to hypothesize additional group factors beyond verbal reasoning and spatial thinking, which constituted a more complete model of cognitive ability. Vernon's model was an important reconciliation of Spearman's two-factor model and Thurstone's primary abilities. Additionally, Vernon's model provided empirical support for the verbal–nonverbal dichotomy of cognitive abilities, which was popular as a result of the publication of the Wechsler scales of intelligence (Wechsler, 1949).

A more direct hierarchical test of the nature of cognitive abilities was completed by Gustafsson (1984), who administered a battery of 16 tests to 1,000 sixth-grade students, and then utilized factor analysis to test the fit of several competing models. He reported that the model that best fit the data was a third-order *g* factor that reigned over the three group ability factors. Gustafsson found that the fluid reasoning factor was nearly identical to the third-order general ability factor; this finding has perpetuated the theory that fluid reasoning is largely a proxy for *g*.

A major breakthrough in applied psychometrics occurred with the publication of John Carroll's (1993) *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Carroll assembled a collection of over 400 datasets of factor-analytic studies of cognitive abilities; reanalyzed them by utilizing varimax rotations of the principal factor matrices; and followed up with the Schmid–Lieman procedure (Schmid & Lieman, 1957), which further orthogonalized the factors for a more parsimonious interpretation of the resulting factor structure. Carroll concluded that a three-tier model best fit the data. This model later became known as the three-stratum model. In Carroll's model, *g* or general ability was placed at the apex of the model and was labeled stratum III. The next level, or stratum II, included such broad abilities as fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Glm), broad visual perception (Gv), broad auditory perception (Ga), broad retrieval ability (Gr), broad cognitive speediness (Gs), and reaction time/decision speed (Gt). Over 70 narrow cognitive abilities, organized according to their loadings on the broad factors, made up stratum I.

Carroll's three-stratum model was widely embraced by the scientific community and represented a major paradigm shift in the study of cognitive abilities. The most significant contribution of the model was that it provided the field with a standardized taxonomy to categorize and describe individual cognitive tasks. Many commentators consider Carroll's work the greatest accomplishment in all of applied psychology. Burns (1994, p. 35) stated, for example, "It is simply the finest work of research and scholarship I have read and is destined to be the classic study and reference work of human abilities for decades to come." In the 25 years since its publication, Carroll's work has yet to be seriously challenged.

THE ASCENDANCY OF CHC THEORY

The Birth of CHC Theory

In the late 1990s, Kevin McGrew negotiated the merger of Carroll's three-stratum theory with Cattell and Horn's Gf-Gc theory; he was thus instrumental in the ascendancy of the consolidated CHC model in the field of cognitive assessment. Although the two models were merged, there was not necessarily complete agreement between their creators. Horn refused to accept the validity evidence provided by Carroll for a general ability factor. Horn and Noll (1997) warned that "the problem for the theory of general intelligences is that the factors are not the same from one study to another. . . . The factors represent different mixture of measures, not one general intelligence" (p. 68). Horn (1986) very clearly argued against a single or unitary factor of intelligence, despite widespread opinion to the contrary. He thought that the evidence conclusively indicated several distinct intellectual abilities, each with differing genetic and environmental determinants, different developmental trajectories or courses of development, and different implications for understanding human cognition and achievement. There were also differences between the originators of the two models regarding the number of broad factors, as well as which broad factors were relevant. For instance, Carroll (2003) concluded that there were data to support 10 broad factors, but argued that Gq (quantitative reasoning) was a narrow ability subsumed under Gf and not a stratum II broad factor. He considered quantitative ability to be "an inexact, unanalyzed popular concept that has no scientific meaning unless it is referred to the structure of the abilities that compose it" (Carroll, 1993, p. 627).

Despite these differences, the two theories were consolidated into one theory with three strata: an optional broad general ability or *g* factor; nine broad-ability factors (crystallized knowledge or Gc, fluid reasoning or Gf, visual–spatial processing or Gv, auditory processing or Ga, short-term memory or Gsm, long-term storage and retrieval or Glr, processing speed or Gs, quantitative knowledge and reasoning or Gq, and reading–writing or Grw); and approximately 89 narrow abilities. In the past several years, Kevin McGrew has become the de facto standard bearer of research with CHC theory, and his classifications of CHC abilities (e.g., McGrew, 2005) have become the standard framework for discussing CHC theory in the empirical literature, although only seven broad CHC abilities constitute the predominant focus of much of the empirical research on human cognitive abilities. While recent work has demonstrated that some of the broad abilities are much more complex than previously thought (McGrew & Evans, 2004), the initial goal of CHC research was to refine the model into a more accurate and parsimonious summary of human cognitive abilities (McGrew, 2009; Wasserman, 2012).

In the years following the consolidation of the Carroll and Cattell–Horn models, CHC theory has had a visible impact on the development of new and revised individually administered intelligence tests (Keith & Reynolds, 2010). It has become the dominant interpretive framework for measures of intellectual functioning, and, according to Schneider and McGrew (2012, p. 109), "CHC theory has attained the status as the consensus psychometric model of the structure of human cognitive abilities." Despite the widespread representation of CHC within the cognitive testing landscape, the Woodcock–Johnson series has been the only test battery founded exclusively on CHC theory, and the only contemporary test to assess all of the nine broad-ability factors. Other tests, such as the Stanford–Binet Intelligence Scales, Fifth Edition (Roid, 2003); the Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004); the Differential Ability Scales—Second Edition (Elliott, 2007); and the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014), only provide representations of a few broad factors. The broad abilities of fluid reasoning (Gf), crystallized knowledge (Gc), visual–spatial processing (Gv), and short-term memory (Gsm) are widely represented. However, auditory processing (Ga) and long-term storage and retrieval (Glr) are underrepresented within most existing cognitive measures (Flanagan, Ortiz, & Alfonso, 2013).

Beyond CHC

In 2012, Schneider and McGrew proposed changes to CHC theory, which McGrew labeled as going "beyond CHC." McGrew and Schneider posited 16 cognitive domains grouped into five functional areas. The first functional area, cognitive knowledge, is composed of Gc, Grw, Gq, and Gkn. Cognitive operations is the second functional area and is made up of Gf, Glr, Gv, and Ga. The third functional area is cognitive efficiency and control, consisting of Gsm and Gs. The fourth functional area is sensory functions and consists of visual, auditory, and tactile (Gh), kinesthetic (Gk), and olfactory (Go) sensations. The final area is motor functions, consisting of strength, finger dexterity, and manual dexterity (Gp and Gps). Similar to all of the reconceptualizations discussed in this chapter, McGrew and Schneider's reconceptualization of CHC is theoretical; there is no research that currently would support the hypothesized changes, although the authors refer to a synthesis of the research literature in the past 10–15 years as supportive of their proposed changes.

Schneider and McGrew (2012) further reconceptualized the CHC broad and narrow abilities into five expanded domains, labeled motor (Gp), perception (Gv, Gk, Ga, Gh, Go), controlled attention (Gf, Gsm), knowledge (Gc, Gq, and Grw with a greater Gkn), and speed (Gps, Gt, Gs, and Glr). Subsequently, McGrew (2016) has suggested that the broad-ability domain of Glr may have been conceptualized incorrectly in the CHC literature since 1997. McGrew posits that Glr should be separated into two broad abilities: Gl (learning efficiency) and Gr (retrieval fluency). Learning efficiency is defined as "the ability to learn, store, and consolidate new information in long-term memory" (McGrew, 2016). Retrieval fluency is defined as the rate and facility with which individuals can generate and regain verbal and nonverbal information or ideas stored in long-term memory (McGrew, 2016). McGrew often refers to his current overview of CHC abilities as CHC model v2.3. Additionally, it has been suggested (e.g., Schneider & McGrew, 2012) that a joint neuropsychological and CHC perspective might be the new frontier for understanding cognitive constructs and assessment of cognitive performance and abilities. For the latest revisions and refinements to CHC theory, see Schneider and McGrew (Chapter 3, this volume).

A joint neuropsychological and CHC perspective has been conceptualized by Miller (2013) and articulated in his integrated school neuropsychology/CHC conceptual model. Miller's model is distinctive, as it uses neuropsychological, neuroanatomical, and neuroassessment research to theorize the model's components. In Miller's conceptual model, tasks are classified according to four broad classifications (basic sensorimotor functions; facilitators and inhibitors; basic cognitive processes; and acquired knowledge), and are then further segmented into second- and third-order classifications that denote the broad and narrow CHC constructs being assessed by various tasks. Miller (2015) and Miller, McGill, and Bauman Johnson (2016) have delineated the neuropsychological applications of the WJ IV, WISC-V, WISC-V Integrated, and Wechsler Individual Achievement Test—Third Edition specifically, as well as in relation to Miller's conceptual model. Further supporting a joint neuropsychological/CHC perspective is work by Flanagan, Alfonso, Ortiz, and Dynda (2010) and Flanagan and colleagues (2013), who present an integrated interpretive framework based on psychometric, neuropsychological, and Lurian perspectives, and provide a neurocognitive demand task analysis of the major test batteries using this framework. Flanagan and colleagues (2010, 2013) posit that specific neuropsychological domains correspond well with eight broad CHC abilities—fluid reasoning (Gf), comprehension–knowledge (Gc), processing speed (Gs), short-term memory (Gsm), long-term storage and retrieval (Glr), quantitative knowledge (Gq), reading and writing ability (Grw), and general knowledge ability (Gkn)—and fit well within a cross-battery conceptual framework.

The WJ IV, published in 2014 (Schrank, McGrew, & Mather, 2014) as a substantial revision of the WJ III, further confuses the issue of the structure of the CHC model. The WJ IV consists of three complementary batteries: the Tests of Cognitive Abilities, the Tests of Achievement, and the Tests of Oral Language. According to the authors of the WJ IV, the nine original CHC factors (Gf, Gc, Gv, Ga, Gsm, Gs, Glr, Gq, and Grw) are still measured, although there are slight reconceptualizations of some factors (such as Gsm, which has been renamed Gwm). However, the factor structure of the WJ IV based on reported factor analyses in the technical manual (McGrew et al., 2014) has recently been questioned (Dombrowski, McGill, & Canivez, 2017). The technical manual reported the results of exploratory cluster, factor, and confirmatory factor analyses. These analyses were conducted "during the early stages of WJ IV data collection" with subsections of the normative group and on the completed normative sample of 7,416 individuals (McGrew et al., 2014, pp. 149–150). The authors chose a model-generating approach to their analysis of the data, and they note that it was a major component in the multistage structural validity procedures utilized. Three different exploratory methods—cluster analysis, exploratory principal-components analysis, and multidimensional scaling analysis—were applied.

Dombrowski and colleagues (2017) have criticized the exploratory and confirmatory analyses described in the technical manual, noting that significant amounts of the data were imputed; that some analyses were extrapolated; and that the methods chosen for the analyses were not appropriate, incomplete, or less sophisticated. McGrew and colleagues (2014) note that the obtained structural models varied by methodology. In general, two consistent factor structures emerged. The first factor structure consisted of five factors (Gc, Gs, Grw, Gq-Gf, and Gwm). The second consisted of three factors that they labeled auditory–linguistic (Ga, Gc, and select Gwm tasks), visual–figural (Gv, Gf, Glr), and quantitative–numeric (Gq and select Gwm tasks). Confirmatory factor analyses were also conducted, and two factor structures were found that fit the data. From one of these analyses, the three-stratum hierarchical model emerged, including the general or *g* factor, broad-ability factors, and narrow abilities. From the other, McGrew and colleagues extracted a two-stratum factor structure consisting of the nine broad-ability factors and narrow abilities. Dombrowski and colleagues used the correlation matrices found in the technical manual to conduct several exploratory factor analyses on two age groups (9–13 and 14–19), using the cognitive tests of the WJ IV. Dombrowski and colleagues identified a four-factor solution as having the most parsimonious fit, but from their professional point of view, they suggest that the WJ IV is best viewed primarily as a measure of *g*, as it accounts for the majority of total and common variance. Such disparate and contradictory findings are confusing and create questions as to which latent theoretical structures should be applied interpretively.

A FUNCTIONAL REFINEMENT OF CHC THEORY

We suggest a more practical reconceptualization of CHC theory and functional nomenclature to describe CHC latent cognitive factors. We refer to this model as the *functional CHC model* (*F-CHC*; see Figure 32.1). We recommend grouping the cognitive abilities represented by CHC factors into three broad conceptual domains: acquired knowledge, thinking abilities, and cognitive efficiency. Additionally, a review of the CHC literature and research regarding broad and narrow abilities contributing to CHC factors suggests that each broad ability or CHC factor can be reduced to two primary narrow abilities, without loss of significant information needed for clinical utility.

Functional CHC nomenclature		**Scientific CHC nomenclature**	
Broad abilities	**Narrow abilities**	**Broad abilities**	**Narrow abilities**
Acquired knowledge		Acquired knowledge	
(Gc) Comprehension–knowledge	(Gc-VA) Verbal ability	Gc: Crystallized intelligence	LD: Language development LS: Listening ability VL: Lexical knowledge
	(Gc-K) Factual knowledge		KO: General verbal information
(Grw-R) Broad reading	(Grw-RS) Reading skills	Grw: Reading and writing	RD: Reading decoding
	(Grw-RC) Reading comprehension		RC: Reading comprehension RS: Reading speed
(Grw-W) Broad writing	(Grw-WS) Writing skills		SG: Spelling ability EU: English usage
	(Grw-WC) Writing composition		WA: Writing ability WS: Writing speed
(Gq) Broad mathematics	(Gq-C) Calculation	Gq: Quantitative knowledge	A3: Mathematical achievement
	(Gq-Ap) Applied math		KM: Mathematical knowledge
(Gp) Psychomotor abilities	(Gp) Handwriting		
Thinking abilities		Thinking abilities	
(Gv) Visual–spatial processing	(Gv-PP) Pictorial processing	Gv: Visual processing	CS: Closure speed
	(Gv-SP) Spatial processing		VZ: Visualization SR: Spatial relations SS: Spatial scanning
(Ga) Auditory processing	(Ga-SD) Sound discrimination	Ga: Auditory processing	US: Speech sound discrimination UR: Resistance to auditory stimulus distortion
	(Ga-Ph) Phonetics		U3: Sound discrimination PC: Phonetic coding
(Glm) Learning–memory	(Glm-IR) Immediate recall	Gl: Learning efficiency	MA: Associative memory MM: Meaningful memory M6: Free-recall memory MV: Visual memory
	(Glm-MR) Memory retrieval	Gr: Retrieval fluency	FI: Ideational fluency FA: Associational fluency FE: Expressional fluency SP: Sensitivity to problems/alternative solution fluency FO: Originality/creativity NA: Naming facility FW: Word fluency LA: Speed of lexical access FX: Figural flexibility

(continued)

FIGURE 32.1. Comparison of functional CHC (F-CHC) and scientific CHC nomenclatures.

Functional CHC nomenclature		Scientific CHC nomenclature	
Broad abilities	**Narrow abilities**	**Broad abilities**	**Narrow abilities**
(Gr) Reasoning	(Gr-CR) Contextual reasoning	Gf: Fluid reasoning	RQ: Quantitative or numerical reasoning
	(Gr-ID) Inductive/deductive reasoning		I: Induction RG: General sequential reasoning
Cognitive efficiency		Cognitive efficiency	
(Gcm) Conscious memory	(Gcm-MS) Memory span	Gsm: Short-term memory	MS: Memory span
	(Gcm-WM) Working memory		MW: Working memory WM: Working memory capacity AC: Attentional control
(Gs) Cognitive processing speed	(Gs-PS) Perceptual speed	Gs: Processing speed	P: Perceptual speed AC: Attention and concentration
	(Gs-TS) Thinking speed		R9: Rate of test taking N: Number facility

FIGURE 32.1. *(continued)*

In the F-CHC nomenclature, the first domain, acquired knowledge, consists of comprehension–knowledge (Gc), reading and writing (Grw), and mathematics (Gq). In the extant CHC perspective, comprehension–knowledge (Gc) is composed of four narrow abilities: language development (LD), listening ability (LS), general verbal information (KO), and lexical knowledge (VL). The F-CHC nomenclature combines LD and LS into an ability called verbal ability (Gc-VA), and KO and VL into an ability called factual knowledge (Gc-K). In the F-CHC nomenclature, the reading and writing ability factor (Grw) is split into two broad abilities, reading (Grw-R) and writing (Grw-W), as these two areas are perceived by most people (including educators) as separate abilities. Reading consists of two abilities, reading skills (Grw-RS) and reading comprehension (Grw-RC); writing consists of writing skills (Grw-WS) and writing composition (Grw-WC). Finally, quantitative knowledge/reasoning (Gq) has been relabeled as mathematics, to better conceptualize the nature of this latent factor. Calculation (Gq-C) and applied math (Gq-AP) are the two primary abilities composing Gq in this reconfiguration.

Within the F-CHC nomenclature, the second domain, thinking abilities, consists of visual–spatial processing (Gv), auditory processing (Ga), learning–memory (Glm), and reasoning (Gr). In current CHC theory, Gv is composed of five narrow abilities: visualization (VZ), closure speed (CS), spatial relations (SR), spatial scanning (SS), and visual memory (VM). Within the F-CHC model of Gv, these five abilities are collapsed into pictorial processing (Gv-PP) and spatial processing (Gv-SP). For auditory processing (Ga), the narrow abilities of speech sound discrimination (US), resistance to auditory stimulus distortion (UR), sound discrimination (U3), and phonetic coding (PC) are reconceptualized as sound discrimination (Ga-SD) and phonetics (Ga-Ph).

Long-term storage and retrieval (Glr) is renamed in F-CHC as learning–memory (Glm), to better reflect the cognitive abilities being measured in this domain. Glr is consistently misinterpreted by clinicians as long-term memory, or long-term retrieval, when in reality it provides measures of learning, memory, and the ability to efficiently and effectively retrieve what has been encoded in memory or learned. In the extant CHC model, Glr has 13 narrow abilities, which are difficult to conceptualize and measure. In the F-CHC reconceptualization of the Glm factor, the two primary abilities are immediate recall (Glm-IR) and memory retrieval (Glm-MR).

Finally, fluid reasoning (Gf) has been renamed as simply reasoning (Gr), to better reflect the skills measured in this domain. The concept of *fluid* is not really meaningful in any context and historically seems to have been used as a synonym for novel problem solving. In the extant CHC nomenclature, Gf has three narrow abilities: quantitative reasoning (RQ, often misinterpreted as Gq), induction (I), and general sequential reasoning (RG). In the F-CHC nomenclature, the two primary narrow abilities for Gf are contextual reasoning (Gr-CR) and inductive/deductive reasoning (Gr-ID). Contextual reasoning essentially replaces quantitative reasoning. The Number Series subtest on the WJ IV COG would be a good example of a subtest ostensibly measuring quantitative reasoning, but in reality providing a comprehensive measure of contextual reasoning and abstract thinking. Sternberg (1990) has long argued that contextual reasoning and abstract thinking abilities are required in many areas of cognitive functioning, including reading or solving scientific equations.

The third domain, cognitive efficiency, includes conscious memory (Gcm), formerly Gsm/Gwm, and cognitive processing speed (Gs). Gsm has been a problematic factor within the CHC framework for many years. Initially called short-term memory (Gsm), it was reconceptualized recently in the WJ IV as short-term working memory (Gwm); however, neither term clearly represents what is being measured within this factor. Gsm incorporates memory span (MS), working memory (WM), working memory capacity (MW), and attention/concentration (AC). Within the F-CHC nomenclature, Gsm is renamed conscious memory (G*cm*) and is composed of memory span (Gcm-MS) and working memory (Gcm-WM). As currently conceptualized in CHC theory, processing speed (Gs) has four narrow abilities: perceptual speed (P), attention/concentration (AC), rate of test taking (R9), and number facility (N). In the F-CHC model, processing speed is renamed cognitive processing speed (Gs), and the primary abilities are perceptual speed (Gs-PS) and thinking speed (Gs-TS).

CONCLUSION

The CHC theory of cognitive functioning is theoretical, but has a solid research base and is supported by over 20 years of additional psychometric research. CHC theory has now been applied in some form or another to several contemporary measures of intellectual functioning. As such, it has been referred to as the consensus psychometric/theoretical model guiding interpretation of cognitive ability measures (see, e.g., Schneider & McGrew, 2012). In order for a good theory or model to stand the test of time, it must evolve as new knowledge is gained, while still retaining its foundation or core. For any revision or refinement of the model to be considered valid, it too must be subjected to rigorous research, with each suggested addition, subtraction, or substitution carefully weighed and considered.

The evolutionary advances proposed by McGrew and others, as well as changes in perspective given to current CHC constructs within measures such as the WJ IV and WISC-V, would suggest a need for practitioner's to re-conceptualize interpretation of CHC latent abilities. A good deal of attention in the WJ IV focuses on cognitive complexity (e.g., constructs that are cognitively complex and difficult to measure; tasks that are cognitively complex and thus measuring more than one latent ability).

CHC theory and its variants have also suffered from "descriptive messiness" of the latent constructs. For example, the construct of short-term memory (Gsm) has always been problematic and not reflective of the breadth and depth of our understanding of memory functions. For the most part, the CHC broad ability of Gsm was referring to short-term memory capacity or short-term working memory capacity, and most tasks measured this latent ability in the auditory domain only. With the advent of the WJ IV, an attempt was made to address the limitations of Gsm—primarily by renaming the factor and calling it Gwm, to better reflect the fact that tasks were measuring short-term working memory capacity. However, there is strong research to suggest that the neurocognitive constructs of short-term memory capacity, short-term working memory capacity, and working memory can be distinguished from each other, as well as, adequately independently measured in both auditory and visual domains. This descriptive messiness has now been extended by McGrew's concern that Glr (long-term storage and retrieval) has probably been misconstrued for the last 20 years, as well as his view that the importance of Ga (auditory processing) has been underrepresented.

For the average clinician attempting to apply the CHC framework to assessment and interpretation, the complexity of CHC theory and the human cognitive abilities it purports to describe, as well as the use of instruments supposedly measuring these abilities, is often challenging in and of itself. The need to possess in-depth understanding of latent neurocognitive factors, multiple narrow abilities, and variations in how tasks measure both latent abilities and narrow abilities, as well as proposed theoretical advances, can seem overwhelming to the average clinician. Many clinicians already find it difficult to describe and/or explain to clients current conceptualizations of CHC broad and narrow abilities and how they are measured by various tasks, as well as what all these things mean for their or their children's functioning. The effectiveness of communicating assessment results to consumers (clients, parents, teachers, etc.) is a direct function of the language and vocabulary being used. Our proposed F-CHC nomenclature provides a more parsimonious structure that is consistent with recent research (neuropsychological, cognitive, achievement), more functional for understanding the CHC theoretical constructs, more practical for the average clinician, and more understandable to consumers of the information.

REFERENCES

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence, 22,* 227–257.

Burns, R. B. (1994). Surveying the cognitive terrain. Reviewed work(s): *Human cognitive abilities: A survey of factor-analytic studies* by John B. Carroll. *Educational Researcher, 23*(3), 35–37.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about 10 broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). New York: Pergamon Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* New York: Elsevier.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment. 29*(4), 394–407.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179–203.

Horn, J. L. (1965) *Fluid and crystallized intelligence: A factor analytic study of structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign. (UMI No. AAT 6507113).

Horn, J. L. (1986). Intellectual ability concepts. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–77). Hillsdale, NJ: Erlbaum.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57,* 253–270.

Horn, J. L. & Noll, J. G. (1997). Human cognitive abilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Jensen, A. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York: Guilford Press.

McGrew, K. S. (2009). CHC theory and the Human Cognitive Abilities Project: Standing on the shoulders of the giants of psychometric intelligence research [Editorial]. *Intelligence, 37,* 1–10.

McGrew, K. S. (2016). The CHC model of human cognitive abilities—a proposed revision (v2.3): Has Glr been incorrectly conceptualized since 1997? Retrieved from *www.iqscorner.com/2016/06/the-chc-model-of-human-cognitive.html*

McGrew, K. S., & Evans, J. (2004). *Internal and external factorial extensions to the Cattell–Horn–Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise.* St. Cloud, MN: Institute for Applied Psychometrics.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock–Johnson IV technical manual.* Rolling Meadows, IL: Riverside.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III technical manual.* Itasca, IL: Riverside.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D. C. (2015). Integration of the WJ IV, WISC V, WISC V Integrated, and WIAT III into a school neuropsychological assessment model. *School Psychology Forum: Research in Practice, 9*(3), 1–18.

Miller, D. C., McGill, R., & Bauman Johnson, W. L. (2016). Neurocognitive applications of the Woodcock–Johnson IV. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 355–388) London: Academic Press/Elsevier.

Noll, J. G., & Horn, J. L. (1998). Age differences in processes of fluid and crystallized intelligence. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 263–282). Mahwah, NJ: Erlbaum.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

Schneider, W. J., & McGrew, K. S., (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293. Retrieved from *http://psychclassics.yorku.ca/Spearman*.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Spearman, C., & Jones, L. W. (1959). *Human ability: A continuation of "The Abilities of Man."* London: Macmillan.

Sternberg, R. J. (1990). Wisdom and its relations to intelligence and creativity. In R. J. Sternberg (Ed.), *Wisdom: Its nature, origins, and development* (pp. 142–159). New York: Cambridge University Press.

Thorndike, R. L., Lay, W., & Dean, P. R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology, 20,* 364–369. Retrieved from *www.jstor.org/stable/1413366*

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of the mind*. Chicago: University of Chicago Press.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear age effects and structural models. *Psychological Bulletin, 122,* 231–249.

Vernon, P. E. (1950). *The structure of human abilities*. London: Methuen.

Wasserman, J. D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 3–54). New York: Guilford Press.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition*. Bloomington, MN: Pearson.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery—Revised*. Allen, TX: DLM Teaching Resources.

CHAPTER 33

The Emergence of Neuropsychological Constructs into Tests of Intelligence and Cognitive Abilities

Daniel C. Miller
Denise E. Maricle

PERSPECTIVES ON NEUROCOGNITIVE ASSESSMENT

Baron (2004) has pointed out that "intelligence tests are not neuropsychological instruments" (p. 114). Historically, tests of intelligence were designed empirically to be predictive of academic achievement, and the lens through which they were viewed was very narrow. However, in the last 25 years there has been a rapid advance in theoretically driven intelligence tests, with the Lurian and Cattell–Horn–Carroll (CHC) theories being the primary approaches used in intelligence test development, construction, and interpretation. Consequently, intelligence tests today are being used differently, with a broader lens; less attention is being given to the concept of *g* (general intellectual ability) as a predictor of academic achievement, and more emphasis is being placed on combinations of cognitive constructs as predictors of academic achievement.

The Lurian and CHC theories provide unique perspectives on the assessment of cognitive abilities. The Lurian perspective is grounded in clinical neuropsychology with extensive empirical evidence stemming from strong brain–behavior research. Currently, Lurian theory provides the foundation for the Cognitive Assessment System—Second Edition (CAS2; Naglieri, Das, & Goldstein, 2014) and can be used as an alternative interpretive approach for the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004). In contrast to Lurian theory, CHC theory is largely based on a factor-analytic cross-battery approach. The Woodcock–Johnson IV (WJ IV) Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014), Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b), and Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a) serve as the preeminent instruments that operationalize CHC theory. CHC theory has also been applied to the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003), the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007), and the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014), as well as being the preferred interpretive model for the KABC-II. CHC theory has thus become the "default" theoretical perspective in the majority of tests of cognitive ability.

Although there is significant psychometric research supporting CHC theory, one of its weaknesses is that there is no empirical evidence for the direct linkage between neuroanatomical functions and the cognitive constructs posited by the theory. As Baron (2004) has correctly pointed out, current intelligence tests have typically not been validated with respect to brain function. We agree with Baron that intelligence tests are not neuropsychological instruments. Presently, neuroimaging studies linking brain function to concurrent measures of commonly used tests of cognitive abilities are not available. Perhaps it is best to view intelligence tests as packaged samples of behavior that (1) may be starting points for generating hypotheses about possible deficits in neuropsychological processing, and (2) may also be interpreted from a neuropsychological perspective.

Miller (2007, 2010, 2013; Miller & Maricle, 2012) has introduced the school neuropsychological conceptual model as a way of organizing cross-battery assessment data on school-age children. Miller (2013) has since expanded his conceptual model to further integrate neuropsychological constructs with CHC theory, and it is now called the *integrated school neuropsychology/Cattell–Horn–Carroll conceptual model* (integrated SNP/CHC model; see Figure 33.1). The purposes of this model are (1) to facilitate clinical interpretation by providing an organizational framework for assessment data; (2) to strengthen the linkage between assessment and evidence-based interventions; and (3) to provide a common frame of reference for evaluating the effects of neurodevelopmental disorders on neuropsychological processes. The complete integrated SNP/CHC model includes the integration of academic achievement and social-emotional functioning with the major neuropsychological assessment components (see Miller, 2013, for a complete review); however, in this chapter we focus only on the neurocognitive portions.

Flanagan, Alfonso, Ortiz, and Dynda (2010) have discussed the possible integration of the Lurian, CHC, and SNP models as a means of providing a common lens through which to examine neurocognitive constructs. For example, they have shown how neuropsychological constructs such as concept formation or working memory can be classified according to the three perspectives. Thus Flanagan and colleagues provide a much-needed framework for translating and integrating the concepts, principles, and nomenclature of the three models. Flanagan, Ortiz, and Alfonso (2013) have further integrated neuropsychological constructs from the major pediatric neuropsychological tests with tests of cognitive and academic achievement in the third edition of their *Essentials of Cross-Battery Assessment*.

Historically, clinical neuropsychological assessment has attempted to link the cognitive and behavioral manifestations of neuropsychological processing with known brain structures or functions; it has relied on both quantitative and qualitative aspects of performance. Traditional neuropsychological tests consist of specifically designed tasks used to measure psychological functions known to be linked to particular brain structures or pathways. The tests are typically used to assess impairment after an injury or illness known to affect neurocognitive functioning, or are used in research to compare neuropsychological abilities across experimental groups. In contrast, tests of cognitive abilities have traditionally focused almost exclusively on quantitative measures of cognitive performance, while ignoring qualitative behaviors. Baron (2004) has stated that this limitation prevents the clinical detection of meaningful performance patterns, such as strategy selection, analysis of error patterns, and response latency.

In response to such limitations, test publishers have begun to gather the prevalence rates of observable qualitative behaviors and to provide practitioners with useful base rate data. Starting with the original CAS (Naglieri & Das, 1997), followed by the Wechsler Intelligence Scale for Children—Third Edition as a Process Instrument (Kaplan, Fein, Kramer, Delis, & Morris, 1999), the Wechsler Intelligence Scale for Children—Fourth Edition Integrated (WISC-IV Integrated; Wechsler et al., 2004), and the WISC-V Integrated (Wechsler & Kaplan, 2015) test authors and publishers have provided bridges between the fields of cognitive and neuropsychological assessment by including qualitative behaviors in tests of cognitive functions.

An emerging trend in both cognitive and neuropsychological assessment is the development of computer-based assessment (CBA). Technological innovation is driving the use of computer-based assessment devices, such as tablets for the adaptation of examiner-administered tests with scoring and interpretation. CBA ranges from stand-alone computer-administered versions of established examiner-administered tests, such as the Wisconsin Card Sorting Test (Heaton & PAR Staff, 2003) or the WISC-V, to fully web-integrated applications (such as CNS Vital Signs; see below). CBA is viewed as less time- and resource-intensive than traditional examiner-administered tests, and thus more cost-effective.

Currently the assumption is being made that if an established examiner-administered test is adapted to computer-based administration, the normative and psychometric data provided by

	Cognitive processes				**Facilitators–inhibitors**		
Sensory–motor	Visual–spatial	Auditory	Learning and memory	Executive functions	Attention	Working memory	Speed, fluency, and efficiency
Lateral preference Sensory functions: • Auditory acuity • Visual acuity • Kinesthetic sensation and perception • Olfactory • sensation and perception Fine motor functions: • Coordinated finger–hand movements Visual–motor integration skills Gross motor skills	Visual–spatial perception: • Visual discrimination and spatial localization • Visual–motor constructions Visual–spatial reasoning: • Recognizing spatial configurations • Visual gestalt closure • Visual–spatial analysis with and without mental rotations	Sound discrimination Auditory/ phonological processing	Rate of new learning: • Verbal • Visual • Paired associative Immediate memory: • Verbal • Visual • Verbal–visual associative Delayed recall: • Verbal • Visual • Verbal–visual associative	Cognitive flexibility (set shifting) Concept formation Problem solving, planning, and reasoning: • Planning • Deductive and inductive reasoning • Sequential reasoning • Quantitative reasoning Response inhibition	Selective/focused attention: • Auditory • Visual Sustained attention: • Auditory • Visual • Auditory and visual Attentional capacity: • Attention span for numbers, letters, or visual sequences • Attention span for words or sentences • Attention span for stories	Verbal working memory Visual working memory	Performance fluency: • Psychomotor • Perceptual • Figural • Naming • Rate of test taking • Oral–motor Retrieval fluency: • Word • Semantic Acquired knowledge: • Reading • Writing • Math Speed–accuracy interactions

FIGURE 33.1. Miller's integrated SNP/CHC neuropsychological assessment model.

the two versions are equivalent. However, Bauer and colleagues (2012) clearly state that even when a traditional examiner-administered test is programmed for computer administration, it "becomes a new and different test" (p. 366) and is not just the existing test in a slightly different format. Equivalency, reliability, and validity data for computer-based assessment are scarce.

Publishers have long offered computer- program or web-based scoring for numerous cognitive, academic, behavioral, and personality assessments. More recently, publishers are moving to computer-based or online platforms for the administration of previous examiner administered assessments. For example, Pearson offers the Q-global Web-Based Administration, Scoring, and Reporting (online system), Q-interactive (tablet-based system), and Q Local Scoring and Reporting Software (desktop-based system) options for administering, scoring, and reporting the results of more than 30 instruments—including cognitive measures such as the WISC-V, achievement measures such as the Wechsler Individual Achievement Test—Third Edition (WIAT-III), and personality measures such as the Minnesota Multiphasic Personality Inventory–2—Restructured Form (MMPI-2-RF). Psychological Assessment Resources (PAR) uses PARiConnect (*www.pariconnect.com*), which is an online assessment, scoring, and reporting platform for numerous instruments that PAR publishes. Similarly, Multi-Health Systems (MHS) has an Online Assessment Center (*www.mhs.com/infocenter.aspx?gr=mhs&prod=service&id=Overview*), where users can administer, score, and obtain reports on the instruments that MHS publishes.

Integrated neuropsychological assessments designed for computer-based administration include the Cambridge Neuropsychological Test Automated Battery (CANTAB; *www.cambridgecognition.com*), the Automated Neuropsychological Assessment Metrics—Version 4 (ANAM 4; *vistalifesciences.com*), CNS Vital Signs (*www.cnsvs.com*), MicroCog: Assessment of Cognitive Functioning Windows Edition 2004 (Powell et al., 2004), NeuroTrax: Innovative Science for Brain Health (*www.neurotrax.com*), and ImPACT Concussion Assessment and Management (*www.impact.com*).

The CANTAB advertises itself as including the world's most validated and sensitive touchscreen tests of cognitive functioning. The publisher cites over 1,600 peer-reviewed papers in a bibliography on the website (see URL above). The CANTAB is the most recent version of the Cambridge Neuropsychological Test Automated Battery developed at Cambridge University by neuroscientists Sahakian and Robbins in the 1980s. Areas covered include memory, executive functions, attention, decision making, and social cognition. For example, subtests included in the memory section are Paired Associates Learning, Delayed Matching to Sample, Graded Naming Test, Pattern Recognition Memory, Spatial Recognition Memory, Verbal Recognition Memory, Spatial Span and Spatial Working Memory. Reviews suggest that the CANTAB subtests demonstrate strong reliability and validity across a variety of age ranges and disability groups (Luciana, 2003).

The ANAM 4 is a library of 22 computer-based cognitive assessments developed by the U.S. military. It provides measures of attention, concentration, reaction time, memory, processing speed, and decision making, but is not considered to include a measure of intelligence. The ANAM 4 is intended to measure an individual's neurocognitive status at a point in time, as well as changes in cognitive status over time. The tasks of the ANAM 4 are sensitive to cognitive changes associated with injury (e.g., trauma, blast), illness, exposure or risk factors (e.g., toxins, fatigue), and interventions (e.g., medications). ANAM 4 is based on decades of clinical and laboratory research, and is referenced in more than 300 peer-reviewed independent research studies. It has the most comprehensive military research and clinical application record of any cognitive assessment technology, with researchers from many federal agencies contributing to its development (see URL above).

CNS Vital Signs touts itself as a world leader in the development and design of neurocognitive and behavioral assessment technologies and tools. A customizable battery can be developed from 10 normed (and 26 un-normed) neuropsychological tests and more than 50 evidence-based rating scales. CNS Vital Signs tests are suitable for clients ages 8–89 and can be customized for more than 50 languages. Reviews of CNS Vital Signs (Gualtier & Johnson, 2006) suggests strong validity and reliability for the 7 subtests of the Brief Clinical Battery. Numerous research studies are listed on the CNS Vital Signs website (see URL above).

The technological revolution is inevitable, and it can be predicted that many currently examiner-administered tests will migrate to computer-based platforms for administration, scoring, and interpretation or reporting of results. It is imperative that researchers examine the equivalency, reliability, and validity of these computer-based formats before computer-based measures of cognition and neuropsychological functioning enter widespread clinical use. Although technology is being systematically integrated into the field, it is still necessary to have a conceptual or theoretical framework to aid in clinical interpretation.

THE INTEGRATED SNP/CHC MODEL

The emerging subspecialization of school neuropsychology bridges the gap between traditional psychoeducational approaches and clinical neuropsychological approaches, allowing for both quantitative and qualitative assessment of performance. Miller's (2013) integrated SNP/CHC model provides a framework for school psychologists to integrate quantitative and qualitative performance from commonly used tests of cognitive abilities and to interpret performance from a neuropsychological perspective. See Miller (2015) and Miller, McGill, and Bauman Johnson (2016) for a complete classification of all subtests from the WJ IV, WISC-V, WISC-V Integrated, and WIAT-III into the integrated SNP/CHC model. The bulk of this chapter is structured to provide information about how neuropsychological constructs are measured in traditional neuropsychological tests and current tests of cognitive ability. Within the integrated SNP/CHC model, tasks from the various instruments are classified on the basis of their underlying neurocognitive demands, the theoretical perspectives from which each originates, and psychometric data from cross-battery research.

One of the limitations of this chapter is that we are only looking at the conceptual overlap between traditional neuropsychological instruments and tests of cognitive functions. In order to fully assess all of the neuropsychological constructs within the integrated SNP/CHC model, practitioners would have to administer a broader array of instruments, such as specialized or targeted tests of learning and memory or of sensory–motor functions. The following constructs covered in common by traditional neuropsychological instruments and tests of cognitive ability are discussed: sensory–motor functions; cognitive processes (visual–spatial, auditory, learning and memory, and executive functions); and facilitators–inhibitors (attention, working memory, and speed and efficiency of processing).

Sensory–Motor Functions

One of the major contrasts between intelligence testing and neuropsychological assessment is related to the assessment of sensory–motor functions. Neuropsychological evaluations typically assess for sensory functions of vision, hearing, and the sense of touch, as well as fine and gross motor functions, whereas the major tests of cognitive ability have traditionally not included these neuropsychological functions as part of their core batteries. Only two neuropsychological constructs within the sensory–motor domain have been integrated into current tests of cognitive ability: motor sequencing and visual–motor integration (see Table 33.1).

Traditional neuropsychological assessment has multiple examples of tests designed to measure motor sequencing actions with both the dominant and nondominant hands. These measures stem from Alexander Luria's original investigations of motor functions (Christensen, 1975), and all re-

TABLE 33.1. Sensory–Motor Functions Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
Motor sequencing	• Finger Sequencing Test • Fist–Edge–Palm Test • NEPSY-II: Fingertip Tapping (dominant and nondominant hand combined), Imitating Hand Positions, and Manual Motor Sequences • Oseretskii Test of Reciprocal Coordination • TOMAL-2: Manual Imitations	KABC-II: Hand Movements
Visual–motor integration	• VMI • NEPSY-II: Design Copying • Rey Complex Figure Test and Recognition Trial	DAS-II: Copying

Note. For full names of abbreviated tests in this and later tables, see chapter text.

quire the placement of the hand in three or more successive positions after either a verbal command or visual modeling from the examiner. Traditional neuropsychological measures of motor sequencing include the Finger Sequencing Test (Welsh, Pennington, & Groisser, 1991), the Fist–Edge–Palm Test (Christensen, 1975), and the Oseretskii Test of Reciprocal Coordination (Buchanan & Heinrichs, 1989) (see Baron, 2004, for a review).

Various sensory–motor tasks (e.g., Fingertip Tapping, Imitating Hand Positions, and Manual Motor Sequences) have also been included in the NEPSY-II (Korkman, Kirk, & Kemp, 2007), which is a comprehensive neuropsychological battery designed for school-age children. The Test of Memory and Learning—Second Edition (TOMAL-2; Reynolds & Voress, 2007) also includes a Manual Imitations test. Among tests of cognitive ability, only the KABC-II includes a motor sequencing task called Hand Movements.

One of the classic neuropsychological measures of visual–motor integration is the Rey–Osterrieth Complex Figure Test (ROCF), developed in the 1940s by Rey (1941) and Osterrieth (1944). The test has been revised and restandardized multiple times (e.g., the Rey Complex Figure Test and Recognition Trial; Meyers & Meyers, 1995), but the purpose of the test remains the same: to assess for visual–spatial constructional ability. The ROCF requires copying a complex figure drawing with an added delayed-recall component. Other two-dimensional copying tasks that may be used to supplement tests of cognitive ability are the Beery–Buktenica Developmental Test of Visual–Motor Integration, Sixth Edition (VMI; Beery, Buktenica, & Beery, 2010) and the Design Copying test from the NEPSY-II.

Only one current test battery of cognitive ability includes a direct measure of visual–motor integration: the Copying test from the DAS-II. Although other current tests of cognitive ability do not include measures of visual–motor integration, it is common practice in psychoeducational assessment to include an additional test like the Beery VMI, to rule out the presence of any sensory–motor deficit that could explain a learning difficulty. For example, before a specific learning disability can be diagnosed, the Individuals with Disabilities Education Improvement Act of 2004 (IDEA; 2004) requires that any sensory–motor impairment be ruled out as a primary causal factor.

Since many childhood disorders are known to have associated sensory–motor deficits (see Decker & Davis, 2010, for a review; see also Decker, Strait, Roberts, & Ferraracci, Chapter 24, this volume), clinicians must not rely on using tests of cognitive ability alone. Tests of cognitive ability do not include other important sensory–motor constructs that should be measured, such as speed and accuracy of motor output, both of which have a direct impact upon a learner's achievement (Miller, 2013). When the referral questions raise serious concerns about sensory–motor functions, an appropriately trained clinician is encouraged to include the sensory–motor subtests from the NEPSY-II or to administer the sensory–motor portions of the Dean–Woodcock Neuropsychological Battery (Dean & Woodcock, 2003). The latter include eight measures of sensory functioning, nine measures of motor functioning, and one measure of lateral dominance or preference.

Cognitive Processes

In the integrated SNP/CHC model (Miller, 2013), four core sets of cognitive processes/functions have been delineated: visual–spatial processes, auditory processes, learning and memory processes, and executive functions. These four cognitive processes are influenced by the basic sensorimotor functions and are enhanced or inhibited by the facilitators and inhibitors, respectively. In what follows, the neuropsychological and cognitive ability tests designed to measure these broad constructs are highlighted.

Visual–Spatial Processes

In the integrated SNP/CHC model (Miller, 2013), visual–spatial processes are influenced both by basic sensory–motor functions and by facilitators–inhibitors (discussed later in this chapter). The tests of visual–spatial processes are conceptually divided into two categories: (1) tests that require visual–spatial perception and (2) tests that require visual–spatial reasoning (see Table 33.2).

Visual–Spatial Perception

Visual–motor three-dimensional construction abilities are measured by both neurocognitive and cognitive instruments (see Table 33.2). The Block Construction test on the NEPSY-II is an example of a neuropsychological measure that requires three-dimensional construction of blocks. Similar visual–motor, constructional tasks on tests of cognitive ability include Pattern Construction on the DAS-II; Triangles on the KABC-II; Block De-

TABLE 33.2. Visual–Spatial Processes Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
	Visual–spatial perception	
Visual–motor constructions	NEPSY-II: Block Construction	• DAS-II: Pattern Construction • KABC-II: Triangles • WISC-V: Block Design and Block Design—No Time Bonus • WISC-V Integrated: Block Design Process Approach • WPPSI-IV: Block Design • WPPSI-IV: Object Assembly
	Visual–spatial reasoning	
Recognizing spatial configurations	Judgment of Line Orientation	• DAS-II: Matching Letter-Like Forms • KABC-II: Block Counting • WISC-V: Visual Puzzles • WJ IV COG: Visualization
Visual gestalt closure	HVOT	• KABC-II: Gestalt Closure • RIAS-2: What's Missing
Visual–spatial analysis with and without mental rotations	• NEPSY-II: Geometric Puzzles • WRAVMA: Matching	• SB5: Nonverbal Visual–Spatial Processing

sign and Block Design—No Time Bonus on the WISC-V; Block Design Process Approach on the WISC-V Integrated; and Block Design and Object Assembly on the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV; Wechsler, 2012).

Visual–Spatial Reasoning

In the integrated SNP/CHC model (Miller, 2013), visual–spatial reasoning tasks are divided into three groups, based on the neurocognitive demands of the task: (1) tasks requiring the recognition of spatial configurations, (2) tasks requiring part-to-whole analysis–synthesis or visual gestalt closure, and (3) tasks requiring visual–spatial analyses with and without mental rotations (see Table 33.2). These three neurocognitive constructs are shared between neuropsychological measures and tests of cognitive ability.

The Judgment of Line Orientation test (Benton, Hamsher, Varney, & Spreen, 1983) is a classic neuropsychological test designed to measure a person's ability to match the angle and orientation of line in space, or recognition of spatial configurations. Tests of cognitive ability designed to measure recognition of spatial configurations include Matching Letter-Like Forms from the DAS-II; Block Counting from the KABC-II; Visual Puzzles from the WISC-V; and Visualization from the WJ IV COG.

The Hooper Visual Organization Test (HVOT; Hooper, 1958) is a classic neuropsychological test designed to measure visual analysis and synthesis, conceptual reorganization, and mental rotation. The HVOT consists of sets of line drawings of familiar objects that have been divided into fragments. The examinee is asked to reassemble each set of fragments mentally and then name the object. Tests of cognitive ability designed to measure similar neurocognitive constructs include Gestalt Closure from the KABC-II and What's Missing from the Reynolds Intellectual Assessment Scale, Second Edition (RIAS-2; Reynolds & Kamphaus, 2015).

The third neurocognitive construct shared between neuropsychological measures and tests of cognitive ability is visual–spatial analysis with and without mental rotation. Examples of neuropsychological tests that measure this neurocognitive ability include the Geometric Puzzles test from the NEPSY-II and the Matching test from the Wide Range Assessment of Visual–Motor Abilities (WRAVMA; Adams & Sheslow, 1995). One test of cognitive ability that measures this same construct is the Nonverbal Visual–Spatial Processing test from the SB5.

Auditory Processes

In the integrated SNP/CHC model (Miller, 2013), auditory processes are considered as core cognitive processes, which are influenced both by basic sensory–motor functions and by facilitators–inhibitors (again, the latter are discussed later in this chapter). The tests of auditory processes are conceptually divided into two categories: (1) tests that require basic sound discrimination skills and (2) tests that require auditory/phonological processing abilities (see Table 33.3).

There are two traditional neuropsychological tests designed to measure basic auditory discrimination skills: the Seashore Rhythm Test from the Halstead–Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1985) and the Wepman's Auditory Discrimination Test (Wepman & Reynolds, 1986) (see Table 33.3). The Sound Awareness test on the WJ IV OL battery is also designed to measure basic auditory discrimination skills.

In regard to auditory/phonological processing, the Phonological Processing test on the NEPSY-II is a contemporary neuropsychological measure of both higher-order phonemic awareness and phonemic manipulation (e.g., blending or deleted sounds). Tests of cognitive ability that measure this same auditory processing construct include Phonological Processing on the DAS-II; Nonword Repetition and Phonological Processing on the WJ IV COG; and Segmentation and Sound Blending on the WJ IV OL.

Learning and Memory

Learning and memory tasks have been classified in many ways. Conceptualizations of memory and learning have recently been shifting away from traditional information-processing models to models based on neuroscience. The integrated SNP/CHC model (Miller, 2013) classifies tests of memory and learning from neuropsychological tests and cognitive measures into three broad categories: (1) rate of learning, (2) immediate memory, and (3) delayed recall and recognition (see Figure 33.1). The constructs of immediate memory and of delayed recall and recognition are shared between neuropsychological tests and tests of cognitive ability.

Rate of Learning

Rate of learning is measured by the change in the number of correctly recalled words from a fixed list of words repeated over multiple trials. The major tests of cognitive ability do not include tests that are designed to measure rate of learning. However, several neuropsychological batteries or comprehensive tests of memory and learning provide assessments of rate of learning. Examples include the Word Pairs—Learning and Word Lists—Learning scores from the Children's Memory Scale (CMS; Cohen, 1997); the List Memory Learning Effect score from the NEPSY-II; and the Word Selective Reminding score from the TOMAL-2 (Reynolds & Voress, 2007). The California Verbal Learning Test—Children's Version (Delis, Kramer, Kaplan, & Ober, 1994) is a targeted rate-of-learning measure that is often administered by both neuropsychologists and school psychologists.

TABLE 33.3. Auditory Processes Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
	Sound discrimination	
Sound discrimination	• Seashore Rhythm Test • Wepman's Auditory Discrimination Test	• WJ IV OL: Sound Awareness
	Auditory/phonological processes	
Auditory/phonological processes	• NEPSY-II: Phonological Processing	• DAS-II: Phonological Processing • WJ IV COG: Nonword Repetition • WJ IV COG: Phonological Processing • WJ IV OL: Segmentation • WJ IV OL: Sound Blending

Immediate Memory

Immediate memory can be further subdivided according to the neurocognitive demands of a particular task. The first subdivision of immediate memory tasks takes into consideration the sensory input modality required to complete the task. As a result, these tasks can be subdivided into verbal immediate memory, visual immediate memory, and verbal–visual associative memory categories. In addition, the verbal and visual immediate memory tasks can be further subdivided according to the use of contextual cues. For example, a simple memory-for-digits task is a verbal immediate memory task without contextual cues, whereas a memory-for-stories task is a verbal immediate memory task with contextual cues.

Verbal Immediate Memory with No Contextual Cues. Verbal immediate memory tasks with no contextual cues are common to both neuropsychological tests and tests of cognitive ability (see Table 33.4). Strings of numbers, letters, or unrelated words are often used as stimuli in these types of tasks, and the examinee is asked to recall increasingly longer spans of stimuli. Neuropsychological measures of verbal immediate memory that do not include contextual cues include Word Lists—Immediate Recall and Word Pairs—Immediate Recall on the CMS; Word List Interference—Repetition on the NEPSY-II; and Digits Forward, Letters Forward, and Word Selective Reminding on the TOMAL-2.

The major tests of cognitive ability that include measures of verbal immediate memory tasks without contextual cues are Word Series on the CAS2; Recall of Digits—Forward on the DAS-II; Number Recall, Word Order, and Word Order (without color interference) on the KABC-II; Digit Span—Forward on the WISC-V; and Memory for Words on the WJ IV COG.

Verbal Immediate Memory with Contextual Cues. Verbal immediate memory tasks with contextual cues are also found on neuropsychological tests and tests of cognitive ability, but to a lesser degree than the verbal immediate memory tasks without contextual cues (see Table 33.4). Sentences of increasing length and stories with increasing complexity are used as stimuli in these types of tasks. Neuropsychological measures of verbal immediate memory tests that include contextual cues are Stories—Immediate Recall on the CMS; Narrative Memory and Sentence Repetition on the NEPSY-II; Memory for Stories on the TOMAL-2; and Sentence Memory and Story Memory on the Wide Range Assessment of Memory and Learning, Second Edition (WRAML2; Sheslow & Adams, 2003).

Two major tests of cognitive ability that include measures of verbal immediate memory with added contextual cues are the Verbal Memory test on the RIAS-2; the Story Recall test on the WJ IV COG; and Sentence Repetition on the WJ IV OL.

Visual Immediate Memory with No Contextual Cues. The types of stimuli used for visual immediate memory tasks with no contextual cues vary, but can include abstract designs, faces, objects, or pictures (see Table 33.4). Examinees are typically shown visual stimuli for a brief exposure and then asked to motorically reproduce the details of what was seen, or are asked to match, nonverbally or verbally, a newly presented visual stimulus with that previously seen.

Neuropsychological tests that require visual immediate memory for abstract designs include Memory for Designs on the NEPSY-II; Abstract Visual Memory and Visual Sequential Memory on the TOMAL-2; and Design Memory on the WRAML2. Tests of cognitive ability that measure a comparable construct include Figure Memory on the CAS2 and Recall of Designs on the DAS-II.

Neuropsychological tests requiring visual immediate memory for numbers, faces, objects, or pictures (see Table 33.4) include Faces—Immediate Recall on the CMS; Memory for Faces—Immediate Recall on the NEPSY-II; and Facial Memory on the TOMAL-2. The major tests of cognitive ability that include visual immediate memory for numbers, faces, objects, or pictures are Recognition of Pictures on the DAS-II; Face Recognition on the KABC-II; Nonverbal Memory on the RIAS-2; Object Memory on the Universal Nonverbal Intelligence Test—Second Edition (UNIT2; Bracken & McCallum, 2016); Coding Recall on the WISC-V Integrated; and Picture Recognition on the WJ IV COG.

Neuropsychological tests requiring visual immediate memory for spatial locations (see Table 33.4) include Dot Locations—Immediate Recall on the CMS and the Memory for Locations and Visual Selective Reminding tests on the TOMAL-2. The major tests of cognitive ability that include visual immediate memory for spatial locations are Spatial Memory on the UNIT2 and Spatial Span on the WISC-V Integrated.

TABLE 33.4. Learning and Memory Processes Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
	Immediate memory	
Verbal immediate memory for numbers, letters, or words (no contextual cues)	• CMS: Word Lists—Immediate Recall and Word Pairs—Immediate Recall • NEPSY-II: Word List—Interference Repetition • TOMAL-2: Digits Forward, Letters Forward, and Word Selective Reminding	• CAS2: Word Series • DAS-II: Recall of Digits—Forward • KABC-II: Number Recall, Word Order, and Word Order (without color interference) • WISC-V: Digit Span—Forward • WJ IV COG: Memory for Words
Verbal immediate memory for sentences or stories	• CMS: Stories—Immediate Recall • NEPSY-II: Narrative Memory and Sentence Repetition • TOMAL-2: Memory for Stories • WRAML2: Sentence Memory and Story Memory	• RIAS-2: Verbal Memory • WJ IV COG: Story Recall • WJ IV OL: Sentence Repetition
Visual immediate memory for abstract designs	• NEPSY-II: Memory for Designs • TOMAL-2: Abstract Visual Memory and Visual Sequential Memory • WRAML2: Design Memory	• CAS2: Figure Memory • DAS-II: Recall of Designs
Visual immediate memory for numbers, faces, objects, or pictures	• CMS: Faces—Immediate Recall • NEPSY-II: Memory for Faces—Immediate Recall • TOMAL-2: Facial Memory	• DAS-II: Recognition of Pictures • KABC-II: Face Recognition • RIAS-2: Nonverbal Memory • UNIT2: Object Memory • WISC-V Integrated: Coding Recall • WJ IV COG: Picture Recognition
Visual immediate memory for spatial locations	• CMS: Dot Locations—Immediate Recall • TOMAL-2: Memory for Locations and Visual Selective Reminding	• UNIT2: Spatial Memory • WISC-V Integrated: Spatial Span
Visual immediate memory with contextual cues	• CMS: Family Pictures—Immediate Recall • WRAML2: Picture Memory	• UNIT2: Symbolic Memory
Verbal–visual associative learning	• NEPSY-II: Memory for Names—Immediate Recall • TOMAL-2: Object Recall and Paired Recall • WRAML2: Sound–Symbol	• DAS-II: Recall of Objects—Immediate • KABC-II: Atlantis and Rebus • WISC-V: Immediate Symbol Translation • WJ IV COG: Visual–Auditory Learning
	Delayed recall and recognition	
Verbal–visual associative delayed recall	• NEPSY-II: Memory for Names—Delayed Recall • WRAML2: Sound–Symbol—Delayed	• DAS-II: Recall of Object—Delayed • KABC-II: Atlantis—Delayed and Rebus—Delayed • WISC-V: Delayed Symbol Translation • WJ IV COG: Visual–Auditory Learning—Delayed

Visual Immediate Memory with Contextual Cues. There are only a few neuropsychological and cognitive tests designed to measure visual immediate memory with added contextual cues. Neuropsychological measures requiring visual immediate memory with added contextual cues include Family Pictures—Immediate Recall on the CMS and Picture Memory on the WRAML2. The Symbolic Memory test on the UNIT2 is the only test on a major cognitive test battery to measure visual immediate memory with added contextual cues.

Verbal–Visual Associative Learning. Verbal–visual associative learning tasks have been included on both neuropsychological and cognitive ability tests in the past decade. Each of the tasks requires the examinee to learn to associate a verbal label with either a picture, object, symbol, or face, often with corrective feedback. Neuropsychological tests requiring verbal–visual associative learning (see Table 33.4) include Memory for Names—Immediate Recall on the NEPSY-II; Object Recall and Paired Recall on the TOMAL-2; and Sound–Symbol on the WRAML2. The major tests of cognitive ability designed to measure verbal–visual associative learning include Recall of Objects—Immediate on the DAS-II; Atlantis and Rebus on the KABC-II; Immediate Symbol Translation on the WISC-V; and Visual–Auditory Learning on the WJ IV COG.

Delayed Recall and Recognition

Tests of cognitive ability do not typically include measures of delayed recall or recognition. A clinician who needs to assess delayed recall or recognition will need to use stand-alone tests of memory and learning (e.g., the WRAML2, TOMAL-2, or CMS) or the Memory and Learning tests from the NEPSY-II.

Verbal–Visual Associative Delayed Recall. All of the tests mentioned in the section above on verbal–visual associative learning, except for the Object Recall and Paired Recall tests on the TOMAL-2 and Visual-Auditory Learning on the WJ IV, have a delayed-recall portion of their tests (see Table 33.4). The purpose of these tests is to assess long-term memory, or the degree of consolidation, for paired verbal–visual stimuli.

Executive Functions

Among all neurocognitive constructs, the construct of executive functions has generated the most attention, interest, and research. Practitioners and researchers often equate executive functions with intelligence (Blair, 2006; Friedman et al., 2006). As a result, test publishers have included a broad array of executive function tasks on the major tests of cognitive ability.

However, Maricle, Johnson, and Avirett (2010) point out that there is not "a mutually agreed upon list of cognitive components which comprise executive functions" (pp. 599–600). The SNP model (Miller, 2013) classifies tests of executive functions from neuropsychological and cognitive measures into four broad categories: (1) cognitive flexibility (set shifting); (2) concept recognition and generation; (3) planning, problem solving, and reasoning; and (4) response inhibition (see Table 33.5).

Classic neuropsychological tests requiring the executive functions of cognitive flexibility or set shifting include the Halstead–Reitan Category Test (Reitan & Davidson, 1974); the Wisconsin Card Sorting Test (Grant & Berg, 1993); Color–Word Interference—Condition 4, Verbal Fluency—Condition 3, and Design Fluency—Condition 3 from the D-KEFS (Delis, Kaplan, & Kramer, 2001); and the Inhibition (switching condition) from the NEPSY-II. There are no similar cognitive flexibility measures specifically included on major tests of cognitive ability.

Concept Recognition and Generation

Concept recognition or generation tasks typically require the examinee to classify objects or pictures into groups that share a common attribute (e.g., the same color or shape). The goal of these tasks is to identify as many classifications as possible; this requires divergent thinking and concept formation. Neuropsychological tests requiring concept recognition or generation (see Table 33.5) include Sorting and Twenty Questions on the D-KEFS and Animal Sorting on the NEPSY-II. Tests of cognitive ability that include measures of concept recognition or generation are Picture Similarities and Verbal Similarities on the DAS-II; Similarities on the WISC-V; Similarities Multiple Choice on the WISC-V Integrated; and Similarities on the WPPSI-IV.

Planning, Problem Solving, and Reasoning

Despite the multiple classification schemas developed for executive functions, most researchers and theorists agree that executive functions include measures of planning, problem solving, and reasoning. More recent neuropsychological tests

TABLE 33.5. Executive Functions Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
Concept recognition and generation	• D-KEFS: Sorting and Twenty Questions • NEPSY-II: Animal Sorting	• DAS-II: Picture Similarities and Verbal Similarities • WISC-V: Similarities • WISC-V Integrated: Similarities Multiple Choice • WPPSI-IV: Similarities
Planning, problem solving, and reasoning	• D-KEFS: Tower, Color Word Interference—Condition 4, Proverbs, Verbal Fluency—Condition 3, and Word Context • NEPSY-II: Clocks and Inhibition (Switching Condition)	• CAS2: Matrices and Planned Connections • DAS-II: Matrices, and Sequential and Quantitative Reasoning • KABC-II: Pattern Reasoning, Rover, and Story Completion • RIAS-2: Guess What, Odd-Item Out, and Verbal Reasoning • SB5: Verbal Fluid Reasoning and Nonverbal Fluid Reasoning • UNIT2: Analogic Reasoning, Cube Design, and Mazes • WISC-V: Comprehension, Figure Weights, Matrix Reasoning, and Picture Concepts • WISC-V Integrated: Comprehension Multiple Choice and Figure Weights Recall • WPPSI-IV: Comprehension, Matrix Reasoning, Picture Concepts • WJ IV ACH: Number Matrices • WJ IV COG: Concept Formation, Analysis–Synthesis, Number Series
Response inhibition	• D-KEFS: Color Word Interference—Condition 3 (Inhibition) • NEPSY-II: Inhibition—Condition 2 and Statue • Stroop Color–Word Test	• CAS2: Expressive Attention

measuring planning, problem solving, or response inhibition include Tower, Color Word Interference—Condition 4 (Switching), Proverbs, Verbal Fluency—Condition 3 (Switching), and Word Context on the D-KEFS; and Clocks and Inhibition (Switching Condition) on the NEPSY-II.

Tests of cognitive ability that are designed to measure similar executive function constructs are Matrices and Planned Connections on the CAS2; Matrices, and Sequential and Quantitative Reasoning, on the DAS-II; Pattern Reasoning, Rover, and Story Completion on the KABC-II; Guess What, Odd-Item Out, and Verbal Reasoning on the RIAS-2; Verbal Fluid Reasoning and Nonverbal Fluid Reasoning on the SB5; Analogic Reasoning, Cube Design, and Mazes on the UNIT2; Comprehension, Figure Weights, Matrix Reasoning, and Picture Concepts on the WISC-V; Comprehension Multiple Choice and Figure Weights Recall on the WISC-V Integrated; Comprehension, Matrix Reasoning, and Picture Concepts on the WPPSI-IV; Number Matrices on the WJ IV ACH; and Concept Formation, Analysis–Synthesis, and Number Series on the WJ IV COG.

Response Inhibition

Response inhibition is the ability to withhold responding to distractor stimuli while focusing on target stimuli. Response inhibition tasks typically involve the ability to inhibit a response after a particular response set has been established. For example, in the classic Stroop test, the examinee is asked to name the color of the ink that a color word (e.g., *red*, *green*, or *blue*) is printed in, rather than reading the word itself.

The neuropsychological tests designed to measure response inhibition are Color Word Interference—Condition 3 (Inhibition) on the D-KEFS; the Inhibition—Condition 2 (Inhibition) and Statue tests on the NEPSY-II; and the numerous versions of the Stroop Color–Word Test (e.g., Golden & Freshwater, 2002). The only measure of cognitive ability that includes a response inhibition task is Expressive Attention on the CAS2, which is modeled after the Stroop test.

Facilitators–Inhibitors of Cognitive Processes

In the integrated SNP/CHC model, Miller (2013) introduced a broad classification called *facilitators–inhibitors,* which included the second-order classifications: (1) allocating and maintaining attention; (2) working memory; and (3) speed, fluency, and efficiency of processing. These three processes are thought either to facilitate or to inhibit other cognitive processes and acquired knowledge skills. For example, learning and memory can be helped or hurt by good or poor attentional skills, respectively. The quality of a person's working memory skills has a major positive or negative impact on many aspects of academic achievement or acquired knowledge (Dehn, 2015). Processing speed, fluency, and efficiency also influence other cognitive processes and acquired knowledge in a positive or negative manner. In what follows, we discuss how neuropsychological tests and tests of cognitive ability assess these facilitators and inhibitors.

Attentional Processes

Attentional processes and executive functions are often intertwined, and some neuropsychological tests such as the NEPSY-II combine these processes into a single domain for interpretation. In Miller's (2013) integrated SNP/CHC model, attentional processes are interpreted separately from executive functions. Consistent with prevailing neuropsychological theories of attention (see Riccio, Reynolds, & Lowe, 2001, for a review), the integrated SNP/CHC model views attention as multidimensional, with two major subcomponents: (1) selective/focused attention, and (2) sustained attention. A third subcomponent included in the model (Miller, 2013) is attentional capacity, which is a duplication of immediate memory tasks with the interpretive focus placed on the attentional load factor of the tasks. Miller (2010) has stated, "While attention is most probably multidimensional, many of the tests that are designed to measure attention . . . do not isolate the subcomponents of attention very well. Many of the common tasks of attention measure, as one unit, multiple subcomponents of attention such as selective and sustained attention" (pp. 95–96).

It is understood that some aspect of attentional processing is a basic requirement to perform almost any cognitive task. As an example, it is difficult to perform well on a memory task if attention is lacking during the encoding process. However, several current tests of cognitive ability are specifically designed to measure one or more of the attentional subcomponents (see Table 33.6).

Selective and/or Sustained Attention

Traditional neuropsychological measures of selective and sustained attention include the d2 Test of Attention (Brickenkamp & Zilmer, 1998) and the Ruff 2 & 7 Selective Attention Test (Ruff & Allen, 1996). Both of these tests require the examinee to choose target stimuli quickly from a visual array. The NEPSY-II Auditory Attention and Response Set tests measure aspects of selective, sustained, and shifting attention. Continuous-performance tests (CPTs) measure sustained attention and are used by both neuropsychologists and school psychologists in the diagnosis of attention-deficit/hyperactivity disorder (ADHD). Examples of CPTs include the Conners Continuous Performance Test 3rd Edition (Conners, 2014) and the Integrated Visual and Auditory Continuous Performance + Plus Test (Sandford & Turner, 1993–2010).

TABLE 33.6. Attentional Processes Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
Selective and sustained attention	• d2 Test of Attention • Ruff 2 & 7 Selective Attention Test • Continuous-performance tests • NEPSY-II: Auditory Attention and Response Set	• CAS2: Number Detection and Receptive Attention • WJ IV COG: Pair Cancellation

Several tests of cognitive ability have included measures of selective and sustained attention. The CAS2 has two tests, Number Detection and Receptive Attention, and the WJ IV COG includes the Pair Cancellation test.

Attentional Capacity

Attentional capacity has a direct relationship with the cognitive capacity or load required on memory tasks. As the stimuli to be recalled increase in length (e.g., in number of digits or letters), and as the semantic loading increases from words to sentences to stories, there are concurrent changes in the attentional demands of the tasks. Figure 33.2 illustrates the relationship of attentional capacity to memory tasks. Attentional capacity, or immediate short-term memory for numbers, letters, or visual sequences, is measured on common tasks in both neuropsychological tests and tests of cognitive functions.

Two classic neuropsychological visual span tests are the Corsi Block Span test (Milner, 1971) and the Knox Cube Test (Arthur, 1947; Knox, 1914). Only one cognitive ability test, the CAS2, includes a Visual Digit Span test. Two tests of memory and learning for children include measures of digit or visual sequence spans: Digits Forward on the TOMAL-2 and the Finger Windows and Number/Letter tests on the WRAML2. Digit span tests are routinely included in tests of cognitive ability, such as the Recall of Digits—Forward test on the DAS-II; the Number Recall test on the KABC-II; and the Digit Span test on the WISC-V.

Attentional capacity and contextual cues increase with memory for words and sentences. Among neuropsychological assessment tasks, the NEPSY-II and the WRAML2 both have sentence memory tests. There are similar memory-for-words tasks on three tests of cognitive ability: Word Series on the CAS2; Word Order (without interference) on the KABC-II; Memory for Words on the WJ IV COG; and Sentence Repetition on the WJ IV OL.

Finally, attentional capacity is maximized with increased cognitive load and semantic content with memory for stories. The three major standalone tests of memory and learning—the CMS, the TOMAL-2, and the WRAML2—all contain tests that assess memory for stories. The WJ IV COG is the only major test of cognitive ability that has a Story Recall test.

Working Memory

Working memory can be assessed according to the input modality, verbal or visual. Table 33.7 presents examples of tests designed to measure verbal working memory included in neuropsychological batteries (e.g., the NEPSY-II) or tests of memory and learning (e.g., the CMS, TOMAL-2, or WRAML2). All of these measures have in common the neurocognitive processing demand of active manipulation of information in immediate memory, which is the core requirement of a working memory task. An example of a working memory task is recalling a string of numbers or letters in reverse order after hearing them presented sequentially.

	Neuropsychological Measures	Neuropsychological Measures in Tests of Cognitive Abilities	
Memory for numbers, letters, or visual sequences	• Corsi Block Span test • Knox Cube Test • TOMAL-2: Digits Forward and Letters Forward • WRAML2: Finger Windows, Number/Letter	• CAS2 Brief: Successive Digits • DAS-II: Recall of Digits—Forward • KABC-II: Number Recall • WISC-V: Digit Span • WISC-V Integrated: Spatial Span Forward	↓ Increase in Attentional Capacity and Cognitive Load
Memory for words and sentences	• NEPSY-II: Sentence Repetition • WRAML2: Sentence Memory	• KABC-II: Word Order (without interference) • WJ IV COG: Memory for Words • WJ IV OL: Sentence Repetition	
Memory for stories	• NEPSY-II: Narrative Memory • TOMAL-2: Memory for Stories • WRAML2: Story Memory	• WJ IV COG: Story Recall	

FIGURE 33.2. Relationship of attentional capacity to memory tasks in the integrated SNP/CHC model.

TABLE 33.7. Working Memory Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
Verbal working memory	• CMS: Numbers Backwards and Sequences • NEPSY-II: Word List Interference—Recall • TOMAL-2: Digits Backwards and Letters Backwards • WRAML2: Verbal Working Memory	• DAS-II: Recall of Digits—Backward and Recall of Sequential Order • KABC-II: Word Order (with color interference effect) • SB5: Verbal Working Memory • WISC-IV: Arithmetic, Digit Span—Backward, and Letter–Number Sequencing • WJ IV COG: Numbers Reversed, Object–Number Sequencing, and Verbal Attention
Visual working memory	• WRAML2: Symbolic Working Memory	• SB5: Nonverbal Working Memory

The major tests of cognitive ability that include verbal working memory tests are Recall of Digits—Backward and Recall of Sequential Order on the DAS-II; Word Order (with color interference effect) on the KABC-II; Verbal Working Memory on the SB5; Arithmetic, Digit Span—Backward, and Letter–Number Sequencing on the WISC-V; and Numbers Reversed, Object–Number Sequencing, and Verbal Attention on the WJ IV COG.

There are very few tests designed to measure visual working memory. The WRAML2 provides a measure of visual working memory with its Symbolic Working Memory subtest. The Nonverbal Working Memory test (Block Tapping) on the SB5 is the only test on a major cognitive ability battery designed to measure visual working memory. The SB5 Nonverbal Working Memory test starts out measuring immediate visual memory capacity, but then shifts to measuring visual working memory at later levels, which makes overall classification of the test difficult.

Speed, Fluency, and Efficiency of Processing

In the current conceptualization of the integrated SNP/CHC model, speed, fluency, and efficiency of processing are classified together as a type of facilitator–inhibitor. This overall factor is conceptualized as having four major second-order subcomponents: (1) performance fluency, (2) retrieval fluency, (3) acquired knowledge fluency, and (4) accuracy as a function of speed (see Figuree 33.3).

Performance Fluency

Performance fluency is defined as "the ability to quickly perform simple, repetitive tasks" (Miller, 2013, p. 399). There are several types of performance fluency: (1) psychomotor fluency, (2) perceptual fluency, (3) figural fluency, (4) naming fluency, and (5) oral–motor fluency.

Psychomotor fluency tasks require rapid motor output. Table 33.8 presents examples of neuropsychological tests designed to measure psychomotor fluency, including Trail Making—Condition 5 (Motor Speed) from the D-KEFS and the Visuomotor Precision test (Total Completion Time) from the NEPSY-II. The only test from a cognitive ability battery designed to assess performance fluency is Coding Copy from the WISC-V Integrated.

Perceptual fluency is defined as the ability to quickly distinguish similar but different visual patterns and maintain attention under timed conditions (Horn & Blankson, 2012). There are no widely used neuropsychological measures that are specifically designed to measure perceptual fluency. However, Table 33.8 lists many tests from cognitive ability batteries that are designed to measure perceptual fluency.

Figural fluency refers to the ability to connect dots with unique line patterns while following rules and performing under time constraints. Two neuropsychological tests measure this construct: the Design Fluency tests on the D-KEFS and the NEPSY-II. There are no comparable tests on batteries of cognitive ability that measure figural fluency.

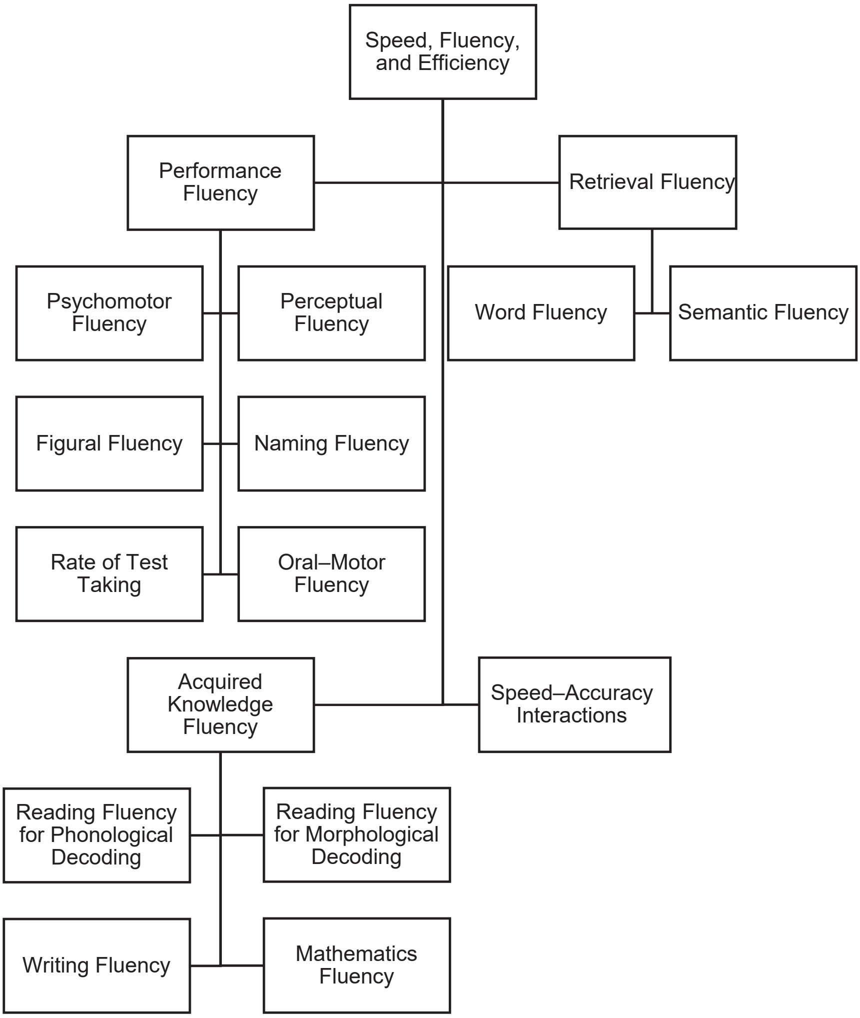


FIGURE 33.3. Speed, fluency, and efficiency of processing in the integrated SNP/CHC model.

Naming fluency is often referred to as rapid automatized naming (RAN). These tasks require an examinee to name common objects, colors, word, or letters quickly. The Rapid Automatized Naming and Rapid Alternating Stimulus Tests (RAN/RAS; Wolf & Denckla, 2005) are considered to be the gold standards in neuropsychological tests designed to measure naming fluency. Other neuropsychological tests that measure naming fluency include Color Word Interference—Conditions 1 and 2 from the D-KEFS and Speed Naming from the NEPSY-II. Several cognitive ability tests are also designed to measure naming fluency, including Rapid Naming on the DAS-II; Naming Speed Literacy and Naming Speed Quantity on the WISC-V; and Rapid Picture Naming on the WJ IV COG.

Oral–motor fluency tasks require the examinee to repeat words that are not real words quickly. The NEPSY-II has two tests, Oral Motor Sequences and Repetition of Nonsense Words, that require oral–motor fluency. Several achievement test batteries include tests that are designed to measure oral–motor fluency, but tests of cognitive ability do not include these types of measures.

Retrieval Fluency

Retrieval fluency is the ability to recall information quickly and accurately from long-term mem-

TABLE 33.8. Speed, Fluency, and Efficiency of Processing Measured on Traditional Neuropsychological Tests and Current Tests of Cognitive Abilities

Neuropsychological construct(s)	Traditional neuropsychological measures	Examples of neuropsychological measures in tests of cognitive abilities
	Performance fluency	
Psychomotor fluency	• D-KEFS: Trail Making—Condition 5 (Motor Speed) • NEPSY-II: Visuomotor Precision—Total Completion Time	• WISC-V Integrated: Coding Copy
Perceptual fluency		• CAS2: Planned Number Matching • DAS-II: Speed of Informational Processing • WISC-V: Cancellation • WPPSI-IV: Cancellation • WISC-V: Coding and Symbol Search • WPPSI-IV: Animal Coding and Bug Search • WJ IV COG: Number–Pattern Matching
Figural fluency	• D-KEFS: Design Fluency • NEPSY-II: Design Fluency	
Naming fluency	• D-KEFS: Color Word Interference—Conditions 1 and 2 • NEPSY-II: Speed Naming • RAN/RAS	• DAS-II: Rapid Naming • WISC-V: Naming Speed Literacy and Naming Speed Quantity • WJ IV COG: Rapid Picture Naming
Oral–motor fluency	• NEPSY-II: Oral Motor Sequences and Repetition of Nonsense Words	
	Retrieval fluency	
Word and semantic fluency	• D-KEFS: Verbal Fluency—Condition 1 (Letter Fluency) and Condition 2 (Category Fluency) • NEPSY-II: Word Generation—Initial Letter Total and Semantic Total	• WJ IV COG: Retrieval Fluency

ory. According to Miller (2013), the efficiency of retrieval strategies is what makes retrieval fluency an executive function, rather than just a measure of long-term memory. On neuropsychological tests, retrieval fluency is measured by using either verbal or nonverbal stimuli. Typically, verbal retrieval fluency tasks require the examinee to recall words that start with a particular letter or fit within a specific semantic category (e.g., food, pieces of furniture); nonverbal fluency tests usually require the examinee to generate unique designs or patterns quickly, using structured or unstructured visual arrays.

The neuropsychological tests that measure verbal retrieval fluency are Verbal Fluency (Conditions 1 and 2) on the D-KEFS and Word Generation on the NEPSY-II (see Table 33.8). Nonverbal retrieval fluency measures include the Design Fluency test on both the D-KEFS and the NEPSY-II. The only test of cognitive ability with a measure of retrieval fluency is the WJ IV COG.

Acquired Knowledge Fluency

Acquired knowledge fluency represents the automaticity of processing for rapid reading, writing, and math calculations. Neuropsychological and cognitive ability measures do not contain these types of tests, which are typically found on tests of achievement.

Speed–Accuracy Interactions

The interaction between speed and accuracy is an important consideration when assessment results are interpreted. Any time a test generates separate scores for completion time and accuracy (either the number of errors or the number correct), the speed–accuracy interaction can be examined. The best response on a test is average to above completion time with good accuracy. Some examinees rush through items, and accuracy suffers. Other examinees have learned to slow down to improve accuracy. Finally, some examinees with the most impairment are very slow and inaccurate. The following neuropsychological tests offer these types of comparisons: the D-KEFS Color Word Interference Test and the NEPSY-II Speeded Naming, Visuomotor Precision, and Inhibition tests. Perhaps the closest test on a cognitive ability battery designed to measure this same construct would be the WISC-V Block Design test with and without the time bonus.

CONCLUSIONS AND FUTURE DIRECTIONS

The major premise of this chapter—that the integration of neurocognitive constructs into tests of cognitive ability is somehow a new or emerging development—is false. In fact, many of these neurocognitive constructs have been found all along in tests of cognitive ability. Early researchers in cognitive assessment (e.g., Binet, Norsworthy, Terman, Wechsler, and Woodcock) were interested in measuring such cognitive constructs as processing speed, memory, or executive functions, but advances in psychometrics and test development led to an emphasis on empirical measurement within the field of cognitive assessment. As a result, there was an interpretive overemphasis on a singular global score as the best measure of intellectual functioning or the best predictor of academic achievement. Other developments within the field of psychology were also influential in the evolution of cognitive assessment. For example, as psychology moved away from the medical model, training programs in school, counseling, and clinical psychology gave less emphasis to the biological bases of behavior. For the most part, training programs in neuropsychology were the exceptions in this trend. In addition, because of the natural expansion in the knowledge base within the field (as well as a certain amount of territorialism/protectionism), psychology became more specialized or fractured, and the transfer of knowledge between related disciplines became less common. In cognitive assessment, this led to a narrow lens through which cognitive abilities were viewed and interpreted.

In actuality, there is far more integration between the disciplines of neuropsychology and cognitive assessment than many professionals realize. For example, many current tests of cognitive ability are composed of tasks that measure commonly identified neurocognitive constructs. In addition, current trends in interpretation of intellectual abilities are more theoretically driven than in the past and are being influenced by research on cognitive constructs of intelligence, which lends itself to a common nomenclature with the field of neuropsychology. Finally, the surge of interest in the neurosciences resulting from remarkable findings in brain science has generated a corresponding surge of research in the applicability of neuropsychology to cognitive assessment, bringing the field back to its historical origins.

To progress as a discipline, the next major advance in the integration of neuroscience and cognitive assessment has to be the use of neuroimaging techniques to validate associations between cognitive constructs and neurological functioning or brain structure, and then to validate the applicability of the various tasks or subtests being used to measure these neurocognitive constructs. This research should lead to more definitive conclusions about the connection between theoretical cognitive constructs and brain functioning, as well as to better identification of neurocognitive deficits with more targeted and applicable measures. Assessing neurocognitive functioning more accurately and parsimoniously should ideally lead to greater applicability for recommendations and interventions to address the manifestation of neurocognitive deficits in real-world settings.

In the meantime, best practice in assessment would suggest that individuals conducting cognitive and neurocognitive assessments need to (1) realize that current cognitive instruments are limited in their comprehensiveness, measuring only limited aspects of cognitive or neurocognitive constructs, and interpret them with this in mind; (2) understand that current cognitive instruments are best used as screeners for neurocognitive functioning, and that further assessment with more targeted measures (in the use of which they must be trained) may be required in the presence of suspected neurocognitive deficits; (3) recognize that a strong foundation or knowledge base in the neurosciences is necessary to use and interpret cognitive measures correctly from a neuropsychological perspective; and (4) appreciate the need for a theoretically based interpretive framework (Luria's theory, CHC theory, the integrated SNP/CHC model, Flanagan and colleagues' integrative interpretive model) within which to draw their conclusions.

The disciplines of cognitive assessment and neuropsychology have been, are, and will continue to be intertwined. Cognition, by definition, is brain-based; therefore, the assessment of cognition needs to be a brain-based endeavor. As the

21st century progresses, these two fields will evolve together, and perhaps psychologists in the future will express surprise that they were ever considered separate fields of study.

REFERENCES

Adams, W., & Sheslow, D. (1995). *Wide Range Assessment of Visual–Motor Abilities*. Wilmington, DE: Wide Range.

Arthur, G. (1947). *A point scale of performance tests (Rev. Form II)*. New York: Psychological Corporation.

Baron, I. S. (2004). *Neuropsychological evaluation of the child*. New York: Oxford University Press.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Neuropsychology, 27*(3), 362–373.

Beery, K. E., Buktenica, N. A., & Beery, N. A. (2010). *Beery–Buktenica Developmental Test of Visual–Motor Integration, Sixth Edition*. Minneapolis, MN: Pearson.

Benton, A. L., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Blair, C. (2006). Towards a revised theory of general intelligence: Further examination of fluid cognitive abilities as a unique aspect of human cognition. *Behavioral and Brain Sciences, 29,* 145–153.

Bracken, B. A., & McCallum, R. S. (2016). *Universal Nonverbal Intelligence Test—Second Edition*. Austin, TX: PRO-ED.

Brickenkamp, R., & Zilmer, E. (1998). *d2—Test of Attention*. Ashland, OH: Hogrefe & Huber.

Buchanan, R. W., & Heinrichs, D. W. (1989). The Neurological Evaluation Scale (NES): A structured instrument for the assessment of neurological signs in schizophrenia. *Psychiatry Research, 27,* 335–350.

Christensen, A.-L. (1975). *Luria's neuropsychological investigation: Text*. New York: Spectrum.

Cohen, M. J. (1997). *Children's Memory Scale*. San Antonio, TX: Psychological Corporation.

Conners, C. K. (2014). *Conners Continuous Performance Test 3rd Edition*. North Tonawanda, NY: Multi-Health Systems.

Dean, R. S., & Woodcock, R. W. (2003). *Dean–Woodcock Neuropsychological Battery*. Itasca, IL: Riverside.

Decker, S., & Davis, A. (2010). Assessing and intervening with children with sensory–motor impairments. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 673–692). Hoboken, NJ: Wiley.

Dehn, M. J. (2015). *Essentials of working memory assessment and intervention*. Hoboken, NJ: Wiley.

Delis, D., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System: Examiner's manual*. San Antonio, TX: Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *California Verbal Learning Test—Children's Version*. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition*. San Antonio, TX: Harcourt Assessment.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172–179.

Golden, C. J., & Freshwater, S. M. (2002). *Stroop Color–Word Test*. Odessa, FL: Psychological Assessment Resources.

Grant, D. A., & Berg, E. A. (1993). *Wisconsin Card Sorting Test: Manual*. Odessa, FL: Psychological Assessment Resources.

Gualtier, C. T., & Johnson, L. G. (2006) Reliability and validity of a computerized neurocognitive test battery: CNS Vital Signs. *Archives of Clinical Neuropsychology, 21*(7), 623–643.

Heaton, R. K., & PAR Staff. (2003). *Wisconsin Card Sorting Test Computer Version 4—Research Edition*. Odessa, FL: Psychological Assessment Resources.

Hooper, H. E. (1958). *Hooper Visual Organization Test: Manual*. Los Angeles: Western Psychological Services.

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73–98). New York: Guilford Press.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. §1400 et seq. (2004).

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). *WISC-III PI: Manual*. San Antonio, TX: Psychological Corporation.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition*. Circle Pines, MN: American Guidance Service.

Knox, H. A. (1914). Mental defectives. *New York Medical Journal, 99,* 215–222.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment*. San Antonio, TX: Harcourt Assessment.

Luciana, M. (2003). Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). *Journal of Child Psychology and Psychiatry, 44*(5), 649–663.

Maricle, D. E., Johnson, W., & Avirett, E. (2010). Assessing and intervening in children with executive function disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 599–640). Hoboken, NJ: Wiley.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial.* Odessa, FL: Psychological Assessment Resources.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment.* Hoboken, NJ: Wiley.

Miller, D. C. (2010). School neuropsychological assessment and intervention. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 81–100). Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D. C. (2015). Integration of the WJ IV, WISC-V, WISC-V Integrated, and WIAT-III into a school neuropsychological assessment model. *School Psychology Forum: Research in Practice, 9*(3), 251–268.

Miller, D. C. & Maricle, D. E. (2012). The emergence of neuropsychological constructs into tests of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 800–819). New York: Guilford Press.

Miller, D. C., McGill, R. J., & Bauman Johnson, W. L. (2016). Neurocognitive applications of the Woodcock–Johnson IV. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 355–388). San Diego, CA: Academic Press.

Milner, B. (1971). Interhemispheric differences in localization of psychological processes in man. *British Medical Bulletin, 27,* 272–277.

Naglieri, J. A., & Das, J. P. (1997). *Das–Naglieri Cognitive Assessment System.* Itasca, IL: Riverside.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition.* Austin, TX: PRO-ED.

Osterrieth, P. A. (1944). Le test de copie d'une figure complex: Contribution a l'étude de la perception et de la mémoire. *Archives de Psychologie, 30,* 286–356.

Powell, D., Kaplan, E., Whitla, D., Weintraub, S., Catlin, R., & Funkenstein, H. (2004). *MicroCog: Assessment of Cognitive Functioning Windows Edition 2004.* San Antonio, TX: Harcourt Assessment.

Reitan, R. M., & Davidson, L. A. (Eds.). (1974). *Clinical neuropsychology: Current status and applications.* Washington, DC: Winston.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation.* Tucson, AZ: Neuropsychological Press.

Rey, A. (1941). L'examen psychologique dan les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 286–340.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Reynolds, C. R., & Voress, J. K. (2007). *Test of Memory and Learning—Second Edition.* Austin, TX: PRO-ED.

Riccio, C. A., Reynolds, C. R., & Lowe, P. A. (2001). *Clinical applications of continuous performance tests: Measuring attention and impulsive responding in children and adolescents.* New York: Wiley.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Ruff, R. M., & Allen, C. C. (1996). *Ruff 2 & 7 Selective Attention Test.* Odessa, FL: Psychological Assessment Resources.

Sandford, J. A., & Turner, A. (1993–2010). *Integrated Visual and Auditory Continuous Performance + Plus Test.* Richmond, VA: BrainTrain.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock–Johnson IV Tests of Oral Language.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Sheslow, D., & Adams, W. (2003). *Wide Range Assessment of Memory and Learning, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition.* San Antonio, TX: Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Wechsler, D., & Kaplan, E. (2015). *Wechsler Intelligence Scale for Children—Fifth Edition Integrated.* Bloomington, MN: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Morris, E., Kramer, J. H., & Maerlender, A. (2004). *The Wechsler Intelligence Scale for Children—Fourth Edition Integrated: Technical and interpretative manual.* San Antonio, TX: Harcourt Assessment.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7,* 131–149.

Wepman, J. M., & Reynolds, W. M. (1986). *Wepman's Auditory Discrimination Test—Second Edition.* Los Angeles: Western Psychological Services.

Wolf, M., & Denckla, M. B. (2005). *Rapid Automatized Naming and Rapid Alternating Stimulus Tests.* Austin, TX: PRO-ED.

CHAPTER 34

Enhancing the Identification of Reading Disorders with Neuropsychological Tests

Marlene Sotelo-Dynega

A survey regarding the cognitive assessment practices of school psychologists (Sotelo-Dynega & Dixon, 2014) found that fewer than 40% of those sampled *always* interpreted the general/global ability score, and that "the majority of respondents reported that they typically used more than one test of intelligence or cognitive abilities during evaluations, mostly to follow up on hypotheses and/or aberrant score performances based on the results of the intelligence tests that were administered" (p. 1036). These data suggest that school psychologists may have begun to move away from traditional psychoeducational assessment methods that involve the interpretation of global ability scores, and toward more specific, diagnostic testing. According to Decker, Hale, and Flanagan (2013),

> Research has demonstrated that cognitive processes in many academic learning tasks are surprisingly complex. Most academic learning tasks require general cognitive processes such as attention, language, and memory. As such, low scores on any academic task could be the result of numerous factors. Not surprisingly, academic learning often involves numerous cognitive processes as well as cultural learning opportunities. Administering specific narrowly defined cognitive tasks helps to systematically isolate the underlying problem. The purpose of administering the tests is not to get an IQ score, but to help understand learning problems . . . and develop effective interventions. . . . This is an important distinction from viewing testing for the purpose of predicting achievement . . . (p. 303)

In fact, Decker and colleagues add that "school psychologists are in a unique position to take the lead role in the translation of brain research, cognitive neuroscience, and neuropsychological research related to neurodevelopmental disabilities into educational practice" (p. 303).

To bridge the gap between the literature and educational practice, school psychologists will have to expand upon their traditional, standard-battery-based testing approaches to include more comprehensive cognitive, neuropsychological, and diagnostic academic achievement measures. In 2010, a panel of experts on learning disabilities published a white paper that summarized their consensus opinions regarding the identification of specific learning disabilities (SLD), along with citations to support the rationales for their beliefs (Hale et al., 2010). Hale and colleagues (2010) stated that

> a strengths and weaknesses model makes good empirical, clinical, and legal sense because it ensures children identified with SLD demonstrate one or more processing deficits that interfere with academic achievement, the core characteristic of SLD . . . Not only does this processing strength and weaknesses approach make sense for SLD identification purposes, but processing assessment could also lead to more

> effective individualized interventions for children who do not respond adequately to intensive interventions in an RTI approach . . . (p. 229)

Considering that traditional assessment methods typically involve the use of a standard testing battery (i.e., the same tests are used regardless of the reason for a child's referral), a practitioner may sometimes have to use a more comprehensive cognitive or neuropsychological battery to identify a student's processing weaknesses. Although the most recent versions of the most popular intelligence/cognitive tests (the Wechsler scales, the Woodcock–Johnson IV [WJ IV], etc.) are more comprehensive than ever before, they are not necessarily exhaustive and lack comprehensive measures of neuropsychological processes. Therefore, once a detailed reason for testing is established, the practitioner should select an individualized battery of tests to assess the abilities or processes that have been found to be associated with the reported weaknesses. For example, if the child is having difficulty concentrating, then measures of attentional processes should be added to the battery. Although this approach to test selection is an ideal practice, is it practical or even reasonable to suggest that practitioners engage in more testing, given that school psychologists already spend nearly 50% of their time engaged in assessment (Castillo, Curtis, & Gelley, 2010a, 2010b; Fagan & Wise, 2007; Sotelo-Dynega & Dixon, 2014)? The response to this question of course depends on the practitioners themselves, their professional goals, and the available funding. To actualize this shift in practice, many school psychologists will have to request that additional tests be purchased and will have to seek additional training in the administration and interpretation of any new measures; both of these require additional financial resources.

To help facilitate this transition for practicing school psychologists, the remainder of this chapter describes a user-friendly framework that can be used with students who present with specific reading issues.

FRAMEWORK FOR IMPROVING THE IDENTIFICATION OF READING DISORDERS

Background Information

When a student is referred for an evaluation to rule out a reading disorder, the first step is typically to confirm that the reported reading problem exists (Flanagan, Alfonso, & Mascolo, 2010). Data obtained from the student's cumulative records, teacher and parent reports, and response to intervention (RTI) are critical pieces of information for the school psychologist to review in the identification process and can help to inform test selection and diagnostic decision making. With these data, the school psychologist can generate hypotheses regarding which cognitive or neuropsychological constructs may be related to the reported academic weaknesses, and can begin to build an individualized assessment battery. For example, if the student is struggling with decoding, deficits in phonological processing may be suspected, and therefore phonological and auditory processing should be assessed.

Unfortunately, some schools may not have formal pre-referral procedures in place. Therefore, the exact academic weaknesses may be unknown at the time of referral and may require a deeper investigation of the student's background and academic history information. Regardless of the background data available at the time of the referral, the practitioner is encouraged to complete a thorough background history with the student's parents. The Behavior Assessment System for Children, Third Edition (Reynolds & Kamphaus, 2015), includes a Structured Developmental History; this is an excellent tool that comprehensively covers an abundance of relevant domains. Information from a thorough background interview can assist in generating hypotheses in the absence of concrete academic data.

Another useful tool for hypothesis generation is Miller's (2012) Neuropsychological Processing Concerns Checklist for School-Aged Children and Youth, Third Edition (NPCC-3). The NPCC-3 is a multipage checklist that is completed by the student's parents and teachers. Each respondent is asked to rate the severity of the student's functioning in several neuropsychological and academic domains: sensory–motor, visual–spatial, and auditory processes; learning and memory, executive functions, attention, and working memory; speed, fluency, and efficiency of cognitive processing; language, reading, and writing; and mathematics. Furthermore, the NPCC-3 provides the respondent with the opportunity to describe specific concerns in each of these domains. The information that can be obtained from this checklist can provide the practitioner with highly specific, qualitative information regarding the student's strengths and weaknesses that can guide test selection, hypothesis development, and approaches to intervention.

This instrument is not only invaluable for practitioners looking to incorporate neuropsychological testing into their assessment practices, but also for those who have received limited pre-referral data regarding the students that they will be evaluating.

Depending on the data gathered during this first step of the assessment, the use of additional rating scales may be warranted. For example, if the student is exhibiting symptomatology consistent with a diagnosis of attention-deficit/hyperactivity disorder (ADHD), the practitioner should most likely incorporate data from the administration of one or several relevant behavior rating scales. Whenever possible, it is best to obtain information about a student's behavior from several informants (e.g., parents, classroom teacher, reading specialist) representing different environments (e.g., home, classroom, resource room). Generally, the practitioner wants to assess whether or not the behavior is consistent across each setting. If the behavior is affecting the student's learning experience, a functional behavioral assessment may be warranted.

Assessment Planning

Although the suggestions listed in the previous step are relatively similar to those utilized during the traditional psychoeducational assessment process, assessment planning will most likely differ significantly. An interpretive approach developed by Flanagan and colleagues (2013) can be used to guide this phase of the evaluation process. The approach depicted in Figure 34.1 incorporates the "Cattell–Horn–Carroll (CHC) theory of human cognitive abilities, neuropsychological, and Lurian perspectives" (p. 129) to guide practitioners in the development of a comprehensive assessment battery for students presenting with a variety of issues that affect learning. Although a detailed description of this approach is beyond the scope of this chapter (the reader is encouraged to refer to the Flanagan et al. book), it important to note that all of its components are essential in comprehensively assessing students, particularly those with reading difficulties. CHC theory is a "taxonomy of cognitive abilities . . . [and] a set of theoretical explanations of how and why people differ in [these] abilities" (Schneider & McGrew, 2012, p. 99), and it has been described as "the best validated model of human cognitive abilities" (Ackerman & Lohman, 2006, p. 140). Furthermore, the inclusion of neuropsychological perspectives arranged according to Luria's (1970) theory is important, considering the "wealth of research on the neurobiological bases of childhood learning disorders . . . [and] increased emphasis on the identification of processing disorders among children diagnosed with a SLD" (Miller, 2013, p. 1), as well as the "empirical research that strongly supports Luria's clinical documentation of the three functional units" (Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005, p. 6).

For the purposes of this chapter, each facet of Flanagan and colleagues' (2013) framework is briefly described, in order to provide an empirical rationale for assessing each of these domains. The primary aim is to highlight the most salient research currently available in each domain regarding the relationships between CHC and neuropsychological constructs and reading issues among school-age children. Practitioners must be careful not to "over-test" the students they are working with. Therefore, they are advised to gather sufficient background information (using the tools described in the previous section) to develop specific hypotheses to guide test selection. Once an initial set of tests is scored, a practitioner may have to reframe the hypotheses according to the findings and administer additional tests.

Lurian Block 1

According to Luria (1973), "human mental processes are complex functional systems . . . that take place through the participation of groups [blocks] of concertedly working brain structures, each of which makes its own particular contribution" (p. 43). The first block, as described by Luria, includes the functions of the brainstem. Some of the main tasks of the brainstem are to "regulate the individual's ability to maintain arousal and attention" (Kaufman et al., 2005, p. 6). The neuropsychological domains included in this block are attention, sensory–motor functions, and speed/efficiency; adequate functioning in these domains is essential for learning, as described in detail below.

Attention

Without attention, learning cannot take place. Students have to be alert, aware, attentive, and available to learn. Although multiple behavioral, emotional, and wellness-related factors can affect an individual's attentional capabilities and should be investigated and ruled out, brain-based attentional issues must also be considered. In a review of

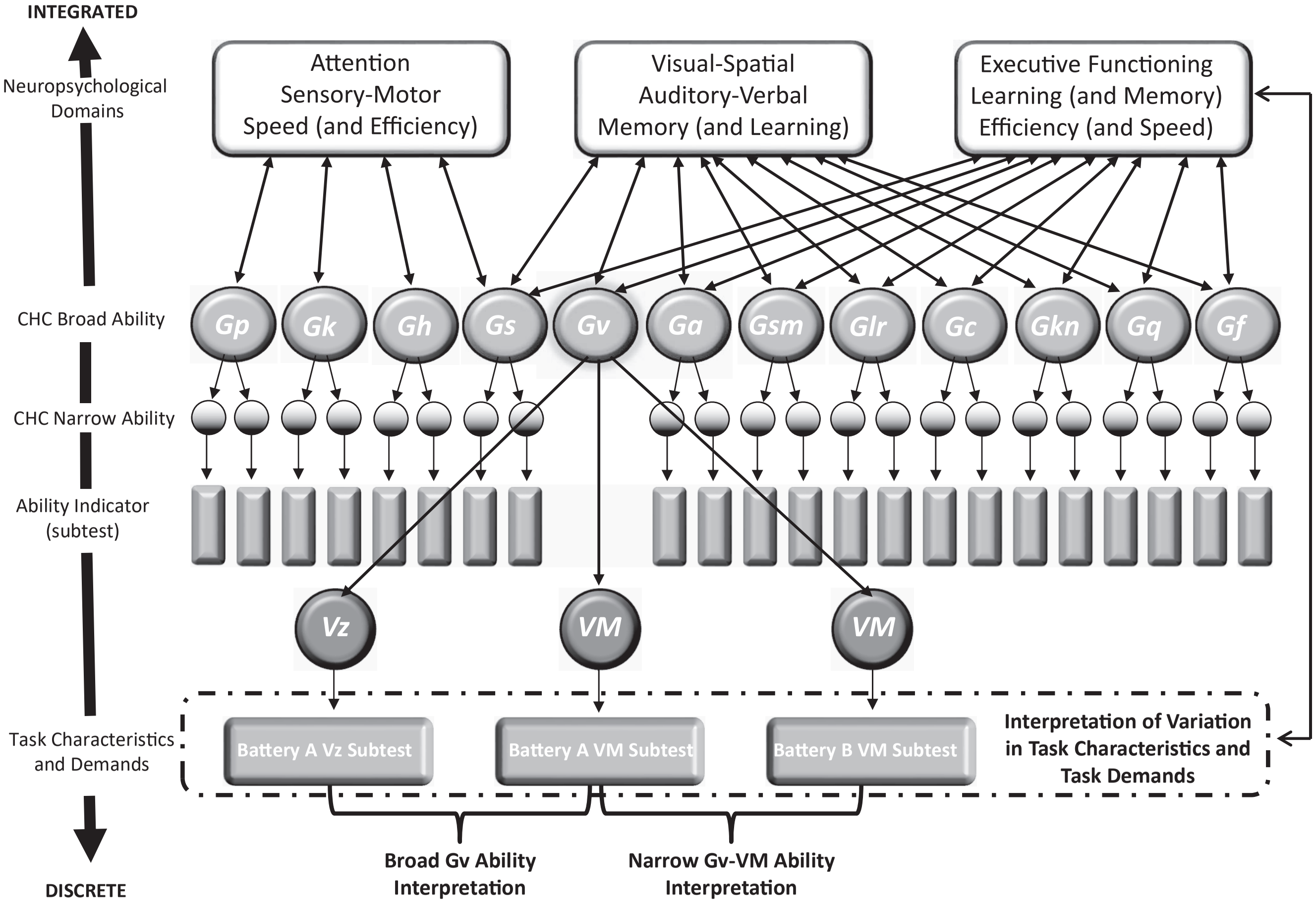


FIGURE 34.1. Integration of psychometric, neuropsychological, and Lurian perspectives for interpretation. From Flanagan, Ortiz, and Alfonso (2013, p. 129). Copyright © John Wiley & Sons, Inc. Reprinted by permission.

the literature on the neurobiology of reading and dyslexia, Shaywitz and Shaywitz (2008) conclude that "attentional mechanisms play a critical role in reading and that disruption of these attentional mechanisms play[s] a causal role in reading difficulties" (p. 1343). These mechanisms can affect every facet of the reading process, including the development of phonological and orthographic reading skills, fluency, and comprehension. Therefore, whenever attentional issues seem to exist for a student, a school psychologist should investigate whether such issues are a potential cause for the student's reading problem. The practitioner can gather this information via the procedures listed earlier in the discussion of background information, as well as via direct behavioral observation of the student's behavior. If necessary, the practitioner can then confirm or rule out attentional issues via standardized measures of attention.

Sensory–Motor Functions and Speed/Efficiency

The next domains to assess include the sensory and motor functions of the brain, as well as the individual's ability to process simple information quickly. Sensory and motor functions "serve as a baseline for all of the higher-order processes (e.g., visual–spatial processing, language skills, memory and learning)" (Miller, 2013, p. 261). For example, if a student has difficulties with visual acuity, or has a fine motor weakness, then all other tasks that involve those abilities will most likely be affected. Although practitioners are encouraged to assess sensory and motor functions, the results should be either confirmed or ruled out by specialists in each of these domains (e.g., ophthalmologists, audiologists, occupational or physical therapists).

According to Miller (2013), tests that measure "speed, fluency, and efficiency of processing [are] classified as a type of facilitator/inhibitor . . . almost all cognitive and behavioral tasks require some aspects of processing speed to increase the automaticity of responses. However, fast processing speed is not always a desired outcome as evidenced by the child who rushes through an assignment to get it done, but makes multiple errors in completing the task" (p. 399). Therefore, when assessing a student's processing speed, a practitioner should qualitatively evaluate how the student achieved the score: Were the responses quick and accurate, quick and inaccurate, slow and accurate, or slow and inaccurate? Considering that processing speed tasks (particularly measures of perceptual speed) have been found to be related to slower reading speed, by qualitatively analyzing how a student earns scores on these tasks, the school psychologist can generate hypotheses as to how the student approaches similar speeded tasks in the academic environment.

Tables 34.1, 34.2, and 34.3 have been developed to provide the reader with a concise, (although not exhaustive) summary of the literature regarding the relations between specific narrow cognitive/neuropsychological domains and reading-related issues. The purpose of these tables is to allow the reader to gauge the importance of including these constructs as part of a comprehensive school-based assessment when evaluating students with reading issues. Although the majority of narrow domains listed in these tables and their relationships to reading issues are supported by at least one study, I have also included some logical but not empirically supported hypotheses as to how low scores on measures of these narrow abilities might affect the reading process. These hypotheses are listed without citations.

Table 34.1 highlights the broad and narrow domains that are associated with Luria's Block 1 and the related domain-specific reading issues that can be manifested when weaknesses are present in those domains. Flanagan and colleagues (2013), Feifer (2010), and Miller (2013) provide various test suggestions for many of these constructs. The reader is urged to select measures that are current and psychometrically sound. Furthermore, if weaknesses are identified, best practices suggest that low scores be confirmed as representing true weaknesses by administering follow-up tests measuring the same abilities (Flanagan et al., 2013). If a suspected weakness is confirmed, the practitioner should also explain how this weakness has been manifested, or can potentially be manifested, in the academic setting (Miller, 2013).

Lurian Block 2

Luria's second block is responsible for "receiving, analyzing and storing information" (Luria, 1973, p. 67) that is projected onto the occipital, temporal and parietal lobes. Cognitive and neuropsychological abilities and processes that help determine the effectiveness and efficiency of this block include visual processing, auditory processing, short-term memory, and learning ability.

Visual Processing

According to Schneider and McGrew (2012), "visual processing can be defined as the ability to make use of simulated mental imagery (often

TABLE 34.1. Lurian Block 1 Domains and Possible Domain-Specific Reading Issues When Weaknesses Are Present

Broad domain	Narrow domain	Possible domain-specific reading issues when weaknesses are present
Attention	Selective/focused attention	• Difficulty remaining focused on reading tasks • Slower reading rate (Casco, Tressoldi, & Dellantonio, 1998) • Letter sequence errors and confusions (Yap & van der Leij, 1993) • Irregular-word reading difficulties (Valdois, Bosse, & Tainturier, 2004)
	Sustained attention	• Difficulty sustaining attention to task when reading lengthier passages • Weaknesses in reading comprehension (Stern & Shalev, 2013) • Decoding weaknesses (Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014)
	Attentional capacity	• Difficulties with reading comprehension (Engle, Carullo, & Collins, 1991; Leather & Henry, 1994; Seigneuric & Ehrlich, 2005)
Sensory functions	Auditory acuity	• Possible hearing and/or auditory processing issues • Phonologically based reading issues (e.g., reading decoding)
	Visual acuity	• Possible visual and/or visual processing issues • Visually/orthographically based reading issues (e.g. irregular word reading)
	Tactile sensation/ perception	• Difficulties with fine motor activities, and/or hyper- or hyposensitivity to certain sensations that could affect the student's attention and availability to learn (Miller, 2013)
Motor functions	Fine motor functions	• Limited vocabulary and reading speed (Viholainen et al., 2006)
	Visual–motor functions	• Impaired general reading performance (Kavale, 1982; Keogh & Smith, 1967; Mazzola Sortor & Taylor Kulp, 2003; Taylor Kulp, 1999)
	Visual scanning/ tracking	• Difficulty reading words printed on a line (Miller, 2013) • Text-reading issues (De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Diller, 1974)
Speed/ efficiency	Perceptual speed	• Slow reading speed (Flanagan, Ortiz, & Alfonso, 2013; Flanagan, Ortiz, Alfonso, & Mascolo, 2006) • Weaknesses in basic reading skills (Johnson, Humphrey, Mellard, Woods, & Swanson, 2010; McGrew & Wendling, 2010; Pennington, 2009) • Weaknesses in reading comprehension (McGrew & Wendling, 2010; Tiu, Thompson, & Lewis, 2003)
	Rate of test taking	• Weaknesses in reading comprehension (McGrew & Wendling, 2010)

in conjunction with currently perceived images) to solve problems" (p. 129). To date, there is no evidence to support the idea that visual processing constructs are related to the process of reading. The absence of evidence supporting a relationship between reading and visual processing may very well be the result of how we define the constructs subsumed by visual processing and how we measure them. According to Feifer (2010), "while reading may begin as a visual process, it ends as a linguistic one" (p. 492). Therefore, constructs such as orthographic processing may be more useful in the assessment of reading disorders. Orthographic processing is described as a broad term given to a group of constructs that enable (or inhibit) the acquisition of orthographic knowledge (Apel, 2011). Orthographic knowledge "represents information that is stored in memory that tells us how to represent spoken language in written form" (Apel, 2011, p. 592). Orthographic processing skills are critical in the acquisition of sight word vocabulary and reading fluency. Unfortunately, little is currently known about the cognitive/neuropsychological processes underlying the acquisition of these skills, which play such a large role in the reading process.

Auditory Processing

Auditory processing involves the "ability to detect and process meaningful nonverbal information in sound" (Flanagan et al., 2013, p. 404). This ability allows us to develop associations between letter sounds (phonemes) and their respective graphemes, which then leads to the process of decoding, or sounding out the words that we read. As presented in Table 34.2, extensive research has demonstrated that multiple constructs subsumed by auditory processing (not all listed in the table), particularly phonemic awareness, are highly related to the development of reading. As previously mentioned, considering that the majority of referrals school psychologists will receive are related to reading issues, it would be a disservice to any referred student to omit assessment of this construct during an evaluation. Unfortunately, the majority of cognitive measures available today

TABLE 34.2. Lurian Block 2 Domains and Possible Domain-Specific Reading Issues When Weaknesses Are Present

Broad domain	Narrow domain	Possible domain-specific reading issues when weaknesses are present
Visual–spatial processing	Visual–spatial perception	No literature available
	Visual–spatial reasoning	No literature available
Auditory processing	Sound discrimination	Weaknesses in basic reading skills (McGrew & Wendling, 2010)
	Auditory processing	• Weaknesses in general reading (Flanagan, Ortiz, Alfonso, & Mascolo, 2006) • Weaknesses in basic reading skills (McGrew & Wendling, 2010) • Weaknesses in reading comprehension (McGrew & Wendling, 2010) • Weaknesses in word reading (Compton, Olson, DeFries, & Pennington, 2002; Olson, Forsberg, Wise, & Rack, 1994; Hulslander et al., 2004; Kamhi & Pollock, 2005; National Reading Panel, 2000; Wagner & Torgesen, 1987)
Memory (and learning)	Verbal immediate memory	• Weaknesses in basic reading skills (Baddeley, 2006; De Jong, 2006; Hutton & Towse, 2001). • Weaknesses in reading comprehension (Just & Carpenter, 1992; Turner & Engle, 1989)
	Visual immediate memory	• Weaknesses in reading comprehension (McGrew & Wendling, 2010)
	Verbal long-term memory	No literature available
	Visual long-term memory	No literature available
	Verbal–visual associative memory	No literature available
	Working memory	• Difficulties with decoding (Palmer, 2000) • Weaknesses in reading comprehension (Cain, Oakhill, & Bryant, 2004; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; De Beni, Borella, & Carretti, 2007; De Jong, 2006; Goff, Pratt, & Ong, 2005; Hulme & Mackenzie, 1992; Just & Carpenter, 1992; Linderholm & Van Den Broek, 2002; Seigneuric & Ehrlich, 2005; Seigneuric. Ehrlich, Oakhill, & Yuill, 2000; Swanson, 2010; Swanson, Howard, & Saez, 2006)

do not include measures of this ability (Flanagan et al., 2013). Therefore, practitioners are encouraged to use a battery that includes this construct, to supplement their preferred assessment battery with measures of this construct, or to consult with a speech–language pathologist to ensure that they have assessed a student's functioning in this domain.

Memory and Learning

According to Miller and Blasik (2010), "memory and learning are very similar and can be easily confused . . . Memory is the ability to recall past events and information. Learning is a process of linking memories with new experience; therefore, memory is essential to learning and memory is dependent on learning" (p. 641). Dehn (2010) adds that "there is no learning without memory, and there is no memory without learning" (p. 4). Therefore, to thoroughly assess a student's functioning in this domain, the three main aspects of memory and learning must be evaluated: encoding, consolidation, and retrieval of information.

Encoding involves the pairing of new information obtained via short-term and working memory with already learned information that has been stored in long-term memory. Foundational abilities that include short-term and working memory must therefore be evaluated to determine whether or not the means by which information enters the brain are functioning as expected. As presented in Table 34.2, abilities such as verbal and visual immediate memory, as well as working memory, have been shown to be related to the development of basic reading skills and comprehension. Consolidation occurs when the new memories are stored in the brain, and retrieval involves efficient access to the memories that have been consolidated (Dehn, 2010). An individual's consolidation skills can be assessed via both immediate and delayed measures of verbal and visual associative memory. Typically, these types of tasks require the examinee to pair at least two nonrelated stimuli and to recall them immediately and after some time has elapsed, respectively. Unfortunately, empirical data to support the relationship of consolidation skills with the process of reading are lacking. Finally, the retrieval of information (discussed in more detail in the section on Lurian Block 3) involves the efficiency with which an individual is able to access the information that has been consolidated or stored in long-term memory. A lack of efficiency in any aspect of this process—encoding, consolidation, or retrieval—can significantly disrupt the learning process, which in turn can directly affect the different aspects of the reading process.

Lurian Block 3

According to Luria (1973), Block 3 involves the functions of the frontal lobes that include "programming, regulation and verification of activity" (p. 79). Some of the primary functions of this block are referred to as a group of skills known as *executive functions*. In addition to executive functions, other abilities/processes that are measured for this block include learning and memory, and crystallized knowledge.

Executive Functions

According to McCloskey and Perkins (2013), "executive functions can be viewed as an overarching developmental cognitive neuropsychological construct that is used to represent a set of neural mechanisms that are responsible for cueing, directing, and coordinating multiple aspects of perception, emotion, cognition, and action (Gioia, Isquith, Guy, & Kenworthy, 1996; McCloskey, Perkins, & Van Divner, 2009; Stuss & Alexander, 2000)" (p. 8). As presented in Table 34.3, various executive functions have been found to be related to different aspects of the reading process, but they primarily have to do with the more complicated aspects of reading that require higher-order thinking (e.g., reading comprehension). For example, measures of fluid reasoning typically involve assessing an individual's ability to solve problems with verbal or visual information. If this ability is not intact, the student may demonstrate considerable difficulties when asked to respond to questions about written text, especially if the questions are inferential in nature and cannot be immediately identified in the text.

Learning and Memory

As previously mentioned, learning and memory work simultaneously. When the abilities subsumed under this domain are classified, those related to memory processes are more in line with the brain functions associated with Lurian Block 2, whereas those related to the assessment of the retrieval of previously learned information are better classified in Lurian Block 3. Specifically, measures of abilities such as retrieval fluency and naming fa-

TABLE 34.3. Lurian Block 3 Domains and Possible Domain-Specific Reading Issues When Weaknesses Are Present

Broad domain	Narrow domain	Possible domain-specific reading issues when weaknesses are present
Executive functioning	Cognitive flexibility	• Problems related to language processing (Deák, 2000, 2003; Jacques & Zelazo, 2005) • Weaknesses in general reading ability (Arlin, 1981; Briggs & Elkind, 1973; Canter, 1975; Cohen, Hyman, & Battistini, 1983; Elkind, Larson, & Van Doorninck, 1965; Reiter, Tucha, & Lange, 2005) • Weaknesses in reading comprehension (Cartwright, 2002, 2006, 2007; Cartwright, Marshall, Dandy, & Isaac, 2010)
	Concept formation	No literature available
	Planning	• Weaknesses in reading comprehension (Keeler, 1995; Palincsar & Brown, 1984; Pearson & Fielding, 1991; Pressley, 2000; Reiter et al., 2005; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Tierney & Cunningham, 1984). • Weaknesses in single-word reading (Sesma et al., 2009)
	Fluid reasoning	• Weaknesses in reading comprehension (Flanagan, Ortiz, Alfonso, & Mascolo, 2006; McGrew, 1993; McGrew & Wendling, 2010).
	Response inhibition	• General reading weaknesses (Altemeier, Abbott, & Berninger, 2008; Palmer, 2000) • Weaknesses with reading comprehension (De Beni & Palladino, 2000; Savage, Cornish, Manly, & Hollis 2006; Savage, Lavers, & Pillay, 2007)
Learning and memory	Retrieval fluency	No literature available
	Naming facility	• Weaknesses in word reading (Ackerman & Dykman, 1993; Blachman, 1984; Bowers, 1995; Bowers & Swanson, 1991; Compton, 2000; Manis, Doi, & Bhadha, 2000; Manis, Seidenberg, & Doi, 1999; McBride-Chang & Manis, 1996; Meyer, Wood, Hart, & Felton, 1998; Scarborough, 1998; Schatschneider, Francis, Fletcher, & Foorman, 2002; Vellutino et al., 1996; Wagner, Torgesen, & Rashotte, 1994; Wolf, Bally, & Morris, 1986)
Crystallized knowledge	Language development	• Difficulties with reading comprehension (Catts, Fey, Zhang, & Tomblin, 2001; Dehn, 2008; Leach, Scarborough, & Rescorla, 2003; Nation, Marshall, & Snowling, 2001; Nation & Snowling, 1998, 1999; Was & Woltz, 2006)
	General verbal information	• Difficulties with basic reading skills (McGrew & Wendling, 2010) • Difficulties with reading comprehension (McGrew & Wendling, 2010)

cility (also known as rapid automatized naming) assess how efficiently the individual is able to retrieve information that has already been encoded and consolidated in long-term memory. Although there is no literature available to support the relation between ideational fluency and reading, a great deal of work has shown that those individuals with deficits and weaknesses in naming facility struggle with word reading. Typically, tasks that assess naming facility require the examinee to rapidly identify the names of letters, numbers, colors, or common objects. When asked to identify the name of a stimulus, students with reading impairments are often able to do so, but at a much slower pace than nonimpaired readers (Wolf, Grieg Bowers, & Biddle, 2000). This difficulty mimics the difficulties that poor readers experience when trying to automatically recognize a printed word. In this case, it is not necessarily true that the individuals have difficulties with the foundational aspects of word reading; rather, they struggle to automatically retrieve the names of words that have already been encoded and consolidated in their long-term memory.

Crystallized Knowledge

Crystallized knowledge is defined as "the depth and breadth of knowledge and skills that are valued by one's culture" (Schneider & McGrew, 2012, p. 122). The assessment of crystallized knowledge typically involves the measurement of language development and general verbal information. Language development entails a "general understanding of spoken language at the level of words, idioms, and sentences" (Flanagan et al., 2013, p. 389), and general verbal information is defined as "the breadth and depth of knowledge that one's culture deems essential, practical, or otherwise worthwhile for everyone to know" (Flanagan et al., 2013, p. 389). As presented in Table 34.3, individuals with weaknesses in these areas tend to struggle with tasks that require reading comprehension. Logically, to comprehend what is being read, a reader must first have an adequate understanding of language, and must also be able to draw from personal experiences to make sense of the content and respond to questions about it. Fortunately, all of the major cognitive tests available today include measures of crystallized knowledge (Flanagan et al., 2013); therefore, it will not be a challenge for school psychologists to incorporate this construct into their assessment batteries.

Administration, Scoring, and Interpretation of Tests

Administration

Considering that the majority of constructs included in Flanagan and colleagues' (2013) framework derive from CHC theory, school psychologists can easily assess many of these factors through the complete administration of one of the newly released cognitive batteries and their accompanying achievement tests. Practitioners must be cognizant of what the tests are actually measuring, however. For example, if a broad construct (e.g., visual processing, auditory processing) is to be adequately assessed, it is necessary for that construct to consist of at least two qualitatively different narrow indicators (Flanagan et al., 2013). In the event that a construct is not adequately assessed as part of a test battery, the practitioner is advised to supplement this battery with measures from another test. The reader is encouraged to refer to Flanagan and colleagues (2013) for detailed guidelines regarding how to organize a comprehensive battery that is consistent with CHC theory and how to interpret scores on clusters that are derived from different test batteries.

Once the practitioner has comprehensively evaluated the student's academic achievement and CHC-based cognitive abilities, additional tests may be necessary. For example, if the student has been referred for reading comprehension difficulties, measures of attention, memory and learning, and executive functioning may be warranted, according to the research summarized in Tables 34.1, 34.2, and 34.3. Typically, neuropsychological assessment involves the inclusion of tests representing all of the neuropsychological domains. The assessment strategy presented herein may actually be more practical for use by school psychologists because it will enhance the amount of information available to these practitioners without having to engage in cumbersome, unnecessary testing.

Scoring and Interpretation

Scoring procedures should not differ from the recommendations provided by test authors for each of the measures used in the assessment. Although test authors typically provide guidelines for score interpretation, it is my own preference to utilize an interpretive framework that takes into consideration whether a student's performances are within normal limits, or normatively strong or weak. Such information provides the practitioner with the ability to differentiate weaknesses from deficits. Table 34.4 displays a suggested guide to score interpretation that can be applied to all tests. All scores that are normatively weak (i.e., standard scores below 85) should be considered as deficits, and scores that are below average but within normal limits (i.e., standard scores of 85–89) should be considered to be weaknesses. This differentiation is important when practitioners are interpreting results because weaknesses warrant attention even when they do not constitute a disorder per se. Therefore, "because research demonstrates that the relationship between the cognitive dysfunction and the manifest learning problems are causal in nature . . . , data analysis at this level should seek to ensure that identified weaknesses or deficits on cognitive and neuropsychological tests bear an empirical relationship to those weaknesses or deficits on achievement tests identified previously" (Flanagan et al., 2013, p. 252). Therefore, when interpreting test results, the examiner must connect the results obtained from cognitive and neuropsychological tests to the achievement test results and to the data obtained from the real-world learning environment. Furthermore, suggestions for interventions, accommodations, and program

TABLE 34.4. Suggested Guide to Score Interpretation

Standard score/*scaled score*	Description of performance	Normative range
131 or above/*17 and above*	Upper extreme	Normative strength
116–130/*14–16*	Above average	Normative strength
85–115/*7–13*	Average	Within normal limits
70–84/*4–6*	Below average	Normative weakness
69 or below/*3 or below*	Lower extreme	Normative weakness

modifications should be made accordingly for all identified deficits and weaknesses.

Linking Results to Recommendations

Among the most important aspects of a comprehensive assessment are the recommendations made by the school psychologist. Regardless of the outcomes of the evaluation, the fact that the student was referred for an evaluation suggests that one or more issues have been occurring in the academic or home environment that warrant comprehensive assessment. Therefore, it is the school psychologist's responsibility to make suggestions aimed at improving the reason for the student's referral issue(s). Mascolo, Alfonso, and Flanagan (2014) present a very practical method for tailoring interventions, which they call a "systematic method of analyzing assessment results for tailoring interventions" (SMAARTI). The SMAARTI "involves the organization, analysis, and synthesis of assessment data to aid in understanding the cognitive basis of students' learning difficulties. Based on multiple data sources, the steps of the SMAARTI assist in identifying various methods of tailoring intervention" (pp. 4 and 6). Mascolo and colleagues provide the reader with guidelines regarding when (and how) to make suggestions for modifications, accommodations, remediation, and compensation.

CONCLUSIONS

Although a great deal of currently available research supports the relationships between specific cognitive and neuropsychological constructs and reading skills, significantly more research is necessary to elucidate relationships between the numerous abilities measured by tests commonly available to practitioners. School psychologists are in a unique position to make a difference for children with reading difficulties in the manner in which they assess, identify, and intervene in these difficulties. The literature summarized in this chapter reveals the need for school psychologists to go beyond traditional psychoeducational assessment practices by including additional measures of less commonly measured cognitive and neuropsychological constructs. By following the simple framework described herein, practitioners will be able to comprehensively assess students that are referred to them for reading-related issues in a way that is consistent with the research literature and limits the amount of testing conducted. School psychologists who engage in these practices will be better able to determine *why* these students are struggling with reading, which will lead to more individualized suggestions for intervention.

REFERENCES

Ackerman, P. L., & Lohman, D. F. (2006). Individual differences in cognitive functions. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 139–161). Mahwah, NJ: Erlbaum.

Ackerman, P. T., & Dykman, R. A. (1993). Phonological processes, confrontation naming, and immediate memory in dyslexia. *Journal of Learning Disabilities, 26,* 597–609.

Altemeier, L. E., Abbott, R. D., & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia. *Journal of Clinical and Experimental Neuropsychology, 30*(5), 588–606.

Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in Schools, 42*(4), 592–603.

Arlin, P. (1981). Piagetian tasks as predictors of reading and math readiness in grades K–1. *Journal of Educational Psychology, 73,* 712–721.

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*(5), 325–346.

Baddeley, A. (2006). Working memory: An overview. In S. J. Pickering (Ed.), *Working memory and education* (pp. 3–33). Burlington, MA: Elsevier.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76,* 610–622.

Bowers, P. G. (1995). Tracing symbol naming speed's unique contributions to reading disability over time. *Reading and Writing: An Interdisciplinary Journal, 7,* 189–216.

Bowers, P. G., & Swanson, L. B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology, 51*(2), 195–219.

Briggs, C., & Elkind, D. (1973). Cognitive development in early readers. *Developmental Psychology, 9,* 279–280.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, vernal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

Canter, A. (1975). *A developmental study of the relationships between cognitive abilities and early reading achievement.* Unpublished doctoral dissertation, University of Minnesota.

Cartwright, K. B. (2002). Cognitive development and reading: The relation of reading-specific multiple classification skill to reading comprehension in elementary school children. *Journal of Educational Psychology, 94,* 56–63.

Cartwright, K. B. (2006). Fostering flexibility and comprehension in elementary students. *The Reading Teacher, 59,* 628–634.

Cartwright, K. B. (2007). The contribution of graphophonological-semantic flexibility to reading comprehension in college students: Implications for a less simple view of reading. *Journal of Literacy Research, 39,* 173–193.

Cartwright, K. B., Marshall, T. R., Dandy, K. L., & Isaac, M. C. (2010). The development of graphophonological–semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *Journal of Cognition and Development, 11*(1), 61–85.

Casco, C., Tressoldi, P. E., & Dellantonio, A. (1998). Visual selective attention and reading efficiency are related in children. *Cortex, 34*(4), 531–546.

Castillo, J. M., Curtis, M. J., & Gelley, C. (2010a). School psychology—Part 1: School psychologists' professional practices and implications for the field. *Communiqué, 40*(7), 1, 28–30.

Castillo, J. M., Curtis, M. J., & Gelley, C. (2010b). School psychology—Part 2: School psychologists' professional practices and implications for the field. *Communiqué, 40*(8), 4–6.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*(1), 38–50.

Cohen, S. A., Hyman, J. S., & Battistini, E. E. (1983). Effects of teaching Piagetian decentration upon learning to read. *Reading Improvement, 20,* 96–104.

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4,* 219–258.

Compton, D. L., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2002). Comparing the relationships among two different versions of alphanumeric rapid automatized naming and word level reading skills. *Scientific Studies of Reading, 6,* 343–368.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3,* 422–433.

De Beni, R., Borella, E., & Carretti, B. (2007). Reading comprehension in aging: The role of working memory and metacomprehension. *Aging, Neuropsychology, and Cognition, 14,* 189–212.

De Beni, R., & Palladino, P. (2000). Intrusion errors in working memory tasks: Are they related to reading comprehension ability. *Learning and Individual Differences, 12,* 131–145.

De Jong, P. F. (2006). Understanding normal and impaired reading development: A working memory perspective. In S. J. Pickering (Ed.), *Working memory and education* (pp. 33–60). Burlington, MA: Academic Press.

De Luca, M., Di Pace, E., Judica, A., Spinelli, D., & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia, 37*(12), 1407–1420.

Deák, G. O. (2000). The growth of flexible problem solving: Preschool children use changing cues to infer multiple word meanings. *Journal of Cognition and Development, 1,* 157–191.

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 271–327). San Diego, CA: Academic Press.

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools, 50*(3), 300–313.

Dehn, M. J. (2008). *Working memory and academic learning: Assessment and intervention.* Hoboken, NJ: Wiley.

Dehn, M. J. (2010). *Long-term memory problems in children and adolescents: Assessment, intervention, and effective instruction.* Hoboken, NJ: Wiley.

Diller, L. (1974). *Studies in cognition and rehabilitation in hemiplegia (Rehabilitation* Monograph*).* New York:

New York University Medical Center Institute of Rehabilitation Medicine.

Elkind, D., Larson, M., & Van Doorninck, W. (1965). Perceptual decentration learning and performance in slow and average readers. *Journal of Educational Psychology, 56,* 50–56.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84,* 253–262.

Fagan, T. K., & Wise, P. S. (2007). *School psychology: Past, present, and future* (3rd ed.). Bethesda, MD: National Association of School Psychologists.

Feifer, S. G. (2010). Assessing and intervening with children with reading disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 483–505). Hoboken, NJ: Wiley.

Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2010). Integrating multiple data sources and multiple data gathering methods. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability assessment* (pp. 233–298). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *Achievement test desk reference: A guide to learning disability identification.* Hoboken, NJ: Wiley.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (1996). *Behavior Rating Inventory of Executive Function.* Lutz, FL: Psychological Assessment Resources.

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing, 18,* 583–616.

Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223–236.

Hulme, C., & Mackenzie, S. (1992). *Working memory and severe learning difficulties.* Hove, UK: Erlbaum.

Hulslander, J., Talcott, J., Witton, C., DeFries, J., Pennington, B., Wadsworth, S., et al. (2004). Sensory processing, reading, IQ, and attention. *Journal of Experimental Child Psychology,* 88(3), 274–295.

Hutton, U. M., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory,* 9(4–6), 383–394.

Jacques, S., & Zelazo, P. D. (2005). Language and the development of cognitive flexibility: Implications for theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 144–162). New York: Oxford University Press.

Johnson, E. S., Humphrey, M., Mellard, D. F., Woods, K., & Swanson, H. L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disabilities Quarterly, 33*(1), 3–18.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kamhi, A. G., & Pollock, K. E. (2005). *Phonological disorders in children: Clinical decision making in assessment and intervention.* Baltimore: Brookes.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment.* Hoboken, NJ: Wiley.

Kavale, K. (1982). Meta-analysis of the relationship between visual perceptual skills and reading achievement. *Journal of Learning Disabilities, 15*(1), 42–51.

Keeler, M. H. (1995). Strategic organization and reading comprehension deficits in middle school children. *Dissertation Abstracts International: Section B. The Sciences and Engineering, 55*(9), 4123B.

Keogh, B. K., & Smith, C. E. (1967). Visuo-motor ability for school prediction: A seven year study. *Perceptual and Motor Skills, 25,* 101–110.

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late emerging reading disabilities. *Journal of Educational Psychology, 95,* 211–224.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58,* 88–111.

Linderholm, T., & Van Den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94,* 778–784.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222*(3), 66–79.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology.* Harmondsworth, UK: Penguin Books.

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33,* 325–333.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading, 3,* 129–157.

Mascolo, J. T., Alfonso, V. C., & Flanagan, D. P. (2014). A systematic method of analyzing assessment results for tailoring interventions (SMAARTI). In J. T. Mascolo, V. C. Alfonso, & D. P. Flanagan, *Essentials of planning, selecting, and tailoring interventions for unique learners* (pp. 3–55). Hoboken, NJ: Wiley.

Mazzola Sortor, J., & Taylor Kulp, M. (2003). Are the results of the Beery–Buktenica Developmental Test of Visual–Motor Integration and its subtests related to achievement test scores? *Optometry and Vision Science, 80*(11), 758–763.

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal, 8,* 323–339.

McCloskey, G., & Perkins, L. A. (2013). *Essentials of executive functions assessment.* Hoboken, NJ: Wiley.

McCloskey, G., Perkins, L. A., & Van Divner, B. R. (2009). *Assessment and intervention for executive function difficulties.* New York: Routledge.

McGrew, K. S. (1993). The relationship between the Woodcock–Johnson Psycho-Educational Assessment Battery—Revised Gf-Gc cognitive clusters and reading achievement across the lifespan. *Journal of Psychoeducational Assessment Monograph Series: Woodcock–Johnson Psycho-Educational Assessment Battery—Revised,* 39–53.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Longitudinal course of rapid naming in disabled and nondisabled readers. *Annals of Dyslexia, 48,* 91–114.

Miller, D. C. (2012). *Neuropsychological Processing Concerns Checklist for School-Aged Children and Youth, Third Edition (NPCC-3).* Hickory Creek, TX: KIDS.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D. C., & Blasik, J. L. (2010). Assessing and intervening with children with memory and learning disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 641–671). Hoboken, NJ: Wiley.

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language and Cognitive Processes, 16,* 241–259.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69,* 996–1011.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70,* 1–13.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its applications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243–277). Baltimore: Brookes.

Palincsar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Palmer, S. (2000). Phonological recoding deficit in working memory of dyslexic teenagers. *Journal of Research in Reading, 23,* 28–40.

Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 815–860). New York: Longman.

Pennington, B. (2009). *Diagnosing learning disorders: A neuropsychological framework* (2nd ed.). New York: Guilford Press.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah, NJ: Erlbaum.

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia, 11,* 116–131.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children, Third Edition, Structured Developmental History.* Bloomington, MN: Pearson.

Savage, R., Cornish, K., Manly, T., & Hollis, C. (2006). Cognitive processes in children's reading and attention: The role of working memory, divided attention, and response inhibition. *British Journal of Psychology, 97,* 365–385.

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review, 19,* 185–221.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48,* 115–136.

Schatschneider, C., Francis, D. J., Fletcher, J. M., & Foorman, B. R. (2002). The relationship between rapid naming and phonological awareness in the prediction of early reading skills. *Journal of Learning Disabilities, 35,* 245–256.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Seigneuric, A., & Ehrlich, M. (2005). Contributions of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18,* 617–656.

Seigneuric, A., Ehrlich, M., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 13,* 81–103.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H.,

& Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15*(3), 232–246.

Shaywitz, S. E., & Shaywitz, B. A. (2008). Paying attention to reading: The neurobiology of reading and dyslexia. *Development and Psychopathology, 20*(4), 1329–1349.

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*(10), 1031–1045.

Stern, P., & Shalev, L. (2013). The role of sustained attention and display medium in reading comprehension among adolescents with ADHD and without it. *Research in Developmental Disabilities, 34*(1), 431–439.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63*(3), 289–298.

Swanson, H. L. (2010). Does the dynamic testing of working memory predict growth in nonword fluency and vocabulary in children with reading disabilities? *Journal of Cognitive Education and Psychology, 9*(2), 139–165.

Swanson, H. L., Howard, C. B., & Saez, L. (2006). Do different components of working memory underlie different subgroups of reading disabilities? *Journal of Learning Disabilities, 39,* 252–269.

Taylor Kulp, M. (1999). Relationship between visual motor integration skill and academic performance in kindergarten through third grade. *Optometry and Visual Science, 76,* 159–163.

Tierney, R. J., & Cunningham, J. W. (1984). Research on teaching reading comprehension. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 609–656). New York: Longman.

Tiu Jr., R. D., Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities, 36*(5), 424–436.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004), The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*(4), 339–363.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Viholainen, H., Ahonen, T., Lyytinen, P., Cantell, M., Tolvanen, A., & Lyytinen, H. (2006). Early motor development and later language and reading skills in children at risk of familial dyslexia. *Developmental Medicine and Child Neurology, 48*(5), 367–373.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Was, C. A., & Woltz, D. J. (2006). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language, 56,* 86–102.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study of average and impaired readers. *Child Development, 57,* 988–1000.

Wolf, M., Grieg Bowers, P., & Biddle, K. (2000). Naming speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 4, 387–407.

Yap, R., & van der Leij, A. (1993). Rate of elementary symbol processing in dyslexics. In S. F. Wright & R. Groner (Eds.), *Facets of dyslexia and its remediation* (pp. 337–347). Amsterdam: Elsevier.

CHAPTER 35

Incorporating Recent Advances in Understanding Word-Reading Skills into Specific Learning Disability Diagnoses

The Case of Orthographic Mapping

David A. Kilpatrick

BACKGROUND

In the spring of 1988, I took a graduate course in neuropsychology from Larry Lewandowski at Syracuse University. After describing the inadequate nature of the neuropsychological tests available at the time, he stated that the best tool we had for conducting a neuropsychological evaluation was between our ears. He indicated that the more we knew about the functioning of the brain, the better equipped we would be to select and interpret cognitive and neuropsychological tests and subtests to address the questions we would want to answer. This chapter is written in the spirit of Lewandowski's comment. It may differ from other chapters in this volume, in that only a portion of it is devoted to some critical "nuts-and-bolts" issues of diagnosing specific learning disabilities (SLD) in reading. However, it provides important background information for understanding the very nature and causes of word-level reading difficulties. Such a knowledge base will help practitioners conduct more informed evaluations and make more valid decisions regarding SLD in word-level reading.

Why a Special Chapter on Word-Level Reading?

The reason for this rather in-depth chapter on the nature of word-level reading problems is quite simple: Approximately 80–85% of students diagnosed with SLD have reading disabilities, and the vast majority of these students have word-level reading issues (Fletcher, Lyon, Fuchs, & Barnes, 2018; Lerner & Johns, 2012). The term *word-level reading disabilities* refers to SLD in basic reading and/or reading fluency. Far and away the most common type of SLD, word-level reading disabilities can occur alone or may be comorbid with SLD in reading comprehension, writing, or math. Given the sheer magnitude of SLD in reading, and the vast empirical research that has been conducted to investigate such problems, it seems judicious to examine word-reading difficulties in a fair degree of depth.

The Psychology of Reading

The late Robert Crowder was a cognitive psychologist at Yale University. In his book *The Psychology of Reading* (1981), he mentions an interesting insight regarding the relationship between the field of psychology and research on reading. He said that the cognitive revolution from the 1950s to 1970s involved studying perception, input, storage, and retrieval. Various stimuli were used to examine these phenomena, including letters of the alphabet, written words, sentences, and paragraphs. By the 1970s, it became recognized that the field of cognitive psychology had unwittingly amassed hundreds of studies on reading. The goal had been to understand complex cognitive processes; the reading-related stimuli were only the raw materials for studying those cognitive processes (Crowder, 1981).

An examination of the departmental affiliations in the empirical research reports on reading acquisition and reading disabilities indicates that the lion's share of scientifically oriented research articles come from departments of psychology around the globe. Reading research is, however, a highly interdisciplinary pursuit: It is shared by such fields as speech pathology, linguistics, special education, general education, literacy, neurology, and pediatrics. Nonetheless, the field of psychology has arguably made the largest volume of contributions to the hundreds of scientifically oriented reports on reading acquisition and reading disorders that appear every year (Kilpatrick, 2015).

Reading Research versus Classroom Practice

Most unfortunately, this large and heavily grant-funded body of research has not made inroads into the teaching of reading in our nation's K–12 schools. It has been pointed out by numerous sources that "a chasm exists between classroom instructional practices and the research knowledgebase on literacy development" (American Federation of Teachers, 1999, p. 7; see also Joshi, Binks, Graham, et al., 2009; Joshi, Binks, Hougen, et al., 2009; Kilpatrick, 2015; Moats, 1994, 2009; Seidenberg, 2017). One attempt to close this gap between research and practice was the implementation of *response to intervention (RTI)*. RTI was prompted by federal grant initiatives on reading that yielded highly encouraging findings in terms of preventing and correcting reading problems (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; National Reading Panel, 2000; Torgesen et al., 2001; Vellutino et al., 1996). The original intent of RTI was to scale up that research, so that all at-risk and struggling readers in the United States would benefit from these highly effective approaches. Most unfortunately, however, the implementation of RTI focused on the processes, frameworks, universal screenings, and progress monitoring for RTI, while the actual instructional and intervention practices that were so highly successful in those seminal studies were never adequately communicated (Kilpatrick, 2015; Seidenberg, 2017). Teachers have been charged with using "research-based" or "evidence-based" instructional practices, without knowing what those practices were (Seidenberg, 2017). As a result, a recent federal report indicated that RTI is having little or no impact on the students it is designed to serve (Balu et al., 2015).

It appears that school psychologists, like teachers, are not likely to be incorporating the findings from the reading research into their professional practice. A study published in *School Psychology Review* indicated that school psychologists, by and large, are not familiar with some of the most important findings from empirical studies of reading acquisition and reading difficulties/disabilities (Nelson & Machek, 2007). The present chapter is intended to provide information about some important recent advances related to word-level reading difficulties that practitioners should consider when evaluating and diagnosing students who display such difficulties.

READING WORDS VERSUS LEARNING WORDS

One of the most plausible reasons for the limited effects of RTI appears to be that educators continue to teach reading the way they have always taught it, but now they do so within an RTI framework (Kilpatrick, 2015; Seidenberg, 2017). There is no evidence that the highly effective general and remedial instructional techniques used in the federal grant initiatives that prompted RTI have been widely incorporated into our schools, while there is evidence that they are not (e.g., Joshi et al., 2009). This would account for the null results found in the recent federal report (Balu et al., 2015). An important problem related to this is that neither of the two dominant reading approaches used in schools over the last 40 years properly distinguishes between *reading* words and *learning* words. They focus on the former without adequately addressing the latter. This is problematic because the most highly successful outcomes in the reading research literature appear to have facilitated the ability to learn and remember words, not just to read and identify them (Kilpatrick, 2015).

The most common approach to teaching reading in the United States in recent decades has been the three-cueing-systems approach. This approach was made popular by the whole language movement in the 1980s and 1990s and now forms the foundation of balanced instruction, Reading Recovery, and the Leveled Literacy Intervention (commonly known as LLI). The three-cueing-systems model teaches students to read words and sentences by using three types of cues: (1) the *context* of the sentence or passage; (2) the *linguistic* features of the words (grammar and syntax), and (3) the *grapho-phonic* features (i.e., letters and sounds) of the word (Goodman, 1996). The second most common approach to reading instruction has been phonics. The phonics approach encourages students to use knowledge of letter–sound relationships to "sound out" unfamiliar words. The phonics approach has been shown to yield superior results to the three-cueing-systems approach, particularly for weak readers (Bond & Dykstra, 1967; Brady, 2011; Ehri, Nunes, Stahl, & Willows, 2001; National Reading Panel, 2000; Moats, in press; Share, 1995).

The problem with both of these approaches is that they focus on developing students' abilities to *identify* words. This is not the same as *learning* those words. A word can be read or identified without actually being learned or remembered. When this happens, a word that has been correctly identified via phonic decoding or contextual guessing may not be remembered when encountered in the next paragraph, and most likely not remembered when encountered the next day. By contrast, skilled readers are very adept at remembering the words they read. From second grade on, skilled readers learn newly encountered words after only one to four exposures (Cunningham, 2006; Reitsma, 1983; Share, 2004b; see more below). From then on, those newly learned words are recognized as familiar when they are encountered, and that recognition is instantaneous and effortless (Ehri, 2005). There is no need to sound out such words, nor is any guessing involved. Currently, there does not appear to be an instructional methodology used in schools that takes account of the empirical research that has occurred on printed-word learning (but see Kilpatrick, 2016). Yet, as mentioned, it appears that the studies that displayed highly successful intervention results all helped students develop the ability to learn words rather than simply read words.

It must be pointed out here that most children will learn to read, "no matter how unhelpful the instruction" (Liberman & Liberman, 1990, p. 54). For approximately two-thirds of students, this distinction between reading/identifying words and learning/remembering words is of very little consequence. These students acquire the ability to learn words as a result of being exposed to literacy activities. The situation is quite different for struggling readers. For them, the distinction between reading words and learning words is of great significance. A large portion of the bottom third of readers are not able to learn words efficiently, regardless of which of the two dominant teaching approaches they receive.

ASSESSMENT CONCERNS REGARDING WORD-LEVEL READING

In the same way that conventional approaches to reading instruction do not adequately distinguish between reading words and learning words, neither do most standardized tests that involve isolated-word reading. Nationally normed, word-level reading assessments evaluate students' ability to read words. They do not directly evaluate their ability to learn words. Additionally, such assessments confound two different aspects of word reading: identification and recognition (Kilpatrick, 2015). *Word identification* refers to the ability to correctly read a given word, regardless of its prior familiarity. *Word recognition* presumes that a word is already familiar (Aaron et al., 1999; Harn, Stoolmiller, & Chard, 2008; Kilpatrick, 2015). Yet the terms *word identification* and *word recognition* are typically used interchangeably. Some subtests that assess isolated-word reading in reading and achievement batteries are called word identification subtests (e.g., in the Woodcock–Johnson Tests of Achievement), while others are called word recognition subtests (in the Kaufman Test of Educational Achievement). This is despite the fact that these subtests use identical or nearly identical formats. The synonymous use of these terms appears to compromise precision in understanding and addressing two different reading-related skills or processes.

Some of the words on standardized word identification subtests are already familiar to any given student. The pool of words that a student already knows has been referred to as an *orthographic lexicon* or a *sight word vocabulary* (Ehri, 2014; Van den Broeck & Geudens, 2012). Such words are instantly recognized on those subtests; no sounding out or guessing is needed. The size of a given student's orthographic lexicon/sight word vocabulary appears to result from the interaction between the student's ability to remember words and the student's reading experience. The latter factor (reading experience) allows him or her to be exposed to more and more words to be learned.

In addition to tapping into a student's orthographic lexicon, standard word identification subtests evaluate another skill. This second skill involves a student's ability to figure out a word, on the spot, that he or she did not previously know. Students can correctly determine unfamiliar words in isolation by using one or more strategies. One strategy involves guessing based on the first letter and the length of the word. For example, a student may say "lunch" when presented with *laugh,* or "expect" when presented with *expert.* Such guessing will often yield a correct response. A correct response does not mean that the student knows the word. It means that the student made a good guess. A second strategy that can be used to determine a previously unfamiliar word involves reading by analogy (Ehri, 2005). If a student is familiar with the word *since,* he or she can use knowledge of that word when encountering a word like *prince.* A third strategy is what researchers call *phonological recoding* (Share, 1995), which educators call *phonic decoding.*

A fourth strategy for determining an unfamiliar word without the aid of context is called *set for variability* (Kearns, Rogers, Koriakin, & Al Ghanem, 2016; Tunmer & Chapman, 1998, 2012). Essentially, set for variability refers to one's ability to correctly determine a mispronounced word. This applies to reading when a student correctly determines a word despite having mispronounced it, either because it is an irregular word or because it was simply misread. Students with stronger oral vocabularies make better use of set for variability than students with more limited oral vocabularies (Tunmer & Chapman, 2012).

Thus, on standard, context-free word identification subtests from normative achievement batteries, students can read unfamiliar words via four different strategies. This means that on our most popular word-reading subtests, two different aspects related to word-level reading are inherently confounded: the size of the sight vocabulary and the ability to figure out unfamiliar words without the aid of context. While subtests of nonsense-word reading may help us understand a student's ability to sound out an unfamiliar word, we have more difficulty assessing the sight vocabulary. Yet the size and growth of the sight vocabulary are what may give us some clues about a student's efficiency in learning/remembering words.

Language Skills May Confound Word-Reading Assessment and SLD Diagnosis

The advantage that vocabulary skills provide in making use of set for variability implies that the scores on word-reading subtests for those with higher vocabulary skills might tend toward an overestimation of their raw word-level reading capabilities. This has no bearing on the classical IQ–achievement discrepancy; indeed, the scoring pattern is in the wrong direction (i.e., it minimizes any discrepancy between IQ and achievement). However, this phenomenon appears to have implications for identifying readers who are poor at learning words and who may benefit from additional general educational remedial reading help, or even in some cases students who may qualify as SLD. Such students have been called *compensators* (Kilpatrick, 2014, 2015, 2016). A study of compensators indicates that they often dislike and avoid reading, despite average word identification subtest scores on normed tests (Kilpatrick, 2014). It seems that their strong verbal skills combined with even rudimentary phonic decoding skills allow them to correctly identify previously unfamiliar words on word-reading tests via set for variability.

Consider the possible impact of set for variability on interpreting the word identification subtest scores of two third-grade boys. The first student has a verbal IQ (VIQ) of 90, and the other has a VIQ of 113. They both have equal phonic decoding skills, as reflected by a standard score of 83 on a test of nonsense-word reading. Let us say they happen to have prior familiarity with the same number of words on the word identification subtest. They also both received a scaled score of 11 on the Blending Words subtest of the Comprehensive Test of Phonological Processing—Second Edition (CTOPP-2; Wagner, Torgesen, Rashotte, & Pearson, 2013), but a 7 on the Elision subtest.

On the word identification subtest, these two students each instantly recognize the same words that are familiar to both of them. The student with the 90 VIQ goes on to correctly identify three unfamiliar words, using his weak phonic decoding skills and subpar set for variability. His word identification score is 83, or 13th percentile. This score is well within the range of weak reading, according to the National Assessment of Educational Progress (2015), which indicates that about 30% of fourth graders read below a basic level. By contrast, using the same rudimentary phonic skills but applying his strong vocabulary skills to yield better set for variability, the student with the 113 VIQ goes on to correctly identify several additional words, yielding a standard score of 92. The former student may receive Tier 2 remedial instruction for his reading skills. Also, depending on the rest of his profile and his school's criteria/cutoffs for SLD, he may even be considered for an SLD designation. By contrast, the second student has an average word identification score. That student would not be considered as having SLD and likely would not even be considered for Tier 2 remedial services. Yet both students have the same raw word-level reading abilities, although the latter student can mask his inadequacies due to strong vocabulary skills. While the classic IQ–achievement discrepancy unfairly favored those with higher IQs, the issue of compensation has the opposite effect.

This illustration is not intended to suggest that the second student should be designated as having SLD. However, it should be acknowledged that such a student is likely to be a struggling reader who, at minimum, should receive Tier 2 remedial instruction. Yet what if one of the school's criteria is that a student must have a standard score of 85 or lower for an SLD diagnosis, and this student received an 87 on word identification? The first student's 83 makes the cut and the second student's does not, although they have a similar-sized orthographic lexicon and the same level of phonic decoding and phonemic awareness. The student with the VIQ of 113 disqualifies himself because he is better at figuring out words, due to his high vocabulary and correspondingly stronger set for variability. But the trajectory for this student is that he will spend more time having to figure out words, while his skilled fellow students will remember the words they read and will not have to keep applying strategies to the same words. Thus, when diagnosing SLD in word-level reading (basic reading and/or reading fluency), we must acknowledge the nature of the inherent problems with the word identification assessments we routinely use to make such diagnoses.

Word-Reading Strategies and the Orthographic Lexicon May Be Confounded

Reading familiar words involves no strategies. Known words are instantly and automatically accessible (Ehri, 2005; Rayner, Pollatsek, Ashby, & Clifton, 2011), even precognitive, so that they are already available before any conscious strategy could be applied. Word-reading strategies, such as the four strategies referred to earlier—to which we could add the use of contextual guessing—are only necessary when encountering unfamiliar words. The confounding phenomenon of evaluating both the recognition of familiar words and the identification of unfamiliar words on the same subtest means that there is no way to know which of the correctly read words were familiar and which were not. Clinical observation may be somewhat useful here by noting whether the student responded to a given word instantly or not. However, such observations cannot be followed up with normative comparisons to determine what constitutes typical performance. This confounding is not without consequence. Known words are read more quickly than unknown words, and the number of known words appears to be the driving force behind reading fluency. That is, reading fluency appears to be primarily a function of the size of a student's orthographic lexicon or sight vocabulary (Ehri, 2005; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Kilpatrick, 2015; Torgesen, 2004b; Torgesen, Rashotte, Alexander, Alexander, & MacPhee, 2003). Students who know all or most all of the words in a given passage read more fluently than students who know fewer of the words in the passage. A student may have 100% accuracy on a passage, but may have poor fluency because he or she had to phonologically recode or guess at a substantial number of the words. While the student correctly *read* those previously unfamiliar words, their unfamiliarity means that these were not words that had been previously learned.

LEARNING WORDS

Having a pool of known words presumes a previous learning history on each and every one of those familiar words. Like other cognitive and linguistic skills, it also presumes that given equal instruction, opportunity, and effort, there will be individual differences in the ease with which students learn and remember written words (Ehri & Saltmarsh, 1995; Share, 2011; Share & Shalev, 2004). But how do students learn words rather than just identify them? At what point does a given word go from being unfamiliar to familiar, and thus instantly and effortlessly accessible? What cognitive, linguistic, and academic skills and processes contribute to our memory for the words we read? Also, why are there such large individual differences in this skill? For example, in a classic study, Ehri and Saltmarsh (1995) discovered that the word-*learning* skills of typically developing first graders were stronger than those of a comparison group of fourth graders with reading disabilities. Share and Shalev (2004) also showed that children with reading disabilities required more exposures to words before they learned them.

The Nature of Dyslexia

In the reading research literature, significant word-level reading difficulties/disabilities are referred to as *dyslexia* (Fletcher et al., 2018; Hulme & Snowling, 2009; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Although popular understandings of the term *dyslexia* are fraught with 100-year-old misconceptions, from the standpoint of researchers dyslexia simply refers to poor word-level reading despite adequate effort and opportunity, and it cannot be accounted for by blindness, deafness, or severe intellectual impairment (Fletcher et al., 2018; Hulme & Snowling, 2009; Vellutino et al., 2004). For the last three decades, many researchers have conceptualized dyslexia as being the result of a nonword-reading deficit (Rack, Snowling, & Olson, 1992; Share, 1995). In this understanding, word-level reading disabilities are based primarily on poor phonological skills, which make acquiring the alphabetic code of written English very challenging. As a result, sounding out new words is very difficult for those with dyslexia. However, more recently, on empirical, statistical, and design grounds, Van den Broeck and colleagues (Van den Broeck & Geudens, 2012; Van den Broeck, Geudens, & van den Bos, 2010) have shown that poor nonsense-word reading among dyslexic readers appears to be only half the story. Students with word-level reading difficulties also have a weakness in remembering the words they read (Ehri & Saltmarsh, 1995; Share & Shalev, 2004; Van den Broeck & Geudens, 2012; Van den Broeck et al., 2010). In addition to supplementing and extending the notion of dyslexia as a nonword-reading deficit, these findings challenge the older notion that dyslexia can be reliably divided into *phonological* and *surface* subtypes (see more on this below).

Word-Learning Theories

There are several theories designed to explain how words are learned. For simplicity, a distinction is made here between *computational* theories and *cognitive* theories. Computational theories of word learning involve computer programs that simulate both reading words and learning words (Coltheart, 2012; Seidenberg, 2002, 2017). These computational models have yielded rich insights about reading, but they are based only indirectly on behavioral evidence from studies of actual human reading. These models are not considered further in this chapter because the number of trials required for word learning in these models is discrepant with actual data from human readers (surprisingly, typical human readers learn words far more quickly).

Cognitive Theories of Word Learning

The two theories of word learning that have generated the most empirical support are Linnea Ehri's (1992, 2005, 2014; Miles & Ehri, in press) *orthographic mapping theory* and David Share's (1995, 1999, 2011) *self-teaching hypothesis*. Torgesen referred to Ehri's theory of word learning as "the most complete current theory of how children form sight word representations" (Torgesen, 2004a, p. 36). Van den Broeck and Geudens (2012) speak as highly of Share's theory when they say that the "self-teaching model is the most dominant account of the developmental process toward fully specified orthographic representations" (p. 416). This latter quote is not inconsistent with the quote by Torgesen because of the tremendous overlap between these two theories. Indeed, one researcher explicitly says that Share's self-teaching hypothesis "is essentially the same as Ehri's [orthographic mapping] hypothesis" (Apel, 2009, p. 43). While this statement is not technically accurate, it testifies to the large overlap between these models of learning to read.

Both orthographic mapping and the self-teaching model posit a central role for letter–sound knowledge and for phonemic awareness in building the orthographic lexicon. Visual memory plays no role (see below). An important difference between Ehri's and Share's theories is that Ehri's theory provides a specific cognitive mechanism for the process of forming connections between pronunciations and their orthographic representations. Share's theory provides the scenario under which this connection-forming process occurs, without providing specific details about how words are encoded into the orthographic lexicon. Ehri, on the other hand, says little about this learning scenario that Share describes. Rather, her theory presents a more abstract representation of the connection-forming process.

Before this process of orthographic learning is described in detail, a more dominant theory of word learning, even if informal and intuitive, must first be addressed. This is the common view that word learning is based on some form of visual memory process via paired-associate learning. This intuitive theory assumes that learning to read words is a process similar to learning to name familiar objects or people (i.e., visual input, verbal output).

Do We Remember Words Based on Visual Memory?

When we look at a chair and say "chair," or when we see the printed word *chair* and say "chair," the naming activity intuitively appears to be similar, if not identical. In both cases, visual input is used to access a phonological code, which is our memory of the pronunciation of the spoken word *chair*. There are, however, multiple, independent lines of evidence demonstrating that word reading is not based on a visual memory process similar to naming objects in our environment. These various lines of evidence have been presented in detail elsewhere (Kilpatrick, 2015), which are summarized below and in Table 35.1.

First, in 1886, James Cattell tested naming speed for objects versus printed words. He did so using a newly developed millisecond-level timing device. Adults read words like *chair* and *tree,* and Cattell compared their reaction times, measured to 1/1,000th of a second, to visual presentations of a chair or a tree. To his surprise, Cattell found that the reaction times to the printed words were consistently faster than to the actual objects. Thus, by the late 1800s, there already existed evidence challenging our intuitive notion that visual memory and orthographic memory (i.e., printed-word recognition) represent the same process.

Second, in the preface to his 1979 book *Dyslexia,* researcher Frank Vellutino says he began the decade of the 1970s assuming the common view that word reading is based on some sort of visual memory process. By the time he wrote his book, however, he had abandoned that view, based on studies he and others conducted in the 1970s that failed to find the expected visual memory deficiencies in dyslexic readers (Vellutino, 1979).

TABLE 35.1. Summary of Reasons Why We Know Words Are Not Remembered via Visual Memory

1. Reaction times to printed words (e.g., *chair* or *tree*) are faster than reaction times to objects (e.g., a picture of a chair or tree), suggesting that visual memory and orthographic memory are different processes.
2. Persons with poor word-level reading tend to have average visual memories.
3. Despite the finding that their visual memory is equivalent to that of hearing individuals, those who are deaf struggle with remembering the words they read.
4. There is a moderate to strong correlation between word reading and phonological skills, but a very weak correlation between word reading and visual memory.
5. Words are instantly recognized despite their visual presentation (uppercase, lowercase, differing fonts, handwriting styles, etc.), as long as the letters are legible.
6. Neuroimaging studies indicate that there is limited overlap in the areas of the brain responsible for visual memory versus memory for written words.
7. We routinely have "visual memory" failures in forgetting the names of familiar people or even objects, but we have no such failures with remembering familiar words.

Third, if reading were based on visual memory, it becomes very difficult to explain why students who are deaf tend to graduate from high school at about a third-grade reading level (Lederberg, Schick, & Spencer, 2013; Leybaert, 2000). Individuals who are deaf have visual memory skills comparable to those of hearing individuals, so if visual memory were the basis for printed word memory, students who are deaf would learn to read at a rate comparable to that of their hearing peers.

Fourth, the correlation between visual skills and word reading tends to be very low, while the correlation between word reading and phonological processes is substantially higher (Vellutino, 1979; Vellutino et al., 1996; Wagner & Torgesen, 1987). It is difficult to understand why this would be the case if visual skills play a substantial role in word learning.

Fifth, studies done to disrupt readers' visual memories of words have been inconsistent with the visual memory hypothesis. As long as readers are acclimated to unusual forms of print (e.g., strange or ornate fonts or a specific individual's handwriting), readers have instant access to words written in those differing ways, despite having no prior exposure to that unusual visual presentation for any given word (Adams, 1990). One example of this involved disrupting a visual presentation of words using mixed type (e.g., wOrDs WrItTeN LiKe tHiS), which virtually guarantees that the reader has no prior exposure to or visual memory of words printed that way (Adams, 1990). Adams (1990) described studies in which words were presented on a computer screen for 1/20th of a second, in either uppercase, lowercase, or mixed case, followed by a mask (e.g., #####). During debriefing after the study, some research participants indicated that they were unaware of these different presentations, and others even insisted that they had all been presented in lowercase letters only.

A likely reason for this finding can be found in studies that show that readers have an abstract representation of each of the letters of the alphabet in the memory system, irrespective of case and font (Bowers, 2000; Frost, 1998; Van den Broeck & Geudens, 2012). Within the first 1/10th of a second after a word is seen, it appears that the particular letters perceived are translated into their respective abstract representations. Apparently, the memory system then seeks to determine whether that *specific sequence of letters*—regardless of its visual characteristics—is stored in the orthographic lexicon. Thus orthographic memory appears to be based on familiar sequences of letters,[1] not familiar visual input at the letter or word level. If the memory system detects a familiar letter sequence, the left fusiform gyrus (in the left ventral occipito-temporal area) activates and the word is recognized. If it does not, the activation then moves higher in the temporo-occipital area associated with the letter–sound conversion process (Dehaene & Cohen, 2011; Glezer, Kim, Rule, Jiang, & Riesenhuber, 2015; Simos et al., 2002). This helps understand why case and font appear to make little or no difference in recognizing words, as long as the letters are legible to the reader.

Sixth, neuroimaging studies have indicated that the areas of the brain that are activated during visual memory tasks show limited overlap with the areas activated during the reading of familiar words (Dehaene & Cohen, 2011; Simos et al., 2002). This helps explain Cattell's results from over 125 years ago.

Finally, in addition to these various lines of research evidence, there is another phenomenon readers experience that is inconsistent with the visual memory hypothesis of word reading. It is not uncommon for us to have an apparent "visual memory failure"—that is, a failure to retrieve the phonological code associated with some visual input. For example, this occurs when we encounter a familiar person and yet fail to retrieve his or her name. It also happens when we fail to retrieve the name of an object in our line of sight ("Hand me that thingy over there").[2] By contrast, it appears that this same retrieval failure never occurs with orthographically familiar words. Familiar words are consistently retrieved. Only unfamiliar words or words printed illegibly represent challenges to accurate retrieval. This disparity in retrieval failures of people's names and objects' names, but not written words, is difficult to explain with the intuitive theory that memory for written words is based on visual processes similar to those used in object recognition. Input and storage are not the same thing. We input words visually, but store them orthographically, phonologically, and semantically.

In sum, several independent lines of empirical evidence appear to falsify our highly intuitive notion that printed words are stored and retrieved from long-term memory via some form of visual memory process. But if visual memory is not the mechanism by which we remember words, what is?

ORTHOGRAPHIC LEARNING AND MEMORY

In contrast to visual memory, the notion of remembering words via *orthographic memory* has received substantial empirical support (Ehri, 1992, 2005; Rack, Hulme, Snowling, & Wightman, 1994; Share, 1995, 2011). *Orthography* refers to the proper way to represent written words in a given writing system. In alphabet-based writing systems, orthographic memory refers to a memory for the precise letter order that comprises a written word or word part (Ehri, 2005; Van den Broeck & Geudens, 2012). Because orthographic memory involves a specific letter sequence, there is no particular relevance to the visual features of the printed words, such as size, case, font, or whether a word is in print or handwriting. The necessary feature is that the letters in the sequence are legible. The question arises as to how such a highly efficient and largely automatic memory for specific letter sequences occurs within the cognitive system. Recent clues have come from neuroimaging studies (Dehaene & Cohen, 2011; Glezer et al., 2015; Simos et al., 2002). As mentioned previously, when familiar words are viewed, the left fusiform gyrus is activated. However, when unfamiliar words and nonsense words are viewed, areas in the left superior and medial temporal and occipital areas are activated. One recent study even tracked this shift in areas of activation as words became familiar (Glezer et al., 2015). In this chapter, however, the focus is on a cognitive description of the orthographic learning process. The neuroimaging and other neurophysiological data serve largely to confirm or disconfirm cognitive explanations of reading (Anderson & Reid, 2009). Currently, neuroimaging and other neurophysiological data appear to provide important evidence that confirms the cognitive explanation of written-word learning described below.

The Nature of Orthographic Knowledge

Orthographic knowledge is understood on multiple levels. On one level, it refers to a familiarity with what would be considered permissible and nonpermissible letter sequences in a given written language. For example, in English, words do not begin with *ck* or *mb,* but they may end with those letters (e.g., *back, thumb*). On another level, orthographic skills refer to the pronunciation of common subword sequences or orthographic patterns that do not yield to simple letter-by-letter, grapheme–phoneme conversion regularities (e.g., *-ight, -alk, -tion, -ould*). Finally, orthography can refer to the correct spelling of any given word (e.g., *brain,* not *brane*).

Orthography as an Independent Reading-Related Subskill

One issue in this area has been whether orthographic skills represent a separate reading-related subskill in the same way that letter–sound knowledge, nonsense-word reading, phonemic awareness, and rapid automatized naming are considered to be reading-related subskills. Orthographic skills have been commonly assessed in research studies via the word likeness task, the homophone or pseudohomophone task, and exception-word reading (e.g., Olson, Forsberg, Wise, & Rack, 1994).

A word likeness task asks children which of two nonwords is most like a word. For example, given *lmk* and *pif,* the latter is more like a word because it has a common consonant–vowel–consonant (CVC) pattern, while the first item has no vowel. A homophone or pseudohomophone task requires students to identify the correct spelling pattern for a given word, such as which of the following is a flower: *rows, rose,* or *roze*. An exception-word task simply involves having students read words that do not yield correct pronunciations via phonic decoding (e.g., *iron, yacht, rendezvous*).

Before the middle of the last decade, many researchers argued that orthographic skills contribute to word reading above and beyond letter–sound skills and phonemic awareness (e.g., Cunningham, Perry, Stanovich, & Share, 2002; Holmes, 1996). In contrast, other researchers claimed that these orthographic skills are by-products of letter–sound knowledge, phonological skills, and reading experience, and thus are not causal elements in word-reading development (Vellutino, Scanlon, & Tanzman, 1994). This latter view was supported by research in the 1990s showing that orthographic sequences become "unitized" as part of reading development and experience. For example, typical fourth-grade readers and strong second-grade readers read a nonsense word like *nalk* to rhyme with *walk* and *talk,* while weaker-reading second graders read it to rhyme with *talc,* following a more strict application of letter–sound regularities (e.g., Bowey & Underwood, 1996). Presumably, with reading experience, orthographic patterns become familiar to readers.

More recently, there appears to have been a shift in understanding the relationship between orthographic skills and reading development. This shift followed a comprehensive review of the empirical literature by Jennifer Burt (2006). Her review indicated that there were no theoretical or empirical grounds for considering orthographic skills to be a reading-related skill that contributes to the development of reading skills apart from letter–sound skills, phonemic awareness, and reading experience. Rather, orthographic skills appear to represent a point in letter–sound knowledge development that occurs as a result of reading experience and noting patterns that are consistently pronounced (e.g., *-ight, -tion*), even if they are inconsistent with a simple letter-by-letter phonic decoding approach. This view has received further support from longitudinal research. Deacon, Benere, and Castles (2012) found that first-grade reading skills predicted third-grade performance on orthographic skills tasks, while orthographic skills tasks assessed in first grade did not predict third-grade reading skills. Despite this trend in the research, it appears that some authors (e.g., Feifer, 2011, 2014; Mather & Wendling, 2012), while describing subtypes of dyslexia, continue to maintain that orthographic skills independently contribute to reading (for more detail, see below).

HOW ORTHOGRAPHIC LEARNING OCCURS: SELF-TEACHING AND ORTHOGRAPHIC MAPPING

The Self-Teaching Hypothesis

The evidence in support of David Share's self-teaching model of orthographic learning is large and growing (e.g., Bowey & Muller, 2005; Cunningham, 2006; Cunningham et al., 2002; Share, 1999, 2004b, 2011), as is the empirical support for Ehri's orthographic mapping model (e.g., Dixon, Stuart, & Masterson, 2002; Ehri & Saltmarsh, 1995; Laing & Hulme, 1999; Rack et al., 1994). I examine each in turn.

The self-teaching hypothesis begins with a simple and self-evident observation. A skilled adult reader has tens of thousands of written words in his or her orthographic lexicon/sight word vocabulary, yet it is likely that only a few hundred of those words were directly taught by teachers or parents. The essence of the self-teaching model is that children teach themselves most of the words they know, once they have in place adequate or better phonic decoding skills (i.e., phonological recoding; Share, 1995). As students with phonological recoding skills encounter new words, they perform the letter–sound conversion process and phonological blending to identify those words (Share, 1995, 2011). Context and set for variability may assist in the identification of an unfamiliar word, especially irregular or exception words. Regardless, the self-teaching model proposes that the process of tracking through the letter sequence and sounds that constitute a printed word helps establish that letter sequence in long-term memory (Share, 1995).

Numerous studies have shown that from second grade on, an average reader requires only one to four exposures to a new word in order for that word to become established in the orthographic lexicon/sight word vocabulary (Cunningham, 2006; Reitsma, 1983; Share, 1999, 2004b; Share & Shalev, 2004). At first this may seem a bit surprising. However, a moment's reflection on the growth trajectories of early readers independently supports these findings (Adams, 1990; Ehri, 2005; Snow, Burns, & Griffin, 1998). Most children enter first grade knowing dozens of words, yet 2 years later they enter third grade knowing thousands of words. This steep growth trajectory within this limited time frame does not allow the opportunity for students to have dozens of exposures to each of those thousands of words, except for high-frequency words.

A common research paradigm involves having students silently read a narrative—for example, about a fictional city called *Yait*. Some students see the target word only once at the beginning of the passage, after which "this city" or "that city" is used. Other students receive two, four, six, or eight exposures to the target pseudoword in the passage. The frequency of exposure varies both within and across studies. Some studies test students on the newly encountered "word" the following day, or a week later, or even a month later. These tests may include a spelling test; an orthographic choice task using phonologically plausible foils (e.g., "Was the name of the city *Yate, Yait, Yat, Yaet,* or *Yaite?*"); or measuring reaction time (RT) to the words flashed on a computer screen (and comparing that to the RT to the homophonic foil). Performance accuracy is quite high for all types of queries, with spelling accuracy being the weakest. Many words are learned after a single exposure, yet there is an increase in accuracy if the words are encountered two to four times. Beyond four exposures, there is a very limited benefit in terms of performance on the various types of posttests. This learning paradigm mimics the self-teaching situation, in which the silent reading of passages involves encountering new words that need to be phonically decoded. The storage in long-term memory appears to be phonological and orthographic, not visual. One way this has been determined has been through efforts to allow only visual exposure to these new words and suppress phonological recoding (e.g., by having students continuously repeat a nonsense word while reading); such efforts result in very limited accuracy in the posttests (Share, 1999). It therefore appears that processing the letter sequence at a phonological level is the key to establishing an orthographic sequence in long-term memory.

Despite the success of the self-teaching model in accounting for a great deal of empirical findings, it leaves open an important question: Precisely what is it about phonological processing during phonic decoding that allows for the establishment of a very secure orthographic sequence in long-term memory? Ehri's theory of orthographic learning directly addresses that question.

Orthographic Mapping

Ehri has been refining her theory of word learning, recently dubbed *orthographic mapping* (Ehri, 2014), for four decades (Ehri, 1978, 1992, 1998, 2005, 2014; Ehri & Wilce, 1985; Miles & Ehri, in press). Her theory provides an empirically supported explanation of how we remember the words we read. Efficient orthographic learning requires two skills: letter–sound knowledge and phoneme segmentation (Ehri, 1998, 2005). Spoken words are already stored in long-term phonological memory, and the object of orthographic learning is to have a sequence of letters attach to, or bond to, the pronunciation of that spoken word. This is not to be confused with a letter-by-letter phonic decoding of the word. Rather, once it is familiar, the particular sequence of letters becomes *unitized* (Treiman, Sotak, & Bowman, 2001); that is, the whole string of letters as a unit is familiar and instantly activates the word's pronunciation, with no need for letter-by-letter phonic decoding.

Conventional phonic decoding involves a flow of information from letters to sounds, and those sounds are blended together to arrive at a pronunciation. Orthographic mapping benefits from this flow of information, but also proposes an additional flow of information that goes in the other direction—from (1) the oral word's pronunciation, to (2) a segmented representation of the oral word, to (3) the alphabetic characters that align with that segmented pronunciation. This process of associating a known and well-established phonological representation (the word's pronunciation) with a newly encountered stimulus (a letter sequence/printed word) allows for that newly encountered stimulus (the letter sequence) to become bonded in memory with that known phonological representation (the oral pronunciation). In a sense, it represents a flow of information that goes in the opposite direction from phonic decoding. It could be said that phonic decoding goes "from text to brain," while orthographic mapping goes "from brain to text." This is an oversimplification, however, because orthographic mapping involves "reciprocal bidirectional connections" (McKague, Davis, Pratt, & Johnston, 2008, p. 69). Nonetheless, this flow of information from pronunciation to letter sequence—a flow in the opposite direction from that found in phonic decoding—does not appear to be commonly understood outside the niche area of reading research that directly studies orthographic learning. Yet this flow of information is central to Ehri's theory (Ehri, 2005; Kilpatrick, 2015; Miles & Ehri, in press).

The result of this mapping process is a sequence of letters that is instantly familiar, stable, and highly unlikely to be confused with other words that look similar (e.g., *black, block, blank, blink, blind*). The fully specified representation is like a precise URL or "web address" within the memory system that activates the word's pronunciation and meaning the instant it is perceived (Ehri, 2005, 2014). Familiar written words are fully specified letter sequences that gain their familiarity by being bonded to the word's pronunciation at the phoneme/letter level, or in some cases the level of a group of letters (e.g., *-ight*; see more below on irregular words).

Phoneme segmentation and letter–sound knowledge work together to produce this orthographic mapping effect. For example, consider a first-grade girl who encounters the word *red* for the first time. If she is capable of segmenting the spoken word into its individual phonemes, /r/ /e/ /d/ (the letters between the slash marks represent the *sounds* associated with those letters and not the letters themselves), she then has three anchoring points in her long-term memory with which to attach that written letter sequence. She is attaching the new information (the letters in that word) to existing, well-specified information in her phonological long-term memory—namely, the segmented pronunciation of the word *red*. Again, notice that this represents the opposite direction of information flow from that required for phonic decoding. The net effect is that this particular letter sequence quickly becomes familiar because of the student's ability to associate the segmented phonemes in the spoken word's pronunciation to the written sequence designed to represent that spoken pronunciation.

By contrast, consider a first-grade boy who lacks proficient phoneme segmentation skills. When he sees the word *red*, how is he to remember it? If that student cannot pull apart the spoken pronunciation, then he cannot attach the spoken word *red* to that particular letter sequence. Most dyslexic students are able to create a connection between the first sound in the pronunciation and the first letter of the word. But beyond that, there is little opportunity to create a familiar sequence out of the rest of those letters because there is nothing in the child's long-term memory to which that letter string can be reliably anchored. Thus the student must sound it out or guess over and over upon seeing the word. With time and many, many exposures (not the one to four exposures found in typical readers), struggling readers map high-frequency words and other words (Ehri & Saltmarsh, 1995; Share & Shalev, 2004). The net effect for these weak readers, however, is that this orthographic mapping process is so inefficient that their sight word vocabularies grow very slowly relative to those of their peers, and they almost never catch up.

The Problem of Irregular Words

A question that arises is how orthographic mapping works with irregular words. English is the most inconsistent of all the major alphabet-based written languages (Seymour, Aro, & Erskine, 2003). Interestingly, however, the inconsistencies of English spellings create much less of a problem for orthographic mapping than they do for phonic decoding.

Orthographic mapping requires creating connections between pronunciations and print. Students cannot map a word unless they know what the word is—either because they sounded it out, they guessed it correctly, or someone told them what the word is. Orthographic mapping thus works from a starting point in which something is already known and already stored in phonological long-term memory (i.e., the word's pronunciation). By contrast, phonic decoding presumes that the word is not known, and thus does not start with any known anchoring point in long-term memory. Phonic decoding requires sufficient accuracy with the letter–sound sequence and blending to identify the spoken word correctly. Orthographic mapping does not require the same level of consistency as phonic decoding. Consider the irregular word *put*. Once a student knows that the written word he or she is looking at is *put*, it is a simple matter of noticing the association between the sounds in the spoken word and the letters. Two of the sounds in the spoken word *put* attach normally to their respective letters (*p, t*), and only one has an irregular connecting point (*u*). It is as if the student were to say "Oh, *that's* how we spell *put*!"

This type of adjustment to the mapping process for an irregular word is equally true for words that are phonically regular. For example, the word *make* is phonically regular, but requires an adjustment when it is being mapped into orthographic memory because *make* has three sounds but four letters. Knowing the silent-*e* rule presumably helps facilitate the adjusted mapping required for remembering such a word, but it requires an adjustment nonetheless. The same kind of adjustment needs to occur with phonically regular vowel and consonant digraphs (*ch, th, oa, ee*) because multiple letters represent a single sound. Also, such adjustments are routinely required in many multisyllabic words when an unstressed syllable has a vowel reduction, such as in *holiday* or *market*. The adjustments needed to map words to orthographic memory are routine for both regular and irregular words. These common adjustments are not problematic for students skilled in both letter–sound knowledge and phoneme segmentation. Yet they represent a major difficulty for those with the phonological-core deficit of dyslexia, due to their weaknesses in letter–sound skills and/or phonemic awareness.

Integrating Orthographic Mapping and the Self-Teaching Hypothesis

Elsewhere (Kilpatrick, 2015), I have made what may be the first formalized attempt to integrate the self-teaching and orthographic mapping models. On one level, this integration is straightforward. As proposed by the self-teaching model, students read and encounter new words. They perform phonological recoding, which activates the sounds of the letters in working memory. Ehri's theory then explains how the segmentation of that newly identified spoken word allows the reader to bond the segmented phonemes in the word's pronunciation to the printed letter sequence.

On another level, the integration of these two models requires a bit more thought. Throughout their elementary school years, readers add thousands of new words to their orthographic lexicons. However, this process appears to happen in the background, without conscious attention. It is doubtful that readers say with each new encounter of an unfamiliar word (let's say *clap*), "Hey, look how the /k/ sound maps onto the letter *c*, and how the /l/ sound I'm hearing next fits so well with that letter *l*," and so forth. Neither Ehri's nor Share's theory tries to account for how orthographic memory occurs without conscious effort or awareness. The fact that this process occurs is well supported by numerous lines of research. But this research does not explain why we do not seem to remember mapping the thousands of words we know. The *phonemic proficiency hypothesis* (Kilpatrick, 2015; Kilpatrick & Song, 2018) appears to have resolved this question. The phonemic proficiency hypothesis allows for a virtually seamless integration of Ehri's orthographic mapping theory with Share's self-teaching hypothesis, while accounting for the fact that the mapping process is largely outside the conscious awareness of the reader.

The Phonemic Proficiency Hypothesis

A colleague and I have proposed (Kilpatrick, 2015; Kilpatrick & Song, 2018) that phonemic proficiency, which is related to but not identical with phonemic awareness, is a critical aspect of efficient orthographic learning when Ehri's (2004, 2014) orthographic mapping hypothesis is integrated with Share's (1995) self-teaching hypothesis. At the same time, the phonemic proficiency hypothesis incorporates the research on the phonological-core deficit of dyslexia (Fletcher et al., 2018; Hulme & Snowling, 2009; Vellutino et al., 2004) with the orthographic learning theories of Ehri and Share. As mentioned, Ehri (2005) proposes that a phoneme analysis mechanism (i.e., segmenting words into phonemes) is required for orthographic memory. However, for that to occur within the very time-limited context of Share's self-teaching opportunities (correctly sounding out a word takes very little time), phonemic segmentation/analysis must be highly proficient and largely unconscious. The phonemic proficiency hypothesis (Kilpatrick, 2015) suggests that proficient letter–sound skills and proficient phonemic skills both involve automatic processes that are precognitive and do not require conscious awareness.

Letter–Sound Proficiency

Studies have shown that by late first grade, typically developing readers can *instantly* respond to CVC nonsense words, such as *mot, tam,* or *gub* (e.g., Harn et al., 2008). Anyone who has administered the Phonological Decoding subtest from the Test of Word Reading Efficiency—Second Edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012) to an average student at the end of first grade has directly experienced this. Consider what is involved for first graders to respond instantly to a CVC word, such as *mip*. In less than a second, they retrieve the sounds for the letters *m, i,* and *p,* and then blend those three sounds together. It is argued that those children do not use a conscious search process to retrieve those letter sounds, but that they are automatically available. This instant responding illustrates letter–sound proficiency: It involves automatic, unconscious access to the most common sounds of the letters, plus proficient phonological blending that allows those letter sounds to be accurately pronounced as a single, spoken unit (Harn et al., 2008). Due to its greater complexity, those first graders may not be able to respond instantly to the nonsense word *splenk*. But by the end of second grade, average students can do so, given their additional year of development of their letter–sound skills. Those second graders have instant access to letter sounds even when they encounter a complex string of letters. No conscious effort is involved in retrieving those letter sounds.

Phonemic Proficiency

Phonemic proficiency can be viewed as an advanced form of phonemic awareness. *Phonemic awareness* has been generally conceptualized as the ability to be aware of and/or manipulate phonemes within words. It is a latent construct that has been assessed in many ways with a variety of tasks, including segmentation, isolation, categorization, deletion, and substitution (Kilpatrick, 2012a, 2012b). Only recently has any effort been made to examine whether some phonemic awareness tasks are better suited than others for assessing the phonemic substrates of reading (Kilpatrick, 2012a, 2015). It turns out that phoneme manipulation tasks, the most common being phoneme deletion and substitution, correlate more strongly with reading than phoneme segmentation and blending tasks do (Catts, Fey, Zhang, & Tomblin, 2001; Kilpatrick 2012a, 2012b, 2015; Swank & Catts, 1994; Wagner, Torgesen, & Rashotte, 1999). Phoneme manipulation "ranks highly among phonological awareness tasks in predicting reading achievement" (Catts et al., 2001, p. 40).

Interestingly, very little attention has been paid to the speed of phonemic awareness task responses. We (Kilpatrick & Song, 2018) reviewed the very limited pool of studies on this. The findings from these studies indicate that using timed manipulation tasks, researchers discovered that phonemic awareness continues to develop well into third and fourth grade and appears to display continued influence on reading development well beyond first grade (e.g., Vaessen & Blomert, 2010). This contrasts with the common assumption that phonemic awareness plays no substantive role in reading development after early first grade (e.g., O'Connor, 2011). Evidence for a causal role in reading development for these more "advanced" phonemic skills comes from a recent review of the word reading intervention literature (Kilpatrick, 2015; Kilpatrick & Van den Broeck, 2016). Studies that rigorously trained students by using manipulation tasks (phoneme deletion and substitution) produced gains in real-word reading that ranged from 12 to 25 standard score points. By contrast, studies that trained phonemic awareness skills by using the more "basic" phonological awareness skills of phoneme segmentation and/or blending yielded increases in standard scores ranging from 6 to 9 points. Studies that incorporated no phoneme awareness training yielded increases of 0–6 standard score points in word-level reading (Kilpatrick, 2015; Kilpatrick & Van den Broeck, 2016). Noteworthy is the fact that socioeconomic status, age of the students, group size, severity of the problem, and total length of the intervention were evenly distributed across these three groups of studies with varying results. This indicates that these factors cannot explain the disparity in outcomes (cf. Flynn, Zheng, & Swanson, 2012; Torgesen, 2004b; Torgesen et al., 2003).

Phonemic proficiency goes beyond the conventional conceptualization of phonemic awareness and can account for the findings from the intervention research just mentioned. Phonemic proficiency, parallel to letter–sound proficiency, is conceptualized as the automatic, unconscious access to the phonemes in spoken words. This is more appropriately assessed via a manipulation task than a segmentation task. For example, in a segmentation subtest, all of a student's focus is on that task, so it is difficult to determine how automatic are the cognitive processes behind the task responses. However, manipulation tasks are more complex. A second grader with phoneme proficiency can respond in 1 second or less to a request to delete the /l/ from the spoken word *clap*. To do this, the student has to perform four classic phonemic awareness tasks in less than 1 second. First, he or she has to *segment* the word *clap*. Then the student has to perform *phoneme isolation*, which involves locating where the target sound appears on the word ("Is the /l/ in the beginning, middle, end . . ."). Next, he or she has to delete (*manipulate*) the sound. Finally, the student has to *blend* the remaining sounds to produce the correct response. Thus, four traditional phonemic tasks—segmentation, isolation, manipulation, and blending—all occur in 1 second or less. I have contended (Kilpatrick, 2015) that for the student to perform those four operations that quickly, it is likely that access to the phonemes via segmentation does not require conscious effort, but is automatic. This is the essence of phonemic proficiency.

We (Kilpatrick & Song, 2018) have provided some evidence for phonemic proficiency and its role in word learning. In one study, 136 first graders were administered a phoneme manipulation task (a mix of deleting and substituting sounds). Correct responses were coded differently, depending on whether those responses occurred in less than or more than 2 seconds. These students were also administered the *Sight Word Efficiency* subtest from the TOWRE-2 (Torgesen et al., 2012). This subtest consists of a graded word list, and students have 45 seconds to read as many words as possible. The inference is that students with larger sight vocabularies will get higher scores than those with smaller sight vocabularies because it takes longer to sound out a word than to recognize a known word. We found that the correlation between this reading task and phonemic awareness items responded to instantly (i.e., in 2 seconds or less) was $r = +.58$. Yet the correlation between the reading task and the non-instant phonemic awareness responses was $r = +.004$. This suggests that instant access to the sounds in words tells us something about word-reading development that is not captured by correctly responding to a phoneme task without evidence of phonemic proficiency.

We (Kilpatrick & Song, 2018) also examined this phenomenon with 58 typical fifth-grade readers. To evaluate the impact of phonemic awareness on sight vocabulary, we used a reading test that only contained irregular words (from Adams & Huggins, 1985), like *iron, tongue, suede,* and *yacht*. The assumption is that sounding out these words is likely to yield an incorrect response, so the test assesses prior familiarity with those words. The inference is that those with higher scores are likely to know more words in general (i.e., to have a larger sight word vocabulary). The same phonemic task was used as with the first graders. Coincidentally, the correlation between the instant responding on the phonemic awareness task and the reading measure was, again, $r = +.58$. But the correlation with the non-instant responding was $r = –.25$, suggesting that even among a population of typical fifth-grade readers, those with presumably larger sight vocabularies had greater phonemic proficiency than those with presumably smaller (though average) sight vocabularies. Thus, even in a population of typical fifth-grade readers, the degree of phoneme proficiency correlated with the likelihood of identifying phonically irregular words, which is a fairly direct assessment of the orthographic lexicon.

Phonemic Proficiency and Orthographic Learning

Students routinely encounter new and unfamiliar words while reading silently. They use their letter–sound skills and phonological blending to determine the word. This is Share's self-teaching scenario. Once the word is correctly determined, the pronunciation of the word is activated. Students with automatic, unconscious access to the sounds within that word's pronunciation can implicitly map those sounds within the pronunciation to the letter sequence representing that pronunciation, as orthographic mapping theory suggests. Phonemic proficiency allows the mapping process to be unconscious, given that the two subprocesses involved in mapping are unconscious and automatic (i.e., letter–sound proficiency and phonemic proficiency). This explains why most of us would have no recall of consciously making connections between pronunciations and letter patterns while we were learning the tens of thousands of words we know.

The Development of Word-Learning Skills

I have proposed a description of the interaction between the developmental of reading-related phonological skills and word-level reading (Kilpatrick, 2015). This developmental paradigm is presented in Table 35.2. The left side of the table portrays three levels of phonological development, while the right side depicts three levels of word-reading development. It is proposed that the phonological skills directly to the left of the given reading skills represent causal factors for that level of reading. Additionally, each level of reading development has a causal relationship with the next level of phonological development. This reciprocal, causal relationship was first established empirically by Perfetti, Beck, Bell, and Hughes (1987).

Early phonological skill development appears to have a causal relationship with the speed and efficiency with which children develop knowledge of letter names and sounds. The "softer" evidence for a causal relationship is found in studies that examined phonological skills before children learned letter names and sounds. Those with stronger early phonological skills learned letter names and sounds more quickly than those with weaker phonological skills (Cardoso-Martins, Mesquita, & Ehri, 2011; Share, 2004a). Harder causal evidence comes from experimental studies in which children provided with early phonological awareness training outperformed untrained children in the acquisition of letter names and letter sounds (Cardoso-Martins et al., 2011; Williams, 1980).

TABLE 35.2. Developmental Levels of Phonological Awareness and Word Reading

Level of phonological awareness	Level of word-reading skill
1. *Early phonological awareness* Rhyming, alliteration, first sounds, and syllable segmentation	1. *Letters and sounds* Requires simple phonology to learn letter names and letter sounds
2. *Basic phonemic awareness* Blending and segmentation	2. *Phonic decoding* Requires letter sounds and blending
3. *Advanced phonemic awareness* Phonemic proficiency	3. *Orthographic mapping* Requires letter–sound skills and advanced phonemic awareness/proficiency

Learning letter sounds is causally related to the development of basic phoneme-level awareness. We know this from studies of adults who, due to lack of opportunity, never learned to read. These individuals do not naturally develop phoneme-level awareness (Morais, 1991). There then appears to be a causal relationship between the development of phoneme-level awareness and blending, and that of phonic decoding and basic spelling. These basic phoneme-level skills are typically developed by the end of first grade. It is often at that point that phonemic awareness assessments (e.g., Dynamic Indicators of Basic Early Literacy Skills, Aimsweb, easyCBM) and training programs (e.g., Blachman, Ball, Black, & Tangel, 2000) discontinue phonological/phonemic awareness training. This appears to assume that any further phonemic awareness development that occurs after first grade is of no consequence for reading. Yet this is not the case (Ashby, Dix, Bontrager, Dey, & Archer, 2013; Kilpatrick, 2015; Torgesen et al., 2001; Truch, 1994; Vaessen & Blomert, 2010). Indeed, practicing phonic decoding/letter–sound skills and spelling throughout first and second grades appears to make these segmenting and blending skills more automatic and efficient. This suggests a causal factor in the development of more "advanced" phonemic awareness skills. It is these advanced skills, as demonstrated by instant responses to phoneme manipulation tasks, that provide the phonemic proficiency to drive orthographic mapping skills and thus rapidly expand the sight word vocabulary.

Impaired Development in Dyslexia

Students with the phonological-core deficit, which is the basis of dyslexia (Ahmed, Wagner, & Kantor, 2012; Vellutino et al., 2004), do not move smoothly through the levels of phonological development or reading development depicted in Table 35.2. They typically have poor early phonological skills, which is why they lag behind their peers in developing letter–name and letter–sound knowledge. When their letter–sound skills do develop, their phonological systems are not efficient enough for the learning of those letter sounds to prompt the next level of phonological skills, that is, phoneme segmentation and blending. However, even many children with dyslexia will develop these "basic" phonological skills by late second or third grade (recall that typically developing readers have these skills in place by late first grade). With proper instruction, children with dyslexia who have basic segmentation and blending skills can benefit from phonics instruction. However, when these children learn phonics and spelling skills, these skills do not naturally prompt the more "advanced" phonemic skills needed for orthographic mapping. Thus children with the phonological-core deficit only develop the phonological skills to the level they are directly taught. They do not develop those skills via reading instruction, like their typically developing peers.

IS THERE A NEED TO DIAGNOSE SUBTYPES OF DYSLEXIA?

There are three very well-established subtypes of reading disabilities (Fletcher et al., 2018; Gough & Tunmer, 1986; Hulme & Snowling, 2009)—namely, *dyslexia*, *hyperlexia*, and a combined type (traditionally called *garden-variety poor readers*; Gough & Tunmer, 1986). Dyslexia refers to poor word reading despite adequate language skills. Hyperlexia refers to skilled word reading but weak reading comprehension, typically due to oral language comprehension difficulties. The combined type refers to problems in word reading and oral language comprehension. Distinguishing among these three types of reading problems is essential for designing a given student's remedial instruction. Students with dyslexia and hyperlexia do not make good small-group partners. Their strengths and weaknesses in reading have no functional overlap.

Subtypes Based on Rapid Automatized Naming

Although there are three empirically derived subtypes of reading disabilities, with dyslexia being one of them, efforts to subdivide dyslexia into valid subtypes have been problematic on many levels. The most popular subtyping approach in the empirical reading research distinguishes among dyslexic students based on the presence or absence of poor phonemic awareness and poor rapid automatized naming, or RAN (Wolf et al., 2002). The subtypes involve the presence of one or the other or both, the latter being referred to as the *double deficit*. The presumption has been that students with problems in both have more severe word-reading difficulties. However, that may not be the case, as students with a severe single deficit in phonemic awareness can have greater difficulties than students with more moderate problems in both phonemic awareness and RAN (Vukovic & Siegel, 2006). The status of subtypes based on these characteristics is still under investigation. Moreover, there is no clear, empirically based protocol for distinguishing among these subtypes when it comes to planning instruction.

Despite these uncertainties, it is still recommended that practitioners invest the 2–4 minutes of total administration time involved in tests of RAN when evaluating struggling readers. Table 35.3 lists six reasons for including assessments of RAN and working memory (WM) in any evaluation of struggling readers. One of these relates to the fact that the double deficit tends to suggest poorer outcomes from milder interventions (i.e., Tier 2). Research has shown that for some students, skipping Tier 2 of RTI and going directly to Tier 3 provides better outcomes for such students than requiring students to demonstrate poor progress at Tier 2 before trying a more intensive Tier 3 remediation (Al Otaiba et al., 2014).

TABLE 35.3. Rationale for Including Rapid Automatized Naming (RAN) and Working Memory (WM) in Reading Evaluations or Universal Screenings

1. They are good predictors of later reading skills.
2. They are good predictors of how well students will respond to reading interventions.
3. They help evaluators and teachers understand *why* students struggle in reading. This is particularly true when students appear to have adequate phonemic awareness and phonic decoding skills, but still struggle with word identification and fluency.
4. They affect how one interprets a student's larger profile. If either RAN or WM is weak, one can anticipate the need to build stronger phonemic awareness skills in struggling readers. For example, a phonemic awareness scaled score of 9 (37th percentile) on the Elision subtest from the CTOPP-2 may be adequate for students with RAN and WM scores of 10 or higher, but inadequate for students with either a RAN or WM score of 7 or lower. The latter students should receive phonemic awareness instruction to compensate for the negative impact that weak RAN or WM is likely having on reading. Several studies that showed large improvements in phonemic awareness and word reading also showed substantial improvements in RAN and WM performance (Kilpatrick, 2015). This was the case even though RAN and WM were not directly addressed in the intervention.
5. Knowing that a student has a WM weakness in particular can affect the choice of remedial strategies. The classic special educational strategies of multiple repetitions and multisensory tasks are based upon decades of clinical experience with struggling students, a large proportion of whom have WM difficulties. Such strategies are not quite so necessary for students who struggle academically but have average or better WM. Knowing a student's WM skill level can thus influence the selection of intervention techniques.
6. The presence of poor RAN and WM increases the validity of an SLD diagnosis in students with reading problems, given the capacity of weaknesses in these skills to predict future struggles in reading and weaker RTI response.

These six possible advantages can justify the brief assessment time involved in administering RAN and WM subtests in reading evaluations and including them in universal screenings.

Subtypes Based on the Dual-Route Model of Reading

In recent years there has been increased discussion of subtypes of dyslexia based on the dual-route model of reading (Feifer, 2011, 2014; Mather & Wendling, 2012), and even a new reading test battery that, in fair measure, is designed to distinguish among these subtypes (Feifer & Nader, 2015). Before the validity of this popular subtyping model is considered, two broader categories of dyslexia must be distinguished: *acquired dyslexia* and *developmental dyslexia*.

Acquired Dyslexia

Acquired dyslexia refers to a situation in which a skilled reader (typically an adult) loses all or some of his or her reading ability as a result of a stroke, head injury, or other neurological condition. Acquired dyslexia was first described in clinical cases in the early 1970s (Marshall & Newcombe, 1973). Although persons with acquired dyslexia showed a variety of reading related difficulties, some displayed one of three subtypes: *surface dyslexia* (Patterson, Marshall, & Coltheart, 1985), *phonological dyslexia* (Coltheart, 1996), or *deep dyslexia* (Coltheart, Patterson, & Marshall, 1980).

Individuals with surface dyslexia struggle to instantly recognize words that were previously familiar to them, but they can sound out phonically regular words and nonsense words. By contrast, those with phonological dyslexia remember the words they previously learned before the neurological incident, but can no longer read nonsense words or sound out new words. Individuals with deep dyslexia are similar in some respects to those with phonological dyslexia but have more varied symptomatology, including a tendency to make semantic errors, such as reading "truck" for *bus*. These acquired dyslexia subtypes are well-established clinical syndromes, even though most individuals with acquired dyslexia do not fall into these distinct subtypes.

This distinction among these types of dyslexia was instrumental in developing the *dual-route model of reading*. The dual-route model acknowledges that some words are not familiar to the reader and must be read by phonological recoding. This is called reading by the *phonological route*. Other words are familiar to the reader, and these words are read instantly, without conscious effort. This is called reading by the *direct route*. These two routes parallel the deficits found among some of those with acquired dyslexia. Individuals with phonological dyslexia still have access to the direct route but struggle immensely with the phonological route, while those with surface dyslexia display the opposite pattern.

It must be pointed out that the dual-route model describes two different ways of *reading* words, not two different ways of *learning* words. These two routes do not translate into two reading strategies. Familiar words are instantly and effortlessly recognized, so no strategy is involved. By contrast, the phonological route uses the strategy of phonic decoding. But the result is that the word is read/identified, but not necessarily learned. The dual-route model makes no presumptions about how unfamiliar words become familiar. Nor does this model tell us how one becomes skilled with the phonological route. It must be emphasized that the dual-route model long predates the more recent advances in our understanding of orthographic learning, described earlier in this chapter.

The dual-route model is not a useful instructional framework. To be useful instructionally, we need a framework that allows us to understand the development of the skills needed for children to become good orthographic mappers. These skills will allow students to efficiently remember more and more words, and thus read more words via the direct route. We also need to know the best way for struggling readers to develop the skill of accurately sounding out unfamiliar words and thus read via the phonological route when encountering new words. The dual route model provides no answers here. As a result, a healthy skepticism must be applied when one seeks to superimpose the subtypes of acquired dyslexia onto developmental dyslexia.

Developmental Dyslexia

Developmental dyslexia, by contrast, refers to a situation in which an individual has never developed typical reading skills, despite adequate opportunity and effort. Unlike those with acquired dyslexia, individuals with developmental dyslexia have not mastered phonological recoding or orthographic mapping. Is there evidence to suggest that some children can develop orthographic mapping without developing phonological recoding (phonological dyslexia) while other children develop the opposite pattern (surface dyslexia)? Can the subtype distinction found in clinical populations of adults with acquired dyslexia be validly superimposed onto cases of developmental dyslexia in children?

This question has been investigated in the research literature for decades. The consensus among reading researchers is that distinguishing between phonological and surface dyslexia as subtypes of developmental dyslexia is not well supported empirically. Multiple teams of reading researchers have reviewed the studies that attempt to make such a distinction and do not find convincing evidence that such a distinction can or should be made (Ahmed et al., 2012; Fletcher et al., 2018; Hulme & Snowling, 2009; Van den Broeck & Guedens, 2012; Vellutino et al., 2004).

There is no attempt here to provide a review of the vast subtyping literature. However, listed below are some of the major problems with the understanding that the surface and phonological patterns represent valid subtypes of developmental dyslexia.

First, as mentioned, this subtyping scenario superimposes an adult, neuropathology-based model onto children who do not display similar neurological conditions. The dual-route model describes the two “routes” of word identification among *skilled* readers. It does not inform us about how those routes develop, which is precisely what needs to be addressed if we are to properly understand developmental dyslexia. The phonological versus surface subtyping model treats the dual-route theory as a word-learning theory when it is actually a “finished-product” theory; that is, it describes the finished product of skilled reading. As a result, using the dual-route model for understanding developmental dyslexia appears to be inherently problematic.

Second, the evidence in favor of the phonological versus surface dyslexia subtypes has been mixed at best, and those results have often depended on the specific type of research methodology used (for more detail, see below). At worst, after adjustments for the methodology, the distinction between those subtypes virtually disappears (Van den Broeck & Guedens, 2012; Van den Broeck et al., 2010). However, despite the fact that the empirical reading research field remains rather skeptical of a distinction between phonological and surface subtypes of dyslexia, some authors in the areas of school psychology and neuropsychology seem to present this subtyping scenario as if it were a well-established phenomenon, and little or no mention is made of the controversy surrounding its existence (e.g., Feifer, 2011, 2014; Mather & Wendling, 2012).

Third, initial and subsequent attempts to find developmental surface and phonological dyslexia have used chronological-age (CA) controls (Castles & Coltheart, 1993; Heim et al., 2008). Such research designs yield results suggesting that a portion of students fit the phonological and surface subtypes, while most exhibit the mixed type. But critics have pointed out major confounds in using CA control groups (e.g., Stanovich, Siegel, Gottardo, Chiappe, & Sidhu, 1997; Van den Broeck & Guedens, 2012). As a result, there has been a shift to including reading-age (RA) controls—for example, matching fifth graders who are reading at a second-grade level with average second grade readers. For two decades, the reading research field considered this a more valid comparison because it removed some of the confounds associated with the CA matched design. When RA controls are used, fewer students fit the phonological dyslexia subtype, more fit the mixed profile, and the surface dyslexia subtype virtually disappears (Stanovich et al., 1997; Van den Broeck & Guedens, 2012).

More recently, Van den Broeck and colleagues (Van den Broeck & Guedens, 2012; Van den Broeck et al., 2010) have demonstrated that, like the CA control design, the RA control design has significant confounds, and the design itself may *produce* the phonological dyslexia subtype rather than *reflect* an actual subtype. They have pointed out that in subtyping studies, it is most common for dyslexic children in fourth through sixth grades to be compared with RA control second graders on word identification tests. In such matches, it is common to find a substantial portion of dyslexic children to have lower rates of nonsense-word reading than the second-grade typical readers used in the comparison. This is taken as evidence for the phonological dyslexia subtype.

The problem with this design is that it fails to account for the fact that the older dyslexic children have had 2–4 more years of instructional experience and exposure to reading. Such experience allows them to eventually learn many common second-grade-level words and thus receive a score comparable to typical second graders on a word identification test. But their actual phonological skills that underlie reading remain weak, as reflected in their poor nonsense-word reading. In addition, based on the fact that these older students have a larger vocabulary than their younger controls, they have better use of set for variability, discussed above, to respond correctly to words on word identification subtests. As a result, matching a fifth-grade dyslexic reader and a second-grade typical reader with the same word identification raw score confounds age, experience, and set for variability. Such confounds create the pattern of phonological dyslexia because the older children sound out words more poorly than their normative word-reading scores would suggest. The apparent cases of phonological dyslexia in these studies thus seem to be an artifact of the confounded research design.

To address the issue of the CA control and RA control designs, Van den Broeck and colleagues have developed two ingenious and sophisticated designs that avoid these confounds without creating new confounds. With these non confounded designs, the phenomena of phonological dyslexia and surface dyslexia virtually disappear. Rather, these authors argue for a developmental explanation in which different continuous skill levels in phonemic awareness, reading experience, and compensating factors all interact differently at different ages to produce the variability we see among children with dyslexia. It is this variability that has been traditionally interpreted as phonological versus surface subtypes of dyslexia (Van den Broeck & Guedens, 2012; Van den Broeck et al., 2010). It is worth pointing out that the findings of Van den Broeck and colleagues are consistent with the orthographic learning theories of Ehri and Share, described above (Van den Broeck & Guedens, 2012; Van den Broeck et al., 2010). By contrast, the conventional phonological versus surface dyslexia subtyping is not consistent with the orthographic learning literature.

A fourth problem is related to the previous point. The variations in nonsense-word and irregular-word reading performance found among individuals with dyslexia that have formed the basis of the proposed dyslexia subtypes can be better explained via an updated model of word-reading development, such as the one described earlier in this chapter. Students who might be considered to have phonological dyslexia are typically older students who can instantly identify an array of common words that have been mapped via orthographic mapping, albeit after many, many more exposures than would be needed by typically developing readers (Ehri & Saltmarsh, 1995; Share & Shalev, 2004). In studies of dyslexia, these students' pools of mapped words are not at grade level. Yet these students struggle with nonsense-word reading because they have not adequately developed the phonological skills listed on the left side of Table 35.2. Students who are considered to have surface dyslexia can be better understood by recognizing that they have typically received phonic-based instruction and they have only developed to the second level of phonemic awareness development and reading development in Table 35.2. Such students can sound out words, but are not efficient at orthographic mapping, since they have not developed the more advanced phonemic awareness skills needed for efficient orthographic mapping. Consistent with this interpretation is the finding that subtyping studies have found that children with both phonological and surface dyslexia display below-average phonemic awareness skills, indicating that poor phonemic awareness skills are found in those alleged to have surface dyslexia as well as those alleged to have the phonological subtype, but typically to a milder degree (Hulme & Snowling, 2009). This is consistent with Table 35.2 and with the developmental explanation provided by Van den Broeck and colleagues.

Fifth, the dual-route subtyping issue does not seem to reflect more recent work on the nature of the orthographic tasks that gave rise to the notion that orthographic skills should be considered separate reading-related subskills. As described above, more recent research has suggested that orthographic task performance is a by-product of phonological skills, letter–sound skills, and reading experience. It does not appear to be an independent reading-related subskill or to have a causal relationship with early reading development. Yet presentations regarding the phonological and surface dyslexia subtypes appear to assume this older understanding of the nature of orthographic skills.

Sixth, related to the previous point, is that the division of dyslexia into phonological and surface subtypes appears to assume that some form of visual memory plays a significant role in word-level reading. For example, one author describes a child with surface dyslexia as a student who has "an under-reliance upon the orthographical or spatial properties of the visual word form" (Feifer, 2014, p. 157). It is unclear precisely how the term "orthographical" is being used in this statement, but we have no evidence that any spatial or visual word "form" properties are involved in the way skilled readers learn words as unitized wholes. This issue has been addressed extensively above.

More problematic in this regard is the instructional advice that would result from such a notion. After making a major contribution to our knowledge of the neuropsychological substrates of learning new words, Glezer and colleagues (2015) lapse into speculation when they say, "These findings have interesting implications for reading remediation in individuals with phonologic processing impairments because they suggest the possibility that these individuals might benefit from visual word learning strategies to circumvent the phonologic difficulties and directly train holistic visual word representations in the VWFA [visual word form area]" (p. 4971).[3] It is no coincidence that they did not cite a study to support this instructional suggestion because it appears likely that no such study exists. Their suggestion is entirely intuitive and without an empirical basis. By contrast, Truch (1994) reports that out of 281 individuals with phonological-core dyslexia ages 5–55, only one did not make progress in phonological awareness training, and 70% reached an average or above-average level of phonological awareness skills. Truch noted that for those students whose phonics skills were not moving forward, training in advanced phonemic awareness resulted in dramatic gains in both phonic decoding and sight word learning. So Truch provides rather extensive and direct evidence (as do others—e.g., Torgesen et al, 2001; Vellutino et al., 1996) that "phonologic processing impairments," as Glezer and colleagues call them, can be successfully overcome in more than 99.5% of those with such "impairments" (i.e., 1 out of Truch's 281 equals less than 0.5% failure rate among those with phonological-core dyslexia). Thus there is no need to suggest speculative ideas about "visual word learning strategies" that have no demonstrated efficacy and run counter to our current empirical understanding of both normal word-learning skills and effective word-reading intervention.

Conclusions Regarding Dyslexia Subtypes

From the previous considerations, it would seem that the proposed phonological and surface subtypes of dyslexia do not have a well-established empirical foundation. This conventional subtyping model does not reflect research advances in the last 20 years regarding word-reading development, orthographic skills, and the role (or lack thereof) of visual skills in reading; nor does it take account of the research literature on word-reading intervention effectiveness. Despite its recent popularity in the field of school psychology, practitioners should not feel the need to attempt to establish dyslexia subtypes when evaluating students who struggle in word-level reading.

IMPLICATIONS OF RECENT ADVANCES FOR DIAGNOSING READING DISABILITIES

This chapter has presented numerous concepts and research related to understanding the nature of word-level reading difficulties. Several implications that can be drawn from these concepts and research may inform decisions about the presence or absence of a reading difficulty, and whether such a difficulty (if present) rises to the level of being considered an SLD.

One of the key themes of the chapter is that we must not just look at isolated word reading and phonic decoding, but must work from a broader understanding of word-reading development—from letter–sound knowledge, to phonic decoding and spelling, to the size of a student's orthographic lexicon. A student can arrive at a given profile of test scores via multiple routes, and the hope is that the material in this chapter will allow practitioners (1) to examine multiple possibilities to determine the nature of a student's reading struggles and (2) to take a proper perspective on interpreting a profile of test scores.

It will be important to consider the relationship between a student's language skills and his or her word identification subtest performance. It is important to acknowledge that students with stronger vocabulary skills can create the impression that their word reading is stronger than it really is, due to their ability to use set for variability in responding to conventional word-reading subtests. It will be important to put more weight on nonsense-word reading subtests for such students.

Timed tests of real words and nonsense words arguably provide a better indication of a student's sight vocabulary and his or her letter–sound proficiency, two hallmarks of skilled reading. It is much harder to compensate under timed conditions than on untimed reading subtests. Tests like the TOWRE-2, the Test of Silent Word Reading Efficiency—Second Edition, or the timed real- and nonsense-word subtests from the Kaufman Test of Educational Achievement—Third Edition (KTEA-3) can be very valuable in this regard. Personal experience with these tests suggests that the real-word versions of these tests are very useful at the elementary level, but less so at the secondary level. This is because the word difficulty is not challenging enough on the TOWRE-2 Sight Word Efficiency subtest, and the timing is not long enough to get to the more difficult words on the KTEA-3 timed word-reading subtest. At the elementary level, however, this type of subtest may provide the most valid assessment of the size of a student's sight vocabulary (i.e., below-average performance means a limited orthographic lexicon whereas an average or better score suggests an average or larger orthographic lexicon).

The 3-second timing on the word identification subtest from the Wechsler Individual Achievement Test—Third Edition is too long to assess automaticity. Also, the timed sentence-reading tasks found in the Woodcock–Johnson IV Tests of Achievement, the Woodcock Reading Mastery Test—Third Edition, and the KTEA-3 use common words that older students eventually map to memory, so it does not adequately address their fluency or automatic word recognition with grade-level reading material. *Such subtests should not be used to "rule out" a reading difficulty with students beyond about fourth grade.* However, if an older student has a low score on such a subtest, that indicates a very limited sight vocabulary.

RAN and WM should also be considered as part of any evaluation of students who struggle with word-level reading. As described in Table 35.3, such tests are quickly administered and have multiple advantages in understanding a student's reading challenges.

Two skills needed for orthographic mapping are letter–sound proficiency and phonemic proficiency. The former can be assessed with the TOWRE-2 Phonetic Decoding subtest (valuable at all age levels) and the timed nonsense-word subtest from the KTEA-3. However, at this writing, there are no commercially available tests for phonemic proficiency. The CTOPP-2 is highly recommended for assessing phonological awareness, RAN, and phonological WM. It is highly recommended and should be a central component in any assessment of a student with word-level reading difficulties. However, the phonological awareness subtest (Elision) is untimed. Universal screeners have timed phonological awareness tasks, but they do not go beyond the first-grade level of skill and thus do not address phonemic proficiency. The *Phonological Awareness Screening Test (PAST)*[4] is free and is designed to assess phonemic proficiency. It is available from *www.thepasttest.com* or from *kilpatrickd@cortland.edu*.

The practitioner's greatest assessment tool is a strong knowledge base regarding the nature of typical word-reading development and the sources of reading difficulties. With such a knowledge base, evaluators can more appropriately select and interpret tests of word-level reading and related skills (i.e., phonemic awareness, RAN, WM, spelling, and vocabulary). Such evaluations should yield more accurate representations of a student's skills, which should lead to better decisions regarding the next step in addressing the student's reading difficulties. The next step may involve general educational intervention within an RTI/multi-tiered system of support framework, or it may involve a designation of SLD. Regardless of which route is taken, it will be important to incorporate the highly effective reading intervention approaches that prompted the development of RTI in the first place (Kilpatrick, 2015). These approaches allowed a large portion of struggling readers to "catch up" with their typically developing peers. And for most of those who did not catch up, they developed better reading skills than they would have if traditional remedial approaches had been used.

NOTES

1. This must not be misinterpreted to mean that general visual sequencing skills underlie our memory for written words. Only letter sequences based on phonology appear to be involved. Skill at recalling sequences of shapes or even nonpronounceable letter sequences (e.g., *ZNWRT*) do not appear to relate to reading like phonologically based, pronounceable letter sequences.

2. Actually, these examples do not truly represent visual memory failures, which is why the term *visual memory failure* is given in quotation marks. Rather, they represent failures in phonological retrieval. A true visual memory failure would involve failure to recognize something as visually familiar. In other words, rather than just failing to come up with the name of an acquaintance, it would involve not even recognizing the person visually as someone we had ever seen before.

3. It is an unfortunate quirk of reading research history that with the discovery that the left fusiform gyrus area is activated when familiar words are seen, this area was improperly named the *visual word form area*. We have no evidence to suggest that the visual form of the word plays any role in the initial storage or subsequent activation of known words. There is ample evidence to show it is *the precise letter order* that is instantly recognized in known words, as a holistic letter sequence. Thus *bear, BEAR,* bear, BEAR, ***bear, BEAR***, and even *bEaR* all provide the same activation—as a holistic, familiar letter sequence—because they all represent the same letter order, despite their dramatically different visual word forms. Interestingly, Glezer and colleagues (2015) showed in their study that the now familiar sequences were all processed first phonologically before they became unitized, orthographically familiar wholes. There was nothing in their study to suggest that phonology can be bypassed in this learning process, nor is there anything in the broader reading research literature to suggest this.

4. This is not to be confused with another phonological awareness test using the same acronym, PAST, that turns up on Internet searches. This other test, called the Phonological Awareness *Skills* Test. It takes a different approach to phonological awareness assessment and does not assess phonemic proficiency.

REFERENCES

Aaron, P. G., Joshi, R. M., Ayotollah, M., Ellsberry, A., Henderson, J., & Lindsey, K. (1999). Decoding and sight-word naming: Are they independent components of word recognition skill? *Reading and Writing: An Interdisciplinary Journal, 11,* 89–127.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adams, M. J., & Huggins, A. W. F. (1985). The growth of children's sight vocabulary: A quick test with educational and theoretical implications. *Reading Research Quarterly, 20*(3), 262–281.

Ahmed, Y., Wagner, R. K., & Kantor, P. T. (2012). How visual word recognition is affected by developmental dyslexia. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 2. Meaning and context, individuals and development* (pp. 196–215). New York: Psychology Press.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, S., et al. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*(1) 11–27.

American Federation of Teachers. (1999). *Teaching reading IS rocket science.* Washington, DC: Author.

Anderson, M., & Reid, C. (2009). Don't forget about levels of explanation. *Cortex, 45,* 560–561.

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly, 31*(1), 42–52.

Ashby, J., Dix, H., Bontrager, M., Dey, R., & Archer, A. (2013). Phonemic awareness contributes to text reading fluency: Evidence from eye movements. *School Psychology Review, 42*(2), 157–170.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Blachman, B., Ball, E., Black, R., & Tangel, D. (2000). *Road to the code: A phonological awareness program for young children.* Baltimore: Brookes.

Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2*(4), 5–142.

Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin and Review, 7*(1), 83–99.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 92,* 203–219.

Bowey, J. A., & Underwood, N. (1996). Further evidence that orthographic rime usage in nonwords reading increases with word-level reading proficiency. *Journal of Experimental Child Psychology, 63,* 526–562.

Brady, S. A. (2011). Efficacy of phonics teaching for reading outcomes: Indications from post-NRP research. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp. 69–96). New York: Psychology Press.

Burt, J. S. (2006). What is orthographic processing skill and how does it relate to word identification in reading? *Journal of Research in Reading, 29*(4), 400–417.

Cardoso-Martins, C., Mesquita, T. C. L., & Ehri, L. (2011). Letter names and phonological awareness help children to learn letter–sound relations. *Journal of Experimental Child Psychology, 109,* 25–38.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47,* 149–180.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind, 44,* 524–538.

Catts, H. W., Fey, M. E., Zhang, Z., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50.

Coltheart, M. (1996). Phonological dyslexia: Past and future issues. *Cognitive Neuropsychology, 13*(6), 749–762.

Coltheart, M. (2012). Dual-route theories of reading aloud. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology* (pp. 3–27). New York: Psychology Press.

Coltheart, M., Patterson, K., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia.* London: Routledge & Kegan Paul.

Crowder, R. G. (1981). *The psychology of reading: An introduction.* New York: Oxford University Press.

Cunningham, A. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology, 95,* 56–77.

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology, 82,* 185–199.

Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg?: Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition, 122,* 110–117.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262.

Dixon, M., Stuart, M., & Masterson, J. (2002). The relationship between phonological awareness and the development of orthographic representations. *Reading and Writing: An Interdisciplinary Journal, 15*(3–4), 295–316.

Ehri, L. C. (1978). Beginning reading from a psycholinguistic perspective: Amalgamation of word identities. In F. B. Murray (Ed.), *Development of the reading process* (IRA Monograph No. 3, pp. 1–33). Newark, DE: International Reading Association.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning, *Scientific Studies of Reading, 18*(1), 5–21.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393–447.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing: An Interdisciplinary Journal, 7,* 295–326.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20,* 163–179.

Feifer, S. (2011). How SLD manifests in reading. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 21–41). Hoboken, NJ: Wiley.

Feifer, S. (2014). Selecting and tailoring interventions for students with reading difficulties. In J. T. Mascolo, D. P. Flanagan, & V. C. Alfonso, *Essentials of planning, selecting and tailoring intervention: Addressing the needs of the unique learner* (pp. 151–176). Hoboken, NJ: Wiley.

Feifer, S. G., & Nader, R. G. (2015). *Feifer Assessment of Reading.* Lutz, FL: Psychological Assessment Resources.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention* (2nd ed.). New York: Guilford Press.

Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research and Practice, 27*(1), 21–32.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37–55.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*(1), 71–99.

Glezer, L. S., Kim, J., Rule, J., Jiang, X., & Riesenhuber, M. (2015). Adding words to the brain's visual dictionary: Novel word learning selectively sharpens orthographic representations in the VWFA. *Journal of Neuroscience, 35*(12), 4965–4972.

Goodman, K. S. (1996). *On reading.* Portsmouth, NH: Heinemann.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41,* 143–157.

Heim, S., Tschierse, J., Amunts, K., Wilms, M., Vossel, S., Willmes, K., et al. (2008). Cognitive subtypes of dyslexia. *Acta Neurobiologiae Experimentalis, 68,* 73–82.

Holmes, V. M. (1996). Skilled reading and orthographic processing. *Australian Journal of Psychology, 48*(3), 149–154.

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and cognition.* Malden, MA: Wiley-Blackwell.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719–729.

Joshi, R. M., Binks, E., Graham, L., Ocker-Dean, E., Smith, D. L., & Boulware-Gooden, R. (2009). Do textbooks used in university reading education courses conform to the instructional recommendations of the National Reading Panel? *Journal of Learning Disabilities, 42*(5), 458–463.

Joshi, R. M., Binks, E., Hougen, M., Dahlgren, M. E., Ocker-Dean, E., & Smith, D. L. (2009). Why elementary teachers might be inadequately prepared to teach reading. *Journal of Learning Disabilities, 42*(5), 392–402.

Kearns, D. M., Rogers, H. J., Koriakin, T., & Al Ghanem, R. (2016) Semantic and phonological ability to adjust recoding: A unique correlate of word reading skill? *Scientific Studies of Reading, 20*(6), 455–470.

Kilpatrick, D. A. (2012a). Not all phonological awareness tests are created equal: Considering the practical validity of phonological manipulation vs. segmentation. *Communiqué: Newspaper of the National Association of School Psychologists, 40*(6), 31–33.

Kilpatrick, D. A. (2012b). Phonological segmentation assessment is not enough: A comparison of three phonological awareness tests with first and second graders. *Canadian Journal of School Psychology, 27*(2), 150–165.

Kilpatrick, D. A. (2014). *The compensating dyslexics we never hear about: A multiple case study approach.* Paper presented at the Society for the Scientific Study of Reading International Conference, Santa Fe, NM.

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties.* Hoboken, NJ: Wiley.

Kilpatrick, D. A. (2016). *Equipped for reading success: A comprehensive, step-by-step program for developing phoneme awareness and fluent word recognition.* Syracuse, NY: Casey & Kirsch.

Kilpatrick, D. A., & Song, M. S. (2018). *Phonemic proficiency, not simply phonemic awareness, influences orthographic learning.* Manuscript under review.

Kilpatrick, D. A. & Van den Broeck, W. (2016). Effective interventions for reading problems. In W. Van den Broeck (Ed.). *Handbook of dyslexia research: Scientific insights in diagnosis, causes, prevention, and treatment of dyslexia* (pp. 153–170). Leuven, Belgium: Acco.

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of Experimental Child Psychology, 73,* 183–207.

Lederberg, A. R., Schick, B., & Spencer, P. E. (2013). Language and literacy development of deaf and hard-of-hearing children: Successes and challenges. *Developmental Psychology, 49*(1), 15–30.

Lerner, J., & Johns, B. H. (2012). *Learning disabilities and related mild disabilities* (12th ed.) Belmont, CA: Wadsworth/Cengage Learning.

Leybaert, J. (2000). Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology, 75,* 291–318.

Liberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Understanding assumptions and their implications for reading instruction. *Annals of Dyslexia, 40,* 51–76.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2,* 175–199.

Mather, N., & Wendling, B. J. (2012). *Essentials of dyslexia assessment and intervention.* Hoboken, NJ: Wiley.

McKague, M., Davis, C., Pratt, C., & Johnston, M. B. (2008). The role of feedback from phonology to orthography in orthographic learning: An extension of item-based accounts. *Journal of Research in Reading, 31*(1), 55–76.

Miles, K. P., & Ehri, L. C. (in press). Orthographic mapping facilitates sight word memory and vocabulary learning. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Eds.), *Reading problems at school*. New York: Springer.

Moats, L. C. (1994). Missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia, 44,* 81–102.

Moats, L. C. (2009). Still wanted: Teachers with knowledge of language. Introduction to special issue. *Journal of Learning Disabilities, 42*(5), 387–391.

Moats, L. C. (in press). Phonics and spelling: Learning the structure of language at the word level. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Eds.), *Reading problems at school*. New York: Springer.

Morais, J. (1991). Constraints on the development of phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.) *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 5–27). Hillsdale, NJ: Erlbaum.

National Assessment of Educational Progress (2015). *Nation's report card*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Reading Panel. (2000). *Report of the National Reading Panel: Reports of the sub-groups* (NIH Publication No. 00-4754). Washington, DC: National Institute of Child Health and Human Development.

Nelson, J. M., & Machek, G. R. (2007). A survey of training, practice, and competence in reading assessment and intervention. *School Psychology Review, 36*(2), 311–327.

O'Connor, R. E. (2011). Teaching older students to read. In R. E. O'Connor & P. F. Vadasy (Eds.), *Handbook of reading interventions* (pp. 380–404). New York: Guilford Press.

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243–277). Baltimore: Brookes.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (Eds.). (1985). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading.* Hillsdale, NJ: Erlbaum.

Perfetti, C. A., Beck, I., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283–319.

Rack, J., Hulme, C., Snowling, M., & Wightman, J. (1994). The role of phonology in young children's learning of sight words: The direct mapping hypothesis. *Journal of Experimental Child Psychology, 57,* 42–71.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 28–53.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2011). *The psychology of reading* (2nd ed.). New York: Psychology Press.

Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology, 36,* 321–339.

Seidenberg, M. S. (2002). Connectionist models of reading. *Current Directions in Psychological Science, 14*(5), 238–242.

Seidenberg, M. (2017). *Language at the speed of sight: How we reading, why so many can't, and what can be done about it.* New York: Basic Books.

Seymour, P. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143–174.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology 72,* 95–129.

Share, D. L. (2004a). Knowing letter names and learning letter sounds: A causal connection. *Journal of Experimental Child Psychology, 88,* 213–233.

Share, D. L. (2004b). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87,* 267–298.

Share, D. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp. 45–68). New York: Psychology Press.

Share, D. L., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing: An Interdisciplinary Journal, 17,* 769–800.

Simos, P. G., Fletcher, J. M., Bergman, E., Breier, J. I., Foorman, B. R., Castillo, E. M., et al. (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology,* 58, 1203–1213.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stanovich, K. E., Siegel, L. S., Gottardo, A., Chiappe, P., & Sidhu, R. (1997). Subtypes of developmental dyslexia: Differences in phonological and orthographic coding. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia* (pp. 115–141). Mahwah, NJ: Erlbaum.

Swank, L. K., & Catts, H. W. (1994). Phonological awareness and written word decoding. *Language, Speech, and Hearing Services in Schools, 25,* 9–14.

Torgesen, J. K. (2004a, March). *Bringing it all together: From phonemic awareness to fluency.* Paper presented at the CORE Literacy Leadership Summit. Retrieved from *www.fcrr.org/science/pdf/torgesen/core_pafluency.pdf*

Torgesen, J. K. (2004b). Lessons learned from the last 20 years of research on interventions for students who experience difficulty learning to read. In P. McCardle

& V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355–382). Baltimore: Brookes.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*(1), 33–58, 78.

Torgesen, J. K., Rashotte, C. A., Alexander, A., Alexander, J., & MacPhee, K. (2003). Progress toward understanding the instructional conditions necessary for remediating reading difficulties in older children. In B. R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 275–297). Baltimore: York Press.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency—Second Edition.* Austin, TX: PRO-ED.

Treiman, R., Sotak, L., & Bowman, M. (2001). The roles of letter names and letter sounds in connecting print to speech. *Memory and Cognition, 29*(6), 860–873.

Truch, S. (1994). Stimulating basic reading processes using auditory discrimination in depth. *Annals of Dyslexia, 44,* 60–80.

Tunmer, W. E., & Chapman, J. W. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 33–67). Mahwah, NJ: Erlbaum.

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading, 16*(2), 122–140.

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology, 105,* 213–231.

Van den Broeck, W., & Geudens, A. (2012). Old and new ways to study characteristics of reading disability: The case of the nonword-reading deficit. *Cognitive Psychology, 65,* 414–456.

Van den Broeck, W., Geudens, A., & van den Bos, K. P. (2010). The nonword-reading deficit of disabled readers: A developmental interpretation. *Developmental Psychology, 46*(3), 717–734.

Vellutino, F. R. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*(1), 2–40.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 279–332). Baltimore: Brookes.

Vukovic, R. K., & Siegel, L. S. (2006). The double-deficit hypothesis: A comprehensive analysis of the evidence. *Journal of Learning Disabilities, 39*(1), 25–47.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing.* Austin, TX: PRO-ED.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing—Second Edition.* Austin, TX: PRO-ED.

Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72,* 1–15.

Wolf, M., O'Rourke, A. G., Gidney, C., Lovett, M., Cirino, & Morris, R. (2002). The second deficit: An investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 15*(1–2), 43–72.

CHAPTER 36

The Role of Cognitive and Intelligence Tests in the Assessment of Executive Functions

Denise E. Maricle
Erin K. Avirett

The study of executive functions is a popular area of research within the fields of cognition and neuroscience. Over the last 15–20 years, there has been a proliferation of articles, books, and research about executive functions. Psychologists and other professionals have referred to "executive functions" or "executive functioning" as if it is a well-known, singular cognitive construct that is understood and easily measured. In fact, despite the wealth of available information, very little is known about executive functions.

The purpose of this chapter is to provide a review of how the major measures of intellectual functioning can be used in the assessment of executive functions. However, assessment of executive functioning involves more than the choice of tools. In order to evaluate executive functions effectively, an understanding of how such functions are defined and conceptualized, knowledge about the neuroanatomical correlates and developmental trajectories of executive functioning, and familiarity with how executive functions are operationalized and measured are mandatory. Therefore, a brief discussion of definitional issues, theoretical conceptualizations, and neuropsychological underpinnings is provided to assist the reader in understanding the complexity of the topic. Each of the major assessment instruments for broad cognitive functioning is reviewed with regard to its utility in the evaluation of executive functions. Within this chapter, the terms *executive functions*, *executive function*, and *executive functioning* are used interchangeably.

DEFINITIONAL ISSUES

The terminology for executive functions is frequently used inconsistently and interchangeably, with little understanding of, or even mutual agreement as to, what the terms actually imply. The complex reciprocal nature of executive functions makes developing a cohesive definition challenging (Maricle, Johnson, & Avirett, 2010). Researchers cannot agree as to whether executive functions are a single process, or a descriptive term for a collection of cognitive processes. In the fields of school psychology, neuropsychology, school neuropsychology, and cognitive psychology, the prevailing perspective is that executive functions consist of separate but related cognitive processes. Despite this consensus, researchers have not agreed upon the components of executive functioning, although several domains are generally accepted. These domains include self-monitoring and regulation of cognition, emotion, and behavior; initiating, planning, and completing complex tasks; working memory; attentional control (inhibition, sustained attention, shifting attention); and cognitive flexibility (Alvarez & Emory, 2006; Anderson, Levine, & Jacobs, 2002; Baron, 2004; Chan, Shum, Toulopoulou, & Chen, 2008; Cheung, Mitsis, & Halperin, 2004; Goldstein, Naglieri, Princiotta, & Otero, 2014; Hughes & Graham, 2002; Hunter & Sparrow, 2012; Jurado & Rosselli, 2007; Lezak, Howieson, & Loring, 2004; Stuss & Alexander, 2000). Overall, the concept of executive functioning includes the processes of mental flexibility, the ability to filter out interference or distractions, the ability to engage in goal-directed behavior, and the ability to anticipate the consequences of one's actions (Ardila, 2008; Ardila & Surlof, 2007; Hunter & Sparrow, 2012; Knight & Stuss, 2002).

This definitional debate is not an esoteric one. How executive functions are defined is critical to the assessment process because how test authors define the construct determines how it is operationalized and thus measured in a particular instrument. It is also necessary for the examiner to define what elements or components of executive function they are interested in measuring, since executive functioning in and of itself has not been proven conclusively to exist. In addition, no single measure or instrument is able to evaluate all of the proposed executive functions. Thus examiners need to know what aspects of executive functions should be measured and how the instruments or tools they have chosen actually measure these specific aspects. Finally, understanding the limitations posed by definitional differences is important when researchers and clinicians alike are reading and filtering through the wealth of literature published on this challenging topic.

THEORETICAL MODELS

The variety of current models and theories reflects diverse and disparate perspectives on the nature, structure, and role of the executive functions; however, the literature is cluttered with competing claims and datasets that are inadequate for leveraging support for any one theory or model over another. Although the structure and role of executive functions have been debated and conceptualized in a multitude of ways by numerous researchers, no clear consensus has emerged regarding a specific theory or model of executive functioning. Executive function models can be categorized according to the way they conceptualize various functions (Zelazo, Muller, Frye, & Marcovitch, 2003). Two primary perspectives dominate the literature at this time. Executive functions are considered to be either a unitary and hierarchical system with control and monitoring processes, or a set of distinct but interlocking cognitive processes.

In the hierarchical perspective, executive functions are considered a unitary construct. In this view, executive functioning is regarded as metacognitive and is frequently seen as analogous to overall intelligence (Anderson, 2008; Blair, 2006; Friedman et al., 2006; Grafman, 2006; Kane & Engle, 2002). This view depicts executive functioning as the supervisor of other subordinate and narrower cognitive processes. For a recent example of this perspective, one could examine McCloskey's holarchical model of executive functions (McCloskey & Perkins, 2013). Accordingly, executive functioning is more difficult to operationalize, define, and assess in most hierarchical models, due to the complexity of identifying and measuring the managerial metacognitive aspect in conjunction with the varied associated cognitive processes. Within this viewpoint, executive functions constitute a nebulous, overarching entity similar to *g*: We all know it when we see it, but no one can really "define" it. Examples of this perspective can be seen in Luria's theory of cognitive functioning (the theory that underlies the NEPSY-II, the Cognitive Assessment System—Second Edition [CAS2], and one interpretive framework for the Kaufman Assessment Battery for Children—Second Edition [KABC-II]) and in Cattell–Horn–Carroll (CHC) theory (the primary theory underlying the Woodcock–Johnson IV Tests of Cognitive Abilities [WJ IV COG], the Stanford–Binet Intelligence Scales, Fifth Edition [SB5], the Differential Ability Scales—Second Edition [DAS-II], the other interpretive framework for the KABC-II, and other modern measures of cognitive functioning).

Luria described human cognitive processes within the framework of three functional units. In Luria's theory, the first functional unit (block 1) is arousal and attention; the second unit (block 2) codes information using simultaneous and successive processes; and the third functional unit (block 3) is involved in the regulation of executive functioning (planning, strategizing, regulating performance, and solving problems). Luria identified the prefrontal lobes of the brain as primarily responsible for the third functional unit. However, rather than a strict localization perspective, Luria took a more dynamic or systems-based approach. He posited that complex brain functions (such as executive functions) are mediated by a coordinated set of brain structures or physiological processes, and that no specific brain area completely controls a specific function/process. As a result, a variety of disruptions may be seen if a particular brain region is damaged, and compromised functioning may or may not be subsumed or ameliorated by other brain regions (Lewandowski & Lovett, 2008). Luria's work served as one impetus for the study of executive functions and has been used as a blueprint for defining components of human intellectual competence.

CHC theory describes human cognitive processes within the framework of three strata: a general overarching factor, broad cognitive factors, and multiple narrow abilities. CHC theory does not describe a specific and separate element of executive functioning; rather, components of executive functions are integrated primarily into the Gf (fluid reasoning) broad-ability factor (Kane & Engle, 2002) and the narrow-ability factors of induction, general sequential reasoning, and attention and concentration. Drawing from more recent research, Flanagan, Alfonso, Ortiz, and Dynda (2010) and Flanagan, Ortiz, and Alfonso (2013) present an integrated interpretive framework based on psychometric, neuropsychological, and Lurian perspectives, and provide a neurocognitive demand task analysis of the major test batteries using this framework. Defining executive functioning as a global neuropsychological domain represented by metacognition, planning, learning, memory, and cognitive efficiency, Flanagan and colleagues (2010, 2013) posit that the neuropsychological domain of executive functions corresponds well with eight broad CHC abilities: fluid reasoning (Gf), comprehension–knowledge (Gc), processing speed (Gs), short-term memory (Gsm), long-term storage and retrieval (Glr), quantitative knowledge (Gq), reading and writing ability (Grw), and general knowledge ability (Gkn).

Proponents of unitary conceptualizations of executive functions argue that their perspective is the most parsimonious view of executive functions, and that nonunitary models or perspectives are reductionistic and not helpful, given their fractionation of executive functions (Sugarman, 2002). In contrast, researchers from the second major viewpoint see executive functions as a label for a collection of distinct, yet associated, cognitively complex higher-order processes (Anderson, 2001; Ardila et al., 2000; Baron, 2004; Elliott, 2003; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Proponents of this perspective would argue that the factorial evidence clearly supports distinct cognitive processes that correlate moderately with each other. However, as previously stated, researchers cannot reach a consensus as to what cognitive processes constitute these executive functions. Salthouse (2005) has suggested that reasoning and perceptual speed represent the underlying features of executive functions, whereas others have proposed working memory (Baddeley, 1996, 2000), verbal working memory (Sugarman, 2002), or inhibition (Barkley, 2000) as the clear foundation of executive functioning. Miyake, Friedman, Emerson, Witzki, and Howerter (2000) propose an intermediate position. They see executive functions as separate but moderately correlated constructs, and suggest that the executive system is composed of both unitary and nonunitary components. Other researchers have elaborated on this idea and proposed two types of executive functions: metacognitive and behavioral (Fuster, 2001, 2002; Happaney, Zelazo, & Stuss, 2004). According to this perspective, the metacognitive type consists of the usual executive functions and is mediated by the dorsolateral prefrontal cortex (PFC), whereas the behavioral type is responsible for the coordination of cognition and emotion, with inhibition as its primary expression, and is mediated by the ventromedial PFC.

Miller's (2007, 2013) integrated school neuropsychology/CHC conceptual model (integrated SNP/CHC model; see also Miller & Maricle, Chapter 33, this volume) provides an intermediate perspective on neuropsychological constructs such as executive functions; he views them as independent but moderately correlated constructs. Miller's model is unique in that he uses neuropsychological, neuroanatomical, and neuroassessment research to conceptualize the model's components. It is specifically intended to be applied to the neuropsychological development of children and adolescents, and it can be utilized in conjunction with interpretive approaches specific to children, such as Hale and Fiorello's (2004) cognitive hypothesis-testing model. In Miller's conceptual model, tasks are classified according to four broad classifications (basic sensorimotor functions; facilitators and inhibitors; basic cognitive processes; and acquired knowledge). These are then further subdivided into second-order and third-order classifications that respectively represent broad and narrow constructs being measured by the various tasks. Miller's model classifies executive functions as a second-order classification (broad construct) under the cognitive processes domain (Miller, McGill, & Bauman Johnson, 2016).

Miller categorizes executive functions as cognitive flexibility, concept generation, inhibition, behavioral/emotional regulation, and planning, reasoning, and problem solving. He views these executive functions as being strongly related to fluid reasoning abilities or tasks that require novel problem-solving skills. In contrast to other researchers, Miller views other common aspects of executive functioning, such as working memory and attention, as separate neurocognitive constructs and labels them as facilitators or inhibitors of cognitive processes. For example, attentional processes in Miller's model are comprehensively addressed as attentional control constructs involving selective or focused attention and sustained attention, and are considered separate and distinct from immediate memory capacity, working memory, or executive functions.

Understanding these different theoretical perspectives is important because they shape the research being conducted on executive functions and influence the interpretation of the obtained data. Some theories have exerted considerable influence on the field even when there is no supporting evidence. It is also important to note that the primary theoretical models of executive functioning were developed from an adult perspective and have been applied to children and adolescents on a post hoc basis. The problem with adult-based studies is that they examine executive functions from the perspective of the developed brain's responding to an acquired injury, resulting in an executive function deficit. Such studies do not take into account issues of development, such as how an acquired injury affects executive functions in a brain that is still developing or how neurodevelopmental disorders may affect the development of neurological functions such as executive functions (Hunter & Sparrow, 2012). Children often display differential performance, which does not fit many of the adult models currently available.

THE NEUROPSYCHOLOGY OF EXECUTIVE FUNCTIONS

Understanding the neuropsychological underpinnings of executive functioning is also critical to the assessment of executive functions. Despite amazing progress in neuroimaging, the neuroanatomy and neurophysiology of executive functions continue to be debated in the literature. Historically, executive functions have been associated with the PFC, and more specifically with the anterior (front) portions of the PFC. Exactly how the PFC supports executive functions is largely unknown and somewhat controversial (Alvarez & Emory, 2006; Hughes & Graham, 2008; Wood & Grafman, 2003). Attempts to localize executive functions to discrete areas in the PFC with neuroimaging techniques have been inconclusive (Roberts, Robbins, & Weiskrantz, 2002). A one-to-one correspondence between executive functions and the PFC has not yet been documented in the research, and many claims are speculative at best (Alvarez & Emory, 2006). Moreover, some neuroimaging results have implicated posterior, cortical, and subcortical regions in executive functioning, and it has been posited that executive functioning may be a more flexible distributed network than previously thought (Jacobs, Harvey, & Anderson, 2011; Roberts et al., 2002).

The PFC comprises approximately one-third of the brain; maintains intricate connections to the rest of the brain; and continues to mature through synaptogenesis, myelination, and pruning well into early adulthood. Three systems—the dorsolateral PFC, the anterior cingulate circuit, and the orbito-frontal cortex—are thought to be involved in executive functioning. The dorsolateral PFC is associated with most of the "typical" executive functions, including cognitive flexibility and behavioral spontaneity; maintaining and shifting cognitive attention; organization and planning; goal setting; performing dual-task activities; short-term memory; focusing and sustaining attention; inhibition; and fluency (Alvarez & Emory, 2006; Hale & Fiorello, 2004; Romine & Reynolds, 2005). Deficits within the dorsolateral PFC are often associated with attention problems, poor problem solving, and difficulties with self-monitoring and control. The anterior cingulate circuit appears to control the behavioral processes associated with initiation, inhibition, motivation, selective or divided attention, response monitoring, and error detection (van Vreen & Carter, 2002; Zilmer, Spiers, & Culbertson, 2008). Damage to the anterior cingulate circuit often results in difficulties with response inhibition (Miller, 2007, 2013), slow completion time or decision speed, lack of persistence, and difficulty with self-monitoring (Hale & Fiorello, 2004). The orbito-frontal cortex is involved with emotional and social behaviors such as tact, sensitivity, impulsivity, and emotional inhibition (Bradshaw, 2001; Knight & Stuss, 2002). Deficits in the orbito-frontal cortex are associated with emotional dysregulation, aggression, sexual promiscuity, disinhibition, impulsivity, and poor decision making. These three systems together create, support, and coordinate the complex cognitive functions involved in problem solving and decision making, which are the hallmarks of the construct being defined as *executive functions*.

Despite our neuroanatomical understanding of executive functioning, it is important to remember that our knowledge and understanding of how the brain functions when it is processing and performing executive function tasks is still very limited (P. Anderson, 2002). The PFC does not act in isolation; therefore, it is challenging to identify which brain regions contribute to which outcomes on specific measures of executive functions. In other words, a deficit in one area can lead to multiple behaviors or, conversely, one behavior may be the result of multiple underlying impairments. As result, Hunter and Sparrow (2012) have proposed that executive functions are best represented on a spectrum or continuum—given this underlying idea that with multiple executive functions in the equation, an individual may not display a significant deficit in any one skill, but the cumulative effects of mild executive dysfunction may result in severe executive function impairment. It is also important to remember that although a variety of neural correlates have been identified in adults for various components of executive functioning, no research to date has confirmed these findings with children.

DEVELOPMENTAL TRAJECTORIES OF EXECUTIVE FUNCTIONING IN CHILDREN

Until recently, executive functions were thought to emerge in adolescence and early adulthood. The belief was that they played little or no role in typical brain development during infancy and childhood. However, recent research has debunked this early perception and demonstrated the critical role executive functions play in typical brain development across the lifespan (V. Anderson, 2002; Reynolds, 2007). According to numerous researchers (Carlson, 2004; Hughes & Graham, 2008; Lidz, 2003), higher-order cognitive skills are present before they are observable, functional, or testable. In addition, research has demonstrated an interaction between developmental processes and the manifestation of executive functions (Blair, Zelazo, & Greenberg, 2005; Brocki & Bohlin, 2004; Hunter, Hinkle, & Edidin, 2012; Zelazo et al., 2003). Executive functions appear to emerge, develop rapidly, and reach adult levels of performance differentially; therefore, research would suggest that developmental profiles or trajectories will depend on the executive skill(s) being examined (Archibald & Kerns, 1999; Romine & Reynolds, 2005). As a result, it is important for psychologists working with children in any capacity to be aware of and understand the developmental trajectories of the executive functions they seek to measure.

ASSESSMENT OF EXECUTIVE FUNCTIONS

Operationalizing definitions of executive functions in order to assess them is even more difficult and fraught with challenges than is developing theoretical definitions. One of the challenges is obtaining valid, accurate, and reliable measures (Hughes & Graham, 2008), given the perspective that executive functions are not unitary but composed of multiple complex functions. Since there is no universal definition of what constitutes executive functions, or which domains are critical for successful executive functioning, there is no agreement as to what executive function domains should be measured or how these domains should be assessed. A second challenge is that by definition executive function tasks are complex, so task impurity becomes an issue, in that a task may require (and thus be influenced by) multiple cognitive processes. Thus it is difficult to distinguish executive function tasks from other tasks because the integrative simultaneous nature of frontal lobe functioning makes it difficult to parse out the specific cognitive functions being utilized in each type of task in order to create a pure measure of executive functioning (Hughes & Graham, 2002; Maricle et al., 2010; Romine & Reynolds, 2005).

Even when a task can be identified as a relatively valid, reliable, and somewhat pure measure of a particular executive function, the task or an individual's performance on the task is often misinterpreted. A common mistake is to equate deficient performance on an identified task of executive function with frontal lobe or neurological dysfunction. A second common mistake is to assume that each task taps an underlying cognitive process universal to all executive function tasks, or that all tasks of executive function are measuring the same aspects of the construct. Such presumptions lead to overly simplified, narrow, or inaccurate interpretations. Interpreting results from executive function tasks depends on precise specification of task demands; this specificity necessitates a systematic understanding of the components of executive function, as well as an understanding that most executive function tasks require several cognitive skills for successful performance (Flanagan et al., 2010; Hughes & Graham, 2008; Miller, 2007, 2013). Research demonstrates that successful performance on most complex cognitive tasks requires a combination of cognitive skills, such as working memory, attention, concept formation, inhibition, and/or cognitive flexibility. The breakdown in an individual's performance can occur at any stage of cognitive processing, from lower-level skills such as attention to the higher-order skills such as planning (Delis, Kaplan, & Kramer, 2001). Accurate interpretation then requires determining the neurocognitive constructs that contribute to successful or poor performance on a specific task, and then considering those results within the context of the assessment of a specific individual.

The broad range of skills implicated in executive functions has led to the use of many different assessment tools; however, no instrument has yet been developed to measure executive functioning in its entirety. Rather, there are tools available to measure specific components of executive functions. Assessment can be divided into broad overall measures of multiple cognitive constructs, or targeted measures of a specific cognitive construct. Executive functions can be more or less effectively evaluated by using both types of measures. Miller and Hale (2008) note that standardized intellectual measures are psychometrically some of the best tools available to practitioners, and that incorporating these tools into neuropsychological assessment is an essential practice. Thus a discussion of the major norm-referenced comprehensive batteries of intelligence and how they relate to the measurement of executive functions is warranted.

EVALUATION OF EXECUTIVE FUNCTIONS WITH THE MAJOR COGNITIVE BATTERIES

Woodcock–Johnson IV Tests of Cognitive Abilities

The WJ IV COG (Schrank, McGrew, & Mather, 2014) consists of a battery of tests designed to measure cognitive and intellectual abilities. The WJ IV COG was designed to work in concert with the Woodcock–Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a) and the Woodcock–Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b). The WJ IV COG was based on the iteration of the CHC theory of cognitive abilities described by Schneider and McGrew (2012), and McGrew, LaForte, and Schrank (2014). According to Ding and Alfonso (2016), the WJ IV COG was designed to focus on "important" broad and narrow CHC abilities critical for academic success, rather than simply measuring the broad abilities of CHC as did its predecessor, the WJ III. Additionally, some of the WJ IV tasks were designed to be cognitively complex, meaning that several skills are needed to complete these tasks successfully.

CHC theory regards cognitive abilities as multidimensional and dynamic rather than as static domains of function. The iteration of CHC as defined by Schneider and McGrew (2012) posits 16 broad abilities and over 80 narrow abilities. However, CHC theory, and by extension the WJ IV COG, does not view executive functioning as a specific independent cognitive domain; rather, executive functions are seen through the lens of cognitive processes such as fluid reasoning (Gf), working memory (Gwm), and processing speed (Gs) (Maricle & Johnson, 2016).

No research is yet available involving the WJ IV COG's relationship to executive functions. Research on the relationship to executive function constructs in the WJ III COG was also quite limited. The majority of this research focused on the broad CHC factors and the underlying narrow abilities. One exception was a study by Floyd and colleagues (2006) that examined the relationship of the WJ III COG clinical clusters with the Delis–Kaplan Executive Function System (D-KEFS; Delis et al., 2001). Their results suggested that the clinical clusters correlated moderately with the D-KEFS tasks, and that both measures appeared to be assessing a general construct—most likely general intellectual ability because of the strong correlation with the General Intellectual Ability score.

Thus, if using the WJ IV COG to screen for executive functioning or executive dysfunction, the examiner must extrapolate from the CHC factors and the WJ IV COG tasks. On the WJ IV COG, there are three Gf (fluid reasoning) tasks designed to measure problem solving and reasoning: Concept Generation, Analysis–Synthesis, and Number Series. One task on the WJ IV ACH, Number Matrices, could also be considered a Gf task. Ford, Keith, Floyd, Fields, and Schrank (2003) describe Concept Formation as a controlled learning task that measures fluid reasoning and requires rule formation, categorical reasoning, inductive thinking, and logical deduction. Schneider (2016) describes it as a measure of induction or the ability to use limited information to identify rule-based behavior and logical patterns. Analysis–Synthesis is thought to measure deduction, or the ability to work with a set of logical rules to deduce new information. Number Series is a measure of sequential/serial or quantitative reasoning. It is highly related to the Number Matrices task on the WJ IV ACH (Schneider, 2016). A study by Miller, McGill, and Maricle (2017) found that Number Series was the most predictive task for almost every academic skill measured by the WJ IV ACH.

The WJ IV COG delineates Object–Number Sequencing, Numbers Reversed, Verbal Attention, and Memory for Words as tasks that measure immediate short-term memory capacity and working memory (Gwm). Two WJ IV OL subtests, Sentence Repetition and Understanding Directions, are also classified as measuring Gwm skills. Previous research (Floyd, Shaver, & McGrew, 2003) and current speculation would suggest that Gwm tasks also measure aspects of cognitive attention, including attentional capacity/control (Numbers Reversed, Verbal Attention, Sentence Repetition), divided attention (Object–Number Sequencing, formerly Auditory Working Memory) and selective attention (Verbal Attention; Understanding Directions).

The WJ IV COG also provides measures of learning efficiency or long-term storage and retrieval (Glr) and processing speed (Gs). Measures of long-term storage and retrieval are often controlled learning tasks involving associative memory, such as Visual–Auditory Learning, or tests of meaningful memory, like Story Recall. However, Schneider (2016) suggests a classification of Visual–Auditory Learning as a meaningful memory task rather than an associative memory task. Two other measures of Glr, Retrieval Fluency and Rapid Picture Naming, are found on the WJ IV OL and appear to be measuring memory retrieval speed or speed of lexical access. The WJ IV COG measures processing speed (Gs) via the Number–Pattern Matching, Letter–Pattern Matching, and Pair Cancellation tasks. Number–Pattern Matching used to be Visual Matching, and Letter–Pattern Matching is an analogue. Both are designed to measure pure perceptual speed. Pair Cancellation is described as a complex measure that requires the examinee to identify a specific repeating pattern; successful performance requires attention, vigilance, speed of visual scanning, and interference control. Miller and colleagues (2016) suggest that Pair Cancellation also measures sustained, as well as selective or focused, attention.

Stanford–Binet Intelligence Scales, Fifth Edition

The SB5 (Roid, 2003b), the most recent revision of the popular Stanford–Binet series, was constructed on a five-factor hierarchical cognitive model consistent with CHC theory (Roid & Barram, 2004). Using nonverbal and verbal tasks, the SB5 evaluates Fluid Reasoning (Gf), Knowledge (Gc), Quantitative Reasoning (Gq), Visual–Spatial Processing (Gv), and Working Memory (Gsm). Each subtest in the Verbal domain has a counterpart in the Nonverbal domain. Roid (2003a) indicates that the Nonverbal domain measures the general ability to reason, to solve problems, to visualize, and to recall information presented in pictorial, figural and symbolic formats; the Verbal domain measures the general ability to reason, solve problems, visualize, and recall information using spoken or written words or sentences.

Studies reported in the technical manual support the presence of a general factor, two group factors, and five specific factors; however, research conducted by other investigators calls these results into question. Canivez (2008), DiStefano and Dombrowski (2006), Sattler, Dumont, Salerno, and Roberts-Pittman (2008) found either no support or limited support for the group (i.e., two-factor) model or for the specific (i.e., five-factor) model. Canivez also noted issues of subtest migration or cross-loading, wherein subtests theoretically associated with one factor were in fact associated with multiple dimensions/factors. For example, some of the SB5 Nonverbal subtests actually accounted for more variance on the Verbal factor than the Nonverbal factor. Another issue of concern with the SB5 is that the subtests in each domain change as the individual progresses through the levels. Each subtest has one to three unique activities or variations of the task. For example, Verbal Fluid Reasoning is composed of Early Reasoning, Verbal Absurdities, and Verbal Analogies, and each variation taps verbal fluid reasoning differentially and utilizes other cognitive skills to a greater or lesser degree. This convolution is problematic, in that it is difficult to determine whether the skill set or function being measured at one level is the same skill set or function being measured at another level within the same subtest. Given the massive task impurity issues, an examiner must have a good understanding of each task within each subtest, how and when the task changes form, what skills are being measured within each form of the task, and finally how the performance and score of the individual may be affected as a result.

No research is currently available that addresses the specific use of the SB5 for assessing executive functions. The current literature indicates that interpretation of the SB5 should be primarily at the level of *g* or overall general intelligence. However, task analyses do suggest how various tasks within the SB5 may be measuring different aspects of executive functioning. If executive functions are conceptualized from a cognitive construct perspective, then tasks within the SB5 could be parsed into likely measures of different aspects of executive functioning. For example, the Nonverbal Fluid Reasoning task (Object Series/Matrices) requires sequential reasoning and deductive thinking for successful completion of items at the early levels, and then incorporates a classic matrix or pattern reasoning task using inductive thinking at the more advanced levels. Matrix tasks have consistently been found to be good measures of Gf or fluid reasoning, which is considered by many to be one of the hypothesized executive function skills (Tranel, Manzel, & Anderson, 2008). In contrast, only the Verbal Analogies items of the Verbal Fluid Reasoning task constitutes true fluid reasoning or abstract thinking skills, and thus the Verbal Fluid Reasoning task should not be considered an adequate measure of verbal executive functioning. Roid (2003a) would take exception to this characterization, as he believes that the Verbal Fluid Reasoning task measures the ability to solve novel verbal problems, to identify cause-and-effect relationships, to classify according to form and function, and to use inductive reasoning with analogies. Another example can be found in the Nonverbal Working Memory subtest and its Verbal Working Memory counterpart. The Nonverbal Working Memory task measures visual short-term memory capacity and visual working memory capacity through a “shell game” at early levels and a block-tapping task at the more advanced levels. Its Verbal Working Memory counterpart uses memory for sentences and memory for a word in the presence of interference to assess verbal short-term memory and working memory capacity.

Differential Ability Scales—Second Edition

The DAS-II (Elliott, 2007) is a revision of the Differential Ability Scales (DAS; Elliott, 1990), which originated from the British Ability Scales (BAS; Elliott, 1983). According to Dumont, Willis, and Elliott (2009), the DAS-II was based on current knowledge of neuroscience and was designed to reflect cognitive processes that contribute to learning difficulties in children; it was not developed to reflect a unitary model of *g* or general intelligence. The selection of abilities to be assessed by the DAS-II was intended to be consistent with CHC theory.

The technical manual reports varying factor structures for the DAS-II at different age levels. At the earliest ages (2–3), a two-factor model (Verbal and Nonverbal) emerges; at ages 4–5, a five-factor model emerges (Verbal, Nonverbal Reasoning, Spatial, Visual–Verbal Memory, and Verbal Short-Term Memory); at ages 6–12, a seven-factor model emerges (Verbal, Nonverbal Reasoning, Spatial, Verbal Memory, Verbal–Visual Memory, Cognitive Speed, and Auditory Processing); and for ages 6–17, a six-factor model emerges (Verbal, Nonverbal Reasoning, Spatial, Verbal Short-Term Memory, Visual–Verbal Memory, and Cognitive Speed). Sattler, Dumont, Willis, and Salerno (2008) conducted a principal-components factor analysis, using data from the technical manual; they identified for the Early Years Lower Level battery a three-factor solution that they found difficult to define, but that did not support the factor solutions cited in the technical manual. Sattler and colleagues identified a seven-factor solution for the Early Years Upper Level battery and felt that five were clearly identifiable as Verbal, Nonverbal, Spatial, Working Memory, and Processing Speed. For the School-Age battery, a six-factor solution was determined with five identifiable factors (Verbal, Nonverbal, Reasoning/Spatial, Working Memory, Processing Speed, and Visual Memory). Examiners using the DAS-II need to be cognizant of these varying factor structures and what they purportedly measure at different age levels when interpreting results.

No research is available that directly addresses the DAS-II’s relationship to executive functions. According to Dumont and colleagues (2009), the DAS-II assesses seven broad abilities: verbal ability (Gc), spatial ability (Gv), nonverbal reasoning ability (Gf), retrieval (Glr), memory (Gsm), processing speed (Gs), and auditory processing (Ga). The Verbal and Spatial ability clusters are intended to reflect the major information-processing systems used to receive, perceive, remember, and process information through both auditory and visual modalities. Verbal ability as measured by the DAS-II essentially consists of language development and lexical knowledge. Spatial ability primarily consists of tasks requiring spatial relations and visualization, although one task, Recall of Designs, measures visual memory. The tasks of the Nonverbal Reasoning ability cluster (Matrices; Sequential and Quantitative Reasoning; and Picture Similarities) provide adequate to strong measures of inductive thinking and reasoning. The Matrices task assesses the ability to formulate and test hypotheses, inductive reasoning, verbal mediation, and visual perception. The Picture Similarities subtest evaluates the ability to formulate and test hypotheses about relationships and to solve nonverbal problems by attaching representation or conceptual meaning to a pictured object. The Sequential and Quantitative Reasoning task assesses the ability to perceive relationships, draw conclusions, reason inductively, formulate and test hypotheses, use analytic reasoning, and retrieve long-term information. The DAS-II also provides strong measures of another executive functioning construct, that of memory (free recall, span, and working memory), through the broad Retrieval and Memory clusters. The subtests Recall of Designs, Recall of Digits (Forward and Backward), Recall of Objects (Immediate and Delayed), Recall of Sequential Order, and Recognition of Pictures allow for a comprehensive assessment of visual and auditory short-term memory capacity, working memory capacity, and long-term retrieval.

Wechsler Intelligence Scale for Children—Fifth Edition

The Wechsler scales are believed to be the most widely used measures of intelligence throughout the world (Flanagan & Kaufman, 2009); however, there is insufficient evidence with neurological populations to ensure the tests' appropriate application for neuropsychological assessment (Loring & Bauer, 2010; McCrea & Robinson, 2011). The Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014) embodies a fundamental shift from previous editions of the WISC, since it was redesigned to more adequately reflect the "latest research on intelligence, cognitive development, cognitive neuroscience, and neurodevelopment" (Wechsler, 2014, p. 1). Some researchers claim that the Wechsler tests lack a unified theoretical foundation (Coalson, Raiford, Saklofske, & Weiss, 2010; Kaufman, 2010; Raiford & Coalson, 2014), whereas others note that the evolution of the Wechsler tests is consistent with CHC theory (Canivez & Watkins, 2016; Miller & McGill, 2016).

The WISC-V consists of 21 subtests: 13 retained from the WISC-IV, and eight new subtests. The WISC-V organizes these subtests into three categories: primary, secondary, and complementary. The 10 primary subtests are considered to constitute the core WISC-V battery. The Full Scale Intelligence Quotient (FSIQ) is derived from seven of these primary subtests. There are six secondary and five complementary subtests, designed to be used in a supplemental or ancillary capacity. Beyond the FSIQ, a number of composite scores can be derived from various combinations of subtests (14 in total). There are five primary index scores, which represent the primary cognitive constructs being measured: the Verbal Comprehension Index (VCI), the Fluid Reasoning Index (FRI), the Visual Spatial Index (VSI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). Additionally, there are five ancillary indexes and three complementary indexes.

Given its recent publication date, there is little research available regarding the WISC-V, but already there are criticisms of its identified factor structure (Canivez & Watkins, 2016; Canivez, Watkins, & Dombrowski, 2016; Miller & McGill, 2016). The harshest criticism comes from Canivez and various colleagues. They suggest that the WISC-V's reported factor structure is untenable and not supported by the normative data or statistical analyses found in the technical manual. Canivez and colleagues reanalyzed the WISC-V data and found no evidence to support a five-factor solution; rather, their analysis suggests either a one-factor *g* solution or a four-factor solution (the latter was favored). Canivez and colleagues' research on the WISC-V contrasts with research conducted on the WISC-IV (Flanagan & Kaufman, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Sattler & Dumont, 2008), which suggested that the WISC-IV four-factor solution was not appropriate and that a five-factor solution was the most parsimonious. Canivez and colleagues (2016) speculate that the shift in adequate factor solutions may be due to the addition/removal/restructuring of subtests in the WISC-V relative to the WISC-IV. Further research will be needed to identify the appropriate factor structure of the WISC-V.

Given the dissension regarding the factor structure of the WISC-V, a brief review of the indexes and their subtests is provided. The VCI is composed of tasks (Similarities, Vocabulary) that measure verbal abilities utilizing reasoning, comprehension, and conceptualization. The index involves tasks that require comprehension–knowledge, verbal fluid reasoning, and long-term memory (Weiss, Beal, Saklofske, Packiam-Alloway, & Prifitera, 2008). Within the CHC perspective, the VCI comprises a measure of Gc or comprehension–knowledge (Keith et al., 2006). Not included in the VCI but still considered measures of Gc are the subtests of Information and Comprehension. The FRI consists of tasks (Matrix Reasoning, Figure Weights) that measure fluid reasoning skills such as analogical reasoning, problem solving, and abstract categorical reasoning. In the Wechsler model, Picture Concepts and Arithmetic are also classified as Gf tasks. The VSI consists of Block Design and Visual Puzzles, and provides a measure of visual–spatial reasoning, problem solving, and constructional abilities (part–whole relationships and dimension).The WMI is defined as a measure of the ability to concentrate, sustain attention, and exert mental control. It consists of Digit Span (Forward, Backward, Sequencing) and Picture Span. Letter–Number Sequencing does not contribute to the WMI but is also a measure of Gwm. Flanagan and Kaufman (2009) classified the WMI as a narrow-band ability factor (the narrow ability of working memory within the broad Gsm/Gwm factor) in the CHC model. Weiss and colleagues (2008) take exception to this characterization and believe that the WMI tasks are strong measures of verbal short-term and working memory. Keith and colleagues (2006) determined that the WMI appears to be a mixture of Gsm and Gf. Finally, the PSI provides a measure of the speed of mental and graphomotor processing. Coding and Symbol Search comprise the PSI, with Cancellation serving in an ancillary role. Within CHC theory, the PSI tasks are seen as visual scanning and speed-of-processing tasks falling under the Gs factor. None of the subtests on the WISC-V are thought to measure executive functioning unequivocally, and no research is yet available that specifically assesses the contribution of executive functioning to WISC-V tasks.

Miller (2007, 2013; see also Miller et al., 2016, and Miller & Maricle, Chapter 33, this volume) conceptually groups the WISC-V and WISC-V Integrated subtests according to his integrated SNP/CHC model. Miller labels Similarities, Similarities Multiple Choice, Comprehension, Comprehension Multiple Choice, Matrix Reasoning, Picture Concepts, Figure Weights, and Figure Weights Recall as measures of executive functioning. Miller's model places measures of processing speed (Cancellation, Coding, Coding Copy, Symbol Search, Naming Speed—Literacy, Naming Speed—Quantity), and some measures of memory (Arithmetic, Arithmetic Process Approach, Spatial Span—Backward, Digit Span—Forward, Spatial Span—Forward) in the facilitator–inhibitor domain. Other measures of memory (Digit Span—Forward, Spatial Span—Forward, Coding Recall, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation) are classified as measures of cognitive processes under the domain of learning and memory.

Flanagan and colleagues' (2010, 2013) integrated interpretive framework incorporates additional subtests besides those discussed in Miller's model, including Cancellation, Coding, Digit Span, Letter–Number Sequencing, Spatial Span—Forward and Backward, Symbol Search, and Visual Digit Span. These subtests are primarily seen as measures of Gsm (short-term and working memory) and Gs (processing speed), but within the integrated interpretive framework they fall into the broader domain of executive functions.

Kaufman Assessment Battery for Children—Second Edition

Kaufman and Kaufman (2004) deliberately designed the KABC-II to incorporate two distinct theoretical models: the CHC and Lurian models. However, the authors recommend using the CHC model interpretively for most purposes. The KABC-II is composed of 18 subtests grouped into four or five scales, depending on the child's age and the interpretive model chosen. The Lurian model organizes the subtests into four scales: Sequential, Simultaneous, Learning, and Planning. The CHC model organizes the same subtests into short-term memory (Gsm), visual processing (Gv), long-term storage and retrieval (Glr), fluid reasoning (Gf), and crystallized ability (Gc). The KABC-II may be a promising tool for assessing intelligence and cognitive impairments, but research on its applicability within a neuropsychological framework for neuropsychological assessment is limited (Mays, Kamphaus, & Reynolds, 2009).

On the KABC-II, learning ability and planning ability from the Lurian perspective are the domains most applicable to executive functioning. Within the CHC perspective, these constructs equate to the cognitive factors of Glr (long-term storage and retrieval) and Gf (fluid reasoning), respectively. Kaufman, Kaufman, Kaufman-Singer, and Kaufman (2005) consider Planning/Gf to be the domain most closely associated with executive functioning. On the KABC-II, Planning/Gf tasks are designed to require a variety of mental operations to solve novel problems, including cognitive flexibility, inductive and deductive reasoning, hypothesis generation, and impulse control. Tasks that encompass the domain of Learning/Glr also require the integration of several executive functions. Kaufman and colleagues note that Learning/Glr on the KABC-II emphasizes efficiency of the storage and retrieval, not the specific information being stored. They further emphasize that effective paired-associate learning requires considerable attention and planning skills, as well as storage and retrieval skills.

Kaufman and colleagues (2005) delineate narrow cognitive abilities underlying the KABC-II subtests, identifying induction, general sequential reasoning, and associative learning/memory as the primary narrow abilities representative of executive functioning. Three KABC-II subtests are considered measures of induction: Conceptual Thinking, Pattern Reasoning, and Story Completion. The Conceptual Thinking task is only administered to children ages 3 years, 0 months to 6 years, 11 months (3:0–6:11), and involves determining conceptual relationships by identifying the concept that does not fit the relationship parameters. The Pattern Reasoning subtest requires an examinee to complete a logical linear pattern (i.e., the examinee must identify what part of the pattern is missing and complete the missing section with the appropriate choice from among a series of choices). The Story Completion task is similar, in that the examinee completes the missing elements of a pictured story by deducing from the remaining story line what aspects of the story are missing. Two subtests, Rover and Riddles, assess general sequential reasoning. Rover requires the examinee to determine the quickest and shortest path from one point to another by using rules, strategy selection, and visual–spatial thinking. The Riddles task involves solving a verbal puzzle/riddle by using characteristics of concrete or abstract verbal concepts. Associative memory is assessed by two subtests, Atlantis and Rebus, both of which are controlled learning tasks associating a word with a picture/symbol/concept.

Flanagan and colleagues' (2010, 2013) integrated interpretive framework suggests the addition of two subtests, Triangles and Word Order, to those discussed above. However, whereas the Gsm tasks on other instruments are also considered within this framework as being representative of executive functioning, these tasks on the KABC-II are not thus considered. Gsm tasks on the KABC-II are measures of short-term memory span or capacity only, and not of working memory.

The NEPSY-II

Conceptualized from Luria's perspective, the NEPSY-II (Korkman, Kirk, & Kemp, 2007) is an assessment instrument that focuses specifically on the neuropsychological development of children. The original NEPSY was traditionally under the purview of child clinical psychologists and pediatric neuropsychologists. However, the NEPSY-II has recently become popular with school psychologists, who tend to use it more as a measure of cognitive functioning in the tradition of intelligence tests than for its intended neuropsychological purpose, and thus it is included in this discussion. The Lurian tradition requires that assessments identify and distinguish between primary and secondary deficits of cognitive functions. From this perspective, examining simple and complex components of specific domains, and the additional use of qualitative information, provide a more thorough assessment of possible deficits.

The NEPSY-II is purported to measure six domains of cognitive functioning (Attention and Executive Functioning; Language; Sensorimotor; Visuospatial Processing; Memory and Learning; and Social Perception). Korkman and her colleagues (2007) acknowledge that not all cognitive functions constituting a specific domain are assessed. They believe that these cognitive functions do not develop in isolation, but work in concert together, and that broad conclusions based on individual subtests measuring only limited aspects of particular domains should not be drawn.

The Attention and Executive Functioning domain is the most germane to this discussion. Titley and D'Amato (2008) describe the tasks in this domain as a continuum of skills ranging from simple attention to complex self-monitoring. Animal Sorting, Auditory Attention and Response Set, Clocks, Statue, Design Fluency, and Inhibition compose the Attention and Executive Functioning domain of the NEPSY-II. Of these six subtests, four span the widest age range (7–16 years); one is limited to preschoolers (3–6 years); and one is limited to children ages 5–12 years.

Research on the NEPSY-II is currently quite limited; therefore, users must rely on information provided in the technical manual by the test authors. Animal Sorting is a classic card-sorting task similar to the Wisconsin Card Sorting Test (considered the "gold standard" for executive function measures). Card-sorting tasks demonstrate good construct validity as measures of concept generation, which is thought to be a fundamental executive function. Animal Sorting was designed to measure the ability to form basic concepts, categorize, and shift fluently from one concept to another. The Auditory Attention and Response Set subtest consists of two discrete tasks. Auditory Attention requires sustained and selective auditory attention, whereas Response Set adds the element of shifting attention and the ability to maintain set while inhibiting previously learned responses. The Clocks task uses analog clocks to assess planning, organization, visual-perceptual, and visual–spatial skills. Similar tasks have been frequently used for adults with brain injuries, who often exhibit impaired performance. Cohen, Riccio, Kibby, and Edmonds (2000) evaluated the use of a clocks task with children and found that performance improvements were associated with age. Children at age 6 can draw the basic elements of a clock, and by ages 10–12 most children can generate an accurate picture of a clock but may make positioning errors. Adult levels of performance are reached in early adolescence. However, no research appears to have looked at the validity of clocks tasks as measures of either executive function or attention. The Design Fluency task was designed to assess the ability to generate as many unique designs as possible by connecting dots presented in either a structured or random array. It provides a measure of the efficiency and speed of cognitive visual processing. The Inhibition task is a variation of the classic Stroop task (another "gold standard" measure of the executive functions of attentional control and set shifting); it was designed to assess the ability to inhibit an automatic response in favor of a novel response and to shift or switch response types. The task requires the individual to view a series of shapes or arrows and then to name either the shape, the direction, or an alternative response.

To date, there are no published independent research studies examining the Attention and Executive Functioning domain of the NEPSY-II; the available research focuses on the original NEPSY. Since the NEPSY-II replaced three of the original six subtests in this domain with new tasks, conclusions from previous research cannot be brought to bear on the validity of the current battery as a measure of attention, executive function, or a mixture of both. Kemp (2007) has stated that the subtests in the Attention and Executive Functioning domain measure the core components of executive functioning (including strategic planning, cognitive flexibility, and self-regulation), as well as subcomponents such as initiation, fluency, inhibition, and working memory. However, Kemp notes that the domain does not measure all aspects of executive functioning or attention. A major limitation of the Attention and Executive Functioning domain is the lack of empirical research on the validity of the domain. More specifically, the merging of the cognitive constructs of attention and executive function, and the validity of the tasks chosen to measure these cognitive constructs, have received limited attention in the literature.

TARGETED EVALUATION OF EXECUTIVE FUNCTIONS: A BRIEF DISCUSSION

Friedman and colleagues (2006) suggest that current measures of intelligence do not capture the breadth and depth of executive functioning and are best considered as screening tools for executive dysfunction. Evidence of executive function deficits therefore warrants a more targeted evaluation of identified areas of dysfunction. There are numerous instruments that target the assessment of executive functions and related neurocognitive constructs. A complete overview of such assessment instruments is beyond the purview and scope of this chapter; the reader is referred to Maricle and colleagues (2010) for a more comprehensive review.

Targeted assessment instruments measure neurocognitive constructs such as executive functioning, attention, or working memory more specifically and narrowly. As such, they often require more specialized knowledge and training for proper use. However, several targeted instruments of executive functions are commonly used by school psychologists and other child clinicians, including the D-KEFS; the Behavior Rating Inventory of Executive Function, Second Edition (BRIEF 2); and various versions of category, trail-making, tower, and Stroop-like tasks. In addition, numerous measures are used to evaluate related neurocognitive constructs such as memory (the Test of Memory and Learning—Second Edition; the Wide Range Assessment of Memory and Learning, Second Edition; the Children's Memory Scale) or attention (the Test of Everyday Attention for Children; the Conners Continuous Performance Test 3rd Edition). Two targeted assessment instruments thought to measure executive functions, The BRIEF 2 and the D-KEFS are briefly discussed here because they are the most commonly used in psychoeducational, psychological, and neuropsychological evaluations.

Behavior Rating Inventory of Executive Function, Second Edition

The BRIEF 2 (Gioia, Isquith, Guy, & Kenworthy, 2015) is the first revision of the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000). The BRIEF has been frequently used by school psychologists and other professionals to measure executive functions in children and adolescents, and it is assumed that the BRIEF 2 will also be considered a popular measure. Unfortunately, a measure such as the BRIEF 2 is often the *only* measure used to assess executive functions. The BRIEF 2 defines executive functioning as "a multidimensional construct with distinct but interrelated domains of self-regulatory or management functions including the ability to initiate behavior, inhibit the effect of stimuli, select relevant tasks goals, plan and organize means to solve complex problems, monitor and evaluate the success of problem-solving strategies and shift problem-solving strategies flexibly when necessary" (Gioia et al., 2015, p. 3). Gioia and colleagues (2015) go on to say that executive functions are not restricted to cognitive constructs, and that control of emotional response and control of behavioral action also fall under the executive function canopy. The BRIEF 2 is standardized for children ages 5–18 years (parent and teacher report) and children 11–18 years (self-report). The questionnaire for teachers and/or parents utilizes a 3-point Likert scale ("never," "sometimes," "often") to determine how often a child performs a behavior that is thought to be a manifestation of executive dysfunction. The BRIEF 2 provides *T* scores (higher scores are indicative of dysfunction); it yields a global measure of executive functioning, the Global Executive Composite, and three indexes, Behavior Regulation, Emotion Regulation, and Cognitive Regulation. Two subtests (Inhibit and Self Monitor) constitute the Behavior Regulation Index; two subtests encompass the Emotion Regulation Index (Shift and Emotional Control); and five subtests make up the Cognitive Regulation Index (Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials).

The BRIEF 2 is not a significant revision. In order to maintain the integrity of the scale and continuity with previous research, the authors chose not to add new items to the clinical scales or indexes. They did shorten the instrument slightly by eliminating items, and they revised other items for clarity and readability. A significant limitation of the original BRIEF was the adequacy and representativeness of its normative sample, which was obtained from one state (Maryland) and limited in size (1,419 parents and 720 teachers). The BRIEF 2 provides a larger, more diverse normative sample. The subsamples for the parent rating scale (n = 1,400), teacher rating scale (n = 1,400), and self-rating scale (n = 803) approximate the U.S. population on key demographic variables such as gender, race/ethnicity, parental education, and geographical region.

Internal consistency (the degree to which items within a single scale are measuring the same underlying construct) is adequate for the BRIEF 2. More problematic is its test–retest reliability (stability of the measure over time). Rather than address the rather mediocre findings (ranging from .61 to .92), the authors note that the results are within the acceptable standard (>.70) and that *T* scores do not change significantly over the test–retest period. A more significant concern is poor interrater reliability (<.50), in that parents have consistently been found to rate their children as having significantly greater difficulties than teacher raters do. If intrarater reliability (the same rater over time) were to be examined, which it has not been with the BRIEF 2, speculation is that it might be equally problematic.

With regard to the validity of the BRIEF 2, its authors claim that validity was established with the original version, and that since the BRIEF 2 is essentially the same measure, its validity is already established. The authors utilized confirmatory factor analysis to describe the latent factors of the model structure underlying the BRIEF 2. They noted that the analysis showed an acceptable fit for the data, but they seemed to minimize questionable fit statistics and the need to account for significant cross-loading or intercorrelation between factors.

Professionals who choose to use the BRIEF 2 need to realize that it is an indirect measure, and that what is being measured is an adult's perception (or a child/adolescent's own perception, in the case of the self-report version) of a young person's behavioral manifestations of executive dysfunction; the youth's cognitive executive functions are not assessed directly. In fact, the research with the original BRIEF suggests that it does not correlate with direct measures of executive functions (Anderson, Anderson, Northam, Jacobs, & Mikiewicz, 2002; Benjamin, 2004; Mangeot, Armstrong, Colvin, Yeates, & Taylor, 2002; Vriezen & Pigott, 2002). Given the BRIEF 2's similarity to its predecessor, it (like the BRIEF) should be considered a measure of attention or behavior, similar to measures such as the Behavior Assessment System for Children, Second Edition; the Achenbach System of Empirically Based Assessment; or the Conners Behavior Rating Comprehensive Scales. Based on research with the original BRIEF, the consensus of most researchers would be that the BRIEF 2 should never be used as the sole measure of executive functions.

Delis–Kaplan Executive Function System

An instrument that is gaining in popularity among school psychologists is the D-KEFS (Delis et al., 2001). The D-KEFS consists of an assortment of standardized individual measures of executive functioning, which Delis and colleagues combined into a comprehensive battery. There are nine tests, which may be administered in combination or separately: the Word Context Test, Sorting Test, Twenty Questions Test, Tower Test, Color–Word Interference Test, Verbal Fluency Test, Design Fluency Test, Trail Making Test, and Proverbs Test. The tests are thought to measure mental flexibility, inhibition, problem solving, planning, impulse control, concept formation, abstract thinking, and verbal or spatial creativity (Homack, Lee, & Riccio, 2005). The D-KEFS does not yield an overall composite score representing executive functioning; rather, each test is scored and interpreted individually. Each test yields an aggregate score (scaled score and percentile) and relevant process scores (response accuracy, error rate, and response latency).

A primary limitation of the D-KEFS is that its clinical usefulness with children and adolescents is largely unknown, since the majority of its tasks were developed and standardized on adults. Research using the D-KEFS is scarce, with the majority of research having occurred previously on the individual measures incorporated into the D-KEFS rather than on the battery as a whole. Research using the battery, or even the individual tests, with children is very limited. This limitation is a critical one, as the D-KEFS is subject to strong age effects: Children at the youngest ages generally exhibit the lowest scores, and performance on most of the measures is highly influenced by speed of processing until early adolescence (Strauss, Sherman, & Spreen, 2006). A second limitation of the D-KEFS is that it is a challenging instrument to learn to administer, score, and interpret. Clinicians wishing to use the battery should have appropriate training, practice, and supervision with the instrument, as well as an adequate background in neuropsychological assessment, in order to interpret the obtained results appropriately.

Currently, the research indicates that the most defensible way to evaluate executive functioning is to assess executive function components separately, using targeted measures specifically developed for this purpose. The dilemma facing the clinician is to determine which executive functions should be measured. As previously stated, researchers disagree about this, and there are various recommendations in the literature. The integrated SNP/CHC model (Miller, 2007, 2013) suggests the following: concept generation, inhibition, cognitive flexibility, planning, reasoning, and problem solving. Stuss (2009) would propose measuring initiation, planning, sequencing, inhibition, flexibility (shifting), and self-monitoring. Maricle and colleagues (2010) and Maricle and Johnson (2016) recommend evaluating the components of cognitive (mental) flexibility, attention, working memory, and problem solving/reasoning. The primary purposes for any assessment are to provide an understanding of an individual's functioning, as well as his or her specific strengths and weaknesses; to assist in differential diagnosis; and to inform intervention. Thus best practice indicates that the choice of which executive functions to assess should depend on the referral question, potential diagnostic conclusions, and subsequent assessment results.

MANIFESTATION OF EXECUTIVE DYSFUNCTION IN COMMON CHILDHOOD DISORDERS

Instead of asking what executive functions should be measured, astute clinicians might ask why it is important to measure executive functions and what executive functions have to do with an individual's ability to function in a home, school, or employment setting. Evans (2009) states that executive functions enable individuals to deal with problems that arise in everyday life and to cope with new situations; they are the cognitive skills required to identify and achieve personal goals and to modify actions when required. Interest in executive functions has resulted in greater study of executive functions in applied settings, and thus the importance of executive functioning to success in intellectual and social environments has been increasingly recognized (Han, Delis, & Holdnack, 2008). Good executive functioning in children or adolescents rarely warrants concern, but executive dysfunction often results in educational performance deficits that result in referrals for assessment. Executive dysfunction may be the fundamental underlying mechanism or symptom of many childhood disorders.

Delis and colleagues (2007) and Han and colleagues (2008) view executive functioning as a key concept that should be routinely assessed as part of any psychoeducational, psychological, or neuropsychological assessment. Including a component of executive functioning may be appropriate whenever the referral question relates to known or suspected attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), nonverbal learning disability (NVLD), specific learning disability (SLD), traumatic brain injury (TBI), or some other identifiable neurological disorder (e.g., epilepsy or Tourette syndrome) (Barkley, 2000; Brookshire, Levin, Song, & Zhang, 2004; Meltzer & Krishnan, 2007; Ozonoff & Schetter, 2007; Parrish et al., 2007; Shallice et al., 2002; Vaquero, Gomez, Quintero, Gonzalez-Rosa, & Marquez, 2008). To obtain the most informed differential diagnosis during an evaluation, it is necessary to understand how executive dysfunction may be manifested in each of these disorders, but it is also important to note that children with these clinical disorders often exhibit global executive functioning impairments that do not necessarily differentiate clinical groups (P. Anderson, 2002; Ozonoff & Schetter, 2007; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005).

Children and adolescents with ADHD most noticeably demonstrate difficulties with inhibition and attention (Barkley, 2000; Shallice et al., 2002). They often display difficulties with self-monitoring and vigilance, working memory, organizational skills, and planning (Roth & Saykin, 2004). These deficits may result in the obvious problems of behavioral regulation seen in ADHD, such as difficulties with inhibitory control (i.e., frequent impulsive or poorly thought-out behavior), problems with behavioral initiation (i.e., inability to independently initiate tasks like homework or chores), or difficulty in sustaining behavior (i.e., with resulting failures of task completion). Children and adolescents with ADHD fail to engage in planning or problem solving. They may carry out routine tasks, but if faced with novelty or lack of structure, they are at a loss as to what to do (Evans, 2009). They are seen as easily distracted and disorganized, in that they fail to pay attention to rules, complete work haphazardly, and rarely have the materials needed to complete a given task. Unfortunately, while the behavioral manifestations of ADHD appear consistent with impairments of executive function, tests of executive function do not always differentiate children with ADHD from children without ADHD or with other clinical disorders.

Children and adolescents with ASD may have problems with cognitive flexibility, planning, fluently shifting attention to new or novel tasks, appropriately responding to social cues and regulating social interactions, and recognizing or using nonverbal behaviors (Ozonoff & Schetter, 2007). These deficits are often manifested as difficulty in adapting to environmental changes, resulting in the rigidity and need for sameness seen in many children on the autism spectrum. Poor abstract reasoning skills often result in an inability to understand the subtleties of social interactions. Children with ASD often view things very concretely and may miss the figurative meaning of language involved in sarcasm or jokes. Although executive dysfunction is a significant component of ASD, manifestations of executive dysfunction do not define distinctions in ASD subtypes (Verte, Geurts, Roeyers, Oosterlaan, & Sergeant, 2006).

Executive dysfunction is also often associated with SLD, due to cognitive difficulties that interfere with academic learning; these include problems with self-regulation and monitoring, problem solving, retrieval fluency, cognitive flexibility, or organizing and prioritizing stimuli (Meltzer & Krishnan, 2007). Executive function deficits can have a significant impact on the essential academic skills involved in reading, writing, or mathematics. For example, reading decoding may be affected by difficulties with sustained attention and working memory, which are needed in order to attend to each phoneme, manipulate and apply phonemes according to reading rules, and fluently retrieve and remember each phoneme in order to sound out a word (McCloskey, Perkins, & Van Divner, 2009). Writing may be affected by poor graphomotor control, retrieval fluency deficits, or organizational difficulties, resulting in imprecise handwriting skills, strained writing with limited creativity, and disconnected or dysfluent ideas (McCloskey et al., 2009). In math, problems with working memory, inhibition, and retrieval fluency may affect the ability to recall rote math facts quickly and efficiently, to solve problems requiring mental manipulation, or to switch mental sets in order to solve problems using complex math facts (Bull & Scerif, 2001). Executive dysfunction has also been implicated in NVLD because of the problems with cognitive flexibility, set shifting, adapting to novel situations, working memory, self-regulation, and attentional control that are often seen in the population with NVLD (Stein & Krishnan, 2007).

Problems with executive functioning are also frequently seen in children and adolescents experiencing a variety of neurological conditions, such as TBI, tumors, or seizure disorders. With these disorders, impairment in executive function is often correlated with age of onset, as well as severity and location of injury, seizure activity, or tumor. Global impairments in executive functioning, and critical disruptions in attention and processing speed, are common in these children (Brookshire et al., 2004; Parrish et al., 2007; Vaquero et al., 2008). Han and colleagues (2008) note that executive function deficits are frequent sequelae of moderate and severe TBI, and that impairment on a variety of executive function measures commonly occurs, suggesting that comprehensive assessment and follow-up of these functions are warranted for these individuals.

Evaluation of executive functioning is critical to understanding why some children have difficulty with learning and behavior. Han and colleagues (2008) believe that the first step in designing appropriate interventions is to identify the core impairments in executive functioning and to understand how the deficits relate to behavior and learning problems. Various approaches to interventions for executive dysfunction have been developed; however, the evidence base for their effectiveness remains limited, though research in this area is growing. Evans (2009) claims that now is the time to develop and evaluate rehabilitation methods that are clearly set in a theoretical context and can be prescribed on the basis of assessment information.

CONCLUSION

Executive functions appear to be the foundation for human development and the cornerstone for cognition. The study of executive functions is a developing field that is still grappling to describe these functions and their relationship to cognition and behavior. This feat is made more difficult by a disconnection among theory, assessment, and intervention, as well as by limited research on the developmental nature of executive functions in children, the best ways to assess these functions, and effective ways to intervene if executive dysfunction is identified. Nevertheless, child clinicians including pediatric neuropsychologists and school psychologists are increasingly focusing on the assessment of executive functions in their assessments. Thus it is important for these clinicians to have appropriate training in administration, scoring, and interpretation of various measures of executive functions; adequate knowledge about what aspects of executive functions are being measured by various tasks in the different instruments; and an understanding of the neurological substrates and developmental trajectories of various executive functions. These issues, as well as the adequacy of measures of intelligence or targeted measures of neurocognitive constructs for measuring executive functioning, need to be discussed in assessment courses within training programs and within the broader field of child assessment in general.

REFERENCES

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16,* 17–42.

Anderson, M. (2008). The concept and development of general intellectual ability. In J. Reed & J. Warner-Rogers (Eds.), *Child neuropsychology: Concepts, theory and practice* (pp. 112–135). Chichester, UK: Wiley-Blackwell.

Anderson, P. (2002). Assessment and development of executive function during childhood. *Child Neuropsychology, 8,* 71–82.

Anderson, V. (2001). Assessing executive functions in children: Biological, psychological and developmental considerations. *Developmental Rehabilitation, 4,* 119–136.

Anderson, V. (2002). Executive functions in children: Introduction. *Child Neuropsychology, 8,* 69–70.

Anderson, V., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive functions in children with brain disease. *Child Neuropsychology, 8,* 231–240.

Anderson, V., Levine, H. S., & Jacobs, R. (2002). Executive functioning after frontal lobe injury: A developmental perspective. In R. T. Knight & D. T. Stuss (Eds.), *Principles of frontal lobe functions* (pp. 504–507). New York: Oxford University Press.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5,* 115–129.

Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and Cognition,* 68(1), 92–99.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 35,* 31–36.

Ardila, A., & Surloff, C. (2007). *Dysexecutive syndromes.* San Diego, CA: Medlink Neurology.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology,* 49A(1), 5–28.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Barkley, R. A. (2000). Commentary on the multimodal treatment study of children with ADHD. *Journal of Abnormal Psychology, 28,* 595–599.

Baron, I. S. (2004). *Neuropsychological evaluation of the child.* New York: Oxford University Press.

Benjamin, M. L. (2004). *Pilot data on the Behavior Rating Inventory of Executive Function (BRIEF) and performance measures of executive function in pediatric traumatic brain injury.* Unpublished master's thesis, University of Florida.

Blair, C. (2006). Toward a revised theory of general intelligence: Further examination of fluid cognitive abilities as unique aspects of human cognition. *Behavioral and Brain Sciences, 29,* 145–153.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in childhood. *Developmental Neuropsychology, 28,* 561–571.

Bradshaw, J. L. (2001). *Developmental disorders of the frontostriatal system: Neuropsychological, neuropsychiatric and evolutionary perspectives.* Hove, UK: Psychology Press.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6–13: A dimensional and developmental study. *Developmental Neuropsychology, 26,* 571–593.

Brookshire, B., Levin, H. S., Song, J. X., & Zhang, L. (2004). Components of executive function in typically developing and head-injured children. *Developmental Neuropsychology, 25*(1), 61–83.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273–293.

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford–Binet Intelligence Scales—Fifth Edition for children and adolescents. *School Psychology Quarterly, 23*(4), 533–541.

Canivez, G. L., & Watkins, M. W. (2016). Review of the WISC-V: Critique, Commentary and Independent Analysis. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 683–702). Hoboken, NJ: Wiley.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition: Exploratory factor analysis with the 16 primary and secondary subtests. *Psychological Assessment, 28*(8), 975–986.

Carlson, N. R. (2004). *Physiology of behavior* (8th ed.). Boston: Pearson/Allyn & Bacon.

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23,* 201–216.

Cheung, A. M., Mitsis, E. M., & Halperin, J. M. (2004). The relationship of behavioral inhibition to executive functions in young adults. *Journal of Clinical and Experimental Neuropsychology, 26,* 393–403.

Coalson, D. L., Raiford, S. E., Saklofske, D. H., & Weiss, L. G. (2010). WAIS-IV: Advances in the assessment of intelligence. In L. G. Weiss, D. H. Saklosfske, D. L. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives.* Amsterdam: Elsevier/Academic Press.

Cohen, M. J., Riccio, C. A., Kibby, M. Y., & Edmonds, J. E. (2000). Developmental progression of clock face drawing in children. *Child Neuropsychology, 6,* 64–76.

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis–Kaplan Executive Function System.* San Antonio, TX: Psychological Corporation.

Delis, D. C., Lansing, A., Houston, W. S., Wetter, S., Han, S. D., Jacobsen, M., et al. (2007). Creativity lost: The importance of testing higher-level executive functions in school-aged children. *Journal of Psychoeducational Assessment, 25,* 29–40.

Ding, Y., & Alfonso, V. C. (2016). Overview of the Woodcock–Johnson IV: Organization, content and psychometric properties. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 2–30). London: Academic Press/Elsevier.

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford–Binet—Fifth Edition. *Journal of Psychoeducational Assessment, 24,* 123–136.

Dumont, R., Willis, J. O., & Elliott, C. D. (2009). *Essentials of DAS-II assessment.* Hoboken, NJ: Wiley.

Elliott, C. D. (1983). *British Ability Scales.* Windsor, UK: NFER-Nelson.

Elliott, C. D. (1990). *Differential Ability Scales.* San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition.* San Antonio, TX: Harcourt Assessment.

Elliott, R. (2003). Executive functions and their disorders. *British Medical Bulletin, 65,* 49–59.

Evans, J. J. (2009). Rehabilitation of executive functioning: An overview. In M. Oddy & A. Worthington (Eds.), *The rehabilitation of executive disorders: A guide to theory and practice* (pp. 59–73). Oxford: Oxford University Press.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Floyd, R. G., McCormack, A. C., Ingram, E. L., Davis, A. E., Bergeron, R., & Hamilton, G. (2006). Relations between the Woodcock Johnson III clinical clusters and measures of executive functions from the Delis–Kaplan Executive Function System. *Journal of Psychoeducational Assessment, 24,* 303–317.

Floyd, R. G., Shaver, R. B., & McGrew, K. S. (2003). Interpretation of the Woodcock–Johnson III Tests of Cognitive Abilities: Acting on evidence. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 29–46). Boston: Academic Press.

Ford, L., Keith, T. Z., Floyd, R. G., Fields, C., & Schrank, F. A. (2003). Using the WJ III Tests of Cognitive Abilities with attention deficit hyperactivity disorder. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 319–344). Boston: Academic Press.

Friedman, N. R., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172–179.

Fuster, J. M. (2001). The prefrontal cortex—an update: Time is of the essence. *Neuron, 30,* 319–333.

Fuster, J. M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology, 31,* 373–385.

Gioia, G. A., Isquith, P. K., Guy, S., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function.* Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function, Second Edition.* Lutz, FL: Psychological Assessment Resources.

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 3–12). New York: Springer.

Grafman, J. (2006). Human prefrontal cortex: Processes and representations. In J. Risberg & J. Grafman (Eds.), *The frontal lobes: Development, function, and pathology* (pp. 69–91). Cambridge, UK: Cambridge University Press.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Han, S. D., Delis, D. C., & Holdnack, J. A. (2008). Extending the WISC-IV: Executive functioning. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 497–515). Burlington, MA: Elsevier/Academic Press.

Happaney, K., Zelazo, P. D., & Stuss, D. T. (2004). Development of orbitofrontal function: Current themes and future directions. *Brain and Cognition, 35,* 1–10.

Homack, S., Lee, D., & Ricco, C. A. (2005). Test review: Delis–Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology, 27,* 599–609.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions. *Child and Adolescent Mental Health, 7,* 131–142.

Hughes, C., & Graham, A. (2008). Executive functions and development. In J. Reed & J. Warner-Rogers (Eds.), *Child neuropsychology: Concepts, theory, and practice* (pp. 264–284). Chichester, UK: Wiley-Blackwell.

Hunter, S. J., Hinkle, C. D., & Edidin, J. P. (2012). The neurobiology of executive functions. In S. J. Hunter & E. P. Sparrow (Eds.), *Executive function and dysfunction: Identification, assessment and treatment* (pp. 37–64). Cambridge, UK: Cambridge University Press.

Hunter, S. J., & Sparrow, E. P. (Eds.). (2012). *Executive function and dysfunction: Identification, assessment and treatment.* Cambridge, UK: Cambridge University Press.

Jacobs, R., Harvey, A. S., & Anderson, V. (2011). Are executive skills primarily mediated by the prefrontal cortex in childhood?: Examination of focal brain lesions in childhood. *Cortex, 47,* 808–824.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review, 17,* 213–233.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual difference perspective. *Psychonomic Bulletin and Review, 9,* 637–671.

Kaufman, A. S. (2010). Foreword. In L. G. Weiss, D. H. Saklosfske, D. L. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. xiii–xxii). Amsterdam: Elsevier/Academic Press.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, J. C., Kaufman, A. S., Kaufman-Singer, J., & Kaufman, N. (2005). The Kaufman Assessment Battery for Children—Second Edition and the Kaufman Adolescent and Adult Intelligence Test. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 344–370). New York: Guilford Press.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher-order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 30,* 89–119.

Kemp, S. (2007, July). *Introduction to the clinical development and applications of the NEPSY II.* Paper presented at the National School Neuropsychology Conference, Grapevine, TX.

Knight, R. T., & Stuss, D. T. (2002). *Principles of frontal lobe function.* New York: Oxford University Press.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II clinical and interpretive manual.* San Antonio, TX: Harcourt Assessment.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Lewandowski, L. J., & Lovett, B. J. (2008). Introduction to neuropathology and brain behavior relationships.

In R. C. D'Amato & L. C. Hartlage (Eds.), *Essentials of neuropsychological assessment: Treatment planning for rehabilitation* (2nd ed., pp. 31–56). New York: Springer.

Lidz, C. (2003). *Early childhood assessment.* New York: Wiley.

Loring, D. W., & Bauer, R. M. (2010). Testing the limits: Cautions and concerns regarding the new Wechsler IQ and Memory scales. *Neurology, 74*(8), 685–690.

Mangeot, S. D., Armstrong, K., Colvin, A. N., Yeates, K. O., & Taylor, H. G. (2002). Long term executive deficits in children with traumatic brain injuries: Assessing using the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology, 8,* 271–284.

Maricle, D. E., & Johnson, W. (2016). Instructional implications from the Woodcock–Johnson IV Tests of Cognitive Abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 123–151). London: Academic Press/Elsevier.

Maricle, D. E., Johnson, W., & Avirett, E. (2010). Assessing and intervening in children with executive function disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidenced-based intervention* (pp. 599–640). Hoboken, NJ: Wiley.

Mays, K. L., Kamphaus, R. W., & Reynolds, C. R. (2009). Applications of the Kaufman Assessment Battery for Children, 2nd Edition in neuropsychological assessment. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (3rd ed., pp. 281–295). New York: Springer.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*(2), 222–243.

McCloskey, G., & Perkins, L. A. (2013). *Essentials of executive functions assessment.* Hoboken, NJ: Wiley.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties.* New York: Routledge.

McCrea, S. M., & Robinson, T. P. (2011). Visual Puzzles, Figure Weights, and Cancellation: Some preliminary hypotheses on the functional and neural substrates of these new WAIS-IV subtests. *ISRN Neurology, 2011,* PMC3263563.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014) *Technical manual: Woodcock–Johnson IV.* Rolling Meadows, IL: Riverside.

Meltzer, L., & Krishnan, K. (2007). Executive function difficulties and learning disabilities: Understandings and misunderstandings. In L. Meltzer (Ed.), *Executive function in education* (pp. 77–105). New York: Guilford Press.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment.* Hoboken, NJ: Wiley.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: Wiley.

Miller, D. C., & Hale, J. B. (2008). Neuropsychological applications of the WISC-IV and WISC-IV Integrated. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 445–495). Burlington, MA: Elsevier/Academic Press.

Miller, D. C., & McGill, R. (2016). Review of the WISC-V. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 645–662). Hoboken, NJ: Wiley.

Miller, D. C., McGill, R., & Bauman Johnson, W. (2016). Neurocognitive applications of the Woodcock–Johnson IV. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives.* London, UK: Academic Press/Elsevier.

Miller, D. C., McGill, R., & Maricle, D. E. (2017). *WJ IV scholastic aptitude/achievement comparisons by age* (Woodcock–Johnson IV Assessment Service Bulletin No. 8). Rolling Meadows, IL: Riverside.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tests: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Ozonoff, S., & Schetter, P. (2007). Executive dysfunction in autism spectrum disorders: From research to practice. In L. Meltzer (Ed.), *Executive function in education* (pp. 133–160). New York: Guilford Press.

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: Parent-report and cognitive assessment. *Developmental Medicine and Child Neurology, 49,* 412–416.

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV assessment.* Hoboken, NJ: Wiley.

Reynolds, C. R. (2007, July). *Frontal lobe development: What is it and when does it end?* Paper presented at the National School Neuropsychology Conference, Grapevine, TX.

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (2002). *The prefrontal cortex: Executive and cognitive function.* Oxford, UK: Oxford University Press.

Roid, G. (2003a). *Interpretive manual: Expanded guide to the interpretation of the SB5 test results.* Itasca, IL: Riverside.

Roid, G. (2003b). *Stanford–Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Roid, G., & Barram, R. A. (2004). *Essentials of Stanford–Binet Intelligence Scales (SB5) assessment.* Hoboken, NJ: Wiley.

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12,* 190–201.

Roth, R. M., & Saykin, A. J. (2004). Executive dysfunction in attention deficit hyperactivity disorder: Cognitive and neuroimaging findings. *Psychiatric Clinics of North America, 27*(1), 83–96.

Salthouse, T. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19,* 532–545.

Sattler, J. M., & Dumont, R. (2008). Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV): Description. In J. M. Sattler, *Assessment of children: Cognitive foundations* (5th ed., pp. 265–315). San Diego, CA: Jerome M. Sattler.

Sattler, J. M., Dumont, R., Salerno, J. D., & Roberts-Pittman, B. (2008). Stanford–Binet Intelligence Scales—Fifth Edition (SB5). In J. M. Sattler, *Assessment of children: Cognitive foundations* (5th ed., pp. 565–604). San Diego, CA: Jerome M. Sattler.

Sattler, J. M., Dumont, R., Willis, J. O., & Salerno, J. D. (2008). Differential Ability Scales—Second Edition (DAS-II). In J. M. Sattler, *Assessment of children: Cognitive foundations* (5th ed., pp. 605–675). San Diego, CA: Jerome M. Sattler.Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock–Johnson IV Tests of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock–Johnson IV Tests of Oral Language.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Schneider, W. J. (2016). Strengths and weaknesses of the Woodcock–Johnson IV Tests of Cognitive Abilities: Best practice from a scientist-practitioner perspective. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 192–211). London: Academic Press/Elsevier.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.

Shallice, T., Marzocchi, G. M., Coser, S., Del Savio, M., Meuter, R. F., & Rumiati, R. L. (2002). Executive function profile of children with attention deficit hyperactivity disorder. *Developmental Neuropsychology, 21,* 43–71.

Stein, J., & Krishnan, K. (2007). Nonverbal learning disabilities and executive function: The challenges of effective assessment and teaching. In L. Meltzer (Ed.), *Executive function in education* (pp. 106–132). New York: Guilford Press.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York: Oxford University Press.

Stuss, D. T. (2009). Rehabilitation of frontal lobe dysfunction: A working framework. In M. Oddy & A. Worthington (Eds.), *The rehabilitation of executive disorders: A guide to theory and practice* (pp. 3–17). Oxford, UK: Oxford University Press.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63,* 289–298.

Sugarman, R. (2002). Evolution and executive functions: Why our toolboxes are empty. *Revista Espanola de Neuropsicologia, 4*(4), 351–377.

Titley, J., & D'Amato, R. C. (2008). Understanding and using the NEPSY II with young children, children, and adolescents. In R. C. D'Amato & L. C. Hartlage (Eds.), *Essentials of neuropsychological assessment* (2nd ed., pp. 149–172). New York: Springer.

Tranel, D., Manzel, K., & Anderson, S. W. (2008). Is the prefrontal cortex important for fluid intelligence?: A neuropsychological study using matrix reasoning. *Clinical Neuropsychology, 22*(2), 242–261.

van Vreen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Science, 14,* 593–602.

Vaquero, E., Gomez, C. M., Quintero, E. A., Gonzalez-Rosa, J. J., & Marquez, J. (2008). Differential prefrontal-like deficit in children after cerebellar astrocytoma and medulloblastoma tumor. *Behavioral and Brain Functions, 4,* 1–16.

Verte, S., Geurts, H., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2006). Executive functioning in children with an autism spectrum disorder: Can we differentiate within the spectrum? *Journal of Autism and Developmental Disorders, 16,* 351–372.

Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental support on the BRIEF and performance based measures of executive function in children with moderate to severe brain injury. *Child Neuropsychology, 8,* 296–303.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V).* Bloomington, MN: Pearson.

Weiss, L. G., Beal, A. L., Saklofske, D. H., Packiam-Alloway, T., & Prifitera, A. (2008). Interpretation and intervention with WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 3–66). Burlington, MA: Elsevier/Academic Press.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention deficit hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 37,* 1336–1346.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience, 4,* 139–147.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3, Serial No. 274), vii–137.

Zilmer, E. A., Spiers, M. V., & Culbertson, W. C. (2008). *Principles of neuropsychology* (2nd ed.). Belmont, CA: Thomson.

CHAPTER 37

The Role of Cognitive and Intelligence Tests in DSM-5 Diagnosis of Specific Learning Disorder

Erin M. McDonough
Dawn P. Flanagan
Megan C. Sy
Vincent C. Alfonso

Controversies related to the definition of specific learning disability and the methods used to diagnose this condition have long been a foremost concern in assessment-related fields, such as school psychology and neuropsychology (e.g., Decker, Hale, & Flanagan, 2013; Reynolds & Shaywitz, 2009), as well as in the medical and legal fields (e.g., Colker, Shaywitz, Shaywitz, & Simon, 2013). The publication of new criteria for diagnosing a specific learning disorder in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association [APA], 2013) has fueled such controversies (see, e.g., *www.landmarkoutreach.org/about/newsworthy/dyslexia-dropped-dsm-5*). This chapter begins with a brief history of learning disabilities, including the diagnostic criteria for learning disorders in the five editions of the DSM. Next, etiology, associated impairments and comorbidities, course and developmental changes, prognosis, outcomes, and treatment are highlighted. This chapter concludes with a brief discussion of the most salient vagaries inherent in the latest diagnostic criteria for specific learning disorder in the DSM-5, some of which will be debated for some time to come. Note that SLD is used from here on in this chapter to refer to *specific learning disorder* as defined in DSM-5, rather than *specific learning disability* as defined in the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004).

A BRIEF HISTORY OF LEARNING DISABILITIES

Between the years 1800 and 1920, the field of learning disabilities began to emerge as the result of developments in the fields of neurology and reading disabilities. As Franz Joseph Gall explored relationships between head injury and movement and sensation, morality, and intellectual qualities during the early 1800s, the groundwork was laid for localization of brain function, and Broca's area was identified as responsible for speech production (Hallahan & Mercer, 2001). Toward the latter part of the 19th century, Sir William Broadbent identified the relationship between the loss of reading abilities and the loss of the ability to name after the onset of neurological symptoms. Kussmaul's work described what he termed *word-blindness*, which is possibly the first conceptualization of a specific learning disability as a disorder

---

This chapter has been reprinted from McDonough, Flanagan, Sy, and Alfonso (2017).

separate from other potential problems (Hallahan & Mercer, 2001). Similar case studies describing reading disabilities were reported by Hinshelwood and Morgan in the late 1800s and early 1900s. Hinshelwood (1902) also studied congenital and acquired reading disabilities and reported on six cases appearing within a family over two generations. Through these studies, he hypothesized that the primary deficit in these children was visual memory for words and letters, which might now be defined as an orthographic processing deficit (Hallahan & Mercer, 2001; see also Berninger, Abbott, Nagy, & Carlisle, 2010).

Interested in the earlier research of the aforementioned European scientists, several American researchers including Orton, Fernald, Monroe, and Kirk continued the study of learning disabilities, ranging from problems with language and reading to perceptual–motor and attentional disabilities. Orton (1925) determined that many students with reading disabilities had average to above-average scores on the Stanford–Binet IQ test, as Hinshelwood (1902) had also suspected. Orton's work significantly contributed to interventions for reading disabilities and emphasized a multisensory approach.

A former research associate for Orton, Marion Monroe, experimented with his methods for remediation as well as the methods of Fernald and Keller. Developing diagnostic tests and using results to inform instruction, she developed a synthetic phonetic approach and compared it to other methods (see Hallahan & Mercer, 2001). Monroe was the first to suggest that reading disabilities could be identified by examining the discrepancy between actual achievement and expected achievement, the latter of which was an average of the child's chronological age, mental age, and arithmetic grade (Monroe, 1932). Monroe also emphasized error analysis in order to guide instruction. Her work thus laid the foundation for the discrepancy approach and for the response-to-intervention approach to specific learning disability identification (Hallahan & Mercer, 2001; Hallahan & Mock, 2003).

The First Definition of Learning Disability

Working alongside Monroe as an instructor while earning his doctorate, Samuel Kirk studied interventions for reading disabilities and provided what is considered to be the first definition of the term *learning disability*. His definition linked brain dysfunction to behavioral manifestation of academic problems. According to Kirk (1962),

> A learning disability refers to a retardation, disorder, or delayed development in one or more of the processes of speech, language, reading, writing, arithmetic, or other school subject resulting from a psychological handicap caused by possible cerebral dysfunction and/or emotional or behavioral disturbances. It is not the result of mental retardation, sensory deprivation, or cultural and instructional factors. (p. 263)

During the 1970s, Kirk continued to develop assessment techniques to identify learning disabilities in children. His contribution to the field of learning disabilities, including the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1967), and his definition of the disorder provided the foundation for later researchers to explicate the relationship between cognitive functions and academic problems. Integrating Monroe's discrepancy-based definition of learning disabilities and Kirk's brain-based definition, Barbara Bateman (1965) proposed an alternative definition:

> Children who have learning disorders are those who manifest an educationally significant discrepancy between their estimated potential and actual level of performance related to basic disorders in the learning process, which may or may not be accompanied by demonstrable central nervous system dysfunction, and which are not secondary to generalized mental retardation, educational or cultural deprivation, severe emotional disturbance, or sensory loss. (p. 220)

Since Bateman's reintroduction of the concept of discrepancy from cognitive potential, discrepancy has remained a central feature of most definitions, including the response-to-intervention approach, and has at times been a controversial feature of later definitions (see Flanagan & Alfonso, 2011, for a comprehensive review).

Contemporaneously, the federal government began its own efforts to provide a definition through the creation of two task forces; however, these task forces failed to provide a unified definition of learning disabilities. Task Force I provided a definition of *minimal brain dysfunction*, while Task Force II defined *learning disabilities*. The disparate definitions reflected the fact that Task Force I was made up primarily of medical professionals, while Task Force II consisted of educators.

Shortly thereafter, in the late 1960s, the U.S. Office of Education (USOE) formed the National Advisory Committee on Handicapped Children

(NACHC) to provide a definition that could serve as the basis for legislation and funding. Because Kirk chaired this committee, the resulting definition was highly similar to the definition he provided in 1962, and much of the language used in this definition is present in current definitions (e.g., that of IDEA 2004). The NACHC definition of 1968 emphasized a "disorder in one or more of the basic psychological processes" manifesting as disorders of "thinking, talking, reading, writing, spelling, or arithmetic" and excluded deficits related to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to "environmental disadvantage." In 1969, the Children with Specific Learning Disabilities Act was passed, which was included in the Education of the Handicapped Act of 1970 (Pub. L. 91-230), and provided for programs to serve children who were eligible (Hallahan & Mercer, 2001; Hallahan & Mock, 2003).

In 1975, McCarthy attempted to operationalize the definition of the NACHC by providing criteria. The first of these criteria included an IQ greater than 85, as well as sufficient opportunities for learning. The second of McCarthy's criteria included a "significant discrepancy" between academic aptitude and academic achievement, and the third criterion was that the discrepancy must be large enough to warrant specialized intervention (Tylenda, Hooper, & Barrett, 1987). McCarthy's inclusion of an IQ above 85 is closely aligned with some current research-based definitions of specific learning disabilities (e.g., Berninger, 2011; Flanagan, Ortiz, & Alfonso, 2013; Flanagan, Ortiz, Alfonso, & Mascolo, 2002; Hale & Fiorello, 2004; Kavale, Kauffman, Bachmeier, & LeFever, 2008; McDonough & Flanagan, 2016; Ortiz, Flanagan, & Alfonso, 2015).

Toward Consensus on the Definition and Identification of Learning Disability

From about 1975 to 1985, there was relative stability in the field as it moved toward consensus on the definition of learning disabilities, as well as methods of identifying the condition. Applied research was abundant at this time, much of which was funded by the USOE. This research led to empirically validated educational procedures for students with learning disabilities (Hallahan & Mercer, 2001).

Following the adoption of a federal definition of learning disabilities, organizations and researchers focused their efforts on developing and testing intervention programs for individuals with learning disabilities, and the definition of learning disabilities remained relatively stable (Hallahan & Mercer, 2001). Applied research funded by the USOE established a body of empirically validated methods to intervene with children with specific learning disabilities. The definition of *specific learning disability* was further consolidated by the passing of the Education for All Handicapped Children Act (Pub. L. 94-142), which essentially adopted the 1968 definition proposed by Samuel Kirk and the NACHC. This definition has survived conceptually in the federal legislation in its entirety, with only minor text revisions (Herr & Bateman, 2003; Sotelo-Dynega, Flanagan, & Alfonso, 2011).

Public Law 94-142 provided legitimacy—and, equally important, funding—for direct services for children with specific learning disabilities. In addition, the legislation set forth regulations as a way to deal with the absence of prescriptive guidance with regard to the methods used to identify learning disabilities (Sotelo-Dynega et al., 2011). These regulations continued to stipulate the presence of an ability–achievement discrepancy until the most recent reauthorization of this act: IDEA 2004 and its attendant regulations (U.S. Department of Education, 2006), the latter of which stated that ability–achievement discrepancy may be used as a method of identifying children with specific learning disabilities, but cannot be mandated. Shortly after Public Law 94-142 was enacted, DSM-III (APA, 1980) included discrepancy criteria for learning disorders.

Disorders of Learning in DSM

Even in the earliest versions of DSM, there was some recognition of learning difficulties as distinct problems facing some individuals. The first edition (APA, 1952) briefly referred to a "learning disturbance" under the category of "special symptom reactions," as well as to "specific learning defects," including alexia and agraphia. The definition included a reference to "word-blindness" as initially described by Kussmaul nearly 60 years earlier (Loriaux, 2010). Coding of the specific learning defects diagnosis was altered if known to be related to organic brain dysfunction. Despite the work of many contemporary American researchers in the field, the first edition of the DSM's definition of learning problems was limited in detail and scope.

DSM-II was published in 1968, several years following Samuel Kirk's initial definition of specific learning disabilities, and was again listed under a main category of "special symptoms" as a "specific learning disturbance." The definition or specification of symptom presentation in DSM-II was even more limited than that provided in the first edition, despite the proliferation of research and published work in the field of learning disabilities.

However, DSM-III (APA, 1980) devoted significantly more attention to the issue of defining learning disabilities. The DSM-III definition included six disorders within its new category of specific developmental disorders: developmental reading disorder; developmental arithmetic disorder; developmental language disorders, expressive and receptive types; developmental articulation disorder; and mixed specific developmental disorder. This definition also included a statement that while mutually exclusive of pervasive developmental disorders, a specific developmental disorder in reading or arithmetic could coexist with a diagnosis of mental retardation (APA, 1980, p. 94); this apparent contradiction served to confuse the issue of defining learning disabilities.

DSM-III provided definitions and criteria for developmental reading disorder and developmental arithmetic disorder. The essential feature of these disorders was a "significant impairment" in the academic skill, not accounted for by mental age, chronological age, or schooling. Diagnostic criteria for each of the disorders required a significant discrepancy between intellectual functioning as measured by an individually administered IQ test and performance on an academic achievement test measuring relevant domains. DSM-III also recommended a 1- or 2-year delay in measured reading skills for children ages 8–13 years. The discrepancy between a full scale IQ and academic achievement remained central to the definition and identification of these two disorders and was retained in DSM-III-R.

As compared with the diagnosis of developmental reading and arithmetic disorders in DSM-III, DSM-IV (APA, 1994) presented the category of learning disorders, and three specific diagnoses within this overarching category: reading disorder, mathematics disorder, and disorder of written expression. In addition, a diagnosis of learning disorder not otherwise specified was included in DSM-IV and encompassed academic problems that did not meet criteria for one of the aforementioned specific academic disorders; this diagnostic label was intended to be applied to individuals presenting problems across reading, math, and/or written expression, or who did not meet the threshold set forth by the term "substantially below" (i.e., achievement substantially below IQ), provided that the observed level of academic performance significantly interfered with academic achievement. DSM-IV also moved these disorders from Axis II to Axis I.

DSM-IV and its text revision, DSM-IV-TR (APA, 2000), sought to improve the definition of learning disabilities over the previous version. The definition again relied on a discrepancy approach, stating that "Learning Disorders are diagnosed when the individual's achievement on individually administered, standardized tests in reading, mathematics, or written expression is substantially below that expected for age, schooling, and level of intelligence" (APA, 2000, p. 49). DSM-IV-TR went on to define further "substantially below" as a difference of at least two standard deviations between achievement and measured intelligence. DSM-IV and DSM-IV-TR provided separate definitions for reading disorder, mathematics disorder, and disorder of written expression, and the second criterion for each disorder required that the "disturbance" identified in the first criterion (i.e., achievement as measured by individually administered standardized tests substantially below expected levels) must significantly interfere with academic achievement or with completing activities requiring that skill. And finally, the DSM-IV criteria required an impairment above and beyond a sensory deficit if one was present.

As DSM-III and DSM-III-R had done, DSM-IV and DSM-IV-TR stipulated that the academic deficits must be observed on "individually administered standardized tests," thereby assuring a formal evaluation of academic skills in addition to a measure of intelligence. These manuals also referenced the possibility of coexisting specific cognitive processing deficits in an area such as memory, visual–spatial processing, or the like, but did not require that one be present.

In terms of differential diagnosis, DSM-IV-TR required that learning disorders be differentiated from "normal variations in academic attainment," as well as "lack of opportunity, poor teaching, or cultural factors" (APA, 2000, p. 51), and should be beyond what one would expect to be associated with a sensory impairment (if present). In DSM-IV and DSM-IV-TR, as in DSM-III, a diagnosis of a learning disorder was also possible in addition to a diagnosis of mental retardation, provided that the academic achievement was significantly below estimated intellectual functioning. The publication of DSM-5 (APA, 2013), however, represents a significant departure in criteria for learning disorders as compared to previous editions.

DSM-5 DIAGNOSTIC CRITERIA

The introduction of the SLD diagnosis in DSM-5 resulted in considerable changes to the DSM-IV(-TR) diagnostic criteria for learning disorders. The most significant changes were that reading, mathematics, and writing disorders were combined into the one overarching diagnostic category of SLD. Tannock (2013) described the rationale behind this change as reflecting the work group's conceptualization of SLD as a single disorder that renders learning very difficult and effortful. Furthermore, whereas subtypes describe mutually exclusive disorders, specifiers allow for concurrent diagnoses of impairment in multiple academic domains. Therefore, the criteria were also altered to include symptoms of reading, writing, and mathematics impairment. Criterion A of the DSM-5 description of SLD calls for "difficulties learning and using academic skills, as indicated by the presence of at least one of the following symptoms that have persisted for at least 6 months, despite the provision of interventions that target those difficulties" (APA, 2013, p. 66). Symptoms include inaccurate or slow and effortful word reading, difficulty understanding the meaning of what is read, difficulties with spelling, difficulties with written expression, difficulties mastering number sense, number facts, or calculation, or difficulties with mathematical reasoning.

Criterion B (APA, 2013, p. 67) requires that academic skills must be "substantially and quantifiably below" what would be expected based on chronological age, and must cause significant interference with academic or occupational performance or with activities of daily living. Criterion B also requires that the determination regarding academic skills must be based on "individually administered standardized achievement measures and comprehensive clinical assessment." For individuals over age 17 years, a "documented history" of learning impairment may be substituted for the standardized assessment.

Criterion C states that "the learning difficulties begin during school-age years but may not become fully manifest until the demands for those affected academic skills exceed the individual's limited capacities" (APA, 2013, p. 67), thereby allowing for situations in which older individuals can be newly diagnosed with SLD. DSM-5 provides examples of situations of such demands, including timed tests, reading or writing long complicated reports for a tight deadline, and excessively heavy academic loads.

Finally, Criterion D remains similar to rule-out criteria in previous versions of DSM: It stipulates that the learning difficulties must not be better accounted for by "intellectual disabilities, uncorrected visual or auditory acuity, other mental or neurological disorders, psychosocial adversity, lack of proficiency in the language of academic instruction, or inadequate educational instruction" (APA, 2013, p. 67). Criterion D contains the only reference to cognitive ability in its provision that the academic deficits are not secondary to intellectual disability. A note below the four listed criteria states that they are to be met "based on a clinical synthesis of the individual's history" (APA, 2013, p. 67), including developmental, medical, family, and educational histories, as well as school reports and psychoeducational assessment.

The unification of three separate learning disorders into one diagnostic category has necessitated the addition of specifiers in order to identify the area(s) of academic weakness; a coding note indicates that when more than one domain is impaired, more than one specifier can and should be assigned. The coding note also directs the clinician to document the subskills that are impaired as part of the diagnosis. The specifier *with impairment in reading* is appended to the SLD diagnosis when the individual demonstrates significant impairment in one or more subskills, including word-reading accuracy, reading rate or fluency, and/or reading comprehension. *Dyslexia* may be used as an alternative term referring to problems with word-reading fluency or accuracy, decoding, and spelling, and if this term is used, impaired subskills should also be identified. An impairment in writing skills is assigned to the specifier *with impairment in written expression* and refers to impaired spelling accuracy, grammar and punctuation accuracy, and/or clarity or organization of written expression. The specifier *with impairment in mathematics* is used for individuals who demonstrate significantly below average skills in number sense, memorization of arithmetic facts, accurate or fluent calculation, and/or accurate math reasoning. The term *dyscalculia* can be used to refer to difficulties processing numerical information, learning arithmetic facts, and performing accurate or fluent calculations. Like the use of *dyslexia*, the use of the term *dyscalculia* also requires the identification of any additional impaired subskills.

In addition to specifying the domain of impairment, the degree of severity should also be indicated in the diagnosis. The range of severity includes the specifiers *mild, moderate,* and *severe*. The *mild* specifier is applied to an individual who has some difficulties in one or two academic domains but is able to function well when provided with appropriate accommodations or support services. The *moderate* specifier is used when there are marked difficulties in one or more academic domains and the individual is not likely "to become proficient without some intervals of intensive and specialized teaching during the school years" (APA, 2013, p. 68). Accommodations or supportive services may be needed in school, in the workplace, or at home in order for activities involving the academic skills to be completed accurately and efficiently. The *severe* specifier is used to describe impairments in individuals who are unlikely to learn those skills without "ongoing intensive individualized and specialized teaching for most of the school years" (APA, 2013, p. 68). Even with accommodations, an individual with severe SLD may not be able to perform academic tasks with efficiency.

Etiology

There is consensus among experts in the field that there is no known single cause, but rather numerous factors that contribute to the development of learning disabilities. According to DSM-5, SLD is a neurodevelopmental disorder with a biological basis resulting in cognitive abnormalities; academic skills deficits are considered the behavioral signs or manifestations of the disorder. Genetic, epigenetic, and environmental factors interact to impair the brain's ability to perceive or process information efficiently or accurately (APA, 2013). Table 37.1 provides a summary of the etiology of SLD, along with information on impairments that are associated with this condition and its cognitive correlates.

There is evidence to suggest the role of heredity as a significant contributory factor in the etiology of learning disabilities. Such evidence includes the fact that learning disabilities appear to run in families. Family and twin studies demonstrate moderate to high familiality and heritability of learning disabilities, specifically impairments in reading and spelling (Scerri & Schulte-Korne, 2010; Schulte-Korne, 2001). For example, heritability estimate values greater than .60 have been found for manifestations of learning abilities and disabilities (see Cortiella & Horowitz, 2014). Moreover, several genetic linkage studies implicate chromosomes 6 and 15 in the etiology of reading disorders, with chromosome 6 also suspected in the etiology of ADHD (Schulte-Korne, 2001; Willcutt et al., 2002; see Table 37.1 for more information on heritability).

In addition to the role of genetics, prenatal and perinatal factors are considered to be contributory factors in the etiology of learning disabilities. Prematurity, low birth weight, and *in utero* exposure to alcohol or cocaine are all associated with increased risk for learning problems. Loss of oxygen during birth may also result in altered brain development contributing to low academic achievement. Maternal metabolic conditions, including diabetes, hypertension, and obesity, likewise appear to contribute to the development of neurodevelopmental conditions such as autism and learning problems (Krakowiak et al., 2012).

In addition, research from the 1990s and early 2000s supported neurobiological factors in the etiology of learning disabilities. In numerous publications, Shaywitz and Shaywitz (e.g., 2004) describe overactivation of Broca's area with underactivation in the left parieto-temporal and occipito-temporal regions of the brain in children with reading disabilities. The left inferior prefrontal cortex, left angular gyrus, and inferior parietal lobes have been implicated in the development of math-related learning disabilities (Chochon, Cohen, van de Moortele, & Dehaene, 1999; Kesler, Sheau, Koovakkattu, & Reiss, 2011). In addition, the intraparietal sulcus appears to be a crucial region in the development of number sense (Dehaene, Molko, Cohen, & Wilson, 2004).

Furthermore, environmental characteristics are also considered to be etiological factors in the development of learning disabilities. Low socioeconomic status, low maternal education, male gender, and nonwhite race were all found to be related to increased risk for learning disabilities in a kindergarten sample (Resnick et al., 1999). Malnutrition may be another significant environment factor contributing to learning problems (Groce et al., 2014).

Symptom Presentation

One essential feature of SLD, as set forth in Criterion A, is persistent difficulty in learning vital academic skills that are first experienced during formal schooling. While skills like walking and talking typically develop naturally, academic

TABLE 37.1. Specific Learning Disorder (SLD): Etiology, Associated Impairments, and Cognitive Correlates

Subskill	Etiology	Associated impairments/cognitive correlates
Word-reading accuracy	SLD with impairment in reading Several cortical and subcortical structures are frequently implicated, including the planum temporale, temporal lobes, corpus callosum, and cerebellum (e.g., Eckert et al., 2003). More recent work appears to identify dysfunction in a left-hemispheric network that includes the occipito-temporal region, inferior frontal gyrus, and inferior parietal region (Fletcher et al., 2004; Richlan, 2012; Richlan et al., 2009; Shaywitz et al., 2000; Silani et al., 2005). Numerous imaging studies have also found that dysfunctional responses in the left inferior frontal and temporo-parietal cortices play a significant role with regard to phonological deficits (Skeide et al., 2015). Family and genetic factors have long been identified as crucial in dyslexia, with some researchers suggesting that a child with a parent with a reading disability is eight times more likely to have dyslexia compared to the general population (Pennington & Olson, 2005). Certainly, there is converging evidence from family and twin studies demonstrating the heritability and familiality of dyslexia (Grigorenko, 2001). Recently, genetic linkage studies have also identified several susceptibility genes for reading disabilities. These include sites on chromosomes 1, 2, 3, 4, 6, 11, 15, and 18, with one of the most commonly identified genetic locus being on chromosome 6 (Grigorenko, 2005; Paracchini et al., 2007; Scerri & Schulte-Korne, 2010; Scerri et al., 2011; Skeide et al., 2015). Shared environmental factors include language and literacy environment during childhood (Wadsworth et al., 2000), as well as quality of reading instruction.	**Phonological awareness:** The primary cognitive correlate—the metacognitive understanding that words have internal structures based on phonemes (Fletcher et al., 2007; Melby-Lervåg et al., 2012; Willcutt et al., 2013). When this awareness is impaired, word recognition is delayed, and fluency and comprehension skills are consequently affected. **Rapid naming:** Some researchers have found that phonological awareness and rapid letter naming both uniquely predict word recognition skills (Schatschneider et al., 2004; Wagner et al., 1994, 1997). However, a meta-analysis of studies examining the relationship between rapid naming and dyslexia found little evidence to support a central and persistent deficit in naming speed in individuals with the disorder (Vukovic & Siegel, 2006). On the other hand, there are findings to suggest that phonological awareness and rapid naming, although correlated, are distinct variables and contribute uniquely to word recognition (Petrill et al., 2006). **Phonological memory:** Working memory for verbal and sound-based information has also been found to be significantly related to word recognition, although it may not uniquely contribute when phonological processing is accounted for (Melby-Lervåg et al., 2012; Schatschneider et al., 2004; Wagner et al., 1997; Willcutt et al., 2013).
Reading comprehension	Several brain regions are often implicated in reading comprehension. These include the anterior temporal lobe, inferior temporal gyrus, inferior frontal gyrus, inferior frontal sulcus, and middle and superior frontal and temporal regions (Ferstl et al., 2008; Gernsbacher & Kaschak, 2003). More recent research has revealed a relationship between listening and reading comprehension and activation along the left superior temporal sulcus, which has referred to by some as the "comprehension cortex" (Berl et al., 2010). However, broader pathways are also activated in reading comprehension, reflecting increased cognitive demand compared to listening comprehension.	**Oral language:** Difficulties in reading comprehension are frequently associated with deficits oral language in general, including areas such as vocabulary, morphology, and syntax (Catts et al., 1999; Cutting & Scarborough, 2006; Share & Leikin, 2004; Torgesen, 2000; Willcutt et al., 2013). **Listening comprehension:** Several studies have demonstrated that a unique portion of the variance in reading comprehension can be explained by listening comprehension (Cutting & Scarborough, 2006; Kendeou et al., 2009). **Working memory:** Comprehension involves holding words and sentences in awareness,

(continued)

TABLE 37.1. *(continued)*

Subskill	Etiology	Associated impairments/cognitive correlates
Reading comprehension *(continued)*	Genetic factors are said to account for 41–76% of the variance in comprehension (e.g., Betjemann et al., 2008; Harlaar et al., 2007; Petrill et al., 2007). While genetic factors that influence decoding and listening comprehension account for nearly 40% of the variance in reading comprehension, there is little evidence for an independent source of genetic influence on comprehension alone (Harlaar et al., 2010; Keenan et al., 2006). However, estimating the genetic influences on reading comprehension may be particularly sensitive to the type of assessment test used (Betjemann et al., 2011).	while integrating prior knowledge with incoming information (Carretti et al., 2009). Poor comprehenders may have particular difficulty updating/revising information already in working memory (Pelegrina et al., 2014). **Executive functioning:** Several executive functions are involved in reading comprehension, including planning, organization, and self-monitoring (Cutting et al., 2009; Locascio et al., 2010; Sesma et al., 2008). Weaknesses in these executive functions result in difficulties with higher-order comprehension skills such as inferencing, integrating prior knowledge, monitoring comprehension, and adapting to text structure or genre (Fletcher et al., 2007).
Reading rate or fluency	Brain regions activated are similar to the network implicated in word reading, but additional activation is observed in areas involved in eye movement and attention (Jones et al., 2013). There is also evidence for increased activation in the left occipito-temporal region, in particular the occipito-temporal sulcus, which is important for rapid processing of letter patterns (Dehaene & Cohen, 2011; Shaywitz et al., 2004). Some studies have found increased activation in this region when normal reading automaticity is disrupted (Benjamin & Gaab, 2012). There is limited evidence of genetic influences specific to rapid naming and reading, suggesting that RAN may be etiologically distinct from phonological awareness (Byrne et al., 2005; Compton et al., 2001; Petrill et al., 2006). Genetic linkage studies have identified susceptibility genes for fluency, namely chromosome 2 (Raskind et al., 2005).	**Rapid automatized naming (RAN):** While the exact relationship between RAN and reading remains unclear, RAN is believed to be one of the best predictors of reading fluency (Georgiou et al., 2008; Tan et al., 2005). The automaticity required to complete RAN tasks is related to the ability to synthesize and automatize letter sequences/words when reading (Norton & Wolf, 2012). There are also various cognitive processes implicated in rapid naming. These include attention, executive functions (i.e., response inhibition, set shifting), lexical retrieval, and processing speed (Moll et al., 2015). **Orthographic processing:** Processing of orthographic information (i.e., the ability to process units of words based on visual long-term memory representations) is considered critical in automatic word recognition and consequently plays a crucial role in fluency (O'Brien et al., 2011). This ability is often impaired or underdeveloped in some reading-disabled individuals.
Number sense	SLD with impairment in mathematics Researchers differentiate between the basic processing of numerical information and processes involved in math calculation and problem solving, suggesting that these are both structurally and functionally distinct (Ansari, 2010). The intraparietal sulcus in both hemispheres is widely viewed as crucial in processing and representing numerical quantity, although there may be differences in activation as a function of age (Ansari & Dhital, 2006; Ansari et al., 2005; Dehaene et al., 2004; Kaufmann et al., 2006; Kucian et al.,	**Number representation:** Math disorders are associated with weaknesses in fundamental number representation and processing, which are manifested in difficulties with quantifying sets without counting, using nonverbal processes to complete simple numerical operations, and estimating the relative magnitude of sets (Feigenson et al., 2004; Geary, 2013; Geary et al., 2008, 2009, 2012; Halberda et al., 2008; Mazzocco et al., 2011; Rouder & Geary, 2014). **Number comparison:** Several studies have indicated that math difficulties are associated

(continued)

TABLE 37.1. *(continued)*

Subskill	Etiology	Associated impairments/cognitive correlates
Number sense *(continued)*	2008; Mussolin, De Volder, et al., 2010; Price & Ansari, 2013).	with deficient basic number-processing abilities, such as number comparison (Price & Ansari, 2013). These weaknesses are characterized by increased reaction times and error rates on tasks that involve comparing numbers, with particular difficulty when numbers are closer together (Mussolin, Mejias, & Noel, 2010).
Memorization of arithmetic facts	A left-hemisphere network that includes the precentral gryus, inferior parietal cortex, and intraparietal sulcus is often implicated in math fact retrieval (Dehaene, 1992; Dehaene & Cohen, 1997; Dehaene et al., 1999). Further, some researchers believe that rote math facts are retrieved from verbal memory, thereby requiring activation of the angular gyrus and other regions associated with linguistic processes (Dehaene, 1992; Dehaene & Cohen, 1995; Dehaene et al., 1999).	**Long-term retrieval:** Weak or impaired long-term retrieval of facts results in increased error rates in recall (Geary, 1993; Mazzocco et al., 2008). Because fact retrieval mechanisms fail to develop adequately, fluency is impaired, and those with dyscalculia continue to utilize procedural strategies rather than memory-based strategies (Geary et al., 1992, 2000; Hanich et al., 2001; Jordan & Hanich, 2003; Landerl et al., 2004).
Accurate or fluent calculation	Regions of the left fronto-parietal cortex, including the intraparietal sulcus, angular gyrus, and supramarginal gyrus, have been consistently associated with math calculation (Ansari, 2008; De Smedt et al., 2011; Dehaene et al., 2004). However, there is evidence to suggest that math fluency, while related to other skills, may be genetically distinct and may reflect variance above and beyond untimed calculation abilities (Hart et al., 2010; Petrill et al., 2012). The dorsolateral prefrontal cortex has also been found to show increased activation during calculation, implying that executive functioning and working memory may be playing roles in the process (Davis et al., 2009).	**Long-term retrieval:** See above. **Rapid naming:** The rate of access to information in long-term storage is believed to affect calculation fluency (D'Amico & Passolunghi, 2009). Some studies have found that math disorders are associated with deficits in rate of access of numerical information alone (e.g., D'Amico & Guarnera, 2005), while others have demonstrated that rate of access to both numerical and non-numerical information is impaired (e.g., Temple & Sherwood, 2002). **Processing speed:** There is a body of evidence to support the contribution of processing speed to math calculation fluency; however, the relationship remains unclear, as processing speed is often highly related to working memory and general intelligence (Berg, 2008; Bull & Johnston, 1997; Geary, 2011; Mazzocco & Rasanen, 2013; Willcutt et al., 2013).
Accurate math reasoning	As mentioned above, the intraparietal sulcus is often identified as a neural correlate of math disorders. However, it is likely that an entire network of brain regions is implicated, as the intraparietal sulcus plays a role in a variety of cognitive processes involved in math achievement (Szücs & Goswami, 2013). It has been suggested that the parietal network is involved in manipulating numerical quantities (Lemer et al., 2003). Some studies have also found that individuals with dyscalculia have structural abnormalities in the parietal cortex (Rotzer et al., 2008; Rykhlevskaia et al., 2009).	**Working memory:** Because mathematical reasoning relies on concurrently retaining multiple pieces of information while performing one or more procedures or mental operations, working memory is often implicated. Those with math difficulties tend to struggle with holding information in working memory, updating or revising the information, and tracking or monitoring the process, resulting in difficulties in sequencing, increased errors in counting, and other procedural errors (Geary, 2003; Lukowski et al., 2014; Pelegrina et al., 2014; Peng & Fuchs,

(continued)

TABLE 37.1. *(continued)*

Subskill	Etiology	Associated impairments/cognitive correlates
Accurate math reasoning *(continued)*	Prevalence of math disabilities is about 10 times higher in those with family members who have/had math disabilities (Shalev et al., 2001). Twin studies suggest a moderate genetic influence, with some studies finding additive genetic influences shared between math calculation and problem solving and several working memory components (Kovas et al., 2007; Lukowski et al., 2014). Environmental factors, including motivation, emotional functioning (e.g., math anxiety), and suboptimal or inadequate teaching, may also contribute to math difficulties (Szücs & Goswami, 2013; Vukovic et al., 2013). Furthermore, math achievement in particular may be associated with cultural or gender-based attitudes that may be transmitted in the family environment (e.g., Chiu & Klassen, 2010; Gunderson et al., 2011).	2016; Raghubar et al., 2010; Swanson & Jerman, 2006; Willcutt et al., 2013). **Visual–spatial ability:** Visual–spatial skills, such as visual perception, spatial reasoning, and mental rotation, have been found to influence math performance (Gunderson et al., 2012). Weaknesses in these may present as difficulties with representing numbers and aligning numerals, and problems in areas such as geometry or fractions (Geary, 2004; Swanson & Jerman, 2006). **Attention and executive functioning:** Math difficulties often reflect weaknesses in executive functioning skills, such as set shifting and cognitive inhibition (D'Amico & Passolunghi, 2009; van der Sluis et al., 2004; Willcutt et al., 2013). Poor attentional control (i.e., difficulty ignoring irrelevant information and focusing on goal-relevant information) is often observed as well (Geary, 2013).
Spelling accuracy	SLD with impairment in written expression Functional neuroimaging studies have provided substantial evidence for the role of the ventral temporal inferior frontal gyrus and the posterior inferior frontal gyrus in spelling (Rapp et al., 2015; van Hoorn et al., 2013). Other areas that have been identified include the left ventral cortex, bilateral lingual gyrus, and bilateral fusiform gyrus (Planton et al., 2013; Purcell et al., 2014; Richards et al., 2005, 2006). However, many of these regions have also been associated with reading and are not distinct to spelling/writing disorders. There is evidence that links spelling to a region of chromosome 15 (Schulte-Korne, 2001), although this locus has also been reported in dyslexia (Grigorenko, 2005).	**Phonological processing:** Phonological awareness is a significant predictor of spelling achievement (Caravolas et al., 2001; Cornwall, 1992; Holm et al., 2008; Skeide et al., 2015; Yeong et al., 2014). Weaknesses in this area may be manifested as poor segmentation of words into phonemes, poor sequencing of sounds, and omission or addition of sounds (Berninger, 1999). **Orthographic processing/orthographic coding:** Effective spelling involves storing and retrieving commonly occurring letter patterns in visual and motor memory; these skills are often impaired in poor spellers (Caravolas et al., 2001; Ehri, 2014; Yeong et al., 2014). **Motor skills:** Poor spelling is often accompanied by underlying skill deficits in areas such as fine motor control, motor planning, orthographic motor coordination, and visual–motor integration (Christensen, 2004; Daly et al., 2003; Feder & Majnemer, 2007).
Grammar and punctuation	With regard to English grammar, some researchers distinguish between the *mental lexicon* (i.e., memorized associations) and *mental grammar* (i.e., language rules and structure), and posit that each has distinct neural correlates (Pinker, 1994). There is some evidence to support this view, with data indicating that the mental lexicon involves left temporal and temporo-parietal regions, whiles the mental grammar recruits a system that includes left frontal regions (Ullman et al., 2005).	**Long-term memory:** It has been suggested that some components of long-term storage, in particular procedural and declarative memory, may be involved in grammar; however, much of this research has focused on children with language impairments (Conti-Ramsden et al., 2015; Hedenius et al., 2011).

(continued)

TABLE 37.1. *(continued)*

Subskill	Etiology	Associated impairments/cognitive correlates
Clarity of written expression	Neural correlates of writing are less understood, but some studies have suggested that the cerebellum and parietal cortex, particularly the left superior parietal lobe, may be involved (Katanoda et al., 2001; Magrassi et al., 2010). In addition, the frontal lobes have also been implicated and are considered crucial in planning, brainstorming, organizing, and goal setting (Shah et al., 2013). While there is a significant genetic component involved in the development of writing skills, this etiology is often shared with a broad variety of reading and language skills (Olson et al., 2013).	**Working memory:** A substantial body of research has highlighted the role of working memory in written expression, as text generation requires the coordination of multiple processes, such as synthesizing multiple ideas, retrieving grammar rules from long-term storage, and ongoing self-monitoring (Berninger, 1999; Bourke et al., 2013; Hooper et al., 2002; McCutchen, 1996). **Attention and executive functioning:** Various executive functions, including attention, planning, and self-monitoring, have been implicated in written expression (Altemeier et al., 2006; Graham et al., 2013; Graham & Harris, 2005; Hooper et al., 2002; Mason et al., 2011; Reiter et al., 2005; Rosenblum et al., 2009; Troia & Graham, 2002). **Language:** Levels of knowledge of syntax, morphology, semantics, and vocabulary have a significant impact on text generation ability (Dockrell et al., 2009; Fey et al., 2004; Olinghouse & Wilson, 2013). Language impairments are associated with higher rates of grammatical errors, less lexical diversity, and poorer overall content (Fey et al., 2004; Mackie & Dockrell, 2004).

skills require teaching and explicit learning. In the case of SLD, the normal pattern of acquiring academic skills is disrupted. Difficulties mastering basic skills such as reading, math, and writing can impair academic performance in other areas, such as science and social studies. In addition, the learning difficulties must persist for a period of 6 months or longer without evidence of "catching up" to same-grade peers, despite provision of extra support. A review of educational history including school reports, work samples, curriculum-based measures, and/or clinical interview can establish the persistence of learning difficulties.

Another critical component of the DSM-5 diagnosis of SLD is performance in an academic domain that is "well below average" (APA, 2013, p. 69) for the individual's age, as set forth in Criterion B. Clinical indicators are below-average scores in school or average performance with extraordinarily high levels of support; avoidance of activities requiring the impaired academic skill; and psychometric evidence from individually administered, psychometrically sound, norm-referenced, and culturally appropriate academic achievement measures. The authors of the DSM-5 description of SLD note that academic skills are distributed along a continuum; therefore, there is no natural cutoff score to differentiate those who have SLD from those who do not. Attempts to establish cutoff points are arbitrary, but the recommended 1.5 standard deviations below the mean is "needed for the greatest diagnostic certainty" (APA, 2013, p. 69). The authors provide a more "lenient threshold" of 1–2.5 standard deviations below the mean (i.e., standard scores from 85 to 63–85) when learning difficulties are supported by other sources of evidence.

According to DSM-5, the third core feature detailed in Criterion C is that the learning difficulties become observable during the early school years in most individuals. However, the authors note that some individuals may not fully manifest symptoms of the disorder until learning demands surpass the individuals' ability to compensate for "limited capacities." Thus, for example, an adult manifesting low academic achievement in college may be diagnosed with SLD for the first time.

Finally, Criterion D excludes learning difficulties attributable to other more general disorders such as intellectual disability (or intellectual developmental disorder, as it is also termed in DSM-5). DSM-5 considers "normal levels of intellectual functioning" to be an IQ estimate above 70, give or take 5 points to account for measurement error, and such cognitive functioning is considered to differentiate SLD from more general learning problems (APA, 2013, p. 73). In order to be considered to have SLD, an individual must demonstrate "unexpected academic underachievement," which is cited as the defining characteristic of SLD. An individual may be able to maintain adequate academic functioning with intense external supports, extraordinary effort, or compensatory strategies. Individuals who are intellectually gifted may also demonstrate symptoms warranting an SLD diagnosis. Criterion D further excludes general learning difficulties secondary to economic disadvantage, chronic absences, or lack of adequate education, and to neurological, motor, vision, or hearing disorders. The learning difficulties are also "specific" in that they may be restricted to one academic skill or domain. DSM-5 requires a comprehensive assessment involving professionals with expertise in SLD and psychological/cognitive assessment. Diagnosis can only be made after formal schooling has begun, but can be made for an individual at any point during the lifespan, provided that evidence of difficulties during formal schooling is available. The clinical diagnosis must be made as a result of synthesizing medical, developmental, educational, and family history; the history of the academic problem; the impact of the problem on academic, occupational, or social functioning; previous and/or current school reports; work samples; curriculum-based assessments; and previous or current scores from individualized standardized tests of academic achievement (APA, 2013).

Associated Impairments and Comorbidities

Delays in language, attention, or motor skills often, but not always, precede the manifestation of SLD. In addition, individuals with SLD often demonstrate uneven profiles of abilities and perform poorly on psychological tests of cognitive processing. According to DSM-5, "it remains unclear whether these cognitive abnormalities are the cause, correlate, or consequence of the learning difficulties" (APA, 2013, p. 70). While cognitive deficits associated with reading (i.e., dyslexia) are well known, those deficits associated with mathematics disorders or written expression disorders are less well understood. DSM-5 further notes that processing deficits observed in individuals with SLD are seen in individuals with other disorders as well, including attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD).

In fact, the diagnosis of SLD is often comorbid with other types of neurodevelopmental disorders or other mental disorders. For example, children with ADHD are three times more likely than peers without ADHD to have a learning disability, but estimates vary widely, depending on the type of disorder and the criteria used for diagnosing learning disabilities (DuPaul, Gormley, & Laracy, 2013). While diagnosis of other disorders does not preclude the additional diagnosis of SLD, differential diagnosis under those conditions may be more difficult, as other disabilities may account for learning problems. Clinical judgment is needed to determine whether academic skills deficits set forth in Criterion A may be attributable to other disorders. For more information on impairments associated with SLD, see Table 37.1.

Course and Developmental Changes

As stated earlier, symptoms of SLD typically emerge within the developmental period during formal schooling, despite the fact that language or fine motor delays may be evident earlier. The disorder is lifelong, but symptom expression may vary depending on environmental demands on deficient skills, severity of the deficiency, comorbidity, and the supports and interventions provided to the individual. Typically, problems persist into adulthood (APA, 2013).

Age-related manifestations of symptoms of SLD vary. Preschool children may have difficulty speaking clearly, learning letters or numbers, or recognizing their names. Similarly, kindergarten children may have difficulty learning and/or writing the alphabet, writing their names, and so forth. Phonetic analysis may also be challenging for kindergarten students with SLD, and they may have trouble breaking words down into syllables, as well as difficulty with rhyming. As these children progress through the early elementary school years, they may have difficulty with sound–symbol correspondence, sounding out words, recognizing irregular words, or learning math facts. Children with SLD often perform academic tasks with significant effort or inefficiency (e.g., Geary, Hamson, & Hoard, 2000).

As children with SLD enter the late elementary and early middle school years, they may omit syllables in multisyllabic words and may have difficulty remembering information learned in class. In addition, they may demonstrate difficulty with comprehension, spelling, or writing assignments (Feifer, 2011; Mather & Wendling, 2011; Santangelo & Graham, 2014). Adolescents may have mastered basic reading skills such as decoding, but their reading may remain slow and dysfluent. They may demonstrate poor understanding of written text and may have difficulty with math problem solving. Adolescents with poor reading skills may also demonstrate increasing difficulty in content areas such as science and social studies. Adolescents and adults may continue to demonstrate poor spelling and difficulty pronouncing multisyllabic words. Rereading in order to understand the text may often be necessary for older individuals with SLD. Adults may have difficulty with drawing inferences from written text or from numerical information. As a result, they may avoid tasks involving these skills; symptoms of anxiety may also accompany the expression of learning difficulties (APA, 2013).

Prognosis

The risk of SLD is increased by several environmental factors, including prematurity, low birth weight, and prenatal exposure to nicotine. More students with SLDs come from households living in poverty as compared to students from the general population. Students with SLDs are also more likely to be in foster care or homeless. Genetic and physiological factors increase the risk of SLD as well. For example, the risk of SLD in reading and mathematics is 4–8 and 5–10 times higher, respectively, in first-degree relatives of individuals with these disorders compared to individuals without these disorders. Children whose parents have dyslexia or a history of reading difficulties are more likely to experience learning difficulties, literacy problems, or SLD (Cortiella & Horowitz, 2014).

Several factors in the preschool years are predictive of later learning difficulties or SLD. For example, inattention is predictive of later difficulties in reading and math, including failure to respond to effective academic interventions (e.g., Bental & Tirosh, 2007). Also, speech or language impairments, or deficits in specific cognitive processes (such as phonological processing, working memory capacity, and naming facility), in the preschool years predict later SLD (e.g., Scarborough, 2005). Furthermore, comorbidity with other disorders, such as ADHD, is predictive of worse mental health as compared to SLD without comorbidity (e.g., Smith & Adams, 2006).

Outcomes

Outcomes for individuals with SLD depend on many variables, such as type of treatment, severity of the disorder, intensity of intervention, sustainability of treatment, nature and extent of comorbidities, effort and motivation of the individuals, and availability of resources (e.g., parental involvement, home–school collaboration, accommodations). Table 37.2 provides a summary of outcomes for students with specific learning disabilities. It is important to note that poor outcomes may be circumvented by evidence-based interventions that are systematic, intensive, and individualized.

TABLE 37.2. Summary of Academic Performance and School Outcomes for Students with Specific Learning Disabilities

- From 12 to 26% of secondary students with learning disabilities received average to above-average scores on math and reading assessments, compared to 50% of students in the general population.
- From 7 to 23% of secondary students with learning disabilities receive well below-average scores on academic performance, compared to 2% of students in the general population.
- Students with learning disabilities earn lower grades and experience higher rates of course failure in high school than students without learning disabilities.
- About one-third of students with learning disabilities have been retained in a grade at least once.
- One in every two students with a learning disability faced a school disciplinary action such as suspension or expulsion in 2011.
- Sixty-eight percent of students with learning disabilities leave high school with a regular diploma, while 19% drop out and 12% receive a certificate of completion.
- Sixty-nine percent of students with learning disabilities have failed one or more graded courses in secondary school, compared to 47% of students in the general population.
- Black and Hispanic students with disabilities experience much higher rates of school disciplinary actions, higher rates of dropout, and lower rates of graduation.

Note. From Cortiella and Horowitz (2014). Copyright © National Center for Learning Disabilities. Adapted by permission.

Interventions

The primary purpose of intervention for individuals with SLD is *remediation* of skill deficits. Many interventions, particularly for reading, have been subjected to rigorous evaluation and found to be effective, meaning that when implemented with fidelity, they lead to positive outcomes (for reviews, see Cooney, Huser, Small, & O'Connor, 2007; Feifer, 2011, 2014; Flanagan & Alfonso, 2011; Kilpatrick, 2014; Mascolo, Alfonso, & Flanagan, 2014). Information about any of these interventions can be found at the What Works Clearinghouse (*https://ies.ed.gov/ncee/wwc*).

Not surprisingly, then, evidence-based interventions (as compared to interventions and techniques without such support) are often the ones that are used first in either general or specialized instructional settings. In general, it is incumbent upon practitioners to use evidence-based interventions with students who struggle academically. It is also prudent to use comprehensive interventions that can meet students' multiple manifest academic difficulties (e.g., remedial reading programs that contain the five essential components of reading; Feifer, 2011). However, it is clear from the literature that despite their overt relevance, not all comprehensive, evidence-based interventions address the academic needs of every student effectively (e.g., Della Tofallo, 2010; Hale, Wycoff, & Fiorello, 2011).

In a tiered service delivery model, interventions are selected on the basis of universal screening data. For example, students who are at risk for reading difficulties may receive Wilson if their reading difficulties are related primarily to decoding difficulties, or Read 180 if their reading difficulties are related primarily to comprehension difficulties (e.g., Feifer, 2011). When a student does not respond as expected to evidence-based interventions, a comprehensive evaluation is often recommended to gain a better understanding of the nature of and basis for the student's learning difficulties. Our opinion, based on our knowledge of and expertise in evaluation of specific learning disabilities, is that only through a comprehensive and focused evaluation of cognitive abilities/processes and specific academic skills can the intervention process move from *selecting an intervention program* to *tailoring an intervention* to meet the needs of students who do not respond as expected to evidence-based instruction (Mascolo, Flanagan, & Alfonso, 2014).

Selecting Interventions versus Tailoring Interventions

Selecting interventions involves identifying evidence-based interventions that are most often used in standard service delivery models to address manifest academic difficulties that are revealed via progress monitoring (e.g., a particular reading program is selected by a district as a tier 2 intervention for students with reading fluency difficulties). On the other hand, a primary focus on *tailoring* interventions involves understanding a student's pattern of cognitive and academic strengths and weaknesses and how this pattern interacts with the instructional materials used by the student, as well as with classroom instructional factors, environmental factors, and other individual/situational factors that may facilitate or inhibit learning. The goals for tailoring intervention, therefore, are (1) to use information about a variety of factors that are intrinsic and extrinsic to the student to design specific interventions; and (2) to ensure that a student has appropriate access to the curriculum by minimizing or bypassing the adverse effects that cognitive and other weaknesses have on the student's learning. Tailoring interventions may include modification (e.g., instructional, curricular), accommodation, remediation, and compensation. Definitions and examples of these interventions are provided in Table 37.3. For additional information about these interventions and the evidence in support of them, see Mascolo, Alfonso, and Flanagan (2014).

Vagaries in the DSM-5 Criteria for SLD

Historically, clinicians overrelied on IQ scores when making decisions about specific learning disorders and disabilities. A large body of research has shown that the IQ–achievement discrepancy method, when used as a primary or sole criterion for diagnosis of SLD, is invalid (see Stanovich, 2005, for a review). Therefore, the developers of the DSM-5 criteria for SLD should be commended for dropping this criterion. However, it seems clear from existing research that assessment of specific cognitive abilities and processes is an invaluable component in the evaluation of suspected learning disorders and disabilities (see Hale et al., 2010). As such, DSM-5 criteria for the diagnosis of SLD are limited, and when applied in isolation they may obscure information that is necessary to understand the nature of learning problems and

TABLE 37.3. Methods of Tailoring Intervention for Students with Specific Learning Disabilities

Tailoring method	Brief description	Examples
Modification	Changes content of material to be taught or measured; typically involves changing or reducing learning or measurement expectations; may change the depth, breadth, and complexity of learning and measurement goals.	• Reducing the amount of material that a student is required to learn • Simplifying material to be learned • Requiring only literal (as opposed to critical/inferential) questions from an end-of-chapter comprehension check • Simplifying test instructions and content
Accommodation	Changes conditions under which learning occurs or is measured, but does not change or reduce learning or assessment expectations. Accommodations may include timing, flexible scheduling, presentation, setting, and response accommodations.	• Extending time on exams • Assigning a project in advance or allowing more time to complete a project • Aligning math problems vertically, as opposed to horizontally • Providing a separate room to work • Having a student dictate responses to a scribe
Remediation	Techniques or programs used to ameliorate cognitive and academic deficits. Academic interventions typically focus on developing a skill, increasing automaticity of skills, or improving the application of skills. Cognitive interventions typically focus on improving cognitive processes such as working memory capacity and phonological processing. There are many techniques, published programs, and software designed for the purpose of remediation.	• Evidence-based programs listed at What Works Clearinghouse (*https://ies.ed.gov/ncee/wwc*) • Reading programs appearing on the Florida Center for Reading Research website (*www.fcrr.org*) • Techniques and materials from the Reading Rockets website (*www.readingrockets.org*) • CogMed (Pearson) • Spotlight on Listening Comprehension (LinguiSystems)
Compensation	Procedures, techniques, and strategies that are intended to bypass or minimize the impact of a cognitive or academic deficit.	• Teaching the use of mnemonic devices • Teaching the use of organizational aids or techniques • Teaching a student to outline or use graphic organizers

thus may lead to overdiagnosis. Following are a few of the most salient vagaries in DSM-5 criteria for SLD.

First, the DSM-5 definition of SLD relies on below-average achievement, without reference to underlying cognitive processing weaknesses that are known to interfere with basic academic skill acquisition and development. Below-average achievement can be caused by a variety of factors (e.g., attention problems, depression, poor motivation, behavior problems, low IQ), and the failure to include specific cognitive processes as part of the diagnostic criteria makes it difficult to determine the basis for academic underachievement. Essentially, the current definition reflects a "low-achievement" model for defining and diagnosing SLD, with an added caveat regarding poor response to intervention. Data show that states relying on low achievement and poor response to intervention only (e.g., Iowa), as the DSM-5 criteria do, have the highest percentage of students in special education under the category of specific learning disability. For example, while the national average of students with specific learning disabilities in special education is 41.5%, over 60% of the total special education enrollments in Iowa are students with specific learning disabilities (IDEAdata.org; cited in Cortiella & Horowitz, 2014).

Second, a number of terms and phrases in the DSM-5 criteria and diagnostic features are not defined sufficiently. For example, DSM-5 criteria allow for diagnosis of SLD with average achievement scores if those scores are only attainable by "extraordinary effort." Because extraordinary effort is not defined objectively, its meaning is left up to the varied, unstandardized judgments of clinicians relying largely on subjective data. Also, it is not clear what is meant by "exceed the individual's limited capacities," despite the examples given in Criterion C, including "reading or writing lengthy complex reports" and "excessively heavy academic loads." One problem with this criterion is that individuals with *average* cognitive and academic abilities often have difficulties with these tasks and often fail to meet the demands of an excessively heavy academic load. This fact, coupled with the fact that the DSM-5 criteria allow for diagnosis of SLD in adulthood based on self-report or report by others, could potentially allow just about anyone to be diagnosed with SLD—particularly older individuals who struggle in postsecondary educational programs.

Third, and perhaps most alarming, is the definition of "normal levels of intellectual functioning." It is important to know an individual's level of intellectual functioning because SLD is "specific" and therefore not attributable to intellectual disability (intellectual developmental disorder) or global developmental delay, for example. According to the DSM-5 criteria for SLD, normal intellectual functioning is generally estimated by an IQ score of greater than about 70 ± 5 points (or standard scores of 65–75 on standardized tests having a mean of 100 and standard deviation of 15). This score range, which corresponds to the 1st–5th percentiles, can in no way be conceived of as reflecting "normal levels of intellectual functioning." In fact, this score range is associated with cognitive impairment and intellectual disability. This score range is also identical to the range used to define intellectual developmental disorder in DSM-5 (APA, 2013, p. 37).

Interestingly, the DSM-5 SLD criterion for academic skills that are "substantially and quantifiably below those expected for the individual's chronological age" (APA, 2013, p. 67) is "low achievement scores on one or more standardized tests or subtests within an academic domain (i.e., at least 1.5 standard deviations . . . below the population mean for age, which translates to a standard score of 78 or less, which is below the 7th percentile)." Moreover, standard scores in this range are "needed for the greatest diagnostic certainty" (APA, 2013, p. 69). Given the properties of the normal probability curve, this criterion is certainly in line with what can be considered substantially below-expected levels of performance. Therefore, it is neither logical nor accurate to interpret scores that are lower than 78 (i.e., scores of 70 ± 5) as representing "normal" intellectual functioning.

Because the same standard score range is used to define normal intellectual functioning for SLD and impaired intellectual functioning for intellectual developmental disorder, the defining characteristic of SLD—*unexpected underachievement*—is not evident in the DSM-5 criteria, despite mention of this hallmark characteristic in the "Diagnostic Features" section (APA, 2013, p. 69). For example, individuals with IQs in the 70s and commensurate academic achievement scores meet DSM-5 criteria for SLD. However, this type of profile does not reflect unexpected underachievement, but rather *expected underachievement,* which historically has never been a defining characteristic of SLD. Expected underachievement is more consistent with general learning difficulties that are perhaps due to suspected but undiagnosed conditions, such as fetal alcohol spectrum disorders (e.g., Kodituwakku, 2007). Because the DSM-5 criterion for normal intellectual functioning is inaccurate, the definition of SLD is overly sensitive but not sufficiently specific, which will undoubtedly lead to a substantial increase in the number of individuals diagnosed with SLD.

It is beyond the scope of this chapter to discuss all of the potential issues related to diagnosis of SLD brought on by the revised SLD criteria included in DSM-5. It is our contention that failure to include criteria related to cognitive processes obscures the very nature of SLD. There is an abundance of evidence in the cognitive and neuropsychological literature demonstrating how cognition affects achievement (see Fiorello, Hale, & Wycoff, 2012; Fletcher-Janzen & Reynolds, 2008). Most academic learning tasks require general cognitive processes such as attention, language, and memory, and specific processes contribute uniquely to specific academic skills (see Geary, 2011; Shaywitz et al., 2004). Even the authors of the DSM-5 definition of SLD found in their review of predictive validators "quite strong support for the inclusion of cognitive processing deficits in the diagnostic criteria for SLD," especially in reading (Tannock, 2013, p. 18). Academic tasks require numerous cognitive processes. Administering specific, narrowly defined cognitive tasks helps to systematically isolate the underlying learning problems and to differentiate SLD from other conditions that affect learning and achievement (see Decker et al., 2013, for a discussion). The purpose of administering cognitive tests is not necessarily to obtain an IQ score, but to assist in understanding learning problems (Flanagan et al., 2013) and to develop effective interventions (Fuchs et al., 2011).

In short, we believe that approaches to SLD diagnosis that do not consider cognitive processing strengths and weaknesses are not supported by research (see Reynolds & Shaywitz, 2009). We support alternative research-based approaches to SLD diagnosis "because they emanate from the marriage of a collective body of knowledge that has been acquired through research in the fields of neuroscience, pedagogy, assessment, and intervention" (Della Tofallo, 2010, pp. 180–181). These approaches deemphasize IQ in favor of theory-based flexible batteries, which include measures of cognitive abilities and processes that are predictive of specific academic skills and that yield information relevant for instructional planning (see Flanagan & Alfonso, 2011, for a review).

CONCLUSION

This chapter has provided a brief history of the definition of specific learning disability, with emphasis on the criteria included in the five editions of the DSM. The DSM-5 criteria for SLD have been reviewed in detail. Also reviewed is the etiology of SLD, as well as associated impairments and comorbidities, course and developmental changes, prognosis, outcomes, and treatment. This chapter has also highlighted some salient, unexpected changes in the DSM-5 criteria—changes that appear to obscure the true nature of SLD as a disorder in one or more basic psychological processes.

REFERENCES

Altemeier, L., Jones, J., Abbott, R. D., & Berninger V. (2006). Executive functions in becoming writing readers and reading writers: Note-taking and report writing in third and fifth graders. *Developmental Neuropsychology, 29,* 161–173.

American Psychiatric Association (APA). (1952). *Diagnostic and statistical manual of mental disorders.* Washington, DC: Author.

American Psychiatric Association (APA). (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association (APA). (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (APA). (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association (APA). (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience, 9,* 278–291.

Ansari, D. (2010). Neurocognitive approaches to developmental disorders of numerical and mathematical cognition: The perils of neglecting the role of development. *Learning and Individual Differences, 20,* 123–129.

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during non-symbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 18,* 1820–1828.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport, 16,* 1769–1773.

Bateman, B. (1965). An educational view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders* (Vol. 1, pp. 219–239). Seattle, WA: Special Child.

Benjamin, C. F. A., & Gaab, N. (2012). What's the story?: The tale of reading fluency told at speed. *Human Brain Mapping, 33,* 2572–2585.

Bental, B., & Tirosh, E. (2007). The relationship between attention, executive functions and reading domain abilities in attention deficit hyperactivity disorder and reading disorder: A comparative study. *Journal of Child Psychology and Psychiatry, 48*(5), 455–463.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology, 99,* 288–308.

Berl, M. M., Duke, E. S., Mayo, J., Rosenberger, L. R., Moore, E. N., VanMeter, J., et al. (2010). Functional anatomy of listening and reading comprehension during development. *Brain and Language, 114,* 115–125.

Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: Automatic and constructive processes. *Learning Disability Quarterly, 22,* 99–112.

Berninger, V. W. (2011). Evidence-based differential diagnosis and treatment of reading disabilities with and without comorbidities in oral language, writing, and math: Prevention, problem-solving consultation, and specialized instruction. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 203–232). Hoboken, NJ: Wiley.

Berninger, V., Abbott, R., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of Psycholinguistic Research, 39*(2), 141–163.

Betjemann, R. S., Keenan, J. M., Olson, R. K., & DeFries, J. (2011). Choice of reading comprehension test influences the outcomes of genetic analyses. *Scientific Studies of Reading, 15,* 363–382.

Betjemann, R. S., Willcutt, E. G., Olson, R. K., Keenan, J. M., DeFries, J., & Wadsworth, S. J. (2008). Word reading and reading comprehension: Stability, overlap, and independence. *Reading and Writing, 29,* 539–558.

Bourke, L., Davies, S. J., Sumner, E., & Green, C. (2013). Individual differences in the development of early writing skills: Testing the unique contribution of visuo-spatial working memory. *Reading and Writing, 27,* 315–335.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology, 65,* 1–24.

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., et al. (2005). Longitudinal twin study of early literacy development: preschool and kindergarten phases. *Scientific Studies of Reading, 9,* 219–235.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45,* 751–774.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19,* 246–251.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331–361.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20,* 2–17.

Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology, 74,* 551–564.

Chochon, F., Cohen, L., van de Moortele, P. F., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience, 11*(6), 617–630.

Colker, R., Shaywitz, S., Shaywitz, B., & Simon, J. (2013). Comments on proposed DSM-5 criteria from a legal and medical/scientific perspective [Online comment]. Retrieved from *http://dylexia.yale.edu/CommentsDSM5ColkerShaywitzSimon.pdf.*

Compton, D. L., Davis, C. J., DeFries, J. C., Gayan, J., & Olson, R. K. (2001). Genetic and environmental influences on reading and RAN: An overview of results from the Colorado twin study. In M. Wolf (Ed.), *Time, fluency, and developmental dyslexia* (pp. 277–303). Baltimore: York Press.

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology,* 6, 1090.

Cooney, S., Huser, M., Small, S., & O'Connor, C. (2007, October). Evidence-based programs: An overview. Retrieved from *www.uwex.edu/ces/flp/families/whatworks_06.pdf.*

Cornwall, A. (1992). The relationship of phonological awareness, rapid naming, and verbal memory to severe reading and spelling disability. *Journal of Learning Disabilities, 25,* 532–538.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends, and emerging issues.* New York: National Center for Learning Disabilities.

Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia, 59,* 34–54.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277–299.

Daly, C. J., Kelley, G. T., & Krauss, A. (2003). Relationship between visual-motor integration and handwriting skills of children in kindergarten: A modified replication study. *American Journal of Occupational Therapy, 57*(4), 459–462.

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences, 15,* 189–202.

D'Amico, A., & Passolunghi, M. C. (2009). Naming speed and effortful and automatic inhibition in children with arithmetic learning disabilities. *Learning and Individual Differences, 19,* 170–180.

Davis, N., Cannistraci, C. J., Rogers, B. P., Catenby, J. C., Fuchs, L. S., Anderson, A. W., et al. (2009). The neural correlates of calculation ability in children: An fMRI study. *Magnetic Resonance Imaging, 27,* 1187–1197.

De Smedt, B., Holloway, I., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage, 57,* 771–781.

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools, 50*(3), 300–313.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1,* 83–120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double disassociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33,* 219–250.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15,* 254–262.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14,* 218–224.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284,* 970–974.

Della Toffalo, D. A. (2010). Linking school neuropsychology with response-to-intervention models. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 159–184). Hoboken, NJ: Wiley.

Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The impact of specific language impairments on adolescents' written text. *Exceptional Children, 75,* 427–446.

DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications of DSM-5 for assessment and treatment. *Journal of Learning Disabilities, 46*(1), 43–51.

Eckert, M. A., Leonard, C. M., Richards, T. L., Aylward, E. H., Thomson, J., & Berninger, V. W. (2003). Anatomical correlates of dyslexia: Frontal and cerebellar findings. *Brain, 126,* 482–494.

Education for All Handicapped Children Act of 1975, Pub. L. No. 94-142, 20 U.S.C. 1401 et seq. (1975).

Education of the Handicapped Act of 1970, Pub. L. No. 91-230, 20 U.S.C. 1401 et seq. (1970).

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18,* 5–21.

Feder, K. P., & Majnemer, A. (2007). Handwriting, development, competency, and intervention. *Developmental Medicine and Child Neurology, 49,* 312–317.

Feifer, S. (2011). How SLD manifests in reading. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 21–42). Hoboken, NJ: Wiley.

Feifer, S. G. (2014). Selecting and tailoring interventions for students with reading difficulties. In D. P. Flanagan & V. C. Alfonso (Eds.), *Planning, selecting, and tailoring interventions for unique learners* (pp. 151–177). Hoboken, NJ: Wiley.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of numbers. *Trends in Cognitive Sciences, 8,* 307–314.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29,* 581–593.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J., & Zhang, X. Y. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 1301–1318.

Fiorello, C. A., Hale, J. B., & Wycoff, K. L. (2012). Cognitive hypothesis testing: Linking test results to the real world. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 484–496). New York: Guilford Press.

Flanagan, D. P., & Alfonso, V. C. (2011). *Essentials of specific learning disability identification.* Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities.* Boston: Allyn & Bacon.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guilford Press.

Fletcher, J. M., Simos, P. G., Papanicolaou, A. C., & Denton, C. (2004). Neuroimaging in reading research. In N. Duke & M. Mallette (Eds.), *Literacy research methods* (pp. 252–286). New York: Guilford Press.

Fletcher-Janzen, E., & Reynolds, C. R. (Eds.). (2008). *Neuropsychological perspectives on learning disabilities in the era of RTI: Recommendations for diagnosis and intervention.* Hoboken: NJ: Wiley.

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011). Two-stage screening for math word-problem difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities, 44*(4), 372–380.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362.

Geary, D. C. (2003). Learning disabilities in arithmetic. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 199–212). New York: Guilford Press.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A five year longitudinal study. *Developmental Psychology, 47,* 1539–1552.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22,* 23–27.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment, 27,* 265–279.

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology, 54,* 372–391.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77,* 236–263.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning difficulties and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104,* 206–223.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299.

Georgiou, G. K., Parrila R., & Liao, C. H. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing, 21,* 885–903.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology, 54,* 91–114.

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26,* 1–15.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers. *Journal of Special Education, 39*(1), 19–33.

Grigorenko, E. L. (2001). Developmental dyslexia: An update on genes, brains, and environments. *Journal of Child Psychology and Psychiatry, 42,* 91–125.

Grigorenko, E. L. (2005). A conservative meta-analysis of linkage and linkage-association studies of developmental dyslexia. *Scientific Studies of Reading, 9,* 285–316.

Groce, N., Challenger, E., Berman-Bieler, R., Farkas, A., Yilmaz, N., Schultink, W., et al. (2014). Malnutrition and disability: Unexplored opportunities for collaboration. *Paediatrics and International Child Health, 34*(4), 308–314.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48,* 1229–1241.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2011). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles,* 66, 153–166.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455,* 665–668.

Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disabilities Quarterly, 33,* 223–236.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hale, J. B., Wycoff, K. L., & Fiorello, C. A. (2011). RTI and cognitive hypothesis testing for identification and intervention of specific learning disabilities: The best of both worlds. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 173–202). Hoboken, NJ: Wiley.

Hallahan, D. P., & Mercer, C. D. (2001). *Learning disabilities: Historical perspectives.* Paper written for the Office of Special Education Programs (OSEP), U.S. Department of Education, and presented at the OSEP's LD Summit Conference, Washington, DC.

Hallahan, D. P., & Mock, D. R. (2003). A brief history of the field of learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 16–29). New York: Guilford Press.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology,* 93, 615–626.

Harlaar, N., Cutting, L., Deater-Deckard, K., DeThorne, L. S., Justice, L. M., Schatschneider, C., et al. (2010). Predicting individual differences in reading comprehension: A twin study. *Annals of Dyslexia,* 60, 265–288.

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development, 78,* 116–131.

Hart, S. A., Petrill, S. A., & Thompson, L. A. (2010). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences, 20,* 63.

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Verissimo, J., Dye, C. D., et al. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research on Developmental Disabilities, 32,* 2362–2375.

Herr, C. M., & Bateman, B. D. (2003). Learning disabilities and the law. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 57–75). New York: Guilford Press.

Hinshelwood, J. (1902). *Congenital word-blindness with reports of two cases.* London: John Bale, Sons & Danielsson.

Holm, A., Farrier, F., & Dodd, B. (2008). Phonological awareness, reading accuracy, and spelling ability of children with inconsistent phonological disorder. *International Journal of Language and Communication, 43,* 300–322.

Hooper, S. R., Swarz, C. W., Wakely, M. B., de Kruif, R. E., & Montgomery, J. W. (2002). Executive functions in elementary school children with and without problems in written expression. *Journal of Learning Disabilities, 35,* 57–68.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1400 et seq. (2004).

Jones, M. W., Ashby, J., & Branigan, H. P. (2013). Dyslexia and fluency: Parafoveal and foveal influences on rapid automatized naming. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 554–567.

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research and Practice, 18,* 213–221.

Katanoda, K., Yoshikawa, K., & Sugishita, M. (2001). A functional MRI study on the neural substrates for writing. *Human Brain Mapping, 13,* 34–42.

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L. B., et al. (2006). Neural correlates of the number–size interference task in children. *NeuroReport, 17,* 587–591.

Kavale, K. A., Kauffman, J. M., Bachmeier, R. J., & LeFever, G. B. (2008). Response-to-intervention: Separating the rhetoric of self-congratulation from the reality of specific learning disability identification. *Learning Disability Quarterly, 31,* 135–150.

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29,* 79–91.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101,* 765–778.

Kesler, S. R., Sheau, K., Koovakkattu, D., & Reiss, A. L. (2011). Changes in frontal–parietal activation and math skills performance following adaptive number sense training: Preliminary results from a pilot study. *Neuropsychological Rehabilitation, 21*(4), 433–454.

Kilpatrick, D. (2014). Tailoring interventions in reading based on emerging research on the development of word recognition skills. In D. P. Flanagan & V. C. Alfonso (Eds.), *Planning, selecting, and tailoring interventions for unique learners* (pp. 123–150). Hoboken, NJ: Wiley.

Kirk, S. A. (1962). *Educating exceptional children.* Boston: Houghton Mifflin.

Kirk, S. A., McCarthy, J., & Kirk, W. D. (1967). *The Illinois Test of Psycholinguistic Abilities.* Urbana: University of Illinois Press.

Kodituwakku, P. W. (2007). Defining the behavioral phenotype in children with fetal alcohol spectrum disorders: A review. *Neuroscience and Biobehavioral Reviews, 31*(2), 192–201.

Kovas, Y., Haworth, C. M. A., Petrill, S. A., & Plomin, R. (2007). Mathmatical ability of 10-year-old boys and girls: Genetic and environmental etiology of typical and low performance. *Journal of Learning of Disabilities, 40,* 554–567.

Krakowiak, P., Walker, C. K., Bremer, A. A., Baker, A. S., Ozonoff, S., Hansen, R. L., et al. (2012). Maternal metabolic conditions and risk for autism and other neurodevelopmental disorders. *Pediatrics, 129*(5), e1121–e1128.

Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., & Martin, E. (2008). Development of neural networks for exact and approximate calculation: A FMRI study. *Developmental Neuropsychology, 33,* 447–473.

Landerl, K., Bevan, K., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition, 93,* 99–125.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia, 41,* 1942–1958.

Locascio, G., Mahone, E. M., Eason, S. H., & Cutting, L. E. (2010). Executive dysfunction among children with reading comprehension deficits. *Journal of Learning Disabilities, 43,* 441–454.

Loriaux, D. L. (2010). Historical note: Adolf Kussmaul (1822–1902). *Endocrinologist, 20*(3), 95.

Lukowski, S. L., Soden, B., Hart, S. A., Thompson, L. A., Kovas, Y., & Petrill, S. A. (2014). Etiological distinction of working memory components in relation to mathematics. *Intelligence, 47,* 54–62.

Mackie, C., & Dockrell, J. E. (2004). The writing skills of children with SLI. *Journal of Speech, Language, and Hearing Research, 47,* 1469–1483.

Magrassi, L., Bongetta, D., Bianchini, S., Berardesca, M., & Arienta, C. (2010). Central and peripheral components of writing critically depend on a defined area of the dominant superior parietal gyrus. *Brain Research, 1346,* 145–154.

Mascolo, J. T., Alfonso, V. C., & Flanagan, D. P. (2014). *Essentials of planning, selecting, and tailoring interventions for unique learners.* Hoboken, NJ: Wiley.

Mascolo, J. T., Flanagan, D. P., & Alfonso, V. C. (2014). A systematic method of analyzing assessment results for tailoring interventions (SMAARTI). In J. T. Mascolo, V. C. Alfonso, & D. P. Flanagan, *Essentials of planning, selecting, and tailoring interventions for unique learners* (pp. 3–55). Hoboken, NJ: Wiley.

Mason, L. H., Harris, K. R., & Graham, S. (2011). Self-regulated strategy development for students with writing difficulties. *Theory into Practice, 50,* 20–27.

Mather, N., & Wendling, B. J. (2011). How SLD manifests in writing. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification.* Hoboken, NJ: Wiley.

Mazzocco, M. M., Devlin, K. T., & McKenney, S. J. (2008). Is it a fact?: Timed arithmetic performance of children with mathematical learning disabilities (MLD) varies as a function of how MLD is defined. *Developmental Neuropsychology, 33,* 318–344.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82,* 1224–1237.

Mazzocco, M. M., & Rasanen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends in Neuroscience and Education, 2,* 65–73.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8,* 299–325.

McDonough, E. M., & Flanagan, D. P. (2016). Use of the Woodcock–Johnson IV in the identification of specific learning disabilities in school-age children. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 212–252). Burlington, MA: Elsevier.

McDonough, E. M., Flanagan, D. P., Sy, M., & Alfonso, V. C. (2017). Specific learning disorder. In S. Goldstein & M. DeVries (Eds.), *Handbook of DSM-5 disorders in children and adolescents* (pp. 77–104). Cham, Switzerland: Springer International.

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 132,* 322–352.

Moll, K., Göbel, S. M., & Snowling, M. J. (2015). Basic number processing in children with specific learning disorders: Comorbidity of reading and mathematics disorders. *Child Neuropsychology, 21,* 399–417.

Monroe, M. (1932). *Children who cannot read: The analysis of reading disabilities and the use of diagnostic tests in the instruction of retarded readers.* Chicago: University of Chicago Press.

Mussolin, C., De Volder, A., Grandin, A., Schlogel, X., Nassogne, M. C., & Noel, M. P. (2010). Neural correlates of symbolic number comparison in developmental dyscalculia. *Journal of Cognitive Neuroscience, 22,* 860–874.

Mussolin, C., Mejias, S., & Noel, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition, 115,* 10–25.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427–452.

O'Brien, B. A., Wolf, M., Miller, L. T., Lovett, M. W., & Morris, R. (2011). Orthographic processing efficiency in developmental dyslexia: An investigation of age and treatment factors at the sublexical level. *Annals of Dyslexia, 61,* 111–135.

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26,* 45–65.

Olson, R. K., Hulslander, J., Christopher, M., Keenan, J. M., Wadsworth, S. J., Willcutt, E. G., et al. (2013). Genetic and environmental influences on writing and their relations to language and reading. *Annals of Dyslexia, 63,* 25–43.

Ortiz, S. O., Flanagan, D. P., & Alfonso, V. C. (2015). *Cross-battery assessment software system (X-BASS).* Hoboken, NJ: Wiley.

Orton, S. T. (1925). Word-blindness in school children. *Archives of Neurology and Psychiatry, 14,* 581–615.

Paracchini, S., Steer, C. D., Buckingham, L. L., Morris, A. P., Ring, S., Scerri, T., et al. (2007). Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *American Journal of Psychiatry, 165,* 1576–1584.

Pelegrina, S., Capodieci, A., Carretti, B., & Cornoldi, C. (2014). Magnitude representation and working memory updating in children with arithmetic and reading comprehension disabilities. *Journal of Learning Disabilities, 48,* 658–668.

Peng, P., & Fuchs, D. (2016). A meta-analysis of working memory deficits in children with learning difficulties: Is there a difference between verbal domain and numerical domain? *Journal of Learning Disabilities, 49*(1), 3–20.

Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 453–472). Oxford, UK: Blackwell.

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., DeThorne, L. S., & Schatschneider, C. (2006). Genetic and shared environmental effects of serial naming and phonological awareness on early reading outcomes. *Journal of Educational Psychology, 98,* 112–121.

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatscheider, C., & DeThorne, L. S. (2007). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading and Writing, 20,* 127–146.

Petrill, S. A., Logan, J., Hart, S. A., Vincent, P., Thompson, L. A., Kovas, Y., et al. (2012). Math fluency is etiologically distinct from untimed math performance, decoding fluency, and untimed reading performance: Evidence from a twin study. *Journal of Learning Disabilities, 45,* 371–381.

Pinker, S. (1994). *The language instinct.* New York: Morrow.

Planton, S., Jucla, M., Roux, F.-E., & Démonet, J.-F. (2013). The "handwriting brain": A meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex, 49,* 2772–2787.

Price, G. R., & Ansari, D. (2013). Dyscalculia: Characteristics, causes, and treatments. *Numeracy, 6*(1), Article 2.

Purcell, J., Shea, J., & Rapp, B. (2014). Beyond the visual word form area: The orthography–semantics interface in spelling and reading. *Cognitive Neuropsychology, 31,* 482–510.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20,* 110–122.

Rapp, B., Purcell., J., Hillis, A. E., Capasso, R., & Miceli, G. (2015). Neural bases of orthographic long-term memory and working memory in dysgraphia. *Brain, 138,* 1–17.

Raskind, W. H., Igo, R. P., Chapman, N. H., Berninger, V. W., Thomson, J. B., Matsushita, M., et al. (2005). A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2p that contributes to phonological decoding efficiency. *Molecular Psychiatry, 10,* 699–711.

Reiter, A., Tucha, O., & Lange, K. W. (2005*). Executive functions in children with dyslexia. Dyslexia, 11,* 116–131.

Resnick, M. B., Gueorguieva, R. V., Carter, R. L., Ariet, M., Sun, Y., Roth, J., et al. (1999). The impact of low birth weight, perinatal conditions, and sociodemographic factors on educational outcome in kindergarten. *Pediatrics, 104*(6), 1–10.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Prevention and remediation, perhaps, diagnosis, no. *Child Development Perspectives, 3,* 44–47.

Richards, T. L., Aylward, E. H., Berninger, V. W., Field, K. M., Grimme, A. C., Richards, A. L., et al. (2006). Individual fMRI activation in orthographic mapping and morpheme mapping after orthographic or morphological spelling treatment in child dyslexics. *Journal of Neurolinguistics, 19,* 56–86.

Richards, T. L., Berninger, V. W., Nagy, W., Parsons, A., Field. K., & Richards, A. L. (2005). Brain activation during language task contrasts in children with and without dyslexia: Inferring mapping processes and assessing response to spelling instruction. *Educational and Child Psychology, 22,* 62–80.

Richlan, F. (2012). Developmental dyslexia: Dysfunction of a left hemisphere reading network. *Frontiers in Human Neuroscience, 6,* 120.

Richlan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping, 30,* 3299–3308.

Rosenblum, S., Aloni, T., & Josman, N. (2009). Relationships between handwriting performance and organizational abilities among children with and without dysgraphia: A preliminary study. *Research in Developmental Disabilities, 31,* 502–509.

Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., & Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *NeuroImage, 39,* 417–422.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17,* 525–536.

Rykhlevskaia, E., Uddin, L. Q., Kondos, L., & Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: Combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience, 3,* 51.

Santangelo, T., & Graham, S. (2014). Selecting and tailoring interventions for students with written expression difficulties. In J. T. Mascolo, V. C. Alfonso, & D. P. Flanagan (Eds.), *Essentials of planning, selecting, and tailoring interventions for unique learners* (pp. 204–230). Hoboken, NJ: Wiley.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothese with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 3–24). Mahwah, NJ: Erlbaum.

Scerri, T. S., Morris, A. P., Buckingham, L., Newbury, D. F., Miller, L., Monaco, A. P., et al. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biological Psychiatry, 70,* 237–245.

Scerri, T. S., & Schulte-Korne, G. (2010). Genetics of developmental dyslexia. *European Journal of Child Adolescent Psychiatry, 19,* 179–197.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265–282.

Schulte-Korne, G. (2001). Genetics of reading and spelling disorder. *Journal of Child Psychology and Psychiatry, 42,* 985–997.

Sesma, H. W., Mahone, E. M., Levine, T. M., Eason, S. H., & Cutting, L. E. (2008). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15,* 1–15.

Shah, C., Erhard, K., Ortheil, H.-J., Kaza, E., Kessler, C., & Lotze, M. (2013). Neural correlates of creative writing: An fMRI study. *Human Brain Mapping, 34,* 1088–1101.

Shalev, R. S., Manor, O., Kerem, B., Ayali, M., Badichi, N., Friedlander, Y., et al. (2001). Developmental dyscalculia is familial learning disability. *Journal of Learning Disabilities, 34,* 59–65.

Share, D. L., & Leikin, M. (2004). Language impairment at school entry and later reading disability: Connections at lexical versus supralexical levels of reading. *Scientific Studies of Reading, 8,* 87–110.

Shaywitz, B. A., Shaywitz, S. E., Blachman, B., Pugh, K. R., Fulbright, R. K., Skudlarski, P., et al. (2004). Development of left occipitotemporal systems for skills reading in children after a phonologically based intervention. *Biological Psychiatry, 55,* 926–933.

Shaywitz, S. E., Pugh, K. R., Jenner, A. R., Fulbright, R. K., Fletcher, J. M., Gore, J. C. (2000). The neurobiology of reading and reading disability (dyslexia). In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 229–249). Mahwah, NJ: Erlbaum.

Shaywitz, S. E., & Shaywitz, B. A. (2004). Reading disability and the brain. *Educational Leadership, 61*(6), 6–11.

Silani, L. S., Frith, U., Démonet, J. R., Fazio, F., Perani, D., Price, C., et al. (2005). Brain abnormalities underlying altered activation in dyslexia: A voxel-based morphometry study. *Brain, 128,* 2453–2461.

Skeide, M. A., Kirsten, H., Kraft, I., Schaadt, G., Muller, B., Neef, N., et al. (2015). Genetic dyslexia risk variant is related to neural connectivity patterns underlying phonological awareness in children. *NeuroImage, 118,* 414–421.

Smith, T., & Adams, G. (2006). The effect of comorbid AD/HD and learning disabilities on parent-reported behavioral and academic outcomes of children. *Learning Disability Quarterly, 29*(2), 101–112.

Sotelo-Dynega, M., Flanagan, D. P., & Alfonso, V. C. (2011). Overview of specific learning disabilities. In D. P. Flanagan & V. C. Alfonso, *Essentials of specific learning disability identification* (pp. 1–19). Hoboken, NJ: Wiley.

Stanovich, K. E. (2005). The future of a mistake: Will discrepancy measurement continue to make the learning disabilities field a pseudoscience? *Learning Disabilities Quarterly, 28,* 103–106.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research, 76,* 249–274.

Szücs, D., & Goswami, U. (2013). Developmental dyscalculia: Fresh perspectives. *Trends in Neuroscience and Education, 2,* 33–37.

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing in Chinese. *Proceedings of the National Academy of Sciences USA, 102,* 8781–8785.

Tannock, R. (2013). Specific learning disabilities in DSM-5: Are the changes for better or worse? *International Journal for Research in Learning Disabilities, 1*(2), 2–30.

Temple, C. M., & Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *Quarterly Journal of Experimental Psychology, 55A,* 733–752.

Torgesen, K. L. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15,* 55–64.

Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine: Changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities, 35,* 290–305.

Tylenda, B., Hooper, S. R., & Barrett, R. P. (1987). Developmental learning disorders. In C. L. Frame & J. L. Matson (Eds.), *Handbook of assessment in child psychopathology: Applied issues in differential diagnosis and treatment* (pp. 187–217). New York: Plenum Press.

Ullman, M. T., Pancheva, P., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language, 93,* 185–238.

U.S. Department of Education. (2006). Part II. 34 C.F.R. Parts 300 and 301: Assistance to states for the education of children with disabilities and preschool grants for children with disabilities: Final rule. *Federal Register, 71*(156), 46539–46845.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology, 87,* 239–266.

van Hoorn, J. F., Maathuis, C. G., & Hadders-Algra, M. (2013). Neural correlates of paediatric dysgraphia. *Developmental Medicine and Child Neurology, 55,* 65–68.

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology, 38,* 1–10.

Vukovic, R. K., & Siegel, L. S. (2006). The double-deficit hypothesis: A comprehensive analysis of the evidence. *Journal of Learning Disabilities, 39,* 25–47.

Wadsworth, S. J., Olson, R. K., Pennington B. F., & DeFries, J. C. (2000). Differential genetic etiology of reading disability as a function of IQ. *Journal of Learning Disabilities, 33,* 192–199.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Hecht, S. A. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468–479.

Willcutt, E. G., Pennington, B. F., Smith, S. D., Cardon, L. R., Gayán, J., Knopik, V. S., et al. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for ADHD. *American Journal of Medical Genetics (Neuropsychiatric Genetics), 114*(3), 260–268.

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., et al. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disabilities, 46,* 500–516.

Yeong, S. H. M., Fletcher, J., & Bayliss, D. (2014). Importance of phonological and orthographic skills for English reading and spelling: A comparison of English monolingual and Mandarin–English bilingual children. *Journal of Educational Psychology, 106,* 1107–1121.

CHAPTER 38

Should Our Future Include the Integration of Evidence-Based Neuropsychological Services into School Settings?

Elizabeth M. Power
Rik Carl D'Amato

CLINICAL NEUROPSYCHOLOGY IN THE SCHOOLS

The purpose of this chapter is to introduce the uninitiated reader to a neuropsychologically driven, brain-based approach to serving comprehensive medical and neurological disorders in the public schools. This chapter first highlights the growing medical and educational needs of children in the educational enterprise, and discusses why school neuropsychologists are uniquely qualified to become the largest group providing rehabilitation services in the world. The chapter then discusses how school psychologists can use a neuropsychologically driven approach to develop evidence-based interventions (EBIs) within a multi-tiered system of supports (MTSS) model. All brain-based areas are discussed in light of appropriate measures and progress-monitoring tools for MTSS. The chapter suggests types of data that can be collected by multiple levels of service providers, such as teachers, interventionists, and school psychologists.

One of the commitments that U.S. society has made has been to educate all children and to help them live and play successfully in the world. Despite our efforts to help children develop successfully in a complex world, various toxic influences affect them in many ways. Thus many children are in need of unique psychological interventions. All children are unique and respond differently to interventions (Merry, McDowell, Hetrick, Bir, & Muller, 2004). In order to find individualized supports, it is beneficial for children with disadvantages/disabilities to undergo neuropsychological evaluations (D'Amato, 1990). Neuropsychologists conduct comprehensive evaluations in order to determine appropriate supports for individuals who may have difficulty learning or functioning in daily life. School psychologists are also instrumental in choosing interventions for children with disabilities (Witmer, 1896). However, many school psychologists lack appropriate training in neuropsychology (D'Amato, Fletcher-Janzen, & Reynolds, 2005). With a neuropsychological background, school psychologists are well equipped to choose, implement, and monitor the progress of EBIs (Davis, 2011; Fletcher-Janzen, 2005).

We have liberally adapted material from 10 previous articles and book chapters to make this chapter possible. These include D'Amato (1990); D'Amato, Crepeau-Hobson, Huang, and Geil (2005); D'Amato, Rothlisberg, and Leu Work (1999); D'Amato and Wang (2015a, 2015b); Rhodes, D'Amato, and Rothlisberg (2008); Root, D'Amato, and Reynolds (2005); Traughber and D'Amato (2005); Witsken, D'Amato, and Hartlage (2008); and Witsken, Stoeckel, and D'Amato (2008).

Through public education, all children, no matter their backgrounds or socioeconomic classes, are given the chance to learn literacy, mathematics, and other related skills. However, it was not until the passing of Public Law 94-142 in 1975 (see U.S. Department of Education, 2016, for a fuller account) that students with disabilities were granted the right to have free and appropriate public education. When students with disabilities started to attend school, special education classes were created so that these students could have rights similar to those of their peers who were considered typically developing. The U.S. education system has made great strides since the 1970s in offering appropriate public education to students who have disabilities. There are now laws that enforce parents' and students' rights regarding education (again, see U.S. Department of Education, 2016). School staff members work hard to ensure that these students are taught in the least restrictive environments, and that they are educated as much as possible with their typically developing peers in general education classrooms. Despite the improvements in the education system, many problems continue to exist for students with disabilities (D'Amato, Fletcher-Janzen, & Reynolds, 2005). A significant problem that many schools have is finding EBIs for all students with similar difficulties (Witsken, Stoeckel, & D'Amato, 2008). This is the focus of the current chapter.

There is a growing need for a neuropsychological approach within school systems. As society continues to change, the need for services within schools has also grown. Due to changes in society, especially advances in modern medicine, the demand on schools to provide appropriate services for students with a wider, more complex array of difficulties (e.g., prematurity, spina bifida, sickle cell disease) has increased. School psychologists may not be well equipped to deal with the growing number of psychological and behavioral problems within classrooms (D'Amato & Hartlage, 2008). It may be that school psychologists are no longer able to master *all* areas of knowledge in order to operate as effective professionals (D'Amato, Fletcher-Janzen, & Reynolds, 2005). For example, school psychologists are expected to know more about the biological bases of neurodevelopmental disorders. There are an increasing number of medical diagnoses, such as low birth weight (Davis, 2011), that contribute to learning deficits in children of which school psychologists need to be aware. Moreover, new disorders seem to emerge each year (e.g., the Zika virus; Power, D'Amato, & Eusebio, 2017).

What Is School Neuropsychology?

According to the National Academy of Neuropsychology (2001), *clinical neuropsychology* is the applied science of brain–behavior relationships. Clinical neuropsychologists must apply a working understanding of psychology, physiology, and neurology to assess, diagnose, and treat patients with neurological, medical, neurodevelopmental, psychiatric, and cognitive disorders (D'Amato, Fletcher-Janzen, & Reynolds, 2005; Witsken, D'Amato, & Hartlage, 2008). In addition to using assessments of neurocognitive, behavioral, and emotional functioning to form hypotheses regarding a client's central nervous system functioning, neuropsychologists must carefully consider how these factors interact with the individual's psychosocial environment (National Academy of Neuropsychology, 2001; Teeter & Semrud-Clikeman, 2007). Viewed from this perspective, neuropsychological assessment should serve a variety of purposes beyond an initial diagnosis. Assessments must be used to guide treatment decisions by identifying an individual's strengths, weaknesses, and needs; to design individual treatment programs tailored to these findings; to evaluate changing treatment needs; and to monitor treatment effectiveness (Davis, 2011; Kazdin, 2011; Root, D'Amato, & Reynolds, 2005). In sum, an understanding of evidence-based neuropsychological functioning must drive our practice of all types and forms of neuropsychology, including counseling, consulting, assessment, and intervention. This claim should not be surprising, since the brain is the origin of human behavior (D'Amato & Wang, 2015a, 2015b).

The field of *school neuropsychology* is relatively new (D'Amato, Fletcher-Janzen, & Reynolds, 2005). In fact, many neuropsychological assessments have roots in adult neuropsychology (Koziol & Budding, 2011). However, as we have grown to understand pediatric brain functioning, more innovative assessments and interventions have been developed specifically for children. Pediatric neuropsychologists have training in applied psychology, clinical psychology, neurology, development, and educational systems (D'Amato & Dean, 1989). The emphasis within pediatric neuropsychology is often on comprehensive evaluations and the interpretation of results from a neurodevelopmental perspective. Clinical neuropsychologists working in schools should also have knowledge of EBIs and can adapt these specific tools to meet needs of individuals (Traughber & D'Amato, 2005). This expertise may be the most important weapon in these psychologists' arsenal.

A comprehensive neuropsychological evaluation typically involves the collection of an individual's detailed history. This history provides a professional with information, such as medical/health history, developmental milestones, and social functioning (see D'Amato, Fletcher-Janzen, & Reynolds, 2005, Appendix A, p. 859, for an example). The history is also essential for gathering information about an individual's culture. Cultural factors, such as age, gender, education, income, and other related factors, should always be taken into consideration during a neuropsychological evaluation. Such implications assist the clinician in determining outcomes.

Training in Clinical Neuropsychology

A neuropsychological interpretation requires specialized training in brain–behavior relationships (D'Amato, Hammons, Terminie, & Dean, 1992). All individuals in this field fall near one end of the neuropsychology training continuum. At one end, individuals have limited training in clinical neuropsychology, and at the other end they are comprehensively trained in it. An individual who is comprehensively trained in clinical neuropsychology will have completed clinical neuropsychology predoctoral coursework, a predoctoral clinical neuropsychology practicum, and an internship and dissertation in this specialization area (Witsken, D'Amato, & Hartlage, 2008). This individual will go on to complete advanced postdoctoral study for up to 2 years, thus allowing him or her to claim expertise in clinical neuropsychology. Individuals who fall at the other end of the continuum, with limited training, usually have completed a single course as required by the American Psychological Association (Witsken, D'Amato, & Hartlage, 2008).

Although many neuropsychologists are also trained as clinical or school psychologists, the reverse is less common. The use of some clinical neuropsychology assessment procedures in practice does not qualify one as a clinical neuropsychologist. By analogy, although a cardiologist uses a stethoscope, anyone who learns how to use a stethoscope does not automatically become a cardiologist (Hartlage, 1987). The fact that clinical and school psychologists and neuropsychologists frequently use similar tools has led to considerable debate about neuropsychologists' training and licensure requirements. Several professional organizations have arisen to represent clinicians and researchers in neuropsychology. These organizations have contributed to establishing training standards and regulating credentialing in neuropsychology. Among these are the International Neuropsychology Association; Division 40 (Clinical Neuropsychology) of the American Psychological Association; and the National Academy of Neuropsychology. Two major national professional credentialing boards in neuropsychology have emerged to specify and regulate practitioner qualifications. These include the American Board of Professional Psychology, which recognizes clinical neuropsychology as a specialty area of practice within psychology, and the American Board of Professional Neuropsychology, which exclusively recognizes specialists in neuropsychology. Currently, most clinical neuropsychologists have obtained a doctoral degree in psychology with coursework, research, and practicum experiences in neuropsychology, followed by postdoctoral training with a neuropsychology emphasis. In addition to core coursework in general psychology, clinical psychology, neurosciences, and clinical neuropsychology, the International Neuropsychology Association internship guidelines requires at least a minimum number of internship hours under supervision of a board-certified clinical neuropsychologist, with at least 50% of the time devoted to clinical neuropsychology. Practicing neuropsychologists are generally expected to obtain licensure from their respective state psychology licensure boards prior to seeking board certification in clinical neuropsychology. Some school psychology programs offer a major or specialization in clinical neuropsychology (e.g., Chicago School of Professional Psychology, University of Utah, Ball State University), which have allowed a variety of school psychologists to specialize in clinical neuropsychology. However, these specializations do not take the place of advanced postdoctoral training. In fact, comprehensive training in school psychology and clinical neuropsychology is advised (D'Amato, Fletcher-Janzen, & Reynolds, 2005).

Multicultural Considerations

Culture is evident in each portion of a neuropsychological evaluation, such as the gathering of background history, choice of assessments, and interpretation of results. Differences in gender, chronological age, nationality, race, and family have significant implications for clinical assessment (Manly, Jacobs, Touradji, Small, & Stern, 2002). All children come from different families, with different backgrounds, morals, and traditions. No parents/guardians raise their children the same way. Although this diversity may make it difficult to compare the results of an intervention's utility even among similar cultural groups, it necessitates that researchers and clinicians tailor programs for specific individuals. An individual's culture is essential in the design of an intervention plan. Our suggested approach highlights the importance of interactions between an individual and his or her interconnected settings (D'Amato, Fletcher-Janzen, & Reynolds, 2005). A professional must consider the social and cultural context in relation to each individual's neuropsychologically driven presentation (Bronfenbrenner & Morris, 1998).

Using Neuropsychology as an EBI in Schools

Numerous authors have argued that neuropsychological principles and practices can be used to improve instructional models and learner outcomes (Gaddes, 1980; Sousa, 2005). Advocates for the training of school psychologists to be able to integrate clinical neuropsychological perspectives into their practices have cited the following potential applications of neuropsychology within the schools (D'Amato, Fletcher-Janzen, & Reynolds, 2005; D'Amato & Wang, 2015a; Hale & Fiorello, 2004; Root, D'Amato, & Reynolds, 2005; Whitten, D'Amato, & Chittooran, 1992). School practitioners should:

1. Interpret the results of neuropsychological assessments to facilitate intervention planning.
2. Propose recommendations for remediation and compensation-based interventions integrating knowledge of developmental and neuropsychological principles, as well as scientifically validated interventions.
3. Integrate a working understanding of neuropsychological development in consultation efforts with teachers, to promote instructional practices that are aligned with knowledge of neurodevelopment.
4. Act as liaisons with the medical community, to better coordinate and evaluate intervention efforts.
5. Educate school staff and parents about the neuropsychological bases of development, behavior, and learning.
6. Conduct research regarding the efficacy of neuropsychologically based interventions and consultation in schools.
7. Serve as rehabilitation specialists in the schools, helping children return from medical or psychological treatment centers.

The literature documenting the potentially invaluable role that could be assumed by school psychologists trained in neuropsychology has been developing (D'Amato, Fletcher-Janzen, & Reynolds, 2005). Particular attention has been paid to the need for practitioners familiar with educational systems, learning, and brain–behavior relationships to serve as liaisons between medical and school teams (D'Amato, Fletcher-Janzen, & Reynolds, 2005; Sousa, 2005, 2006). Pelletier, Hiemenz, and Shapiro (2004) have advocated for the need for school psychologists to capably interpret and translate data from reports by medical professionals and clinical neuropsychologists into practical, empirically based school interventions. Many clinical neuropsychologists and medical practitioners trained in the medical model may be more accustomed to doling out prescriptive interventions, with little feedback from recipients regarding practical considerations unique to the school environment that may be necessary to facilitate successful outcomes. School psychologists are in a unique position to bridge this gap.

Grounded in the biological bases of behavior, a neuropsychological perspective can give practitioners added breadth and depth to their understanding of the thinking processes underlying behavior; such a viewpoint can be held without compromising the value of other perspectives in explaining observed behavior (Gaddes & Edgell, 1994). A neuropsychological approach also integrates the traditional views of looking for the internal causes of behavior (e.g., psychoanalysis) with the external antecedents and consequences of observable behavior (e.g., behaviorism) to gain a full understanding of each learner (Whitten et al., 1992). For more than a century, research has consistently shown that environmental factors explain approximately 40–60% of behavior, while biogenetic factors also explain approximately 40–60% of behavior (Hartlage & D'Amato, 2008).

School psychologists trained in neuropsychology may be the only professionals within the school system who are equipped to address teachers' concerns and questions about each child's unique neurological needs and expected outcomes (Root et al., 2005). School neuropsychologists may be able to provide inservice training to all school staff members who will be working with a child, to facilitate intervention fidelity. School neuropsychologists may also provide valuable services to children's families by answering their questions and providing them with ongoing information. Furthermore, some families may be coping with additional stress, financial problems, guilt, grief, sadness, or anger related to their children's medical conditions (Root et al., 2005). School neuropsychologists recognize the importance of family systems in facilitating successful outcomes for children and may be in a unique position to assist these families (Gaddes, 1980). Consultation between the school neuropsychologist and medical service providers may also be useful in limiting the costs of assessment and intervention by reducing the likelihood of duplicating services (Root et al., 2005). Increased collaboration can also reduce the likelihood of contradictory information and/or recommendations, which in turn will potentially reduce parents' and school staff members' frustration.

NEUROANATOMY AND THE BRAIN

It is important to understand that the brain is composed of four major lobes: the (1) frontal lobe, (2) parietal lobe, (3) occipital lobe, and (4) temporal lobe (see Figure 38.1). It is important to understand the location and function of these structures because such an understanding should inform interventions for children. While the distinct lobes work in concert with one another, each lobe has a specialized function. Due to advances in technology, such as functional magnetic resonance imaging, it is now apparent that EBIs are successful at changing the brain (Shaywitz, 2003). For example, the brain of a child who cannot read can be scanned. The child then can be taught to read by using an evidence-based reading intervention, as specified by the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004). Next, a brain scan can be performed to show which part of the brain has changed as a result of the intervention (D'Amato, Fletcher-Janzen, & Reynolds, 2005).

Much has been learned about the structure and function of the brain in the last century. Although the cerebral hemispheres act in concert, the right hemisphere seems to be specialized for holistic, spatial, and/or nonverbal reasoning, whereas the left is more involved in verbal, serial, and/or analytic-type tasks (Reynolds, 1981; Walsh, 1978). Similarly, models of cognitive processing have been proposed that demonstrate the specialization of brain structure.

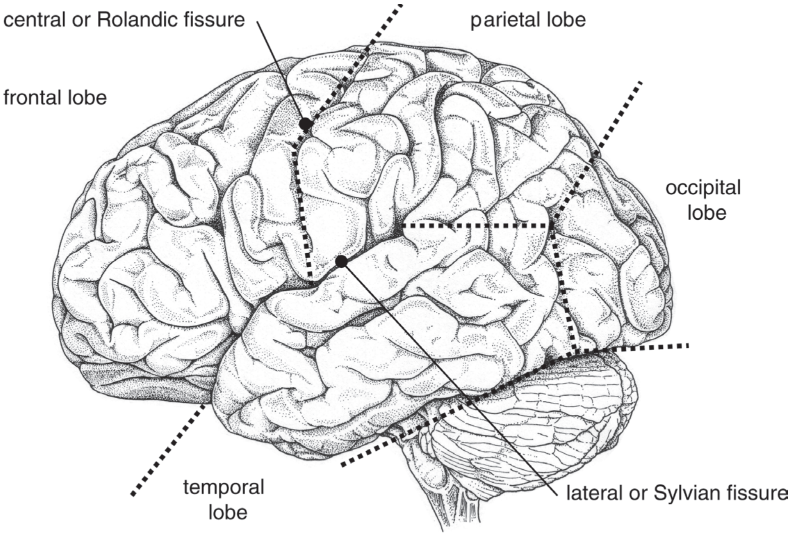


FIGURE 38.1. Lobes of the brain. From D'Amato and Wang (2015a). Copyright © John Wiley & Sons, Inc. Adapted by permission.

Luria and His Functional Integrative Systems

Alexander Luria was a neuroscientist from Russia who helped to pave the way for pediatric neuropsychology (Reynolds & French, 2005). Some areas of Luria's research interests included genetic psychology, aphasia, and neurology. However, it was Luria's neuroscientific approach to the study of human mental processes that contributed the most information to the field of neuropsychology viewed *systemically*. According to Luria, groups of brain areas, working in concert, each make unique contributions to an overall functional system (Kostanaya & Rossouw, 2013). Luria (1973) believed that the basic functions of brain are represented by three functional systems. The first system is responsible for arousal and attention. The second block involves the ability to analyze, code, and store knowledge. The third block is for the application of executive functions for the use of programming behavior. Each block is based on the function of another, and these units cannot carry out one activity completely independently. Luria used his *integrative functional systems* approach to diagnose and find the locations of brain injuries.

Using this innovative, integrative approach, Luria advocated for clinicians' administering a variety of assessments in order to evaluate the sensory, integrative, and generative structures of the brain. Based on the results of these evaluations, clinicians could determine approximate locations of brain injuries, as well as the structures that were affected by the injuries (Luria, 1973). A model such as Luria's hierarchal model of cortical functioning allows neuropsychologists to accurately identify children as having neurodevelopmental disorders (Davis, 2011). Not only can neuropsychologists utilize this model; other professionals, such as school psychologists, can benefit as well. The influence of Luria's model has led to significant development in popular intelligence tests. Both editions of the Kaufman Assessment Battery for Children (K-ABC and KABC-II; Kaufman & Kaufman, 1983, 2004), and both editions of the Cognitive Assessment System (CAS and CAS2; Naglieri & Das, 1997; Naglieri, Das, & Goldstein, 2014), stand firm on a Lurian foundation.

Luria has offered a model of brain organization that has attempted to explain the diversity and complexity of behavior (Luria, 1970, 1973; Reynolds & French, 2005). The elegance of Luria's system lies in its capacity to characterize the brain as an interconnected system of attention, information processing, and action, allowing researchers to visualize the integrated and reciprocal nature of the brain's organization. What is observed after damage or as a result of dysfunction is not the deficit per se, but how the remaining intact areas and subsystems deal with the task or information presented (Morris, 1989). Basic knowledge of how behaviors seem to be interrelated, and of which areas of the brain are purportedly involved in the behavior–behavior associations, gives practitioners an additional way of investigating learning.

Hemispheric Specialization and Learning

One of the simplest ways to understand the brain is to consider how the hemispheres process information in a specialized fashion (see Table 38.1). For example, *simultaneous processing* ability has been affiliated with the right hemisphere because of its holistic nature; it deals with the synthesis of parts into wholes and is often implicitly spatial (Das, Kirby, & Jarman, 1975; see also Naglieri & Otero, Chapter 6, this volume). In contrast, the left hemisphere processes information by using a more *successive/sequential processing* method—considering the serial or temporal order of input (Dean & Anderson, 1997).

Many learners display a preference for either sequential/successive or simultaneous styles of teaching and learning (Luria, 1973). These two modes of cognitive processing are complementary rather than hierarchical. That is, they work together to create an efficient and effective manner of learning. Reynolds and French (2005) also have advocated that the examination of these two styles of processing should be a principal component of understanding how an individual best learns (D'Amato, Crepeau-Hobson, Huang, & Geil, 2005).

Sequential processing involves breaking the stimuli into separate parts, in order to understand what the learner is experiencing. This also involves the serial or temporal order of the stimuli. Input often is organized in a defined order (Davis, 2011). The *Encyclopedia of Clinical Neuropsychology* (Kreutzer, DeLuca, & Caplan, 2011) has defined sequential (successive) processing as follows:

> The perception of stimuli in sequence and the subsequent production of information in a specific arrangement fall under successive processing. . . . Thus,

TABLE 38.1. Description of the Different Styles of Brain Processing by Cerebral Hemisphere

Function	Reference
	Right hemisphere
Processing modes	
Simultaneous	Sperry (1974)
Holistic	Dimond and Beaumont (1974); Sperry (1969)
Visual/nonverbal	Savage and Thomas (1993); Sperry (1974)
Imagery	Seamon and Gazzaniga (1973)
Spatial reasoning	Polzner, Bellugi, and Klima (1990); Sperry (1974)
Nonverbal functions	
Depth perception	Carmon and Bechtoldt (1969)
Melodic perception	Shankweller (1966)
Tactile perception (integration)	Boll (1974b)
Haptic perception	Witelson (1974)
Nonverbal sound recognition	Milner (1962)
Motor integration	Kimura (1967)
Visual constructive performance	Parsons, Vega, and Burn (1969)
Pattern recognition	Eccles (1973)
Memory/learning	
Nonverbal memory	Stark (1961)
Face recognition	Milner (1967) Hecaen and Angelergues (1962)
	Left hemisphere
Processing modes	
Sequential	Sperry, Gazzaniga, and Bogen (1969)
Temporal	Efron (1963); Mills (1977)
Analytic	Eccles (1973); Morgan, McDonald, and McDonald (1971)
Verbal functions	
Speech	Posner, Petersen, Fox, and Raichle (1988); Reitan (1955); Wada (1949)
General language/verbal abilities	Gazzaniga (1970); Smith (1974)
Calculation/arithmetic	Eccles (1973); Gerstmann (1957); Reitan (1955)
Abstract verbal thought	Gazzaniga and Sperry (1962)
Writing (composition)	Hecaen and Marcie (1974); Sperry (1974)
Complex motor functions	Dimond and Beaumont (1974)
Body orientation	Gerstmann (1957)
Vigilance	Dimond and Beaumont (1974)
Learning/memory	
Verbal paired associates	Dimond and Beaumont (1974)
Short-term verbal recall	Kimura (1961)
Abstract and concrete words	McFarland, McFarland, Bain, and Ashton (1978); Seamon and Gazzaniga (1973)
Verbal mediation/rehearsal	Dean (1983); Seamon and Gazzaniga (1973)
Learning complex motor functions	Dimond and Beaumont (1974)

Note. From Davis and Dean (2005). Copyright © John Wiley & Sons, Inc. Reprinted by permission. References can be found in the original chapter; they are not included in the present chapter's References list.

> information can only be comprehended in a temporal, sequential manner, with each piece being dependent on the preceding element. (p. 2262)

For example, when one is teaching reading, the word *cat* can be broken into three segments, with *c* listed on one card, *a* listed on the second card, and *t* listed on the final card. This style of processing is similar to phonetic approaches to reading instruction (Sousa, 2006). Some have used the term *serial* to describe this approach, often seen as residing in the left hemisphere.

The *simultaneous processing* method, in contrast, is holistic and synchronized. Information is considered together, in its entirety, or as a whole. This approach involves integrating the stimuli spatially, organizing them simultaneously, and then integrating them into a whole (Davis, 2011). These functions are seen as residing in the right hemisphere. Kreutzer and colleagues (2011, p. 2301) have defined simultaneous processing as "the process of combining discrete and unconnected stimuli into a single group or whole to assist in comprehension and interpretation. It involves the comprehension of the relationships of and between separate entities and its relation or position to the whole." An example of simultaneous processing is teaching reading by emphasizing the shape of the word *cat*. In this approach, it is important to emphasize how the three letters are linked and shaped. Consequently, people who use this processing style may be able to look at the shape of a word and tell that it is spelled incorrectly, or they may remember where the idea they are searching for can be found on a text page.

It is essential to know that simultaneous and successive processes are neither modality- nor stimulus-specific. That is, although certain functions are processed more efficiently through one process than through the other, any type of information can be processed through either simultaneous or successive means (Reynolds & French, 2005). In terms of teaching and learning, it is important to find out which type of processing is most suitable to assist each student to learn. Some authors have suggested that informal tests and/or observations be used to determine which type of processing is most appropriate. Sousa (2006) has provided an informal hemispheric preference measure and mentions that similar tests are available, although no psychometric data are presented to support any measure. In addition, task demand, genetic predisposition, neurocultural traditions, an individual's level of attention to the task, and the individual's preferred means of completing the task are all factors that can change, depending on the cognitive processing style the individual applies (Luria, 1970, 1973).

NEUROPSYCHOLOGICALLY INFORMED APPROACHES TO ASSESSMENT

Recognizing Biogenetic and Ecological Differences in Disorders

Table 38.2 presents an abbreviated list of the many neuropsychological disorders that have been found to have biogenetic underpinnings (Witsken, D'Amato, & Hartlage, 2008). While clearly related to environmental support systems, the biogenetic bases of these disorders have been recently illuminated (Hartlage & D'Amato, 2008). All school personnel, parents, and teachers who are working to offer school support for biogenetic disorders need to realize that these categories are not unrelated. For instance, while a child with a traumatic brain injury may be supported through cognitive rehabilitation from the school psychologist, it is just as critical to educate those in the child's classroom (including both school personnel and peers) about their role in supporting the child's emotional, academic, and intellectual changes.

Lurian/Eastern versus Actuarial/Western Approaches

Beginning almost a century ago, two distinct approaches were advocated as suitable models to follow in assessing for intervention with children and youth: the *quantitative* approach and the *qualitative* approach. The basic differences between these two approaches relate to how data are collected, organized or aggregated, and linked to interventions related to student performance outcomes. Table 38.3 compares and contrasts these unique approaches to serving children, youth, and families. The quantitative approach, used primarily in the Western world, has focused on the acquisition of formal test data followed by a comparison of scores to normative samples. This approach compares the student's performance on standardized tests, in a variety of brain-related domains, to the performances of same-age peers. This comparison enables the evaluator to determine whether student functioning is below average, average, or

TABLE 38.2. Disorders Found to Have Biogenetic Underpinnings

- Alcoholism
- Alzheimer disease
- Asthma
- Aphasia
- Attention deficit disorder (ADD) and attention-deficit/hyperactivity disorder (ADHD)
- Behavioral/personality disorders
- Cancer
- Dementia not otherwise specified
- Diabetes
- Eating disorders
- Epilepsy
- Fetal alcohol syndrome
- Genetic and chromosomal disorders: phenylketonuria (PKU), Down syndrome
- Hearing/auditory disorders
- HIV/AIDS
- Huntington disease
- Hypertension
- Infants' exposure to prenatal toxins
- Language disorders
- Learning disorders: reading, mathematics, written expression
- Malnutrition
- Migraines/headaches
- Motor skill disorders
- Multiple sclerosis
- Muscular dystrophy
- Parkinson disease
- Perceptual disorders
- Pervasive developmental disorders
- Pick disease
- Prematurity
- Seizure disorders
- Traumatic brain injuries as a result of motor vehicle accidents, pedestrian vehicle accidents, contact/noncontact sports, accidental injuries, abuse, assault
- Vascular disorders
- Vision problems/disorders
- Zika virus

TABLE 38.3. Lurian/Eastern versus Actuarial/Western Approach

Lurian/Eastern approach	North American/Western approach
Theory-driven	No overall a priori theory; data-driven
Attempts to support or confirm a theory	Attempts to disconfirm specific hypotheses
Synthetic	Analytical
Observation-oriented	Evaluation-oriented
Single-case-oriented	Group-comparison-oriented
Describes behaviors	Evaluates behaviors
Subjective	Objective
Looks for patterns of functioning	Looks for differential diagnosis
Qualitative in nature	Quantitative in nature
Flexible	Fixed
Process-oriented	Product-oriented
Focuses on individualized activities	Focuses on multiple tests/procedures
Links behavioral data to functioning	Links psychometric data to diagnosis
Considers the functional system	Considers discrete brain-related areas
Clinical–theoretical	Actuarial–standardized

above average in a number of areas. Standardized tests can offer the practitioner a helpful set of tasks to evaluate important neuropsychological abilities. If an appropriately selected, comprehensive, wide-band neuropsychology battery is administered, most essential areas are likely to be evaluated (Johnson & D'Amato, 2011). For many practitioners who use uniform assessments, such tests are selected because they measure core neuropsychological abilities that should be evaluated within a traditional assessment. From a quantitative view, student data are generated and used primarily for four comparisons: (1) How does the student's performance compare to that of others in the class, in the community, or in the state/nation? (2) What are the student's unique patterns of performance, including strengths and weaknesses? (3) When right-hemispheric abilities (i.e., including sensory–motor processes) are compared to left-hemispheric abilities, do unique patterns emerge that reveal significant processing styles? (4) Does an analysis of displayed signs and symptoms suggest problems related to a specific disorder (e.g., a nonverbal learning disorder) or the potential course of a disorder? A quantitative approach is often used for complex cases in which multiple interventions have not helped students learn. For example, quantitative approaches are often used in an MTSS tier 3 evaluation.

On the other hand, the qualitative approach focuses on the uniqueness of the individual under study and seeks to match the procedures used with the individual's distinctive profile. Glozman (1999) indicated that "Luria's neuropsychological assessment is recognized today by the world's scientific community to be the most comprehensive and flexible method of neuropsychological evaluation available, which is also based on an understanding of the factors underlying complex psychological activities" (p. 23). A few seasoned neuropsychological practitioners have offered models of assessment that examine curriculum-related processing skills to evaluate student difficulties (D'Amato, Rothlisberg, & Leu Work, 1999; Glozman, 1999; Sousa, 2005). Sample tasks that could be used in a neuropsychologically based MTSS (NB-MTSS) model have been offered by Gaddes and Edgell (1994) and Luria (1970). Gaddes and Edgell offer tasks to evaluate auditory processes and aphasic signs in oral speech (17 questions; p. 411), visual processes (13 questions; pp. 411–412), tactile processes (6 questions; p. 412), and motor–expressive processes (6 questions; pp. 412–413). To evaluate auditory processes/abilities, here are six sample questions from Gaddes and Edgell's list (p. 411):

1. Can he recite all the letters of the alphabet?
2. Can he associate all the phonetic sounds of all the letters? . . .
3. Can he name common objects without hesitation?
4. Can he describe the use of common objects?
5. Word fluency: How many nouns can the child produce in one minute?
6. Can he construct a meaningful sentence if given three words?

It is obvious that an analysis such as this can provide a wealth of data that can be used to inform the choice of intervention(s). It is important to use the child's strengths to support his or her needs.

Similarly, Luria (1970) provided a list of evaluation activities to assess the neuropsychological processes that underlie arithmetic (Gaddes & Edgell, 1994). His steps included asking the student (1) to count aloud (to check memory of number in the correct sequence), (2) to recognize quantities, (3) to read and write single digits, (4) to read and write multidigit numbers (to show an understanding of the decimal system), (5) to recognize relative values, (6) to show competence in the basic arithmetic skills, and (7) to attempt more complex calculations (Luria, 1970, as cited in Gaddes & Edgell, 1994, p. 419). If integrated within an NB-MTSS framework, this type of approach is useful at tier 1, when interventions are developed from the curriculum or when informal techniques are used. When used at tier 2 or 3, this approach offers the ability to select or develop a battery based on the needs of the student under study.

Fixed- versus Flexible-Battery Approaches

The use of fixed or flexible batteries in neuropsychology has been described extensively elsewhere (Gaddes & Edgell, 1994; Rhodes, D'Amato, & Rothlisberg, 2008; Riccio & Wolfe, 2003; Teeter & Semrud-Clikeman, 2007). In general, practitioners appear to be divided between those who advocate a fixed-battery approach (i.e., administering a consistent, standardized battery to all students, regardless of their referral questions) and those who advocate a flexible-battery approach (i.e., using a case study analysis focusing on the dynamic clinical interactions and "the process" to better understand each individual). Integrative models of assessment incorporating elements of each approach have become increasingly popular (D'Amato, Fletcher-Janzen, & Reynolds, 2005; Davis, 2011). Selecting aspects of both approaches allows the clinician to emphasize the person being assessed, as well as the individual's fit with his or her environment. Because this approach allows examiners to select instruments unique to each child's needs in response to specific referral questions, it appears particularly useful in school-based practice, where large caseloads and legal requirements governing timely completion of evaluations make it necessary to efficiently select instruments that will be most likely to facilitate treatment planning. Our model advocates the composition of assessments based on both fixed and flexible ideas. For example, at tier screening, all students are evaluated using a fixed number of tasks, but tasks are selected from both quantitative and qualitative areas.

Integrating Comprehensive Ecological Neuropsychological Assessments with Interventions

The primary aim of neuropsychological approaches to assessment and later to intervention is to systematically integrate information collected about the individual's brain integrity with information obtained from a comprehensive review of the environmental systems that are influencing the child. The argument that neuropsychological models look only at the intrinsic aspects of an individual's performance misrepresents the true nature of practice (Gaddes, 1985; Gaddes & Edgell, I994). The practitioner who is using a neuropsychological model of inquiry is much more likely to conduct a more complete assessment than one who neglects the internal components of behavior. Far from ignoring environmental controls on performance, the practitioner with a neuropsychological background will also try to consolidate environmental and biological aspects that are contributing to the problem under study, with an emphasis on understanding *why* the difficulty exists (D'Amato, 1990; Taylor & Fletcher, 1990). Such an assessment allows the examiner to reduce the subjectivity in traditional neurological examinations by conducting assessments that lead to quantifiable standardized scores, thereby increasing the reliability of the assessment, as well as allowing for a more useful baseline for comparisons across time.

A comprehensive, ecologically driven, neuropsychologically based evaluation should include appraisal of all brain-based *systems* and individualized *contexts*, using a variety of *methods* covering various *sources* and *settings* (see Figure 38.2). Systems are brain-based and include communication/language, personality/behavior, sensory/perception, environmental fit, academic ability, motor functions, and cognitive ability. Information gathered should include data about when the problem or problems do and do not occur. Given these parameters, contemporary neuropsychological assessment should be designed to evaluate a variety of ecologies (Leu & D'Amato, 1994). Additionally, parents, teachers, and students should be consulted to determine for whom the behavior causes concern, what the desired changes are, and how individuals in the environment might facilitate change. Indeed, exploration of how the setting may have to change to accommodate students' needs should be considered. The most critical question we ask teachers is this: "What would you like the students to do that they are not currently doing?" (D'Amato, 1985).

WHAT AREAS SHOULD BE ASSESSED FROM A NEUROPSYCHOLOGICAL PERSPECTIVE?

D'Amato and colleagues (e.g., D'Amato et al., 1999) have long advocated for the importance of all ecologically sensitive neuropsychological evaluations to cover the nine major areas displayed in Table 38.4. These include the following areas: perceptual/sensory functions, motor functions, intelligence/cognitive abilities, executive functioning/attention, memory, communication/language skills, academic achievement, personality/behavior/family, and educational/classroom environment.

When an evaluation of the interaction between a child's characteristics and the instructional environment has been completed, an individualized education program based on the child's strengths as well as needs can be developed. Unlike the deficit model, the ecological neuropsychological approach to understanding how the brain processes information should focus on what the child can do, rather than what he or she cannot do (Hartlage & Telzrow, 1983). Such an approach to learning and educational intervention in general is strengths-based, keeping in mind the old adage "Dead tissue will not learn" (Hartlage & Telzrow, 1983; Riccio, Hynd, Cohen, & Gonzalez, 1993). All intervention should focus on the match between the child

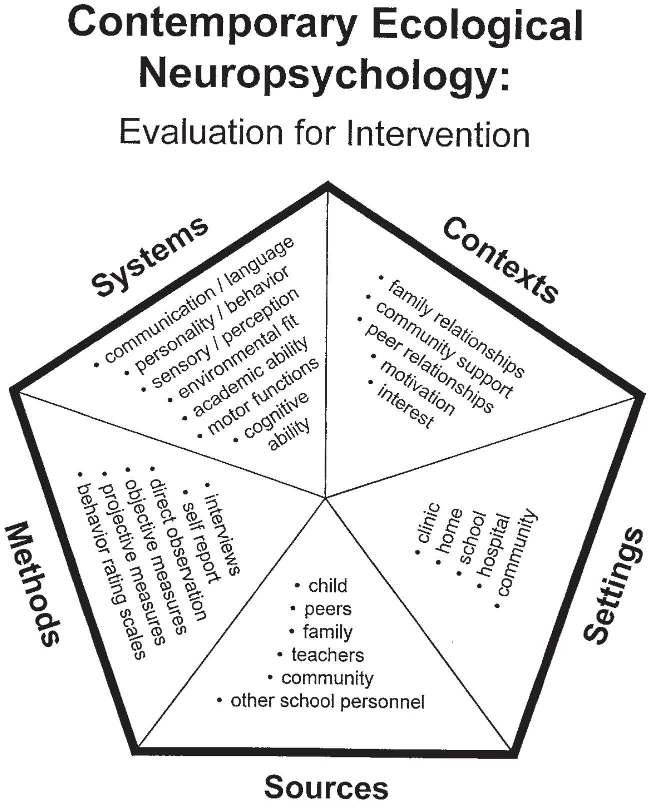


FIGURE 38.2. Components of an ecological neuropsychology evaluation. From D'Amato, Crepeau-Hobson, Huang, and Geil (2005). Copyright © Springer US. Reprinted by permission.

and his or her instructional environment, in addition to providing *remediation* and *compensation* skills for the child (D'Amato & Rothlisberg, 1996; Leu & D'Amato, 1994).

A Learner's Motivation and Neurodevelopment

The identification of a connection between the brain and learning creates the foundation for neuropsychological intervention in teaching and learning in educational settings. As D'Amato and colleagues (1999) have previously stated, offering an EBI should be the intended outcome of any evaluation for a school or university setting. Ornstein and Sobel (1987) have claimed that what a person can learn is also affected and organized by emotions and motivation; that is, the understanding of such aspects of the learner's motivation as attention, relevance, satisfaction, and confidence will determine future learning. Figure 38.3 shows how learner motivation must be understood within the context of attention (A), relevance (R), satisfaction (S), and confidence (C).

Evidence has shown that environmental factors contribute to the growth of the brain as a result of developmental plasticity through numerous activities such as cell death, dendritic branching, pruning, and selecting neuronal connections, just to name a few (Lezak, Howieson, Bigler, & Tranel, 2012; Witsken, D'Amato, & Hartlage, 2008). Furthermore, the neurodevelopment of the brain is influenced by environmental and inherited factors to a considerable degree (Beaumont, 2008; Davis, 2011). This fact also points to the importance of parental involvement during neurodevelopmental years (e.g., reading to toddlers).

Interventions for a student with emotional or learning issues may be directed toward any one area or a combination of the following areas: (1) changing the child (e.g., teaching compensatory or remedial skills); (2) changing the environment (e.g., altering the type of instruction and teaching to the student's strengths); (3) changing the per-

TABLE 38.4. What Areas Should Be Assessed from a Neuropsychological Perspective?

1. Perceptual/sensory
    - Visual
    - Auditory
    - Tactile/kinesthetic
    - Integrated
2. Motor functions
    - Strength
    - Speed
    - Coordination
    - Lateral preference
3. Intelligence/cognitive abilities
    - Verbal functions
        - ○ Language skills
        - ○ Concepts/reasoning
        - ○ Numerical abilities
        - ○ Integrative functioning
    - Nonverbal functions
        - ○ Receptive perception
        - ○ Expressive perception
        - ○ Abstract reasoning
        - ○ Spatial manipulation
        - ○ Construction
        - ○ Visual
        - ○ Integrative functions
4. Executive functioning/attention
    - Sustained attention
    - Inhibition
    - Shifting set
    - Problem solving
5. Memory
    - Short-term memory
    - Long-term memory
    - Working memory
    - Retrieval fluency
6. Communication/language skills
    - Phonological processing
    - Listening comprehension
    - Expressive vocabulary
    - Receptive vocabulary
    - Speech/articulation
    - Pragmatics
7. Academic achievement
    - Preacademic skills
    - Academic skills
        - ○ Reading decoding
        - ○ Reading fluency
        - ○ Reading comprehension
        - ○ Arithmetic facts/calculation
        - ○ Social studies
        - ○ Language arts
        - ○ Science
        - ○ Written language
8. Personality/behavior/family
    - Adaptive behavior
        - ○ Daily living
        - ○ Development
        - ○ Play/leisure
    - Environmental/social
        - ○ Parental/family
        - ○ School environment
        - ○ Peers
        - ○ Community
    - Student coping/tolerance
    - Family interpersonal style
9. Educational/classroom environment
    - Learning environment fit
    - Peer reactions
    - Community reactions
    - Teacher/staff reactions
    - Classroom dispositions

Note. From D'Amato and Wang (2015a). Copyright © John Wiley & Sons, Inc. Adapted by permission.

ceptions, attitudes, and expectations of significant others toward the child (e.g., the teacher views the child's abilities rather than his or her disabilities); and/or (4) changing the child's perceptions, attitudes, and expectations toward him- or herself (i.e., the child recognizes that he or she is not incapable, but able, when learning is approached differently).

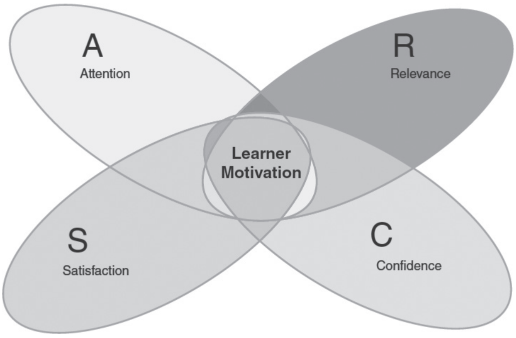


FIGURE 38.3. Visual display of the components of a learner's motivation. From D'Amato and Wang (2015a). Copyright © John Wiley & Sons, Inc. Reprinted by permission.

Problem Solving and EBIs in Neuropsychology

Neuropsychological assessment should be an approach to problem solving in which the goal is to provide both direct and indirect services to children to improve their rehabilitation, mental health, and educational development (D'Amato et al., 1999; Traughber & D'Amato, 2005). Therefore, the role of clinical neuropsychologists, school neuropsychologists, and special educators in the schools should be to make observations, formulate theories, generate and test hypotheses, evaluate data, and draw conclusions from empirical evidence about ways to intervene with children who are unsuccessful in their current educational system (Bray & Maykel, 2016a, 2016b; Rothlisberg, D'Amato, & Palencia, 2003; Stoner & Green, 1992). While special educators have been primarily responsible for remediating skills and developing compensatory strategies for students with learning difficulties (Hallahan & Kauffman, 1991), psychologists in the schools have generally been limited to their role as gatekeepers of special education (Ysseldyke, Reynolds, & Weinberg, 1984). As gatekeepers, school psychologists test, diagnose, and place children in special education programs. The focus has not been on individual differences, unique styles of processing, individual data, evidence-based practices, or providing services based on students' needs (Sattler & D'Amato, 2002a, 2002b). An NB-MTSS approach can be used to provide all of these previously discussed emotional and psychological interventions to help children, teachers, and schools find success. The traditional neuropsychological model has been criticized as a model that stressed pathology and labeling, as well as chronicity and permanence, while ignoring psychosocial factors and potential for change in children identified with specific learning difficulties (Gaddes & Edgell, 1994). As demonstrated throughout this chapter, this criticism is inaccurate and unwarranted.

At one of the most basic levels, all practitioners must understand the *content* and the *task* that present problems for the individual student under study. As Luria has stated, the next step should be a contextual analysis of the subject, as well as of how best to teach the content that has not been learned. This procedure should culminate in an understanding of how this student will best neuropsychologically process information for future learning.

BRIDGING THE GAP BETWEEN NEUROPSYCHOLOGY, SCHOOLING, AND SPECIAL EDUCATION

Clinical neuropsychology does pursue what Martens (1992) has referred to as "high-inference attributions" (i.e., "why" questions) about the relationship between atypical brain development and emotional or learning difficulties. These questions are referred to as high-inference because of the functional relationship between brain dysfunction and emotional or learning difficulties. A pediatric neuropsychologist should combine knowledge about brain–behavior relationships with knowledge gleaned from informal sources (i.e., the student, peers, teachers, and parents), in order to develop a complete picture of both biogenetic and psychosocial/environmental factors (D'Amato et al., 1999; Gaddes & Edgell, 1994).

In summary, ecological neuropsychology seeks to provide an *assessment-for-intervention link* rather than an *assessment-to-placement link*. In 2002, an international Future of School Psychology Conference was held in Indiana and resulted in the development of 54 goals grouped under five priority categories (Sheridan & D'Amato, 2003). Significantly, each priority area included an emphasis on *linking assessment to intervention,* as well as on ecological and systems theories in which students are viewed as part of family and school systems within the context of community and culture (Meyers, 2002). The ecological neuropsychology model is consistent with the mission of the conference and provides all psychologists with a framework to utilize in practice (Sheridan & D'Amato, 2003).

Historically, the educational system has sought to confirm diagnostic hypotheses about children, rather than to focus on the appropriate rehabilitation of children (Epps & Tindal, 1987; Traughber & D'Amato, 2005; Ysseldyke, 1987). Within this framework, an attempt to individualize education and seek effective treatment of children's learning problems has not been efficaciously pursued. To counter this movement, the MTSS framework has been instituted to intervene before students begin to fail. That is, this model does not wait for a student to fail before curriculum adaptations are made, and educators are able to offer interventions early, as needed. So, too, intervention outcomes are also documented and evaluated.

The school psychologist's role within MTSS has often been limited to providing small-group or classwide interventions. This circumstance continues to be surprising, since school neuropsychologists have again become one of the largest groups of providers of rehabilitation services for neuropsychologically impaired children and adolescents. Education has continued to change across the years, but what we know about how the brain processes information has not been integrated into the educational enterprise. IDEA is reauthorized from time to time, and it is only during the reauthorization review period that changes to the law can be made. In the review period for IDEA 2004, both the President's Commission on Excellence in Special Education and the Secretary of the U.S. Department of Education eliminated the learning disability discrepancy model from the federal regulations, which has reduced the demand for neuropsychological evaluations in the schools (Cortiella, 2003). While a discrepancy model may still be used, it is much more likely that alternative approaches like MTSS models will be employed. Yet, in more severe cases of students with learning disabilities, neuropsychological evaluations will still be needed. It would seem that we are not using our brain-based knowledge to guide our public education system.

INTEGRATING TRADITIONAL NEUROPSYCHOLOGICAL ASSESSMENT BATTERIES INTO EVALUATION

This section highlights some of the more widely used neuropsychological measures of cognitive, academic, and behavioral functioning. It is important to note that neuropsychology as conceptualized by most of the early pioneers includes an evaluation of each client's full range of human functioning. Hundreds of assessments exist that claim to be neuropsychological measures, and here we only describe some of the major batteries and tests (Davis, 2011). While many view neuropsychological assessments as a set of specific procedures for a clinical psychologist to administer, many commonly used school psychology measures can be interpreted as representing brain–behavior relationships (Rhodes et al., 2008). Some assessments can be used informally (i.e., they can be administered by general and special education teachers in a classroom setting).

A traditional example, and the most commonly used fixed battery, is the Halstead–Reitan Neuropsychological Battery (HRNB; Johnson & D'Amato, 2011). The HRNB was developed to address Halstead's insight that the then-current measures of intelligence did not account for a biogenetic basis of intelligence and failed to link assessments to brain functioning (Davis, Johnson, & D'Amato, 2005). Years later, research indicated that measures of neuropsychological functioning overlapped by a meager 10% with traditional IQ tests; this finding supported Halstead's early theories that these tests failed to capture the full range of human cognitive functioning (D'Amato, Dean, & Rhodes, 1998; D'Amato, Gray, & Dean, 1988; Sattler & D'Amato, 2002a, 2002b). The current HRNB was designed to differentiate patients with and without brain injuries through 10 subtests that are intended to be used as part of a complete battery, including the age-appropriate Wechsler scale and a comprehensive personality assessment measure. Historically, this was the first widely used, and for many years the most popular, neuropsychological battery in the world. Some have argued that the Luria–Nebraska Neuropsychological Battery (LNNB) is the second most commonly used neuropsychological test battery, although it has received quite mixed reviews (Davis et al., 2005).

Two contemporary, unique batteries have become popular recently and are marketed toward school psychologists. These additions include the NEPSY-II and the Dean–Woodcock Sensory–Motor Battery (DWSMB). Although these batteries may be classified as fixed quantitative approaches, they may also be used as part of a flexible battery. The original NEPSY: A Developmental Neuropsychological Assessment was the first to attempt to measure neuropsychological functioning specifically for children using a Lurian approach, rather than slightly modifying or renorming adult measures (Korkman, Kirk, & Kemp, 1998; Titley & D'Amato, 2008). The NEPSY-II, the second edition of the original NEPSY, was designed to assess neuropsychological development in children and adolescents ages 3–16 (Korkman, Kirk, & Kemp, 2007). To suit a variety of diagnostic needs, examiners may select from subtests organized to assess functioning across six domains: (1) Attention and Executive Functioning, (2) Language, (3) Visuospatial Processing, (4) Sensorimotor, (5) Memory and Learning, and (6) Social Perception. The NEPSY-II does not offer scores based on these six domains.

The DWSMB is an important part of the neuropsychological examination (Davis & D'Amato, 2005). This battery was developed to offer normative data related to the traditional neurological (behavioral) examination (Davis, 2011). This measure may be used alone or in conjunction with other psychological assessment tools. Used in this fashion, the DWSMB can provide useful data, including pathognomonic signs of cerebral dysfunction as well as neuropsychological functioning across sensory and motor domains (Davis & D'Amato, 2005).

The KABC-II (Kaufman & Kaufman, 2004) was among the first cognitive measures to be based on Luria's neuropsychological foundation. This measure includes the areas of learning, sequential and simultaneous processing, planning, and knowledge. No student can be understood without an assessment of his or her academic foundational knowledge, and the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014), allows such an analysis based on performance in reading, mathematics, written language, and oral language. Finally, it is critical not to disregard basic skills, which are the building blocks of how the brain processes information. Thus some component of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) should be included in all initial evaluations (Witsken, Stoeckel, & D'Amato, 2008).

EVIDENCE-BASED NEUROPSYCHOLOGICAL INTERVENTIONS

Given the current legal and educational context, school psychologists must have a working understanding of brain–behavior relationships as well as unique qualities of the educational environment to select EBIs that can be reasonably expected to produce desirable, measurable child outcomes (Traughber & D'Amato, 2005). Developments both within and outside the field of school psychology have encouraged practitioners to use treatment and intervention approaches that have been supported by credible research (Traughber & D'Amato, 2005). One ambitious effort was the creation of the APA Division 16 Task Force on Evidence-Based Interventions in School Psychology. The task force produced a procedural and coding manual containing a comprehensive and rigorous set of criteria for evaluating school-based interventions (Kratochwill & Stoiber, 2002).

However, numerous barriers to implementation of EBIs exist in the practice of school psychology. Kratochwill and Shernoff (2003) have discussed the challenge of promoting EBIs in professional practice. There are a number of groups participating in evaluating literature for EBIs whose various criteria may confound efforts. In addition, the adoption of EBIs into school psychology practice may be hampered by time constraints and limited resources, while some practitioners may follow clinical judgment rather than EBIs. Furthermore, lack of training among psychologists and teachers may prevent the implementation of EBIs in practice.

In addition to promoting EBIs in the field of school psychology, there is a need for shared responsibility, guidelines to support implementation and ensure efficacy, and professional development. Although understanding and applying EBIs to practice remains complex and challenging (especially when related to clinical neuropsychology), this chapter advocates for the use of neuropsychologically based EBIs by means of an MTSS approach (e.g., Traughber & D'Amato, 2005). At the same time, we believe that all practitioners must be reviewing traditional interventions in schools as potential EBIs. Each intervention they carry out should be viewed as a clinical study searching for additional evidence.

How to Measure EBIs

Within the MTSS framework, it is imperative that schools utilize EBIs to address the academic and behavioral needs of children. EBIs are practices that have been proven effective through outcome evaluations. As noted, there is no one encyclopedia or database that lists all EBIs; however, multiple sources, such as the What Works Clearinghouse (WWC; *https://ies.ed.gov/ncee/wwc*) and Kazdin (2011), provide critical features of EBIs that aide in determining effectiveness (Traughber & D'Amato, 2005). According to the WWC, the results of an intervention study can fall into one of three categories: (1) meets WWC standards without reservations, (2) meets WWC standards with reservations, or (3) does not meet WWC standards. In order for a study to meet WWC standards, the study must make clear that it had the following characteristics: groups randomly assigned, low sample attrition, and low/no confounding factors or concerns with outcomes.

The WWC standards have helped to identify many large-scale academic and behavioral interventions for children and youth ages 0–21, as well as school professionals with expertise in these interventions. However, due to the fact that the interventions listed on this website tend to be more expensive and require larger commitments for school districts, it may be more beneficial to refer to criteria such as those listed by Kazdin (2011). According to Kazdin, seven criteria can be used to establish an intervention as evidence-based. Other treatment manuals/coding schemes are reviewed in Traughber and D'Amato (2005).

According to Kazdin (2011), a multitude of disciplines in different countries, organizations, and professional groups see the necessity of EBIs. However, given the fact that not all interventions will meet the standards, it is beneficial to consider the benefits of using a single-case research design.

Using Single-Case Research Designs to Evaluate Neuropsychological Treatment

Single-case research designs allow experimental investigation of one subject (Kazdin, 2011). Single-case designs are used heavily in applied research, such as in schools or physicians' offices; they are used to evaluate programs in settings where typical group designs are not feasible (Kazdin, 2011). Single-case study designs are easily adjusted and altered in response to the child's performance. According to Kennedy (2005), single-case study designs have brought a new source of experimental rigor to the educational setting, have resulted in new and effective strategies for educating children, and have facilitated continued growth into new areas of educational implementation. A single-subject design is ideal for working with children with disabilities through a neuropsychological lens, as well as studying student behavior within the educational setting (Hill, 2015). In addition to their utility within applied settings, single-case research designs have played a significant role in elaborating relationships between the brain and behavior (Kazdin, 2011).

Specifically, a single-case study design allows experimental investigation of one subject (Kazdin, 2011) and involves repeated, systematic variables before, during, and after the introduction of an independent variable (Kratochwill et al., 2010). Thus a single-case study design often develops hypotheses about human behavior, serves as a source for developing intervention techniques, permits the study of unique and rare phenomena, provides a counterinstance of ideas that are considered universal, and has persuasive and motivational value (Kazdin, 2011).

We strongly encourage all doctoral students to consider dissertations that review neuropsychologically driven single-case EBIs. Almost two dozen such dissertations are available from our institution, and three studies are highlighted here to show how unique school neuropsychological participants can be matched with EBIs to demonstrate important results. In our first example, Sands, Fischer, and D'Amato (2016) explored the neuropsychological functioning of a 14-year-old male with comorbid Tourette syndrome and a generalized anxiety disorder. Results of the participant's neuropsychological evaluation were utilized to adapt and implement EBIs, which included cognitive-behavioral therapy (CBT) and the Breathe2Relax application, a stress management tool used for diaphragmatic breathing developed by the National Center for Telehealth. Breathe2Relax can be used in isolation as a stress reduction tool, or in combination with treatment by a health care professional. Results indicated that the participant demonstrated a decrease in anxiety during treatment and posttreatment conditions, as evidenced by self-report ($r = -.90$, slope $= .1021$, $p \leq .001$) and one teacher's standardized rating. This study was based on a dissertation by Fischer (2015).

We (Power, Hill, D'Amato, & Losoff, 2016) used a neuropsychological approach to intervene with a 15-year-old Hispanic male who was receiving special education services through meeting IDEA 2004 eligibility criteria for autism and an intellectual disability. The purpose of this study was to determine the effectiveness of Social Stories™ in increasing the frequency of eye contact, the knowledge of eye contact, and the social-emotional reciprocity of the participant. The results indicated that Social Stories was significantly effective in increasing the frequency of eye contact and knowledge of eye contact, but did not improve underlying neuropsychological deficits such as social-emotional reciprocity. This work is based on a dissertation by Hill (2015).

Skierkiewicz and D'Amato (2017) analyzed data from a single case to evaluate the effectiveness of CBT in reducing anxiety symptoms in a male child who had experienced sexual trauma. This study incorporated manualized CBT principles with neuropsychological theory, which allowed for an analysis of the intersection between behavioral principles and brain–behavior relationships. This study evaluated the effectiveness of Coping CAT, an intervention designed to reduce anxiety in elementary-age children, over the course of 14 individual sessions. Findings were mixed, suggesting continuation of this line of research, but resulting in more than 14 individualized suggestions for future research. This study was based on a dissertation by Skierkiewicz (2016).

Using an NB-MTSS Model

It seems important to offer a multi-tiered approach to providing services and interventions to struggling students at increasing levels of intensity. NB-MTSS can provide academic and behavioral screening with valid assessment measures and ongoing monitoring if improvement has not been made (Semrud-Clikeman, 2005). An NB-MTSS model can be used to help students improve in most academic and behavioral areas. Although reading is a primary focus in this section, other academic areas that should also be evaluated as part of an NB-MTSS model are presented in Table 38.5.

Using an NB-MTSS model will provide a valid evaluation of a child's ability to learn. To understand and predict educational outcomes, a child's ability to process language, comprehend what is heard, and organize and use information will need to be assessed. Semrud-Clikeman (2005) has argued that it is typical for current MTSS models to neglect neuropsychological variables by focusing on the curriculum in isolation. Although both a traditional MTSS model and an NB-MTSS model focus on data from the regular education cur-

TABLE 38.5. Formal and Informal Neuropsychological Areas to Evaluate, or Classroom and Curriculum Data to Collect, in an NB-MTSS Model

Areas that should be evaluated, with relevant references	Evaluation tools
	Reading
Phonemic awareness (e.g., sound comparison, segmentation, blending): D'Amato, Fletcher-Janzen, and Reynolds (2005); Fletcher, Lyon, Fuchs, & Barnes (2007); Joseph (2005); Shaywitz (2003)	DIBELS: Initial Sound Fluency, Phoneme Segmentation Fluency[a] • aimsweb: Test of Early Literacy[a] • IGDI: Alliteration, Rhyming[a] • Rigby Reads[a] • CTOPP-2: Elision, Blending Words[b] • KTEA-3: Nonsense Word Decoding[b] • WIAT-III: Pseudoword Decoding[b] • WJ IV COG: Sound Blending[c] • HRNB: Speech Sounds Perception Test[c] • NEPSY-II: Phonological Processing[c]
Phonological awareness/ phonics (letter names/sounds and word recognition): Fletcher et al. (2007); Shaywitz (2003); Sousa (2005)	DIBELS: Letter Naming Fluency, Nonsense Word Fluency[a] • WJ IV ACH: Letter–Word Identification[b] • WIAT-III: Word Reading, Pseudoword Decoding[b] • WRAT4: Reading/Word Calling[b] • GORT-5[b] • WJ IV COG: Word Attack[c] • TOWRE-2[c] • DAS-II: Phonological Processing[c] • NEPSY-II: Phonological Processing[c]
Vocabulary: D'Amato, Fletcher-Janzen, and Reynolds (2005); Joseph (2005); Shaywitz (2003); Sousa (2005)	DIBELS: Word Use Fluency, Word Naming[a] • IGDI: Picture Naming[a] • Rigby Reads[a] • KTEA-3: Reading Vocabulary[b] • WJ IV ACH: Oral Reading, Sentence Reading Fluency[b] • CREVT-3[c] • CELF-5[c] • DAS-II: Word Definitions[c] • KABC-II: Expressive Vocabulary, Verbal Knowledge[c] • PPVT-4[c] • NEPSY-II: Word Generation[c] • WISC-V: Vocabulary, Word Reasoning[c]
Reading fluency: Fletcher et al. (2007); Shaywitz (2003); Sousa (2005)	Classroom words correct per minute[a] • Informal assessment of words read correct/per minute[a] • aimsweb: Reading CBM[a] • DIBELS: Oral Reading Fluency[a] • WJ IV ACH: Reading Fluency[a] • WIAT-III: Oral Reading Fluency[a] • GORT-5[b] • TOWRE-2[b] • Test of Reading Fluency[b]
Reading comprehension: Fletcher et al. (2007); Shaywitz (2003); Sousa (2005)	DIBELS: Retell Fluency, Daze[a] • AIMSweb: Reading Maze passages[a] • WJ IV ACH: Passage Comprehension, Reading Recall[a] • Rigby Reads[a] • TORC-4[b] • OWLS-2: Reading Comprehension[b] • WIAT-III: Reading Comprehension[b] • KTEA-3: Silent Reading Fluency, Reading Comprehension[b] • GORT-5[b]

(continued)

TABLE 38.5. *(continued)*

Areas that should be evaluated, with relevant references	Evaluation tools
Phonological access (rapid automatic naming): Hale and Fiorello (2004); Joseph (2005); Shaywitz (2003)	Timed naming activities: Naming numbers, letters, animals, foods[a] • IGDI: Picture Naming*a* • CTOPP-2: Rapid Letter Naming, Rapid Number Naming*b* • KTEA-3: Associational Fluency, Object Naming Facility, Letter Naming Facility*b* • WJ IV OL: Rapid Picture Naming*b* • DAS-II: Rapid Naming*c* • NEPSY-II: Speeded Naming*c*
Oral language/listening comprehension: Semrud-Clikeman (2005); Shaywitz (2003); Sousa (2005)	DIBELS: Retell Fluency[a] • WIAT-III: Oral Expression*b* • KTEA-3: Oral Expression*b* • CELF-5*c* • EOWPVT-4*c* • OWLS-2: Oral Expression*b* • WJ IV ACH: understanding directions*b* • WJ IV OL: understanding directions*b* • PPVT-4*c* • NEPSY-II: Comprehension of Instructions, Word Generation*c* • WIAT-III: Listening Comprehension*b* • KTEA-3: Listening Comprehension*b* • OWLS-2: Listening Comprehension*b* • DAS-II: Verbal Comprehension*c* • KABC-II: Expressive Vocabulary, Verbal Knowledge, Riddles*c*
Additional neuropsychological areas to evaluate for reading	
Short-/long-term and working memory: Fletcher et al. (2007); Hale and Fiorello (2004); Shaywitz (2003)	Ability to follow two- and three- part directions[a] • WJ IV ACH: Reading Recall/Reading Recall—Delayed*b* • NEPSY-II: Sentence Repetition, Narrative Memory*c* • HRNB: Speech Sounds Perception Test*c* • DAS-II: Recall of Digits, Recall of Objects*c* • WISC-V: Naming Speed Literacy, Naming Speed Quantity*c* • KABC-II: Number Recall, Hand Movements, Atlantis/Atlantis—Delayed, Rebus/Rebus—Delayed, Word Order*c*
Receptive/expressive language: Shaywitz (2003); Semrud-Clikeman (2005); Sousa (2005)	Observations of conversation: turn taking, tangential conversation, time to process information[a] • PPVT-4*c* • EOWPVT-4, ROWPVT-4*c* • CELF-5*c* • Comprehensive Assessment of Spoken Language*a* • WJ IV ACH: understanding directions, Reading Recall*b* • NEPSY-II: Body Part Naming and Identification*c* • HRNB: Aphasia Screening Test*c*

(continued)

TABLE 38.5. *(continued)*

Areas that should be evaluated, with relevant references	Evaluation tools
Attention/executive functions: Fletcher et al. (2007); Hale and Fiorello (2004); Semrud-Clikeman (2005)	Informal classroom observations (e.g., time on task)[a] • Performance on *N*-backs[a] • WJ IV ACH: Understanding directions[b] • WJ IV COG: Numbers reversed, planning[c] • WISC-V: Working Memory subtests[c] • NEPSY-II: Animal Sorting, Inhibition, Auditory Attention, Response Set[c] • HRNB: Category Test[c] • Tower of London[c] • D-KEFS[c] • Task of Executive Control[c] • Conners CPT 3[c] • Trail Making Test/Stroop Test[c] • TEA-Ch[c] • CAS2: Planning subtests[c] • KABC-II: Rover, Word Order, Pattern Reasoning, Story Completion[c]
Visual–motor functioning: D'Amato, Fletcher-Janzen, and Reynolds (2005); Fletcher et al. (2007); Hale and Fiorello (2004)	Classroom observations: tracing; copying from board or from paper on desk[a] • Bender Gestalt-II[c] • Beery VMI-6[c] • Rey–Osterrieth Complex Figure Test[c] • WISC-V: Visual Spatial subtests[c] • KABC-II: Triangles, Block Counting, Gestalt Closure[c] • DTVP-2[c] • MVPT-3[c] • BOT-2[c] • NEPSY-II: Arrows, Design Copying[c] • HRNB: Finger Tapping Test[c] • HRNB: Trails A[c]
	Math
Math computation: D'Amato, Fletcher-Janzen, and Reynolds (2005); Fletcher et al. (2007)	Classroom problems correct per minute[a] • aimsweb: Math Computation[a] • Review of assignments/homework[a] • KeyMath3: Basic Concepts, Operations[b] • KTEA-3: Math Computation, Math Fluency[b] • WJ IV ACH: Calculations[b] • WIAT-III: Numerical Operations[b] • WRAT4: Arithmetic[b] • WISC-V: Arithmetic[b]
Math problem solving: D'Amato, Fletcher-Janzen, and Reynolds (2005); Fletcher et al. (2007)	Classroom exercises correct per minute[a] • aimsweb: Math Concepts and Applications[a] • Review of assignments/homework[a] • WJ IV ACH: Applied Problems[b] • WIAT-III: Math Reasoning[b] • KTEA-3: Math Concepts and Applications[b] • KeyMath3: Applications[b] • WISC-V: Arithmetic[c]
	Additional neuropsychological areas to evaluate for math
Attention/executive functions	See this category for reading
Short-/long-term and working memory	See this category for reading

(continued)

TABLE 38.5. *(continued)*

Areas that should be evaluated, with relevant references	Evaluation tools
	Writing
Handwriting: D'Amato, Fletcher-Janzen, and Reynolds (2005); Fletcher et al. (2007)	Classroom work samples[a] • aimsweb: Written Expression CBM[a] • WJ IV ACH: Writing Fluency, Writing Samples[b] • KTEA-3: Written Expression, Writing Fluency[b] • WIAT-III: Alphabet Writing Fluency, Sentence Composition, Essay Composition[b] • TOWL-4[b] • OWLS-2[b] • NEPSY-II: Design Copying[c]
Spelling: Fletcher et al. (2007)	Classroom words correct per timing[a] • aimsweb: Spelling CBM[a] • KTEA-3: Spelling[b] • WJ IV ACH: Spelling[b] • WIAT-III: Spelling[b] • WRAT-IV: Spelling[b]
Written composition: Fletcher et al. (2007)	Classroom timed work sample[a] • aimsweb: Written Expression CBM[a] • WIAT-III: Essay Composition[b] • KTEA-3: Written Expression[b] • TOWL-4[b] • OWLS-II [b] • NEPSY-II: Design Copying[c]
	Additional neuropsychological areas to evaluate for writing
Attention/executive functions	See this category for reading
Short-/long-term and working memory	See this category for reading
Language	See this category for reading

Note. Abbreviations: DIBELS, Dynamic Indicators of Basic Early Literacy Skills; IGDI, Individual Growth and Development Indicators; CTOPP-2, Comprehensive Test of Phonological Processing—Second Edition; WJ IV COG, Woodcock–Johnson IV Tests of Cognitive Abilities; HRNB, Halstead–Reitan Neuropsychological Battery; NEPSY-II, NEPSY—Second Edition; WJ IV ACH, Woodcock–Johnson IV Tests of Achievement; WJ IV OL, Woodcock–Johnson IV Tests of Oral Language; WIAT-III, Wechsler Individual Achievement Test—Third Edition; WRAT4, Wide Range Achievement Test 4; GORT-5, Gray Oral Reading Tests, Fifth Edition; PPVT-4, Peabody Picture Vocabulary Test, Fourth Edition; WISC-V, Wechsler Intelligence Scale for Children—Fifth Edition; EOWPVT-4, Expressive One-Word Picture Vocabulary Test, Fourth Edition; ROWPVT- 4, Receptive One-Word Picture Vocabulary Test, Fourth Edition; CELF-5, Clinical Evaluation of Language Fundamentals—Fifth Edition; TOWL-4, Test of Written Language, Fourth Edition; Beery VMI-6, Beery–Buktenica Developmental Test of Visual–Motor Integration—Sixth Edition; Bender Gestalt-II, Bender Visual–Motor Gestalt Test, Second Edition; OWLS-2, Oral and Written Language Scales, Second Edition; TORC-4, Test of Reading Comprehension, Fourth Edition; KeyMath 3; BOT-2, Bruininks–Oseretsky Test of Motor Proficiency, Second Edition; MVPT-3,Motor-Free Visual Perception Test, Third Edition; DTVP-2, Developmental Test of Visual Perception, Second Edition; KABC-II, Kaufman Assessment Battery for Children—Second Edition; CREVT-3, Comprehensive Receptive and Expressive Vocabulary Test—Third Edition; TOWRE-2, Test of Word Reading Efficiency—Second Edition; DAS-II, Differential Ability Scales—Second Edition; D-KEFS, Delis–Kaplan Executive Function System; Conners CPT-3, Conners Continuous Performance Test 3rd Edition; TEA-Ch, Test of Everyday Attention for Children; CAS2, Cognitive Assessment System—Second Edition; KTEA-3, Kaufman Test of Educational Achievement, Third Edition.

[a] Can be measured via classroom-focused, curriculum-related assessments or the DIBELS (usually tier 1).

[b] Can be measured via standardized academic assessments (usually tier 2).

[c] Can be evaluated via psychological or neuropsychological examinations (usually tier 3).

riculum, the NB-MTSS model analyzes *why* skill weaknesses exist and evaluates the specific brain-based abilities that are needed for those specific skills. In the usual MTSS model, most summaries of student data (e.g., words pronounced correctly) are couched in terms of skill deficits, whereas the neuropsychological approach focuses on processing both strengths and weaknesses. It is important to ask how a student may best learn by using appropriate processing strengths with a focus on working around the observed processing weakness. A standard curriculum analysis does not offer information on the student's ability to generalize learning or to complete more inferential or abstract tasks, a potential problem as children grow older. Thus an NB-MTSS evaluation should consist of an assessment of specific neuropsychological processes (e.g., attention/executive functions, short-/long-term memory), as well as standardized information from the curriculum. During the initial weeks of school, screening should be completed for all children using data collected from the regular education curriculum. Most tier 1 students who display problems in learning will most likely display no neurological impairment (Lyon, Fletcher, & Barnes, 2003).

At the completion of screening, children's strengths and needs should be identified so that they can be grouped according to neuropsychological abilities. Targeted interventions can then be offered in reading groups with instructional measures that are related to reading abilities. Progress monitoring should take place every 2 weeks. If students continue to display significant difficulties after monitoring, teachers are asked to make instructional modifications and to document these reading interventions. At the completion of 6 weeks of monitoring, children who have been identified within the program as not making adequate progress should be identified and recommended for an NB-MTSS tier 2 service evaluation. Decisions concerning the need for additional services should be based on students' responses to additional EBIs (Lyon et al., 2003; Traughber & D'Amato, 2005).

A slightly different approach to providing services to students in tiers 1 and 2 has been detailed by Semrud-Clikeman (2005). Her model indicates that it is important to evaluate the needs of students so that the services offered are sure to succeed. Thus she has suggested that students screened at tier 1 who display extremely significant discrepancies (e.g., more than one standard deviation but less than two standard deviations below expectations) should move directly into a tier 2 intervention. Table 38.6 presents measures that can be used at all levels, in all areas, by all school personnel.

Tier 1 Services

We are proposing a three-tiered NB-MTSS model. In this model, tier 1 offers specialized neuropsychological screening at the beginning of first grade for all children. The majority of tier 1 screening takes place using classroom-based or curriculum-focused qualitative measures (e.g., phonemic awareness, vocabulary, reading comprehension). These measures are usually administered to students by classroom teachers. Specifically, a screening at this level based on a neuropsychological approach includes measures tapping phonological access and retrieval (see Table 38.6 for sample tasks). Another area supported by research that may need to be screened at this level (and is typically absent from traditional MTSS screening measures) is that of oral language skills (e.g., oral vocabulary, listening comprehension).

We recognize that research has not yet delineated how to match students' unique learning needs with appropriate levels of intervention. Some researchers have advocated for moving students to a level based on the magnitude of the students' impairment (e.g., two standard deviations below the mean), whereas others support determining intervention intensity based on students' outcomes following intervention (e.g., failure to demonstrate improvement following two interventions). However, research indicates that students who present with severe learning challenges (e.g., a suspected traumatic brain injury or cognitive disability) may warrant immediate referral to tier 3, which is the completion of a comprehensive neuropsychological evaluation.

Tier 2 Services

Individuals who have been identified at tier 1 but have not benefited from targeted interventions from the regular education teacher should be referred for a tier 2 evaluation. Tier 2 evaluations are conducted by the special education teacher or school psychologist, who works in concert with the regular education teacher to collect information from the student and classroom. Typically, a mixture of qualitative and quantitative data is collect-

TABLE 38.6. Sample Subtests from Six Neuropsychological and Psychological Measures

NEPSY-II

Attention and Executive Functioning
- Animal Sorting
- Auditory Attention and Response Set
- Clocks
- Design Fluency
- Inhibition
- Statue

Language
- Body Part Naming and Identification
- Comprehension of Instructions
- Oromotor Sequences
- Phonological Processing
- Repetition of Nonsense Words
- Speeded Naming
- Word Generation

Sensorimotor
- Fingertip Tapping
- Imitating Hand Positions
- Manual Motor Series
- Visuomotor Precision

Social Perception
- Affect Recognition
- Theory of Mind

Visuospatial
- Arrows
- Block Construction
- Design Copying
- Geometric Puzzles
- Picture Puzzles
- Route Finding

Memory and Learning
- List Memory
- List Memory Delayed
- Memory for Designs
- Memory for Designs Delayed
- Memory for Faces
- Memory for Faces Delayed
- Memory for Names
- Memory for Names Delayed
- Narrative Memory
- Sentence Repetition
- Word List Interference

Halstead–Reitan Neuropsychological Battery

Motor
- Finger Tapping Test
- Grip Strength Test
- Tactual Performance Test
- Marching Test

Abstract Reasoning
- Category Test
- Trails B

Language
- Reitan–Indiana Aphasia Screening Test

Alertness and Concentration
- Progressive Figures
- Speech Sounds Perception Test
- Rhythm Test

Sensory
- Fingertip Writing
- Finger Localization Test
- Tactile Perception Test
- Auditory Perception Test
- Visual Perception
- Tactile Form Recognition Test
- Sensory-Perceptual Examination

Visual
- Trails A

Visual–Spatial
- Matching V's and figures, concentration square, and star
- Target Test

Reason
- Color Form Test

Dean–Woodcock Sensory–Motor Battery (DWSMB)

Sensory Tests
- Lateral Preference Scale
- Near-Point Visual Acuity
- Visual Confrontation
- Naming Pictures of Objects
- Auditory Acuity

Tactile Examination
- Palm Writing
- Object Identification
- Finger Identification
- Simultaneous Localization (hands only and hand/cheek)

Motor Tests (Subcortical)
- Gait and Station
- Romberg Test
- Coordination Test (finger to nose and hand/thigh)

Motor Tests (Cortical)
- Construction Test (cross and clock)
- Mime Movements
- Left/Right Movements
- Finger Tapping
- Expressive Speech
- Grip Strength

(continued)

TABLE 38.6. *(continued)*

Kaufman Assessment Battery for Children—Second Edition (KABC-II)		
Learning/Glr • Atlantis • Rebus • Atlantis—Delayed • Rebus—Delayed *Sequential/Gsm* • Word Order • Number Recall • Hand Movements	*Simultaneous/Gv* • Triangles • Face Recognition • Conceptual Thinking • Pattern Recognition • Rover • Block Counting • Gestalt Closure	*Planning/Gf* • Pattern Reasoning • Story Completion *Knowledge/Gc* • Riddles • Expressive Vocabulary • Verbal Knowledge
Kaufman Test of Educational Achievement, Third Edition (KTEA-3)		
Reading • Letter and Word Recognition • Nonsense Word Decoding • Reading Comprehension • Reading Vocabulary *Reading Fluency* • Word Recognition Fluency • Decoding Fluency • Silent Reading Fluency	*Math* • Math Concepts and Applications • Math Computation • Math Fluency *Written Language* • Written Expression • Writing Fluency • Spelling	*Oral Language* • Listening Comprehension • Oral Expression • Associational Fluency *Language Processing* • Phonological Processing • Letter Naming Facility • Object Naming Facility
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)		
• Phonemic Awareness • Alphabetic Principle • Accuracy and Fluency with Connected Text • Reading Comprehension	• Vocabulary • Initial Sound Fluency • Letter Naming Fluency • Phoneme Segmentation Fluency • Nonsense Word Fluency	• Oral Reading Fluency • Retell Fluency • Word Use Fluency

ed again, with a focus on classroom-linked, curriculum-related measures. Tier 2 evaluations build on the areas that have been evaluated in tier 1. At tier 2, more neuropsychological data (see Table 38.6) will be collected to help understand why the student has not demonstrated improvement following specialized instructional interventions. Semrud-Clikeman (2005) has called for the evaluation of neuropsychological constructs such as working memory, attention, executive functioning, processing speed, and auditory processing ability. Screening children on neuropsychological variables that predict instructional success will help to identify those children who are at risk of not responding to an intervention later in the school year.

Tier 2 interventions will incorporate small-group EBIs tailored to the student's individual needs. Students not responding to the prescribed interventions will be assessed using more comprehensive measures of various neuropsychological constructs to determine why they are not responding to appropriate interventions. Again, any child performing more than one standard deviation below expectations should be closely monitored, and children scoring two standard deviations below expectations should be referred directly to tier 3 for a comprehensive neuropsychological evaluation.

Subsequent to the tier 3 evaluation, practitioners should be able to identify effective interventions for most students. Before this point, comprehensive neuropsychological assessment should generally be unnecessary. A neuropsychological evaluation does become appropriate, however, when school-related problems are not only resistant to intervention but also continue to defy standard analysis. When uncertainty regarding the nature of and solutions to individuals' problems persists, and little or no progress is made, a neuropsychological evaluation can generate unique hypotheses and help resolve the question of how best to intervene (D'Amato et al., 1999).

Tier 3 Services

Students can progress to a tier 3 evaluation in three different ways: (1) Students who have completed interventions in tiers 1 and 2 but have not made adequate progress are automatically referred to tier 3; (2) students displaying significant disabilities (e.g., cognitive problems, emotional problems, suspected autism) when assessed at tier 1 should be sent directly to tier 3; and (3) students who perform more than two standard deviations below expectations on the measures used to collect information about various neuropsychological constructs should also be referred to tier 3. At tier 3, students complete a wide-range neuropsychological evaluation as detailed in Table 38.6, as well as additional evaluations from other school team members (e.g., speech, physical, and/or occupational therapy) as needed. If a student qualifies for special education services, an individualized education program is developed, and services are provided to the student. If the student does not qualify for special education services after all evaluations are completed, the student moves back to tier 2 to receive additional evidence-based, curriculum-focused instructional interventions. The NB-MTSS model is unique in that neuropsychological knowledge, skills, and processes drive the special education service model and link neuropsychological processing to the classroom curriculum.

There is a serious, dramatic shortage of brain-driven EBIs that can be utilized in the schools. The global need for school psychologists to have additional training in neuropsychology has been repeatedly discussed within this chapter. This training will provide school psychologists with detailed knowledge of brain–learner–environment interactions, which will be instrumental in developing EBIs that are linked to the individual needs of students. School neuropsychology should no longer be viewed as a wave of the future. It is part of a new *Zeitgeist* in education and psychology that will only bring about positive changes in the educational enterprise and related fields that serve children, youth, teachers, and parents.

REFERENCES

Beaumont, J. G. (2008). *Introduction to neuropsychology* (2nd ed.). New York: Guilford Press.

Bray, M., & Maykel, C. (Eds.). (2016a). Holistic school psychology: Understanding the mind–body connection to address student concerns—Part I [Special issue]. *International Journal of School and Educational Psychology 4*(1), 1–96.

Bray, M., & Maykel, C. (Eds.). (2016b). Holistic school psychology: Understanding the mind–body connection to address student concerns—Part II [Special issue]. *International Journal of School and Educational Psychology, 4*(2), 97–142.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental process. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 992–1028). New York: Wiley.

Cortiella, C. (2003). *The discrepancy dilemma: Congress ponders a better way to identify learning disabilities*. Retrieved June 4, 2004, from *http://schwablearning.org/articles.asp?r = 703*

D'Amato, R. C. (1985, April). *Teacher assistance teams: A problem solving system that creates interventions*. Presented at the 17th annual convention of the National Association of School Psychologists, Las Vegas, NV.

D'Amato, R. C. (1990). A neuropsychological approach to school psychology. *School Psychology Quarterly*, *5*, 141–160.

D'Amato, R. C., Crepeau-Hobson, F. C., Huang, L. V., & Geil, M. (2005). Ecological neuropsychology: An alternative to the deficit model for conceptualizing and serving students with learning disabilities. *Neuropsychology Review, 15*(2), 97–103.

D'Amato, R. C., & Dean, R. S. (1989). The past, present, and future of nontraditional school psychology. In R. C. D'Amato & R. S. Dean (Eds.), *The school psychologist in nontraditional settings: Integrating clients, services, and settings* (pp. 185–209). Hillsdale, NJ: Erlbaum.

D'Amato, R. C., Dean, R. S., & Rhodes, R. L. (1998). Subtyping children's learning disabilities with neuropsychological, intellectual, and achievement measures. *International Journal of Neuroscience, 96*, 107–125.

D'Amato, R. C., Fletcher-Janzen, E., & Reynolds, C. R. (Eds.). (2005). *Handbook of school neuropsychology*. Hoboken, NJ: Wiley.

D'Amato, R. C., Gray, J. W., & Dean, R. S. (1988). Construct validity of the PPVT with neuropsychological, intellectual, and achievement measures. *Journal of Clinical Psychology, 44*, 934–939.

D'Amato, R. C., Hammons, P. A., Terminie, T. J., & Dean, R. S. (1992). Neuropsychological training in American Psychological Association-approved and non-approved school psychology programs. *Journal of School Psychology, 30*, 175–183.

D'Amato, R. C., & Hartlage, L. C. (Eds.). (2008). *Essentials of neuropsychological assessment: Rehabilitation planning for intervention* (2nd ed.). New York: Springer.

D'Amato, R. C., & Rothlisberg, B. A. (1996). How education should respond to students with traumatic brain injuries. *Journal of Learning Disabilities, 29*, 670–683.

D'Amato, R. C., Rothlisberg, B. A., & Leu Work, P. H. (1999). Neuropsychological assessment for intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (3rd ed., pp. 452–475). New York: Wiley.

D'Amato, R. C., & Wang, Y. Y. (2015a). Using a brain-based approach to collaborative teaching and learning with Asians. *New Directions in Teaching and Learning, 143,* 41–60.

D'Amato, R. C., & Wang, Y. Y. (2015b). Using an ecological clinical neuropsychology approach to understanding and intervening with men with learning disorders and attention deficit/hyperactivity disorders. In C. Zaroff & R. C. D'Amato (Eds.), *The neuropsychology of men: From neurodevelopment to evidence-based interventions* (pp. 85–110). New York: Springer.

Das, J. P., Kirby, J. R., & Jarman, R. F. (1975). Simultaneous and successive syntheses: An alternative model for cognitive abilities. *Psychological Bulletin, 82,* 87–103.

Davis, A. S. (2011). (Ed.). *Handbook of pediatric neuropsychology*. New York: Springer.

Davis, A. S., & D'Amato, R. C. (2005). Evaluating and using contemporary neuropsychological batteries: The NEPSY and the Dean–Woodcock Neuropsychological Assessment System. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 264–287). Hoboken, NJ: Wiley.

Davis, A. S., & Dean, R. S. (2005). Lateralization of cerebral functions and hemispheric specialization: Linking behavior, structure, and neuroimaging. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 120–142). Hoboken, NJ: Wiley.

Davis, A. S., Johnson, J., & D'Amato, R. C. (2005). Evaluating and using long standing neuropsychological batteries: The Halstead–Reitan and the Luria–Nebraska Neuropsychological Batteries. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 236–263). Hoboken, NJ: Wiley.

Dean, R. S., & Anderson, J. L. (1997). Lateralization of cerebral functions. In A. M. Horton, D. Wedding, & J. Webster (Eds.), *The neuropsychology handbook* (2nd ed., pp. 183–214). New York: Springer.

Epps, S., & Tindal, G. (1987). The effectiveness of differential programming in serving students with mild handicaps: Placement options and instructional programming. In M. C. Wang, M. Reynolds, & H. J. Walberg (Eds.), *Handbook of special education: Research and practice* (Vol. 1, pp. 213–248). New York: Pergamon Press.

Fischer, C. (2015). *Neuropsychological approach to treating comorbid Tourette syndrome in the school setting.* ProQuest. Retrieved from *http://gradworks.umi.com/37/18/3718848.html*

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford Press.

Fletcher-Janzen, E. (2005). The school neuropsychological examination. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 172–212). Hoboken, NJ: Wiley.

Gaddes, W. H. (1980). *Learning disabilities and brain function: A neuropsychological approach*. New York: Springer-Verlag.

Gaddes, W. H. (1985). *Learning disabilities and brain function: A neuropsychological approach* (2nd ed.). New York: Springer-Verlag.

Gaddes, W. H., & Edgell, D. (1994). *Learning disabilities and brain function: A neuropsychological approach* (3rd ed.). New York: Springer-Verlag.

Glozman, J. M. (1999). Quantitative and qualitative integration of Lurian procedures. *Neuropsychology Review, 9*(1), 23–32.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Hallahan, D. P., & Kauffman, J. M. (1991). *Exceptional children*. Englewood Cliffs, NJ: Prentice-Hall.

Hartlage, L. C. (1987). Neuropsychology: Definition and history. In L. C. Hartlage, M. J. Asken, & J. L. Hornsby (Eds.), *Essentials of neuropsychological assessment* (pp. 1–11). New York: Springer.

Hartlage, L. C., & D'Amato, R. C. (2008). Understanding the etiology of neurological and psychiatric disorders. In A. MacNeil Horton, Jr. & D. Wedding (Eds.), *The neuropsychology handbook* (3rd ed., pp. 87–108). New York: Springer.

Hartlage, L. C., & Telzrow, C. F. (1983). The neuropsychological basis of educational interventions. *Journal of Learning Disabilities, 16,* 521–528.

Hess, R. S., & D'Amato, R. C. (1998). Assessment of memory, learning, and special aptitudes. In A. S. Bellack & M. Hersen (Series Eds.) & C. R. Reynolds (Vol. Ed.), *Comprehensive clinical psychology encyclopedia: Vol. 3. Assessment* (pp. 239–265). Oxford, UK: Pergamon/Elsevier Science.

Hill, A. (2015). Social Stories™ as a neuropsychological and educational intervention with an adolescent diagnosed with an autism spectrum disorder. *ProQuest.* Retrieved from *http://gradworks.umi.com/37/15/3715073.html.*

Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), Pub. L. No. 108-446, 20 U.S.C. 1400 et seq. (2004).

Johnson, J. A., & D'Amato, R. C. (2011). Evaluating and using the Halstead–Reitan Neuropsychological Battery: Is it our future or our past? In A. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 353–366). New York: Springer.

Joseph, L. M. (2005). Understanding and implementing neuropsychologically based literacy interventions. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynold (Eds.), *Handbook of school neuropsychology* (pp. 738–757). New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement, Third Edition*. Bloomington, MN: Pearson.

Kazdin, A. E. (2011). *Single-case research design: Methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Pearson/Allyn & Bacon.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment manual*. San Antonio, TX: Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Clinical and interpretative manual*. San Antonio, TX: Harcourt Assessment.

Kostanaya, M. I., & Rossouw, P. (2013). Alexander Luria—life, research and contribution to neuroscience. *International Journal of Neuropsychotherapy, 1*(2), 47–55.

Koziol, L. F., & Budding, D. (2011). Pediatric neuropsychological assessment: Theoretical models of test interpretation. In A. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 443–455). New York: Springer.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D., et al. (2010). *Single case designs technical documentation*. Retrieved from *http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf*

Kratochwill, T. R., & Shernoff, E. S. (2003). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly, 18*(4), 389–408.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*(4), 341–389.

Kreutzer, J. S., DeLuca, J., & Caplan, B. (Eds.). (2011). *Encyclopedia of clinical neuropsychology*. New York: Springer.

Leu, P. W., & D'Amato, R. C. (1994). *Right children, wrong teachers?: Using an ecological assessment for placement decisions*. Poster presented at the 26th Annual Convention of the National Association of School Psychologists, Seattle, WA.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222*, 66–78.

Luria, A. R. (1973). *The working brain*. Harmondsworth, UK: Penguin.

Lyon, G. R., Fletcher, J. M., & Barnes, M. C. (2003). Learning disabilities. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 520–586). New York: Guilford Press.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and white elders. *Journal of the International Neuropsychological Society, 8*, 341–348.

Martens, B. K. (1992). The difference between a good theory and a good treatment is a matter of degree. *School Psychology Quarterly*, 7(2), 104–107.

Merry, S. N., McDowell, H. H., Hetrick, S. E., Bir, J. J., & Muller, N. (2004) Psychological and/or educational interventions for the prevention of depression in children and adolescents. *Cochrane Database of Systematic Reviews, 1*, CD003380.

Meyers, J. (2002). Reflections on the futures conference from the perspective of a school psychology educator. Retrieved June 5, 2004, from *http://www.indiana.edu/futures~/reflections_meyer.ppt.*

Morris, R. G. (1989). Synaptic plasticity and learning: selective impairment of learning rats and blockade of long-term potentiation *in vivo* by the *N*-methyl-D-aspartate receptor antagonist AP5. *Journal of Neuroscience, 9*, 3040–3057.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System—Second Edition*. Austin, TX: PRO-ED.

National Academy of Neuropsychology. (2001). Definition of a clinical neuropsychologist. Retrieved August 13, 2007, from *http://nanonline.org/paio/defneuropsych.shtm*

Ornstein, R., & Sobel, D. (1987). *The healing brain: Breakthrough discoveries about how the brain keeps us healthy*. New York: Simon & Schuster.

Pelletier, S. L., Hiemenz, J. R., & Shapiro, M. B. (2004). The application of neuropsychology in the schools should not be called school neuropsychology. *The School Psychologist, 58*, 17–24.

Power, E., D'Amato, R. C., & Eusebio, E. C. (2017). *Understanding and program planning for children diagnosed with the Zika virus*. Poster presented at the 125th Annual Convention of the American Psychological Association, Washington, DC.

Power, E., Hill, A., D'Amato, R. C., & Losoff, R. (2016). *Social Stories™ as a neuropsychological and educational intervention with an adolescent diagnosed with an Autism Spectrum Disorder.* Poster presented at the 124th Annual Convention of the American Psychological Association, Denver, CO.

Reynolds, C. R. (1981). Neuropsychological assessment and the habilitation of learning: Considerations in the search for the aptitude × treatment interaction. *School Psychology Review, 10(3)*, 343–394.

Reynolds, C. R., & French, C. L. (2005). The brain as a dynamic organ of information processing and learning. In R. C. D'Amato, E. Fletcher-Janzen, & C. R.

Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 86–119). Hoboken, NJ: Wiley.

Rhodes, R. L., D'Amato, R. C., & Rothlisberg, B. A. (2008). Utilizing a neuropsychological paradigm for understanding educational and psychological tests. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (3rd ed., pp. 321–348). New York: Plenum Press.

Riccio, C. A., Hynd, G. W., Cohen, M. J., & Gonzalez, J. J. (1993). Neurological basis of attention deficit hyperactivity disorder. *Exceptional Children, 60,* 118–124.

Riccio, C. A., & Wolfe, M. E. (2003). Neuropsychological perspectives on the assessment of children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 205–234). New York: Guilford Press.

Root, K. A., D'Amato, R. C., & Reynolds, C. R. (2005). Providing neurodevelopmental, collaborative, consultative, and crisis intervention school neuropsychology services. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 15–40). Hoboken, NJ: Wiley.

Rothlisberg, B. A., D'Amato, R. C., & Palencia, B. N. (2003). Assessment of children for intervention planning following traumatic brain injury. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (2nd ed., pp. 685–706). New York: Guilford Press.

Sands, R., Fischer, C. C., & D'Amato, R. C. (2016). *Neuropsychological approach to treating comorbid Tourette syndrome in the school setting.* Poster presented at the 124th Annual Convention of the American Psychological Association, Denver, CO.

Sattler, J. M., & D'Amato, R. C. (2002a). Brain injuries: Formal batteries and informal measures. In J. M. Sattler (Ed.), *Assessment of children: Behavioral and clinical applications* (4th ed., pp. 440–469). San Diego, CA: Jerome M. Sattler.

Sattler, J. M., & D'Amato, R. C. (2002b). Brain injuries: Theory and rehabilitation programs. In J. M. Sattler (Ed.), *Assessment of children: Behavioral and clinical applications* (4th ed., pp. 401–439). San Diego, CA: Jerome M. Sattler.

Semrud-Clikeman, M. (2005). Neuropsychological aspects for evaluating learning disabilities. *Journal of Learning Disabilities, 28*(6), 563–568.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level.* New York: Knopf.

Sheridan, S. M., & D'Amato, R. C. (2003). Partnering to chart our futures: *School Psychology Review* and *School Psychology Quarterly* combined issue on the multisite conference on the Future of School Psychology [Article published jointly in *School Psychology Quarterly* and *School Psychology Review*]. *School Psychology Review, 22*(4), 352–357.

Skierkiewicz, A. L. (2016). A neuropsychological study to reduce anxiety in a child exposed to trauma. *ProQuest.* Retrieved from *http://gradworks.umi.com/10/24/10241052.html*

Skierkiewicz, A. L., & D'Amato, R. C. (2017). *Reducing anxiety in a child exposed to sexual trauma.* Poster presented at the 125th Annual Convention of the American Psychological Association, Washington, DC.

Sousa, D. A. (2005). *How the brain learns to read.* Thousand Oaks, CA: Sage.

Sousa, D. A. (2006). *How the brain learns* (3rd ed.). Thousand Oaks, CA: Sage.

Stoner, G., & Green, S. K. (1992). Reconsidering the scientist-practitioner model for school psychology practice. *School Psychology Review, 2*(1), 155–166.

Taylor, H. G., & Fletcher, J. M. (1990). Neuropsychological assessment of children. In M. Hersen & G. Goldstein (Eds.), *Handbook of psychological assessment* (2nd ed., pp. 228–255). New York: Pergamon Press.

Teeter, P. A., & Semrud-Clikeman, M. (2007). *Child neuropsychology: Assessment and intervention for neurodevelopmental disorders* (2nd ed.). New York: Springer.

Titley, J. E., & D'Amato, R. C. (2008). Understanding and using the NEPSY-II Test for neuropsychological assessment for intervention. In L. C. Hartlage & R. C. D'Amato (Eds.), *Essentials of neuropsychological assessment: Rehabilitation planning for intervention* (2nd ed., pp. 149–172). New York: Springer.

Traughber, M. C., & D'Amato, R. C. (2005). Integrating evidence-based neuropsychological services into school settings: Issues and challenges for the future. In R. C. D'Amato, E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 827–857). Hoboken, NJ: Wiley.

U.S. Department of Education. (2016, April 27). The Office of Special Education and Rehabilitative Services celebrates 35 years of the Individuals with Disabilities Education Act (IDEA). Retrieved from *https://www2.ed.gov/about/offices/list/osers/idea35/index.html*

Walsh, K. W. (1978) *Neuropsychology: A clinical approach.* Edinburgh: Churchill Livingstone.

Whitten, J. C., D'Amato, R. C., & Chittooran, M. M. (1992). A neuropsychological approach to intervention. In R. C. D'Amato & B. A. Rothlisberg (Eds.), *Psychological perspectives on intervention: A case study approach to prescriptions for change* (pp. 112–136). New York: Longman.

Witmer, L. (1896). Practical work in psychology. *Pediatrics, 2,* 462–471.

Witsken, D., D'Amato, R. C., & Hartlage, L. C. (2008). Understanding the past, present, and future of clinical neuropsychology. In L. C. Hartlage & R. C. D'Amato (Eds.), *Essentials of neuropsychological assessment: Rehabilitation planning for intervention* (2nd ed., pp. 1–30). New York: Springer.

Witsken, D., Stoeckel, A., & D'Amato, R. C. (2008). Promoting educational change: Using a neuropsychological approach to develop a response-to-intervention program. *Psychology in the Schools, 45*(9), 781–798.

Ysseldyke, J. E. (1987). Classification of handicapped students. In M. C. Wang, M. Reynolds, & H. J. Walberg (Eds.), *Handbook of special education: Research and practice* (Vol. 1, pp. 253–271). New York: Pergamon Press.

Ysseldyke, J. E., Reynolds, M. C., & Weinberg, R. A. (1984). *School psychology: A blueprint for training and practice.* Minneapolis, MN: National School Psychology In-Service Training Network.

CHAPTER 39

Intellectual, Cognitive, and Neuropsychological Assessment in Three-Tiered Service Delivery Systems in Schools

George McCloskey
Jaime Slonim
Deanna Rumohr

From our perspective, intelligence test performance can be interpreted on multiple levels, with each successive level of interpretation representing a greater degree of aggregation and narrowing of the information available for interpretation. As illustrated in Figure 39.1, interpretive levels from bottom to top consist of the task-specific cognitive constructs involved in the performance of each item; the set of item scores of each subtest; the set of subtest scores; the set of specific composite scores; and the global composite score. Each aggregated level of interpretation within the framework masks clinical information that is potentially important to understanding a child's specific pattern of cognitive strengths and weaknesses. Additionally, we believe that two major conceptual perspectives influence contemporary intellectual assessment in the schools; we refer to these here as the *general abilities model* and the *cognitive neuropsychological model.* Each model emphasizes different levels of test interpretation. As shown in Figure 39.1, the general abilities model emphasizes interpretation of a global composite score and/or a set of specific composite scores (e.g., the Full Scale IQ [FSIQ] and the primary index scores of the Wechsler Intelligence Scale for Children—Fifth Edition [WISC-V]). In contrast, the cognitive neuropsychological model primarily emphasizes interpretation at the levels of subtests, items, and task-specific cognitive constructs, with attention to the specific composite level when such composites represent meaningful clinical clusters (Kamphaus, 2001; Reynolds & French, 2005). Meaningful clinical clusters may or may not be aligned with the composites that test developers identify through the use of factor-analytic techniques. In this chapter, the utility of these two models is discussed, with particular emphasis on the extent to which each model does or does not influence assessment at different service delivery tiers, and how each model can or cannot be integrated within the broader context of a three-tiered service delivery model.

The general abilities model is a traditional approach that has been in use since the inception of the intelligence test early in the 20th century (Jensen, 1998; Sarason, 1975). Proponents of the general abilities model often gravitate to the construct of *g* (Spearman, 1904) to explain the importance of assessing general ability (Deary, 2001; Gottfredson, 1997, 1998, 2008; Jensen, 1980, 1998; Kamphaus, Winsor, Rowe, & Kim, 2005; Kuncel, Hezlet, & Ones, 2004; McDermott & Glutting, 1997). Over time, *g* has come to be viewed as the quintessential indicator of overall intellectual ability. It is conceptualized as a stable trait that is

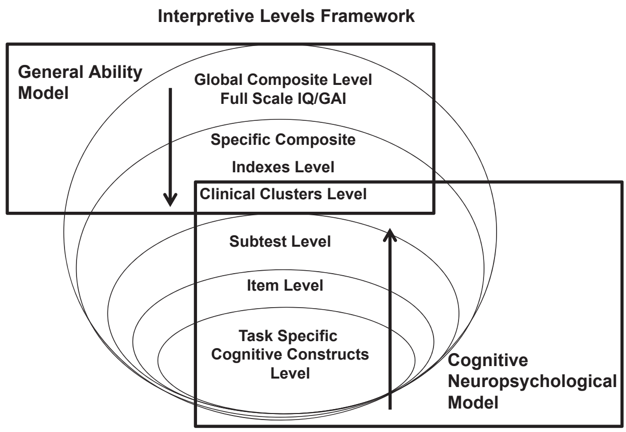


FIGURE 39.1. Interpretive levels coinciding with interpretive models.

not easily modified by educational or environmental interventions (Gottfredson, 1997, 1998, 2008; Jensen, 1998).

The general abilities model has guided the development of the intelligence tests currently available for use with school-age children, including the WISC-V (Wechsler, 2014), the Stanford–Binet Intelligence Scales, Fifth Edition (SB-5; Roid, 2003), the Differential Ability Scales—Second Edition (DAS-II; Elliott, 2007), the Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004), the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014), and several others. A general factor tends to emerge in analyses of subtest score data and is found in almost all intelligence tests that tap into a cognitively complex array of abilities (Saklofske, Prifitera, Weiss, Rolfhus, & Zhu, 2005). The conception of a learning disability as a condition wherein a student's intellectual capacity is significantly greater than the student's academic achievement (i.e., an ability–achievement discrepancy) stems directly from the rationale underlying the general abilities model (Bradley, Danielson, & Hallahan, 2002; Kavale, 2002; Sternberg, Grigorenko, & Bundy, 2001).

More recently, a cognitive neuropsychological perspective has emerged from the basic conceptions embodied in the fields of clinical neuropsychology and cognitive neuroscience. Cognitive neuroscience findings, including those gleaned from functional magnetic resonance imaging studies, have been applied to neuropsychology and have led to advances in the understanding of how children learn and produce academically (Berninger, 1994; Berninger & Richards, 2002; Posner & Rothbart, 2007). Several researchers and clinicians have applied this knowledge to reconceptualize the general methods used in assessment and intervention in educational settings, including approaches to intellectual assessment (Berninger, 1994; Berninger & Richards, 2002; Hale & Fiorello, 2004; Kaplan, Fein, Morris, Kramer, & Delis, 1999; Levine, 1998; Mapou & Spector, 1995; McCloskey, 2009a, 2009b; Miller, 2007; Pennington, 2009; Reynolds & Fletcher-Janzen, 1997; Sattler & D'Amato, 2002; Temple, 1997) and to generate specific cognitive neuropsychological models (Hale & Fiorello, 2004; Levine, 1998; McCloskey, 2009a; McCloskey & Perkins, 2012; McCloskey, Perkins, & VanDivner, 2009; McCloskey, Slonim, Whitaker, Kaufman, & Nagoshi, 2017).

The general abilities model has been articulated in this text and elsewhere (Gottfredson, 1997, 1998, 2008; Kamphaus et al., 2005) as a viable approach to conceptualizing intelligence and as a method of assessment with clinical utility within educational settings. The argument for the viability and clinical utility of a cognitive neuropsychological model has also been expressed in different sources (Allen, Hulac, & D'Amato, 2005; Berninger & Richards, 2002; Hale & Fiorello, 2004; Hale & Miller, 2008; Hale et al., 2008; Kaplan et al., 1999; McCloskey, 2009a, 2009b; McCloskey et al., 2017; Miller, 2007), but further refinements in the articulation of such a model are needed in order for the benefits of this approach to be fully recognized in educational settings. In this chapter, we outline and refine the cognitive neuropsychological model and contrast its use with the general abilities model within the context of the three-tier service delivery system currently used in U.S. schools.

A COGNITIVE NEUROPSYCHOLOGICAL MODEL OF ASSESSMENT

Within a neuropsychological model of cognition, brain function can be defined and assessed by using a set of psychological constructs. These cognitive constructs represent patterns of neural activation within various cortical and subcortical regions of the brain that are involved in the production of perception, thought, action, and the cognitive aspects of emotion (Berninger, 1994; Berninger & Richards, 2002; Dehaene, 2009, 2011; Levine, 1998; Mapou & Spector, 1995; McCloskey, 2009a; McCloskey et al., 2017; Willis, 2005). Assessment of cognition requires the understanding and use of the following specific cognitive constructs, the definitions of which are provided in Box 39.1: *processes*, *abilities*, *knowledge bases*, *skills*, *executive functions*, *strategies*, and *memory time frames of reference* (*initial registration*, *working memory*, and *long-term storage*). Additional terms defined in Box 39.1 that are critical for understanding the interaction of cognitive constructs include *sensory processing*, *motor functioning*, *processing speed*, *learning*, and *achievement*.

BOX 39.1. Operational Definitions within a Cognitive Neuropsychological Model of Assessment

Processes

Processes are narrow-band cognitive constructs responsible for the organization of input leading to the creation of basic mental representations. Processes thereby provide the basic elements of conscious thought used for academic learning and production, and serve as the springboard for academic skill development. Process deficits can impede learning and production, but often can be bypassed or compensated for (at least to some degree) because of their relatively restricted range of operation. For example, students who have difficulty perceiving and discriminating subword sound units can learn how to decode words and read, despite the phoneme discrimination difficulties that result from this highly specific auditory process deficit (Torgesen et al., 1999). In some instances, the effects of process deficits can be significantly reduced if the deficits are addressed during early developmental stages with a good intervention program (e.g., phonemic awareness training for difficulties with subword sound unit discrimination). In these cases, the process deficit may have been due more to underutilization or lack of maturation of intact neural networks than to the presence of damaged neural interconnections. In such cases, early instruction increases the frequency and effectiveness of the use of the process (McCloskey, 2009b). It is possible that basic process deficits resulting from damage to neural networks can be remediated through early childhood intervention as well (McCloskey, 2009b).

Severe basic process deficits can result in learning disabilities involving slowed and/or inconsistent learning and production. Basic processes underlying skill development in one or more academic domains include the following:

- Auditory perception
- Auditory discrimination
- Auditory attention
- Visual perception
- Visual discrimination
- Visual attention
- Kinesthetic perception
- Kinesthetic discrimination
- Kinesthetic attention

Processing

The term *processing* refers to neural activity that involves the coordinated use of one or more processes, almost always in conjunction with the accessing of one or more knowledge bases (described below) and typically under the direction of one or more executive functions (described below). Because of the manner in which processes are involved in creating mental representations, it is not possible to isolate and measure processes without some involvement from other cognitive constructs. *Processing* therefore refers not only to the integrated process of creating mental representations, but also to the tasks used to assess the effectiveness of processes during the act of creating mental representations.

(continued)

Basic Motor Functions

Basic motor functions are the fine motor capacities used in the performance of cognitive assessment tasks. All intellectual and cognitive assessment tasks require some form of motor output in order for the adequacy of task performance to be judged. Variations in the adequacy of motor functioning are important to observe and quantify, in order to understand how strengths or weaknesses in motor functioning may be enhancing or impeding learning of academic skills and production with academic tasks.

Processing Speed and Basic Motor Functioning Speed

The term *processing speed* refers to the speed with which one or more processes can be coordinated and applied in the formation of mental representations, often in conjunction with the accessing of knowledge bases and typically under the direction of one or more executive functions. For all processing speed assessment tasks, processing speed is combined with some form of basic motor function, in order to have an output that can be judged for adequacy of task performance. As a result, all measures of processing speed are also measures of basic motor functioning speed.

Abilities

Abilities are broad-band cognitive capacities that operate on the mental representations initially formed through cognitive processing. Abilities enable extended formulation and use of mental representations during learning and production. They include such integral constructs of thought as receptive and expressive language, complex visual–spatial representation and visualization, reasoning with language, reasoning with visual–spatial representations, reasoning with quantity, and idea generation.

Unlike process deficits, which can be bypassed or compensated for, ability deficits constrain learning and production. In other words, the degree of ability deficit places an upper limit on the quality (i.e., depth and complexity) of learning and production. Given the broad-based effects of ability deficits, compensatory or bypass strategies are typically not very effective in countering ability deficits within relatively short periods of time (McCloskey, 2009b). Severe ability deficits therefore result in cognitive impairments that greatly constrain learning and production, possibly throughout an individual's lifetime (e.g., severe language impairment, severe visual–spatial impairment, severe reasoning impairment, intellectual disability).

It should be noted that the word *ability* also often is used to denote the adequacy of a person's use of any or all cognitive constructs, as in "He has the ability to process phonemes," "She demonstrated the ability to retrieve information from long-term storage," or "He demonstrated the ability to decode words." As is clear from these examples, the use of the word in this sense indicates the individual's facility with the use of the specific cognitive constructs mentioned, rather than indicating the use of a specific cognitive construct that has been operationally defined as an ability. Use of the word *ability* as a term that can be applied to all cognitive constructs results in redundancy when one is discussing specific abilities, as in "He demonstrated the ability to use reasoning abilities."

Knowledge Bases

Knowledge bases are built up gradually through the storage of information during learning and skill acquisition. Once established, a knowledge base can be accessed for retrieval of information that can be used in conjunction with processes to form mental representations and to inform new learning or production. Knowledge bases can range from very basic, narrow forms of knowledge (e.g., the separate phonemes of the English language; how light strikes objects and creates shadows) to very complex forms of knowledge that vary greatly in depth and breadth (e.g., how to factor a polynomial; how to put a car engine [specific make and model] together; the influences of 18th-century classical composers on the early development of rock and roll).

Skills (Basic, Complex, Domain-Specific)

Skills are "knowledge bases under construction" that are acquired through formal or informal educational experiences. The term *skill* can be used in a temporal sense to represent what is being learned in the present

(continued)

moment, to represent what was learned in the past, or to represent what will be learned in the future. The set of skills to be taught often determines the content of instructional lessons. Skills can be further delineated by content as *basic*, *complex*, or *domain-specific*.

Basic Skills

Basic skills are the skills that form the foundation for all additional skill acquisition. The four broad basic skill domains are oral communication (listening and speaking), reading, writing, and mathematics. Each of these basic skill domains consists of many subdomains or subskills:

- Basic oral communication skills include reflective listening, diction and projection of voice, prosody, and rapid speech production.
- Basic reading skills include sight word recognition, phonological awareness, word decoding, and rapid word recognition.
- Basic writing skills include graphomotor letter, number and word formation and copying; word spelling; written sentence structure and formation; and rapid text production.
- Basic mathematics skills include computation procedures, basic quantity problem solutions, and rapid application of computation procedures.

Basic skills are the foci of instruction and learning in early elementary school. Basic skill learning represents the building of a set of general knowledge bases that will enable the application of oral communication, reading, writing, and mathematics to a wide range of subject content areas and to the learning of more complex skills. Skill building is an intermediate state between the immediate experiencing of new information and the retrieval of information from an established knowledge base. Skills that have been mastered form distinct knowledge bases. *Automaticity* refers to the speed with which basic skill knowledge bases can be accessed, and the ease with which information can be retrieved from the knowledge base and then applied.

Basic skill learning and use relies heavily on effective processing of auditory, visual, and kinesthetic stimuli taken in from the educational environment. Process deficits, therefore, can have a significant negative impact on skill acquisition and use.

Basic skill learning and acquisition also rely heavily on the use of multiple executive functions to cue and direct new learning and the construction of a new knowledge base. Basic skill learning, therefore, is likely to be disrupted when executive function deficits are present.

Complex Skills

Complex skills are oral communication, reading, writing, and mathematics skills that enable a person to take the mental representations formed through the use of basic skills, add new layers of representation, and manipulate all the information to produce relatively complex levels of meaning. Complex skills include extended listening and/or speaking for meaning, reading comprehension, extended written text generation, and applied mathematics problem solving.

Complex skill development and use involves the application of one or more basic skills integrated with the use of one or more abilities and the accessing of one or more knowledge bases, all under the direction of multiple executive functions. For example, the skill of complex reading comprehension requires the application of the basic skills of word recognition and/or decoding and reading rate to formulate an accurate basic mental representation of the information on the page. The more complex the grammatical structure of the material that was read, the greater the need for involvement of specific language abilities to enable meaningful representation at a deeper level. If the material being read relates to a specific topic, knowledge bases representing that person's knowledge of the topic will need to be accessed, along with language ability to provide a context for what was read. If the ideas represented by the words are complex, reasoning abilities will need to be engaged to obtain the highest level of meaning possible from the material. The application of these skills, knowledge bases, and abilities requires extensive use of executive functions to coordinate the multitasking that must take place during such complex reading comprehension. In addition to the use of multiple basic skills, knowledge bases, and abilities, application of complex skills very often requires the direction and use of working memory capacities (described below).

Complex skills are the foci of instruction and learning in the upper elementary grades.

(continued)

The development and use of complex skills can be disrupted by inadequate development of, or inadequate use of, basic skills (resulting from the effects of process deficits or lack of direct instruction); by insufficient storage of information in knowledge bases; by constraints imposed by inadequate or underdeveloped abilities; by constraints imposed by inadequate or underutilized working memory capacity; and/or by constraints imposed by inadequate or underutilized executive functions.

Domain-Specific Skills

Domain-specific skills are skills that are developed in specific subject domains and subdomains (e.g., the domain of science and the subdomains of biology, chemistry, and physics). Although basic and complex skills may be involved in learning in these content domains, the skills that are the foci of learning involve the building of knowledge bases related to the specific area of knowledge. The link between learning and knowledge base building in these content areas is apparent in the language used to denote course and learning objectives. Educators speak of increasing a student's knowledge of biology, rather than of increasing or building a student's biology skills. Despite the emphasis on specific content knowledge storage and retrieval, other knowledge bases are acquired (e.g., how to use biology or chemistry laboratory equipment) that are more readily perceived as skills.

Strategies

Strategies are learned and stored or newly generated routines that can be applied to increase an individual's efficiency of learning and/or production. Strategies are ways to chain together in a specific order a combination of processes, abilities, skills, and knowledge retrieved from knowledge bases to enhance learning and production. Strategies also involve the use of executive functions to sequentially or simultaneously cue and direct the use of the cognitive constructs involved in the strategy. Like skills, strategies can be taught and learned in formal or informal educational settings, or can be self-taught. Strategy development, storage, and use are all cued internally by executive functions or by an external mediating source, such as a teacher or a parent.

Executive Functions

Executive functions are a unique category of cognitive constructs defined by their directive role. Executive functions cue and direct the use of other cognitive constructs and coordinate multitasking efforts. They can be used to guide all cognitive constructs in all aspects of mental activity. They are not the processes, abilities, knowledge bases, skills, strategies, or memory states, but rather the cognitive constructs that orchestrate the use of all of these other cognitive constructs—cueing, directing, and coordinating all other elements of thought to produce coherent learning and production. Although executive functions are intricately involved in learning, the teaching process can mediate or substitute for students' inadequate use of executive functions. Executive function deficits are most noticeable in situations requiring independent or unsupervised production.

Memory: Time-Referenced States of Cognitive Functioning

Memory is very different from the other cognitive constructs discussed because different forms of memory actually represent different temporal states of mind in which a person may be using any number of processes, abilities, knowledge bases, skills, or strategies. Memory states are the mind's manifestations of time and space, providing the temporal and spatial contexts—a time signature—for all perception, emotion, cognition, and action.

The three major time-related memory states are these:

- Initial registration of information in the immediate moment: The experience of "now."
- Retrieval from long-term storage: Going back in time to recall previous "immediate" moments.
- Holding and manipulating information in mind: Extending the immediate moment into the future, projecting possible immediate moments into the future, or creating scenarios for future immediate moments.

These memory states enable a person to have a psychological sense of now (immediate

(continued)

memory or initial registration of stimuli), the past (retrieval from long-term storage), and the future (holding and manipulating mental representations beyond the immediate moment, or projecting mental activity various distances beyond the immediate moment).

Memory cannot be experienced or assessed outside the context of one or more processes, abilities, knowledge bases, skills, or strategies. What defines a memory state is what a person is doing at any given point in time.

A *memory deficit* is inadequate use of one or more cognitive constructs within a specific time frame. A person with poor immediate memory has difficulty effectively using processes and knowledge bases to initially register and briefly hold stimuli in mind. A person with long-term retrieval problems has difficulty accessing information stored in knowledge bases. A person with working memory problems has difficulty effectively applying abilities, knowledge bases, skills, or strategies to manipulate information being held in mind.

Although memory deficits are similar to ability deficits, in that both types can have a broad-based impact on learning and production, they are also similar to process deficits, in that memory state deficiencies often can be bypassed or compensated for (at least to some degree).

Lack of memory capacity can greatly obstruct learning and production:

- Poor initial registration constrains how much information can be represented in mind at one time.
- Poor retrieval capacity limits access to knowledge bases.
- Poor working memory capacity constrains how much information can be held in mind, how long that information can be held, and the extent to which the information being held in mind can be manipulated to enable extended states of learning, problem solving, and production.

Learning

Learning is the building of new knowledge bases. Whereas simple learning involves building a knowledge base through basic skill acquisition, complex learning occurs through the use of processes, skills, and strategies along with the accessing of knowledge bases and the application of abilities to create a new knowledge base or to link multiple knowledge bases for greater ease of strategy use. Learning can occur on a continuum from being mediated extensively by others to being self-mediated. The greater the self-mediation, the greater the demand for executive function involvement in the learning process.

Production

Production is the observable motor output or the unobservable mental output resulting from attempts to use processes, skills, and strategies; access knowledge bases; and apply abilities and executive functions to perform a task.

Achievement

Like production, *achievement* is the end result (the product) of the use of processes, skills and strategies; accessing of knowledge bases; and the application of abilities to perform a task. Assessment of achievement, therefore, often is not simply a matter of assessing skill development (basic, complex, or domain-specific). *Production* is typically referred to as *achievement* in formal educational settings, where production is directed at meeting specific academic goals related in some way to skill development.

Category Boundaries

It is important to note that the boundaries of these categories are somewhat amorphous and changeable. At least theoretically, processes, abilities, and executive functions can be taught and learned, thereby becoming skills. Skills can be stored, retrieved, and applied in the immediate moment, making them knowledge bases. The interrelated and overlapping nature of the category definitions, however, should not deter clinicians from making the important distinctions represented by each of these categories.

The cognitive neuropsychological model described in this chapter necessitates a more careful use of many terms that are frequently associated with intellectual assessment. Intellectual capacities have often been described in terms of *abilities*, *processes*, and *skills*, but these terms are typically used interchangeably in discussions of intelligence and in psychological reports describing the results of an intellectual assessment. In the model described herein, the terms ability, process, and skill have distinct, noninterchangeable definitions, and each represents a critical component of cognition. In addition to these three core terms, the role of the more recently introduced concept of *executive functions* is defined, and the more traditional but highly ambiguous constructs of *memory*, *achievement*, *learning*, and *strategy* (and the more generalized meaning of the term *ability*) are clarified. As explained in Box 39.1, effective application of tests of cognition requires not just an understanding of the distinctions among the various categories of cognitive constructs, but also an understanding of how the components of cognition interrelate during the occurrence of learning and production (Floyd, 2005; Saklofske et al., 2005). Figures 39.2–39.5 are conceptual diagrams applying the cognitive neuropsychological model to the primary academic domains of oral language, reading, writing, and math, respectively. These diagrams offer static representations of the cognitive constructs involved in listening/speaking, reading, writing and solving math problems. The brief narrative in Box 39.2 offers a description of the dynamic flow of information and the interplay of the cognitive constructs in Figure 39.3 as they are used in reading.

COMPARING THE TWO MODELS IN SCHOOL-BASED ASSESSMENT

In educational settings, the general abilities model is used to interpret intelligence test performance from the perspective of predicting later academic success (Gottfredson, 1997, 1998, 2008; Sternberg et al., 2001; Weinberg, 1989). From this model's perspective, intellectual assessments are conducted because the results can effectively predict academic achievement, and can therefore be used to identify learning disabilities by distinguishing a child who has the capacity to achieve academically from a child who does not have this capacity (Kavale, 2002; Mather & Wendling, 2005; Sternberg et al., 2001). We have been able to apply the general abilities model of interpretation to individ-

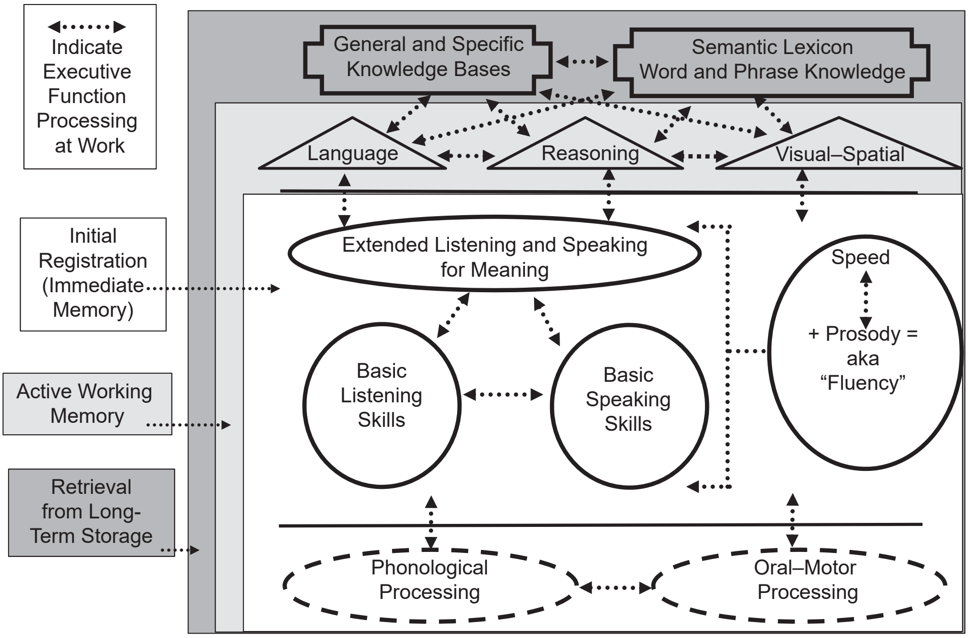


FIGURE 39.2. A cognitive neuropsychological model specifying the cognitive constructs used for listening and speaking. Copyright © 2010 George McCloskey. Reprinted by permission.

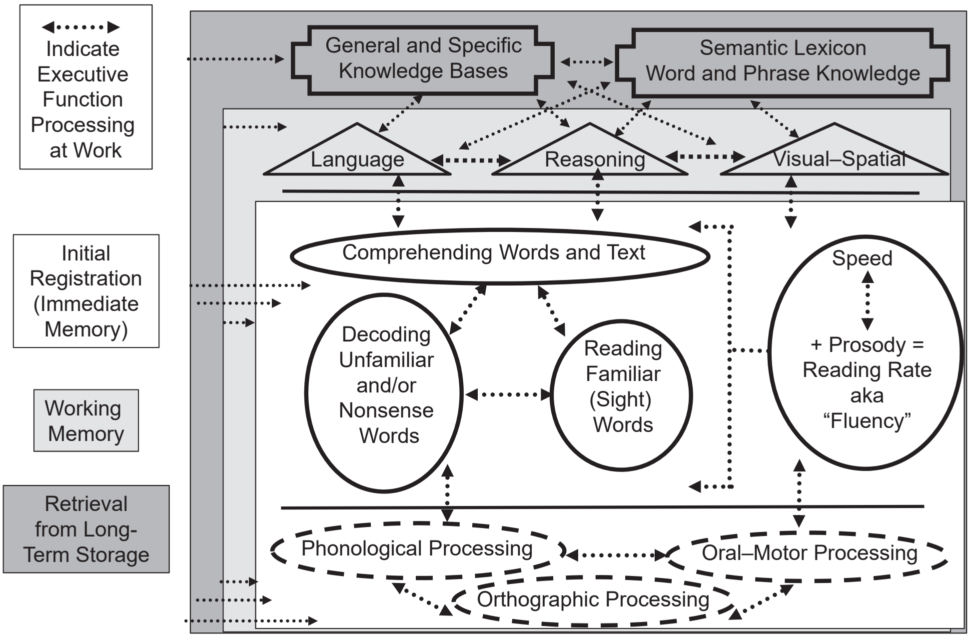


FIGURE 39.3. A cognitive neuropsychological model specifying the cognitive constructs used for reading. Copyright © 2010 George McCloskey. Reprinted by permission.

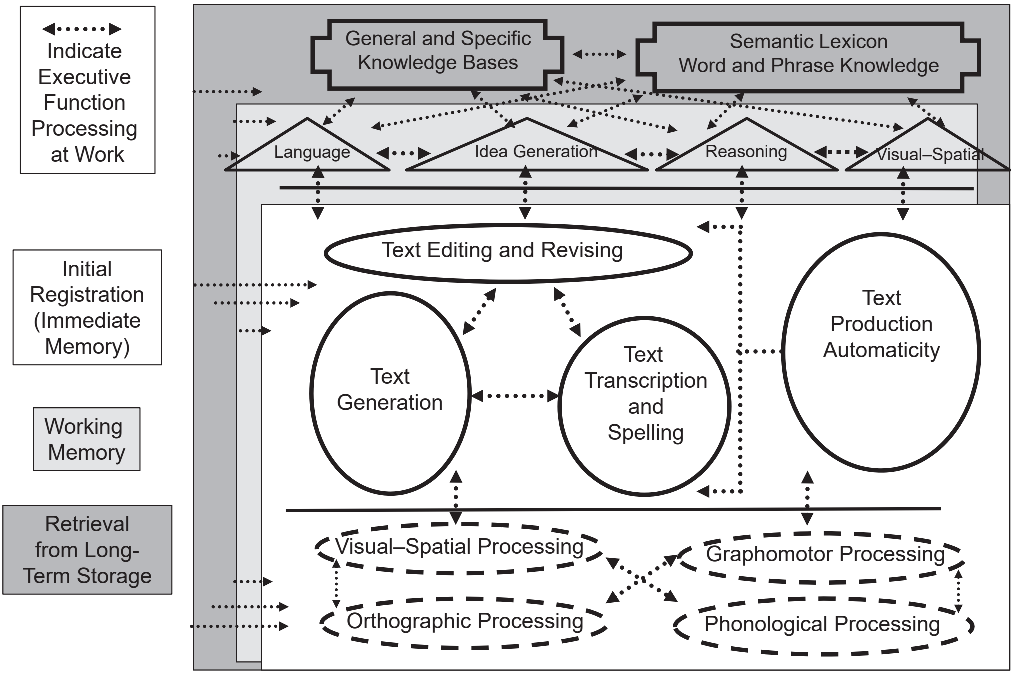


FIGURE 39.4. A cognitive neuropsychological model specifying the cognitive constructs used for writing. Copyright © 2010 George McCloskey. Reprinted by permission.

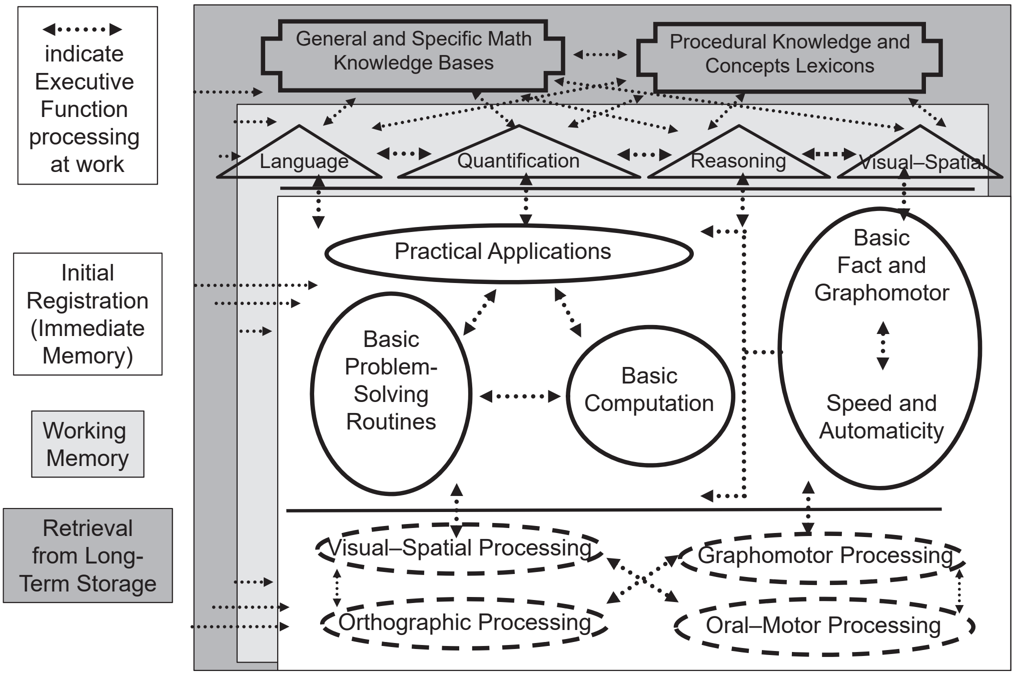


FIGURE 39.5. A cognitive neuropsychological model specifying the cognitive constructs used for doing math. Copyright © 2010 George McCloskey. Reprinted by permission.

ual case results to help convince parents, teachers, and administrators that particular students have more academic potential than previously thought, as indicated by the relatively high general ability scores obtained by these students. We also have been able to apply this model to convince parents, teachers, and administrators of the substantial educational challenges faced by children with limited cognitive resources, as indicated by these students' extremely low general ability scores.

The limitations of the traditional general abilities model become apparent, however, when professionals who administer intelligence tests are considerably more interested in actively helping a child improve his or her academic achievement than in passively reporting current levels of ability in order to predict future levels of academic achievement. When the purpose of assessment is reconceptualized in an effort to characterize a student's pattern of cognitive strengths and weaknesses, to understand how these strengths and weaknesses will affect learning and production, and to understand how instruction can be modified to improve learning and production, the general abilities model falls short. This model also falls short in light of research suggesting that children with learning disabilities are more likely to demonstrate varied than consistent cognitive capacity profiles at the subtest level (Fiorello, Hale, McGrath, Ryan, & Quinn, 2001; Fiorello et al., 2007; Flanagan, Alfonso, Ortiz, & Dynda, 2010; Fletcher-Jansen, 2005; Hale & Fiorello, 2004; Miller, 2007). In cases such as this, global composite level interpretation can mask clinically meaningful information and lead to inaccurate characterizations of cognitive constructs (Hale & Miller, 2008; Hale et al., 2008; Kaplan, 1988; Kaplan et al, 1999; McCloskey, 2009a, 2009b; McCloskey & Maerlender, 2005).

One thing that should be clear from Figures 39.2–39.5 and the discussion of the cognitive neuropsychological model applied to the various academic domains is that listening, speaking, reading, writing, and mathematical thinking are extremely complex in nature. Given the multifaceted nature of all these domains, it is not surprising that diagnostic assessment is a complicated endeavor that requires the use of specialized assessment instrumentation rather than the use of a single intelligence test (even one with multiple subtests) in order to identify the potential cognitive deficits underlying problems in each domain. In cases

BOX 39.2. The Dynamic Interplay of Cognitive Constructs during Reading

Reading does not start with the words on the page. Rather, reading starts in the brain with the accessing of knowledge bases containing what we already know about how to read—that is, our knowledge of how words sound when they are said and what they look like in print. In young children new to reading, these knowledge bases are built up through prereading instruction in phonemic awareness, orthographic awareness, and language experience activities (Berninger & Richards, 2002; Dehaene, 2009; Posner & Rothbart, 2007; Shaywitz, 2003; Uhry & Clark, 2005). Once we are engaged with words in print, these knowledge bases must be accessed and used in conjunction with sensory processes that enable us to see the words on the page (visual processes involved in perceiving and discriminating the shapes on the page in order to create visual mental representations of words) and to hear ourselves say the words (auditory processes involved in perceiving and discriminating the phonemes and syllables in order to create the pronunciations of the words on the page), as well as with basic motor routines that enable us to orally or subvocally say the words. If the orthographic and phonological awareness knowledge bases have been constructed, and the basic processes and motor capacities are in place, these can be used during the instructional process to learn and store new words (sight recognition skills), to analyze and sound out unfamiliar words (decoding skills), and to learn to perform both of these tasks quickly and efficiently (fluency skills). It is important to note that the goal of learning these skills is to enable readers to build part- and whole-word pronunciation knowledge bases, so that they can apply them independently to read any word they encounter in a text.

Quick and efficient recognition of the words in print, however, is only the first stage of reading. Additional cognitive constructs must be used to link the words being read with their individual meaning, and to link the meaning of each word with every other word in a sentence to grasp the meaning of each sentence (Berninger, 1994, 1998; Berninger & Richards, 2002). Turning words in print into meaning requires the use of language abilities, as well as the accessing of knowledge bases that store knowledge of the meanings of words and phrases, and knowledge bases that store knowledge about the topic that the sentence is addressing. Depending on what the words are attempting to communicate, extracting meaning from what is being read also may require the use of reasoning to understand more fully what is being read and/or visual–spatial abilities to “see” what is being read (Dehaene, 2009). Reading also involves the three time frames of reference in varying degrees. The words on the page must be initially registered in the immediate moment; knowledge of words in print and their meanings must be retrieved from long-term storage; and the longer the passage being read, the greater the need to hold and manipulate what is being read in working memory. Finally, all of this mental activity must be cued, directed, and integrated through the use of multiple executive functions working in a coordinated manner (Berninger & Richards, 2002; McCloskey & Perkins, 2012; McCloskey, Perkins, & Van Divner, 2009).

where production is poor within a particular academic domain, the assessment process must determine the extent to which each cognitive construct is capable of performing its role during functioning within that domain.

When contemporary intelligence tests are considered in light of the conceptual model presented in Box 39.1 and the context of diagnostic assessment, it is apparent that such tests consist of multiple tasks, each of which represents a complex amalgam of many of the constructs described in the cognitive neuropsychological model. At the same time, however, the ways in which the tasks are combined are not consistent with the cognitive models depicted in Figures 39.2–39.5. This lack of alignment with the academic domain models greatly reduces or completely nullifies their usefulness in the diagnostic process (Alfonso, Flanagan, & Radwan, 2005). The primary reason for this state of affairs is the fact that the development of contemporary intelligence tests has been dictated primarily by the general abilities model rather than by the cognitive neuropsychological model.

The statistical techniques (predominantly factor analysis) used to organize intelligence test tasks into composites for interpretation based on a general abilities model often obscure the distinctions necessary for effective understanding and interpretation of the cognitive constructs that an individual may use to complete these tasks (Flanagan & Kaufman, 2009; Hale & Miller, 2008; Hale et al., 2008; McCloskey, 2009a, 2009b; McCloskey & Maerlender, 2005; McCloskey et al., 2017). Table 39.1 illustrates the obfuscation created by the general abilities model: It examines the cognitive constructs required to perform tasks that have been aggregated to form a general ability composite. The inadequacy of the general abilities model for diagnostic purposes becomes even clearer when the general ability composites are disaggregated into their component subtests and the additional subtests within each composite domain (Watkins, Glutting, & Youngstrom, 2005). Table 39.2 aligns the subtests of multiple intelligence tests with the

TABLE 39.1. Analysis of the Cognitive Constructs That May Be Involved in the Performance of the Individual Items of Each Subtest within the Verbal Comprehension Index within the Full Scale IQ of the WISC-V

Interpretive level	Score/index/subtest/capacity			
Global composite	Full Scale IQ			
Specific composite	Verbal Comprehension Index			
Subtests	SI[a]	VC[b]	CO[c]	IN[d]
Items and task-specific cognitive constructs	Cognitive constructs	Cognitive constructs	Cognitive constructs	Cognitive constructs
Retrieval of verbal knowledge	XX	XXX	X	XXX
Reasoning with verbal content	XXX		XXX	
Auditory attention	X	X	X	X
Auditory discrimination	XX	XX	XX	XX
Auditory comprehension	X	X	XX	X
Auditory processing speed			XX	
Initial registration of auditorily presented information	X	X	X	X
Working memory	X		X	
Expressive language production	XX	XXX	XXX	XX
EF—Cueing appropriate consideration of the cognitive constructs and mental effort required to perform the task	XX	XX	XX	XX
EF—Directing auditory perception, discrimination, and comprehension	XX	XX	XX	XX
EF—Directing auditory attention	X	X	X	X
EF—Directing retrieval of verbal information	X	XX	X	XXX
EF—Directing reasoning with verbal information	XX		XX	
EF—Directing language expression	XX	XX	XX	X
EF—Directing flexible shifting of reasoning mindset	XXX		XX	
EF—Directing working memory	X		X	
EF—Recognizing and responding to prompts for more information	X	X	X	X
EF—Coordinating the use of multiple cognitive constructs simultaneously	XX	XX	XX	XX
EF—Cueing the inhibition of impulsive responding	X	X	X	X
EF—Cueing the focusing and sustaining of attention to auditory details	X	X	X	X

Note. XXX, primary construct targeted for assessment with the task; XX, secondary construct highly likely to be affecting task performance; X, secondary construct possibly affecting task performance; EF, executive function.

[a] Similarities.

[b] Vocabulary.

[c] Comprehension.

[d] Information.

TABLE 39.2. Alignment of Intellectual Assessment Measures with Academic Skill Domains within a Cognitive Neuropsychological Model

	Intellectual assessment measure				
Cognitive construct	WISC-V/WISC-V Integrated/WAIS-IV	SB5	WJ IV COG/OL	DAS-II	KABC-II
		Academic domain: Listening and speaking			
Processing					
Phonological processing			Phonological Processing Nonword Repetition Sound Blending Sound Awareness Segmentation		
Oral–motor functioning			Retrieval Fluency Rapid Picture Naming		
Knowledge bases					
Word knowledge	Vocabulary Vocabulary MC	Vocabulary	Oral Vocabulary Picture Vocabulary	Word Definitions	Riddles
General knowledge	Information Information MC		General information		Verbal Knowledge
Abilities					
Listening comprehension (receptive language)		Verbal Absurdities	Oral Comprehension Story Recall		Riddles
Oral expression (expressive language)	Vocabulary Comprehension Similarities	Vocabulary Picture Absurdities Verbal Absurdities/Analogies	Story Recall	Word Definitions Verbal Similarities	
Visual–spatial representation of language	Picture Vocabulary MC	Position and Direction	Picture Vocabulary		
Reasoning with verbal information	Similarities Similarities MC Comprehension Comprehension MC	Verbal Absurdities/Analogies		Verbal Similarities	Riddles
Reasoning with verbal/visual–spatial information					
Reasoning with verbal/quantitative information	Arithmetic	Quantitative Reasoning			

Executive functions, strategies, memory time frames of reference					
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing listening/speaking tasks					
Strategies applied while performing listening/speaking tasks					
Initial registration of phonology and words	Digit Span—Forward	Memory for Sentences/Last Word	Nonword Repetition Memory for Words		Number Recall
Holding and manipulating information while listening/speaking	Digit Span—Backward Digit Span—Sequencing Letter–Number Sequencing Sentence Recall	Memory for Sentences/Last Word	Verbal Attention Sentence Repetition Understanding Directions Memory for Words Numbers Reversed		Word Order
Retrieving information from long-term storage while listening/speaking	Vocabulary Vocabulary MC Information Information MC	Vocabulary	Retrieval Fluency Rapid Picture Naming		Verbal Knowledge
<u>Academic domain: Reading</u>					
Processing					
Phonological processing			Phonological Processing Nonword Repetition Sound Blending Sound Awareness Segmentation		
Orthographic processing			Letter–Pattern Matching		
Oral–motor functioning			Retrieval Fluency Rapid Picture Naming		
Knowledge bases					
Word knowledge	Vocabulary		Oral Vocabulary Picture Vocabulary	Word Definitions	Riddles
General knowledge	Information		General information		Verbal Knowledge

(continued)

TABLE 39.2. *(continued)*

	Intellectual assessment measure				
Cognitive construct	WISC-V/WISC-V Integrated/WAIS-IV	SB5	WJ IV COG/OL	DAS-II	KABC-II
Abilities					
Listening comprehension (receptive language)		Verbal Absurdities	Oral Comprehension Story Recall		Riddles
Oral expression (expressive language)	Vocabulary Comprehension Similarities	Vocabulary Picture Absurdities Verbal Absurdities/Analogies	Story Recall	Word Definitions Verbal Similarities	
Visual–spatial representation of language	Picture Vocabulary MC	Position and Direction	Picture Vocabulary		
Reasoning with verbal information	Similarities Similarities MC Comprehension Comprehension MC	Verbal Absurdities/Analogies		Verbal Similarities	Riddles
Reasoning with verbal/visual–spatial information					
Reasoning with verbal/quantitative information	Arithmetic	Quantitative Reasoning			
Executive functions, strategies, memory time frames of reference					
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing reading tasks					
Strategies applied while performing reading tasks					
Initial registration of phonology and/or orthography	Digit Span—Forward	Memory for Sentences/Last Word	Nonword Repetition Memory for Words		Number Recall
Holding and manipulating information while listening and/or while reading	Digit Span—Backward Digit Span—Sequencing Letter–Number Sequencing Sentence Recall	Memory for Sentences/Last Word	Verbal Attention Sentence Repetition Understanding Directions Memory for Words Numbers Reversed		Word Order
Retrieving information from long-term storage while listening/speaking and/or while reading	Vocabulary Vocabulary MC Information Information MC	Vocabulary	Retrieval Fluency Rapid Picture Naming		Verbal Knowledge

Academic domain: Writing					
Processing					
Phonological processing			Phonological Processing Nonword Repetition Sound Blending Sound Awareness Segmentation		
Orthographic processing			Letter–Pattern Matching		
Visual–spatial processing			Retrieval Fluency Rapid Picture Naming		
Graphomotor functioning	Coding				
Knowledge bases					
Word knowledge	Vocabulary		Oral Vocabulary Picture Vocabulary	Word Definitions	Riddles
General knowledge	Information		General information		Verbal Knowledge
Abilities					
Listening comprehension (receptive language)		Verbal Absurdities	Oral Comprehension Story Recall		Riddles
Oral expression (expressive language)	Vocabulary Comprehension Similarities	Vocabulary Picture Absurdities Verbal Absurdities/Analogies	Story Recall	Word Definitions Verbal Similarities	
Visual–spatial representation of language		Position and Direction			
Reasoning with verbal information	Similarities Similarities MC Comprehension Comprehension MC	Verbal Absurdities/ Analogies		Verbal Similarities	Riddles
Reasoning with verbal/visual–spatial information					
Reasoning with verbal/quantitative information	Arithmetic	Quantitative Reasoning			
Executive functions, strategies, memory time frames of reference					
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing writing tasks					

(continued)

TABLE 39.2. *(continued)*

	Intellectual assessment measure				
Cognitive construct	WISC-V/WISC-V Integrated/WAIS-IV	SB5	WJ IV COG/OL	DAS-II	KABC-II
Executive functions, strategies, memory time frames of reference *(continued)*					
Strategies applied while performing writing tasks					
Initial registration of phonology and/or orthography	Digit Span—Forward	Memory for Sentences/ Last Word	Nonword Repetition Memory for Words		Number Recall
Holding and manipulating information while listening, writing, or reading	Digit Span—Backward Digit Span—Sequencing Letter–Number Sequencing Sentence Recall	Memory for Sentences/ Last Word	Verbal Attention Sentence Repetition Understanding Directions Memory for Words Numbers Reversed		Word Order
Retrieving information from long-term storage while writing and reading	Vocabulary Vocabulary MC Information Information MC	Vocabulary	Retrieval Fluency Rapid Picture Naming		Verbal Knowledge
		Academic domain: Mathematics			
Processing					
Phonological processing					
Orthographic processing			Number–Pattern Matching		
Visual–spatial processing					
Oral–motor functioning					
Knowledge bases					
Word knowledge related to math					
General knowledge related to math					
Abilities					
Listening comprehension (receptive language) related to math tasks	Arithmetic	Verbal Absurdities			
Oral expression (expressive language) related to math tasks		Picture Absurdities Verbal Absurdities/Analogies			

Visual–spatial representation of nonverbal information	Visual Puzzles Block Design Block Design MC	Form Patterns	Visualization	Pattern Construction	Triangles Rover
Visual–spatial representation of verbal information		Position and Direction			
Visual–spatial representation of quantity	Quantitative Naming Speed				Block Counting
Reasoning with verbal/visual–spatial information		Verbal Absurdities/ Analogies			
Reasoning with verbal/quantitative information	Arithmetic	Picture Absurdities			
Reasoning with quantitative information	Figure Weights	Quantitative Reasoning	Number Series	Sequential and Quantitative Reasoning	
Reasoning with abstract visual information	Matrix Reasoning Block Design Block Design MC	Matrices	Concept Formation Analysis/ Synthesis	Matrices Pattern Construction	Pattern Reasoning Conceptual Thinking Rover
Executive functions, strategies, memory time frames of reference					
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing math tasks					
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing visual–spatial or quantitative tasks	Figure Weights		Planning		Rover
Strategies applied while performing math tasks					
Initial registration of phonology and/or orthography related to numbers	Digit Span—Forward Spatial Span—Forward				Number Recall
Holding and manipulating information while listening about and/or performing math-related tasks	Arithmetic Digit Span—Backward Digit Span—Sequencing Spatial Span—Backward	Block Span	Numbers Reversed Object–Number Sequencing		Word Order
Retrieving information from long-term storage while listening/speaking about and/or performing math tasks	Arithmetic	Quantitative Reasoning			

Note. WISC-V (Integrated), Wechsler Intelligence Scale for Children—Fifth Edition (Integrated); WAIS-IV, Wechsler Adult Intelligence Scale—Fourth Edition; SB5, Stanford–Binet Intelligence Scales, Fifth Edition; WJ IV COG/OL, Woodcock–Johnson IV Tests of Cognitive Abilities and Oral Language; DAS-II, Differential Ability Scales—Second Edition; KABC-II, Kaufman Assessment Battery for Children—Second Edition.

cognitive construct components of the cognitive neuropsychological model as applied to each of the basic academic domains. As reflected in Table 39.2, the most frequently used intelligence tests in school settings involve very few of the cognitive constructs related to performance in all four academic domains. Because the intelligence tests in use today have been based on a general abilities model rather than a cognitive neuropsychological model, they are not well suited to the diagnostic process at the heart of psychoeducational assessment in the schools, especially as it occurs at tier 3 (Mather & Wendling, 2005).

INTELLECTUAL ASSESSMENT IN THE CONTEXT OF THREE-TIERED SERVICE DELIVERY SYSTEMS

Three-tiered service delivery systems emphasize the need for appropriate instruction, assessment, and intervention at varying levels of intensity. Assessment at tier 1 typically involves the administration of brief progress-monitoring instruments at regular benchmark intervals to all students in general education. These brief assessments may be administered more frequently to students who do not make progress at the expected levels and/or rates. Students who continue to lag behind despite efforts to alter instruction are referred for tier 2 services. At tier 2, progress-monitoring assessment efforts continue, typically on a more frequent basis than at tier 1. In some instances, diagnostic assessments of some processes and skills are conducted to try to pinpoint a student's academic difficulties more specifically, and to identify interventions that might be more likely to enable the student to succeed. Students who continue to struggle at tier 2 for a prolonged period of time despite multiple efforts to alter instructional approaches are referred for a comprehensive assessment before either their assignment to instruction at tier 3 or their return to tier 2 services with a more specific plan for intervention efforts (Berninger, 1998; Berninger, O'Donnell, & Holdnack, 2008).

Although the results of the assessment may not lead to special education placement, a tier 3 assessment is accompanied by a host of requirements and stipulations associated with federal laws governing consideration of a student for special education placement (including receipt of written permission from the parents allowing an intellectual assessment to be conducted, and a specific time frame within which the assessment must be completed after permission has been received). The nature of a tier 3 referral makes it imperative that the student's specific cognitive strengths and weaknesses, and their impact on learning and production, be clearly specified. As noted previously, although an intellectual assessment guided by the general abilities model may be helpful in identifying overall intellectual capacity, the use of global and specific composites does not enable the level of specificity of cognitive strengths and weaknesses required in a tier 3 assessment.

The need for more specific information about a child's pattern of cognitive construct strengths and weaknesses associated with one or more specific academic domains makes the cognitive neuropsychological model a better fit for selection and interpretation of assessments at tier 3. As mentioned previously, however, contemporary intelligence tests are not well suited to the needs of a tier 3 assessment within the context of a cognitive neuropsychological model. This is because modern-day intelligence tests do not sufficiently assess many of the cognitive constructs involved in academic learning and production. Psychologists conducting assessments at tier 3 need to incorporate tasks from a broader array of cognitive test batteries in order to effectively assess all of the cognitive constructs associated with one or more academic domains (Berninger, Dunn, & Alper, 2005; Decker, 2008; Hale, Kaufman, Naglieri, & Kavale, 2006; Volker, Lopata, & Cook-Cottone, 2006). Table 39.3 provides examples of the types of tasks from other cognitive tests that could be incorporated into diagnostic assessment work at tier 3.

The utility of an intellectual assessment has been questioned, even at tier 3 (Reschly & Grimes, 2002). If psychologists persist in relying solely on a general abilities model for interpretation of test results, criticisms of the use of intellectual assessments at tier 3 are likely to continue because of the lack of relevance of the findings to the diagnostic process necessitated at tier 3. Application of a cognitive neuropsychological model at tier 3, however, is likely to steer psychologists away from the use of traditional intellectual assessments in favor of tests that offer more specific information about the cognitive constructs involved in academic learning and production. This will create the potential for more meaningful assessments, which in turn can improve intervention selection (Decker, 2008).

The critical differences between the general abilities model and the cognitive neuropsychologi-

TABLE 39.3. Alignment of Some Specific Cognitive Assessment Measures with the Academic Domain of Reading within a Cognitive Neuropsychological Model

	Intellectual assessment measure			
Cognitive construct	PAL-II Reading/ Writing	FAR	NEPSY-II	D-KEFS
Processing				
Phonological and morphological processing	Rhyming Phonemes Syllables Rimes Are They Related? Does it Fit?	Phonemic Awareness Positioning Sounds Morphological Processing	Phonological Processing Repetition of Nonsense Words	
Orthographic processing	Receptive Coding	Visual Perception Orthographical Processing		
Oral–motor functioning	RAN Letters RAN Words	RAN	Oromotor Sequencing Speeded Naming Part II	CWI Word Reading
Knowledge bases				
Word knowledge				
General knowledge				
Abilities				
Listening comprehension (receptive language)	VWM Sentences: Listening Notetaking A		Understanding Directions Narrative Memory	
Oral expression (expressive language)		Verbal Fluency	Narrative Memory	
Visual–spatial representation of language			Understanding Directions	
Reasoning with verbal information		Semantic Concepts		
Reasoning with verbal/visual–spatial Information			Animal Sorting	
Reasoning with verbal/ quantitative information				
Executive functions, strategies, memory time frames of reference				
Executive functions involved in cueing, directing, and coordinating cognitive capacities while performing reading tasks	Rapid Automatic Switching Sentence Sense		Auditory Attention and Response Set	CWI Inhibition CWI Inhibition/ Switching
Strategies applied while performing language or reading tasks				Twenty Questions
Initial registration of phonology and/or orthography	All phonological and orthographic processing tasks	All phonological and orthographic processing tasks		
Holding and manipulating information while listening and/ or while reading	VWM Letters VWM Words VWM Sentences	Word Recall	Understanding Directions Narrative Memory	
Retrieving information from long-term storage while listening/speaking and/or while reading	VWM Letters VWM Words	Verbal Fluency	Word Generation	Verbal Fluency

Note. PAL-II, Process Assessment of the Learner—Second Edition; FAR, Feifer Assessment of Reading; NEPSY-II, NEPSY—Second Edition; D-KEFS, Delis–Kaplan Executive Function System; RAN, Rapid Automatic Naming; CWI, Color–Word Interference; VWM, Verbal Working Memory.

cal model are most apparent when they are applied to the concept of learning disabilities. As shown in Figure 39.6, current models of learning disabilities—such as those discussed by Hale and Fiorello (2004) and by McCloskey (2009a)—look significantly different, depending on the model applied in the interpretation of intellectual assessment results. In the general abilities model, the contrast between a global composite score obtained from an intellectual assessment and scores from academic skill measures is central to the identification of a learning disability. In contrast, the cognitive neuropsychological model incorporates at most only a few subtest level scores or clinically meaningful composite cluster scores from an intellectual assessment and supplements these with cognitive measures of processing, abilities, knowledge bases, executive functions, memory time frame use, and academic skills measures based on the constellation of academic difficulties exhibited by the student (Berninger et al., 2005).

Applying the General Abilities Model in a Three-Tiered System

When the general abilities perspective is applied across all three levels of a three-tiered system, the focus is on predicting which students are likely to succeed and which are likely to fail. At tiers 1 and 2, the progress-monitoring devices serve as the predictors of the students' later competencies in the academic domains. The need for intellectual assessment at these tiers is typically not even considered, since the progress-monitoring devices are serving as the predictors during the early stages of service delivery. Even at tier 3, the need for an intellectual assessment is questioned when the purpose of such an assessment is the prediction of later academic success. At best, such an assessment at tier 3 offers an indicator of potential for later success from a longer, more reliable intellectual assessment source. In school systems where instruction at tiers 1 and 2 is not particularly strong, many students are referred for what is per-

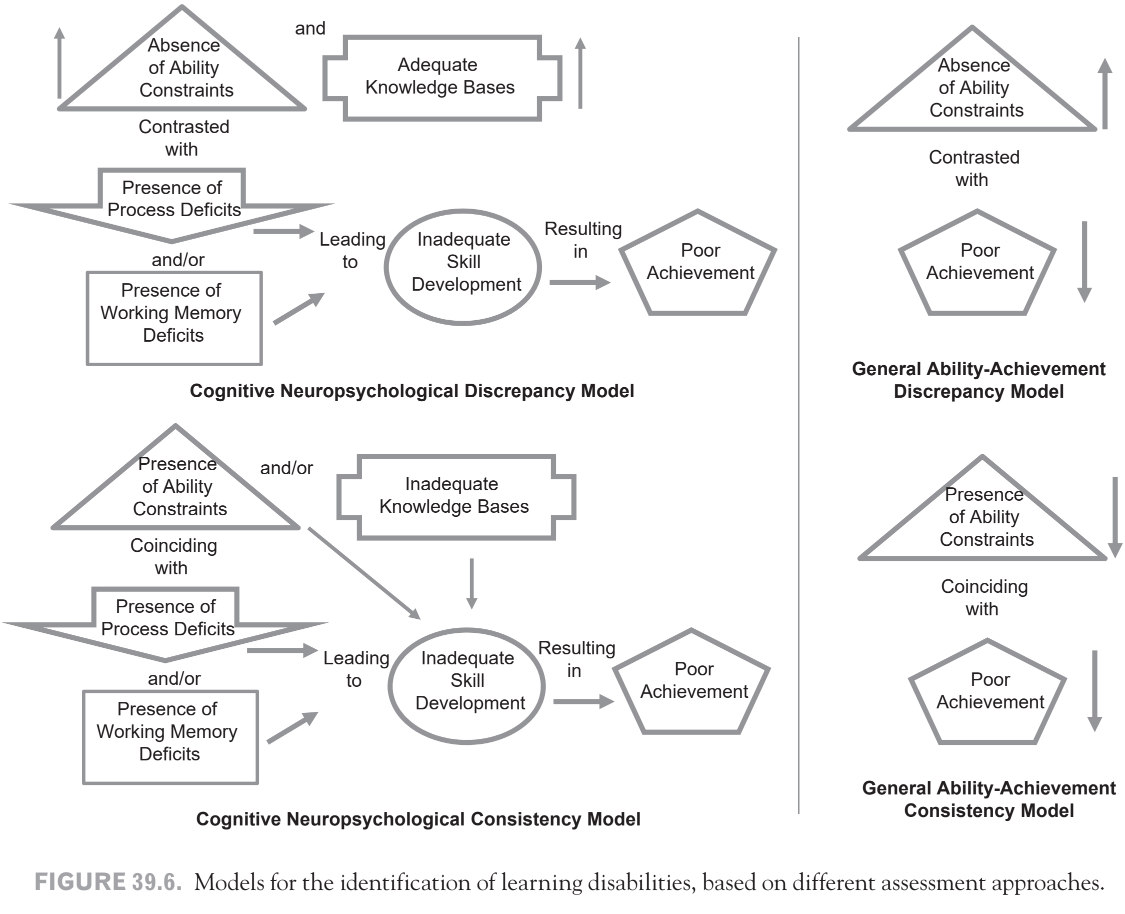


FIGURE 39.6. Models for the identification of learning disabilities, based on different assessment approaches.

ceived as a meaningless intellectual assessment at tier 3, simply because it is mandated as part of the process for consideration for placement in special education and for access to some types of tier 3 services. Whether those services are any better than the supports being provided at tiers 1 and 2 seems irrelevant, as the assessment process merely seems to function as a catalyst for moving a student out of the instructional environments at these tiers and into a tier 3 instructional environment. It should not be surprising that psychologists balk at being involved in such a process and bemoan their role as that of "gatekeepers for special education," with little or no relevance to the educational process other than to specify an arbitrary numerical cutoff for discrepancy between actual achievement and predicted achievement, on the basis of which it will be determined whether the student is or is not placed in a tier 3 special education program (Reschly & Grimes, 2002).

Applying the Cognitive Neuropsychological Model in a Three-Tier System

When a cognitive neuropsychological model is applied, the focus is on characterizing students' cognitive strengths and weaknesses, appreciating how these strengths and weaknesses affect learning and production, and understanding how instruction can be modified to improve learning and production. The need for this type of information is apparent at all three tiers. When assessments directly address these needs, parents, teachers, and administrators grasp the value of a comprehensive assessment for students who are struggling academically—not simply for placement in a tier 3 program, but for identifying instructional practices that are most likely to improve learning and production. Although the services of psychologists who competently use a cognitive neuropsychological model may be desired at all three tiers, it is neither possible nor practical to provide every struggling student with a comprehensive assessment. It may, however, be possible to administer to additional students at tier 2 a brief diagnostic battery that addresses most of the cognitive constructs in the cognitive neuropsychological models associated with the specific academic area of concern. This type of assessment would, at times, necessitate the involvement of psychologists in general education assessment below tier 3. The reason for this involvement rests with the need to administer some tasks, such as those involving reasoning with verbal information, that have traditionally been administered only by a psychologist in the context of a comprehensive intellectual assessment. Such diagnostic assessment activity could open the door for academic consultation services that would have an impact on a larger number of students, providing that the psychologist is well trained in diagnostic assessment and research-based instructional programs (Berninger et al., 2005).

The Contrast between the Two Models

With the emphasis on prediction of academic skill levels based on global intellectual assessment results, the general abilities model has little to offer school-based professionals who must identify a student's specific pattern of cognitive strengths and weaknesses and recommend interventions most likely to increase academic skill proficiency. Conversely, the cognitive neuropsychological model can offer a wealth of information about the extent to which a student can effectively process information during initial registration and create adequate mental representations; apply abilities and access knowledge bases; hold, manipulate, store, and/or retrieve mental representations; and use academic skills—all cued, directed, and coordinated by executive functions in order to demonstrate adequate academic production. How knowledge of these cognitive constructs can affect educational programming for students is illustrated by the six case profiles presented in Table 39.4. All of these elementary-school-age students were referred for evaluations by teachers or parents because of suspected reading difficulties. Comparing and contrasting the six cases provides some important insights related to assessment and intervention for reading problems and their relationship to what has traditionally been defined as intellectual ability:

1. The greater the number of process, ability, knowledge base, executive function, and memory time frame weaknesses identified, the greater the number of reading skill weaknesses, and the poorer the overall level of reading achievement.
2. Although students may be diagnosed with the same type of reading problem, their specific patterns of cognitive strengths and weaknesses may vary greatly, thereby necessitating different intervention plans.
3. The amount of time and energy invested, and the diversity of intervention techniques that

TABLE 39.4. Cognitive Neuropsychological Profiles of Six Students Referred by Teachers or Parents Due to Concerns about Reading

	Cognitive construct levels: S = strength, A = adequate, W = weakness					
Cognitive construct	Case 1: DPD	Case 2: DPD	Case 3: DPD	Case 4: AD	Case 5: AD/DPD	Case 6: AD/DPD
Processes/processing						
Phonological processing	W	W	W	A	W	W
Orthographic processing	A	A	A	A	A	W
Oral–motor functioning	A	A	A	A	A	W
Executive functions						
Executive functions involved in cueing, directing, and coordinating cognitive constructs while performing reading tasks	S	W	W	A	A	W
Knowledge bases						
Word knowledge	S	S	S	W	W	W
General knowledge	S	S	S	W	W	W
Abilities						
Listening comprehension (receptive language)	S	S	S	A	A	W
Oral expression (expressive language)	S	S	S	A	A	W
Reasoning with verbal information	S	S	S	W	W	W
Memory time frames of reference						
Initial registration of phonology (P) and/or orthography (O)	W/P	W/P	W/P	A	W/P	W/P&O
Holding and manipulating information while listening and/or while reading (working memory)	A	A	W	A	A	W
Retrieving information from long-term storage while listening/speaking and/or while reading	S	S	A	W/A	A	W
Strategies						
Strategies applied while performing reading tasks	S	W	W	A	W	W
Reading skills						
Word recognition	A	A/W	W/A	A	A	W
Decoding	W	W	W	A	W	W
Word recognition fluency	A	W	W	A	A	W
Word decoding fluency	W	W	W	A	W	W
Oral reading (passage) fluency	W	W	W	A	W	W
Comprehension	S	S	W/A	W	W	W
Silent reading comprehension/fluency	A	W	W	W	W	W
Reading achievement						
Grade-level group test	A	W/A	W	W	W	W
Grade-level state competency test	A	W/A	W	W	W	W

Note. DPD, developmental phonological dyslexia; AD, ability deficit; AD/DPD, ability deficit/developmental phonological dyslexia.

must be employed will vary in proportion to the number of cognitive construct weaknesses identified.

To illustrate these points, consider the first three cases provided in Table 39.4, all of which display the cognitive construct characteristics of *developmental phonological dyslexia* (DPD) (Berninger & Richards, 2002; Dehaene, 2009; Shaywitz, 2003; Temple, 1997; Uhry & Clark, 2005). Although the processing profiles of these students look similar, the case 1 student exhibits fewer reading skill and reading achievement weaknesses because she is able to effectively employ her well-developed executive functions, reasoning and language abilities, strategies, knowledge bases, and working memory time frame—all of which enable her to compensate for (but not completely eradicate) her word-level reading disability.

The case 2 student exhibits difficulties with more reading skills and inconsistent performance with reading achievement, stemming from the presence of additional weaknesses in the use of executive functions to consistently cue the use of strategies and skills. Note, for example, the additional weaknesses in word recognition fluency and silent reading/comprehension fluency that reflect an inability to balance speed and accuracy, resulting in a quick work pace that is countered by excessive word-reading error rates. Note also that the weakness in use of strategies is not due to an overall lack of knowledge of word- or sentence-level reading strategies, but rather to a failure to cue the use of these learned strategies when reading individual words and sentences (McCloskey et al., 2009; Meltzer, 2010; Swanson, 1993).

The case 3 student exhibits basic reading skill and executive function weaknesses similar to those of the case 2 student, but the additional weaknesses in the use of the working memory time frame experienced by this student are creating difficulties with the more complex skill of reading comprehension, and subsequently are resulting in poorer performance on measures of reading achievement (Berninger & Richards, 2002; Swanson, 1999, 2008).

The critical importance of assessing abilities and knowledge bases that constrain the act of reading—constructs typically classified as intellectual abilities—becomes more evident when the first three cases (involving only DPD) are contrasted with the next three cases, in all of which the students exhibit weaknesses in reasoning with verbal information and poor stores of word knowledge and general knowledge. For this reason, these students are designated in Table 36.4 as exhibiting an *ability deficit (AD)*. The absence of any processing deficits in the case 4 student, combined with adequate use of executive functions to direct basic word reading and effective use of the working memory time frame, have enabled this student to develop basic reading skills at the word level. At the same time, however, the student's ability and knowledge base weaknesses are constraining the development of reading comprehension skills at the sentence and passage levels, and are resulting in poor performance on measures of reading achievement.

The case 5 student exhibits ability and knowledge base weaknesses similar to those of the case 4 student, but these are compounded by a phonological processing weakness that is impacting the development of decoding skills and performance on fluency measures due to poor use of decoding skills in a manner similar to that of the case 1 student. Unlike that student, however, this student is exhibiting weaknesses with reading comprehension and poor performance on reading achievement measures, due to weaknesses in reasoning ability and word knowledge. Note that although this student has not stored an adequate amount of knowledge about words and topics related to school, he can adequately recall information that actually has been stored.

The case 6 student exhibits reasoning ability and knowledge base weaknesses similar to those of the students in cases 4 and 5, but these are joined by weaknesses in receptive and expressive language abilities, phonological and orthographic processing, and oral–motor functioning, as well as weaknesses in executive functions and inadequate use of all three time frames of reference. The consequent effect on reading is evident in the display of weaknesses for all reading skills and extremely poor performance on measures of reading achievement.

In terms of classification, recommendations for intervention, and outlook for improvement, the advantages of the use of a cognitive neuropsychological model over a general abilities model are unequivocal and numerous. The general abilities model would merely specify that an ability–achievement discrepancy exists for the first three students (those with DPD only), and that an ability–achievement consistency exists for the other three students (those with AD or AD/DPD). The cognitive neuropsychological model offers a richer context for understanding the nature of the reading problems in each of these six cases by specifying levels of performance with tasks involving the full array of cognitive constructs involved in the act of reading. As a result, each child's specific pattern of cognitive strengths and weaknesses can effectively be used to specify the nature and number of interventions required, as well as the intensity of the intervention efforts needed to improve reading skills.

In terms of intervention, the case 1 student represents the least degree of difficulty, as supplemental instruction in decoding skills will most likely be sufficient to address the reading skills deficits of this third-grade student (National Reading Panel [NRP], 2000; Shaywitz, 2003; Uhry & Clark, 2005). The case 2 student will also require supplemental instruction in decoding skills, but in addition, the executive function difficulties and failure to use learned strategies for word reading and comprehension will need to be addressed through guided practice in their use when the student is reading sentences and paragraphs. Gradually (likely over the course of 1 or more years), instruction will need to move from guided practice to self-regulated practice in using executive functions to cue and direct word reading and comprehension skill use. Intervention efforts with this third-grade student will be more challenging and will require more time than those for the case 1 student, and progress is likely to be slower (McCloskey et al., 2009; Swanson, 1993).

The case 3 student will benefit from supplemental interventions similar to those provided for the case 2 student, but will require additional instructional strategies to compensate for poor use of the working memory time frame (Swanson, 1999, 2008). Note that the interventions outlined briefly here relate to supplemental instruction and are not intended to be complete replacements for all elements of a balanced literacy curriculum. All three of these students will need to receive instruction focused on vocabulary, comprehension, and fluency development, but the cognitive strengths of these students should enable them to benefit from general education instruction related to these components of the reading curriculum (NRP, 2000). Because of the students' age, the phonological processing deficits they exhibit are not addressed directly. Instead, the decoding instruction provided reflects a compensation for weak phonological processing (Aylward et al., 2003).

The case 4 student (the one with AD only) represents a very different challenge in terms of intervention because the reading skill weaknesses demonstrated by this student do not stem from deficits in processing, executive functions, strategy use, or memory time frame use, but rather are results of the student's ability and knowledge base weaknesses. For the two students with AD/DPD (cases 5 and 6), interventions must take into account deficits in processing, executive functions, strategy use, and memory time frame use, and the associated specific reading skill deficits in a manner similar to that described for each of the three students with DPD (cases 1–3). Like case 4, cases 5 and 6 are much more challenging because these students exhibit concomitant deficits in reasoning ability, word knowledge, and general knowledge that are constraining reading comprehension (Berninger & Richards, 2002). Intervention efforts, therefore, will need to be greater in number and will require greater amounts of time devoted to remediation. Given the severity and number of deficits that each of these students exhibits, intensified intervention efforts are still likely to require long periods of time.

It is important to note that the cognitive construct deficits that distinguish the three students with AD or AD/DPD from the three students with DPD only are the cognitive constructs that have traditionally been associated with intellectual assessment (i.e., cognitive abilities and knowledge bases). The intervention challenges presented by the students in cases 4–6 are at the heart of a fundamental ideological debate about intelligence: Do intelligence test scores based on tasks involving reasoning with verbal information and descriptions of word meanings represent innate, immutable intellectual traits or acquired, malleable cognitive constructs? Some psychologists point out that such scores combine very different cognitive constructs that are better addressed individually (Hale & Fiorello, 2004; Hale & Miller, 2008; McCloskey, 2009a, 2009b; McCloskey et al., 2017). Others assert that vocabulary represents a crystallized knowledge base that can be increased through academic instruction or self-directed learning, but that reasoning with verbal information represents a more fluid ability and is much more difficult to alter through academic instruction (Flanagan & Kaufman, 2009; Lichtenberger & Kaufman, 2009; McGrew, 2000). Strict adherents of a general abilities model ignore the differences between the two tasks and espouse the view that the composite represented by the combination of these tasks represents a core of relatively innate immutable verbal ability, or *g* (Gottfredson, 1998; Jensen, 1973, 1998).

When the issue of intervention is raised, however, proponents of a general ability model and most adherents of a cognitive neuropsychological model tend to believe that tasks that involve reasoning with verbal information and describing the meaning of words are representative of innate, immutable traits. Regardless of which model guides the interpretation of intellectual assessments, reports we have reviewed typically do not provide specific recommendations for interventions focused on improving vocabulary knowledge and/or improving reasoning with verbal information. In other words, these constructs tend to be viewed as innate abilities rather than teachable skills. The tendency to view the tasks that are combined to form verbal ability composites as representative of unitary traits have led some to ignore or deny the role that reasoning with verbal information plays in the development of reading comprehension skills (Fletcher, Lyon, Fuchs, & Barnes, 2007).

Proponents of this viewpoint could correctly point out that no specific, well-researched intervention curriculum has been developed that enables educators to raise the reasoning ability of a student from two standard deviations below the mean to the mean within any reasonable amount of time. This fact fuels the argument that reasoning (whether innate or conditioned) should be discounted as an instructional variable, or even as a source of variability in skill performance, in teaching for acquisition of reading skills. Although no specific intervention program for quickly and dramatically improving reasoning deficits exists at this time, a number of instructional techniques and specific teaching exercises designed to increase reasoning with verbal information have been developed and have been used to improve the academic achievement of students in grades K–12 (Dean, Hubbell, Pitler, & Stone, 2013; Marzano, Pickering, & Pollock, 2001; Sternberg & Grigorenko, 2007). Likewise, teaching strategies and techniques have been developed and used with good results to improve students' vocabulary knowledge and subsequently to raise students' reading achievement levels (Dean et al., 2013; Marzano et al., 2001; NRP, 2000). What these instructional techniques are doing, in essence, is reinforcing the idea that a word knowledge lexicon can be built through academic instruction. Most importantly, such strategies serve to advance the idea that reasoning ability can be reframed as a teachable skill.

In order for improvement in reading comprehension skill acquisition and increased scores on reading achievement measures to occur, intervention efforts must attempt to address the reasoning and word knowledge deficits of the three students with AD (cases 4–6) depicted in Table 39.4, and of all other children with similar cognitive construct weaknesses who are currently enrolled in our K–12 schools. This discussion serves to highlight the most important distinction between the general abilities model and the cognitive neuropsychological model: The general abilities model locks the door of opportunity for further cognitive growth by perpetuating the belief that intelligence is an innate, immutable trait, whereas the cognitive neuropsychological model opens the door of opportunity for further cognitive growth by advancing the belief that intelligence represents a multifaceted set of cognitive constructs that are malleable and teachable given the appropriate investment of energy, time, and effort by all stakeholders—students, parents, and educators.

SUMMARY AND CONCLUSIONS

This chapter has described and contrasted two models that can be used to guide the use and interpretation of intellectual assessments. Although the general abilities model may be useful in specific situations where overall level of ability is a central factor in educational programming, the cognitive neuropsychological model holds much greater promise for fulfilling the important diagnostic role of identifying patterns of cognitive strengths and weaknesses and their relationships to intervention programming, especially for students with specific learning disabilities and AD. Additionally, the cognitive neuropsychological model appears much better equipped to address assessment needs in the context of the three-tiered service delivery systems that have become more prevalent in U.S. schools. The assessment demands in school settings necessitate a shift away from a narrow general abilities model focused on predicting achievement. Such a model avoids addressing or highlighting the specific cognitive constructs that may be constraining or impeding academic production; confines interpretation to one or a handful of composite scores; and limits perspectives on intervention by espousing a model that implies that intelligence is an innate, immutable trait.

Assessment practices in today's schools ought to encompass a much broader perspective. The cognitive neuropsychological model provides a means to accomplish this objective by enabling clinicians to assess a broad array of cognitive constructs, which can then be interpreted within the context of their role in learning and production in specific academic domains. Moreover, the model facilitates the linkage of assessment results to a broad array of intervention efforts. The need for a paradigm shift in assessment practices is reflected in the Learning Disabilities Association of America's white paper on learning disability identification (Hale et al., 2010), which advocates for the assessment and consideration of patterns of cognitive strengths and weaknesses in identification and treatment of learning disabilities.

Although contemporary intellectual assessments are not well suited to meeting these specific assessment needs in the schools, some of the requisite instrumentation is currently available through various specific cognitive tests of processing, abilities, knowledge bases, executive functions, memory time frames of reference, and academic skills. Ideally, the shift to a cognitive neuropsychological model would necessitate the development of new cognitive test batteries that directly address the broad array of cognitive constructs specifically involved in listening/speaking, reading, writing, and math.

Psychologists must make a choice about their role in the use of assessment instruments in the future. To effectively implement a cognitive neuropsychological model such as the one described in this chapter, psychologists working in the schools will need to have a thorough grasp of cutting-edge assessment instrumentation and intervention techniques in reading, writing, and math, in order to diagnose academic problems and to recommend interventions most likely to produce skill growth. Although the field of school psychology generally claims expertise in this academic skill domain (National Association of School Psychologists, 2006), many school psychologists acknowledge a lack of adequate knowledge in these areas, and some other professionals question whether psychologists really have the knowledge to be involved in such work (Kirby, 2009). If psychologists choose to continue emphasizing the use and interpretation of traditional intelligence tests through the lens of the general abilities model, and if they continue to approach AD and knowledge base deficits as immutable and irremediable traits, then they may find it increasingly difficult to break out of a meaningless test-and-place model of assessment. Indeed, they may find that their services are no longer needed, as they are being replaced by a cadre of other professionals competently trained in the use and interpretation of a host of processing, ability, and skill tests that are directly related to learning and production in academic domains but do not have the word *intelligence* in their titles.

Alternatively, psychologists can choose to expand their repertoire of assessment skills, incorporating cognitive measures of processing, skills, knowledge bases, abilities, executive functions, and memory time frames of reference with the occasional handful of intelligence test subtests to provide assessment results that can drive intervention efforts. They can become the leading proponents of a movement to change reasoning ability into a teachable skill, can develop expertise in academic intervention programs, and can attain mastery in linking assessment results with intervention practices. Concurrently, psychologists can encourage test publishers to increase their efforts to develop more germane cognitive neuropsychological test batteries—ones that directly relate to learning and production in the various academic domains.

Regardless of the choices made by psychologists working in the schools, we predict that the future of assessment belongs to cognitive assessment guided by a cognitive neuropsychological model, rather than to intellectual assessment guided by a general abilities model. If the next edition of this text is titled *Contemporary Cognitive Assessment*, then the accuracy of this prediction will be apparent.

REFERENCES

Alfonso, V., Flanagan, D., & Radwan, S. (2005). The impact of the Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 136–181). New York: Guilford Press.

Allen, T., Hulac, D., & D'Amato, R. C. (2005). The pediatric neurological examination and school neuropsychology. In R. C. D'Amato, E. Fletcher-Janzen, & C. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 145–171). Hoboken, NJ: Wiley.

Aylward, E. H., Richards, T. L., Berninger, V. W., Nagy, W. E., Field, K. M., Grimm, A.C., et al. (2003). Instructional treatment associated with changes in brain activation in children with dyslexia. *Neurology, 61,* 212–219.

Berninger, V. W. (1994). *Reading and writing acquisition: A developmental neuropsychological perspective*. Boulder, CO: Westview Press.

Berninger, V. W. (1998). *Process Assessment of the Learner: Intervention guide*. San Antonio, Texas: Psychological Corporation.

Berninger, V. W., Dunn, A., & Alper, T. (2005). Integrated multilevel model for branching assessment, instructional assessment, and profile assessment. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 151–188). Burlington, MA: Elsevier/Academic Press.

Berninger, V. W., O'Donnell, L., & Holdnack, J. (2008). Research supported differential diagnosis of specific learning disabilities and implications for intervention and response to intervention. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 69–110). Burlington, MA: Elsevier/Academic Press.

Berninger, V. W., & Richards, T. L. (2002). *Brain literacy for educators and psychologists*. Boston: Academic Press.

Bradley, R., Danielson, L., & Hallahan, D. P. (2002). *Identification of learning disabilities: Research to practice*. Mahwah, NJ: Erlbaum.

Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. J. (2013). *Classroom instruction that works: Research-based strategies for increasing student achievement* (2nd ed.). Arlington, VA: Association for Supervision and Curriculum Development.

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford, UK: Oxford University Press.

Decker, S. (2008). School neuropsychology consultation in neurodevelopmental disorders. *Psychology in the Schools, 45*(9), 799–811.

Dehaene, S. (2009). *Reading in the brain.* New York: Viking Press.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Elliott, C. D. (2007). *Differential Ability Scales—Second Edition: Administration and scoring manual.* San Antonio, TX: Harcourt Assessment.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence tests results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14,* 2–12.

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2001). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13,* 115–125.

Flanagan, D. P., Alfonso, V., Ortiz, S., & Dynda, A. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. Miller (Ed.), *Best practices in school neuropsychology* (pp. 101–140). Hoboken, NJ: Wiley.

Flanagan, D. P., & Kaufman, A. S. (2009). Introduction and overview. In D. P. Flanagan, A. S. Kaufman, *Essentials of WISC-IV assessment* (pp. 1–52). Hoboken, NJ: Wiley.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guilford Press.

Fletcher-Janzen, E. (2005). The school neuropsychological examination. In R. C. D'Amato, E. Fletcher-Janzen, & C. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 172–212). Hoboken, NJ: Wiley.

Floyd, R. (2005). Information-processing approaches to interpretation of contemporary intellectual assessment instruments. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 203–233). New York: Guilford Press.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents,* pp. 24–29.

Gottfredson, L. S. (2008). Of what value is intelligence? In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 545–564). Burlington, MA: Elsevier/Academic Press.

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33*(3), 223–236.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Hale, J. B., Fiorello, C. A., Miller, J. A., Wenrich, K., Teodori, A., & Henzel, J. N. (2008). WISC-IV interpretation for specific learning disabilities identification and intervention: A cognitive hypothesis testing approach. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 111–172). Burlington, MA: Elsevier/Academic Press.

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools, 43*(7), 753–770.

Hale, J. B., & Miller, D. C. (2008). Neuropsychological applications of the WISC-IV and WISC-IV Integrated. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 445–496). Burlington, MA: Elsevier/Academic Press.

Jensen, A. R. (1973). *Educability and group differences.* London: Methuen.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kamphaus, R. W. (2001). *Clinical assessment of children's intelligence.* Needham Heights, MA: Allyn & Bacon.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 23–38). New York: Guilford Press.

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain functions: Research,* *measurement, and practice* (pp. 125–167). Washington, DC: American Psychological Association.

Kaplan, E., Fein, D., Morris, R., Kramer, J. H., & Delis, D. C. (1999). *The WISC-III as a processing instrument*. San Antonio, TX: Psychological Corporation.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children—Second Edition*. Circle Pines, MN: American Guidance Service.

Kavale, K. A. (2002). Discrepancy models in the identification of learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 369–407). New York: Routledge.

Kirby, M. W. (2009). Why is the school psychologist involved in the evaluation of struggling readers? *Journal of Educational and Psychological Consultation, 19,* 248–258.

Kuncel, N., Hezlet, S., & Ones, D. (2004). Academic performance, career potential, creativity, and performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86,* 148–161.

Levine, M. (1998). *Developmental variation and learning disorders*. Cambridge, MA: Educators Publishing Service.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Hoboken, NJ: Wiley.

Mapou, R. L., & Spector, J. (1995). *Clinical neuropsychological assessment: A cognitive approach*. New York: Plenum Press.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Arlington, VA: Association for Supervision and Curriculum Development.

Mather, N., & Wendling, B. (2005). Linking cognitive assessment results to academic interventions for students with learning disabilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 269–294). New York: Guilford Press.

McCloskey, G. (2009a). Clinical applications I: A neuropsychological approach to interpretation of the WAIS-IV and the use of the WAIS-IV in learning disability assessments. In E. O. Lichtenberger & A. S. Kaufman, *Essentials of WAIS-IV assessment* (pp. 208–244). Hoboken, NJ: Wiley.

McCloskey, G. (2009b). The WISC-IV Integrated. In D. P. Flanagan & A. S. Kaufman, *Essentials of WISC-IV assessment* (2nd ed., pp. 310–467). Hoboken, NJ: Wiley.

McCloskey, G., & Maerlender, A. (2005). The WISC-IV integrated. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 101–149). Burlington, MA: Elsevier/Academic Press.

McCloskey, G., & Perkins, L. A. (2012). *Essentials of executive functions assessment*. Hoboken, NJ: Wiley.

McCloskey, G., Perkins, L. A., & VanDivner, B. (2009). *Assessment and intervention for executive function difficulties*. New York: Routledge.

McCloskey, G., Slonim, J., Whitaker, R., Kaufman, S., & Nagoshi, N. (2017). A neuropsychological approach to interpretation of the WISC-V. In D. P. Flanagan & V. C. Alfonso, *Essentials of WISC-V assessment* (pp. 287–404). New York: Wiley.

McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—or, more illusions of meaning? *School Psychology Review, 26,* 163–175.

McGrew, K. S. (2000). *Clinical interpretation of the Woodcock–Johnson Tests of Cognitive Ability*. Boston: Allyn & Bacon.

Meltzer, L. (2010). *Promoting executive function in the classroom*. New York: Guilford Press.

Miller, D. C. (2007). *Essentials of school neuropsychological assessment*. Hoboken, NJ: Wiley.

National Association of School Psychologists. (2006). *School psychology: A blueprint for training and practice III*. Bethesda, MD: Author.

National Reading Panel (NRP). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development.

Pennington, B. F. (2009). *Diagnosing learning disorders: A neuropsychological framework* (2nd ed.). New York: Guilford Press.

Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain*. Washington, DC: American Psychological Association.

Reschly, D. J., & Grimes, J. P. (2002). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1337–1350). Bethesda, MD: National Association of School Psychologists.

Reynolds, C. R., & Fletcher-Janzen, E. (1997). *The handbook of clinical child neuropsychology* (2nd ed.). New York: Kluwer Academic/Plenum.

Reynolds, C. R., & French, C. L. (2005). The brain as a dynamic organ of information processing and learning. In R. C. D'Amato, E. Fletcher-Janzen, & C. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 86–119). Hoboken, NJ: Wiley.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition: Examiner's manual*. Itasca, IL: Riverside.

Saklofske, D., Prifitera, A., Weiss, L., Rolfhus, E., & Zhu, J. (2005). Clinical interpretation of the WISC-IV FSIQ and GAI. In A. Prifitera, D. Saklofske, & L. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 36–71). Burlington, MA: Elsevier/Academic Press.

Sarason, S. (1975). The unfortunate fate of Alfred Binet and school psychology. *Teachers College Record, 77,* 579–592.

Sattler, J. M., & D'Amato, R. C. (2002). Brain injuries: Theory and rehabilitation programs, In J. Sattler (Ed.), *Assessment of children: Behavioral and clinical applications* (4th ed., pp. 431–439). San Diego, CA: Jerome M. Sattler.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level.* New York: Knopf.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Sternberg, R. J., & Grigorenko, E. L. (2007). *Teaching for successful intelligence to increase student learning and achievement* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill–Palmer Quarterly, 47*(1), 1–41.

Swanson, H. L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology, 56*(1), 87–114.

Swanson, H. L. (1999). Reading comprehension and working memory in learning-disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology, 72*(1), 1–31.

Swanson, H. L. (2008). Neuroscience and RTI: A complementary role. In E. Fletcher-Janzen & C. R. Reynolds (Eds.), *Neuropsychological perspectives on learning disabilities in the era of RTI: Recommendations for diagnosis and intervention* (pp. 28–53). Hoboken, NJ: Wiley.

Temple, C. (1997). *Developmental cognitive neuropsychology.* Hove, UK: Psychology Press.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 1–15.

Uhry, J. K., & Clark, D. B. (2005). *Dyslexia: Theory and practice of instruction* (3rd ed.). Baltimore: York Press.

Volker, M. A., Lopata, C., & Cook-Cottone, C. (2006). Assessment of children with intellectual giftedness and reading disabilities. *Psychology in the Schools, 43*(8), 855–869.

Watkins, M., Glutting, J., & Youngstrom, E. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 251–268). New York: Guilford Press.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition.* Bloomington, MN: Pearson.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44*(2), 98–104.

Willis, W. (2005). Foundations of developmental neuroanatomy. In R. C. D'Amato, E. Fletcher-Janzen, & C. Reynolds (Eds.), *Handbook of school neuropsychology* (pp. 41–60). Hoboken, NJ: Wiley.

Author Index

Subject Index

Note. *f*, *n*, or *t* following a page number indicates a figure, a note, or a table.