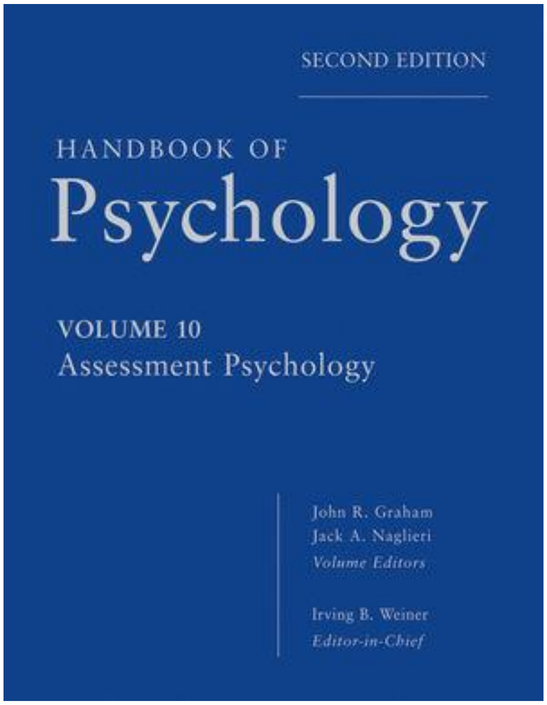

To cite this chapter:

Wasserman, J. D. (2013). Assessment of intellectual functioning. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.), *Handbook of psychology*. *Assessment psychology* (2nd ed., Vol. 10, pp. 451-501). Hoboken, NJ: John Wiley & Sons.

CHAPTER 18

Assessment of Intellectual Functioning

JOHN D. WASSERMAN

> 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, p. 146)

The study of intelligence and cognitive abilities dates back more than a century and has been characterized by some of the best and worst of aspects of science—the development of new methodologies, research breakthroughs, and vigorous scholarly debates as well as bitter rivalries, allegations of academic fraud, and the birth of a commercial testing industry that generates hundreds of million dollars in annual revenue. The assessment of intelligence can understandably elicit strong individual reactions, such as that expressed by progressive journalist Walter Lippmann as part of an exchange with Stanford-Binet author Lewis M. Terman in a prescient series of *The New Republic* magazine articles in the 1920s (Lippmann, 1922a–f, 1923). Intelligence testing gave rise to divisive public controversies at regular intervals throughout the 20th century, most recently after the 1994 publication of *The Bell Curve* by Richard J. Herrnstein and Charles Murray. Even with its more controversial aspects, however, intelligence remains a robust and important scientific construct. As of this writing, the American Psychological Association (APA) database PSYCInfo reports nearly 50,000 scholarly publications with *intelligence* as a keyword, 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).

This chapter describes contemporary approaches to the assessment of cognitive and intellectual functioning with an emphasis on omnibus intelligence tests. Seven major intelligence tests are presented in terms of their history, theoretical underpinnings, standardization features and psychometric adequacy, and interpretive indices and applications. These tests include the Cognitive Assessment System (CAS; Naglieri & Das, 1997a, 1997b), Differential Ability Scales (DAS-II; C. D. Elliott, 2007a, 2007b, 2007c), Kaufman Assessment Battery for Children (KABC-II; Kaufman & Kaufman, 2004), Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003), Stanford-Binet Intelligence Scales (SB5; Roid, 2003a, 2003b, 2003c), Wechsler Intelligence Scale for Children and Wechsler Adult Intelligence Scale (WISC-IV and WAIS-IV; Wechsler, 2003a, 2003b, 2008a, 2008b), and Woodcock-Johnson Tests of Cognitive Abilities (WJ III NU Cog; Woodcock, McGrew, & Mather, 2001a, 2007a). The most common diagnostic applications for intelligence testing are provided. Finally, the status of intelligence assessment as a maturing clinical science is assessed.

DESCRIPTIONS OF THE MAJOR INTELLIGENCE TESTS

This section presents seven of the leading individually administered intelligence tests, along with brief reviews and critical evaluations. The descriptions are limited to intelligence tests that purport to be reasonably comprehensive and multidimensional, covering a variety of content areas; more specialized tests (such as nonverbal cognitive batteries) and group-administered tests (usually administered in large-scale educational testing programs) have been excluded. Students of intellectual assessment will notice considerable overlap and redundancy between many of these instruments, in large part because they tend to measure similar psychological constructs with similar procedures, and in many cases they have similar origins.

The tests are presented in alphabetical order. For each test, its history is briefly recounted followed by a description of its theoretical underpinnings. Basic psychometric features, including characteristics of standardization, reliability, and validity are presented. Core interpretive indices are also described in a way that is generally commensurate with descriptions provided in the test manuals. Emphasis is placed on the interpretive indices that are central to the test, but not the plethora of indices that are available for some tests. Applications, strengths, and limitations of each test are discussed.

Some generalizations may be mentioned at the outset. First, it is apparent that contemporary intelligence measures are much more similar than different. While each instrument has some characteristic limitations, all of them are fairly adequate from a psychometric point of view. They all have satisfactory normative samples, and their composite scales tend to meet at least minimal standards of measurement precision. Second, while a few eschew Spearman's general factor of intelligence (psychometric *g*), most of them end up yielding good overall estimates of *g*. Third, while several different structural models are presented, there is considerable overlap in the constructs being tapped, epitomized by Kaufman and Kaufman's (2004) acknowledgment that their scales can be validly interpreted according to multiple theoretical perspectives, based on the user's theoretical inclinations. Finally, it is becoming increasingly clear that the traditional dichotomy between verbal and nonverbal intelligence overlaps with and is somewhat redundant with the Cattell-Horn-Carroll (CHC) crystallized and fluid ability dichotomy. It is no accident that tests of crystallized ability are overwhelmingly verbal while tests of fluid ability are consistently nonverbal and visual-spatial.

Cognitive Assessment System

The Das-Naglieri Cognitive Assessment System (CAS; Naglieri & Das, 1997a, 1997b) is a cognitive processing battery intended for use with children and adolescents 5 through 17 years of age. The origins of the CAS may be traced to the work of A. R. Luria, the preeminent Russian neuropsychologist whose work has been highly influential in American psychology. Beginning in 1972, J. P. Das initiated a program of research based on the simultaneous and successive modes of information processing suggested by Luria. Ashman and Das (1980) first reported the addition of planning measures to the simultaneous-successive experimental tasks, and separate attention and planning tasks were developed by the end of the decade (Naglieri & Das, 1987, 1988). The work of Luria and Das influenced Alan and Nadeen Kaufman, who published the K-ABC (see section below) in 1983. Jack A. Naglieri, a former student of Kaufman's who had assisted with the K-ABC development, met J. P. Das in 1984 and began a collaboration to assess Luria's three functional systems. Thirteen years and some 100 studies later, the CAS was published.

CAS is available in two batteries: an 8-subtest basic battery and a 12-subtest standard battery. The basic battery can be administered in about 40 minutes while the standard battery requires about 60 minutes. Translations/adaptations of CAS are now available in Dutch, Chinese, Greek, Italian, Japanese, Korean, Norwegian, and Spanish.

Theoretical Underpinnings

The CAS has its theoretical underpinnings in Luria's (1970, 1973, 1980) three functional units in the brain: (1) the first unit regulates cortical tone and alertness and arousal (interpreted by the test authors as *attention*); (2) the second unit receives, processes, and retains information in two basic forms of integrative activity (*simultaneous* and *successive*); and (3) the third unit involves the formation, execution, and monitoring of behavioral *planning*. Luria (1966) credited Russian physiologist Ivan M. Sechenov with the concept of *simultaneous* and *successive* processing, which he introduced in this way:

> The first of these forms is the integration of the individual stimuli arriving in the brain into *simultaneous, and primarily spatial, groups,* and the second in the integration of individual stimuli arriving consecutively in the brain *into temporally organized, successive series.* We shall refer conventionally to these as simultaneous and successive syntheses. (p. 74; emphasis in the original)

Simultaneous processing tends to be parallel or synchronous, in which stimuli are perceived and processed as a whole. Successive processing tends to involve a serial, chainlike progression of processing, in which information is processed in order and each activity is related to those that preceded it. Simultaneous and successive cognitive processes formed the core elements of the dual processing system (Das, Kirby, & Jarman, 1975, 1979) that was later expanded and articulated as PASS theory, using the acronym for planning, attention, simultaneous, and successive processes (Das, Naglieri, & Kirby, 1994).

In 1986 Das defined *intelligence* as “the sum total of all cognitive processes” (p. 55; see also Das, 2004, p. 5). According to Naglieri (1999), “the single most important goal of the Cognitive Assessment System is to encourage an evolutionary step from the traditional IQ, general ability approach to a theory-based, multidimensional view with constructs built on contemporary research in human cognition” (p. 9).

Standardization Features and Psychometric Adequacy

The CAS was standardized from 1993 through 1996 on 2,200 children and adolescents from 5 through 17 years of age, stratified on 1990 census figures. Sample stratification variables included race, ethnicity, geographic region, community setting, parent educational attainment, classroom placement, and educational classification. The standardization sample was evenly divided between males and females, in nine age groups, with $n = 300$ per year for ages 5 through 7 years, $n = 200$ per year for ages 8 through 10 years, and $n = 200$ per 2- or 3- year intervals for ages 11 through 17 years. Demographic characteristics of the standardization sample are reported in detail across stratification variables in the CAS interpretive handbook and closely match the targeted census figures (Naglieri & Das, 1997b).

The reliability of the CAS is generally adequate. Internal consistency was computed through the split-half method with Spearman-Brown correction, and test–retest stability was the basis for estimating the reliability of the Planning and Attention subtests as well as a single Successive Processing subtest (Speech Rate). Stability coefficients were measured with a test–retest interval from 9 to 73 days, with a median of 21 days. Across these two methods of determining score reliability, the average reliabilities for the PASS and Full Scale composite scores across age groups ranged from .84 (Attention, Basic Battery) to .96 (Full Scale, Standard Battery). Average subtest reliability coefficients across age groups ranged from .75 to .89, with a median reliability of .82. A total of 9 of 13 subtests yielded reliability estimates at or above .80. Half (50%) of the composite score reliabilities were at .90 or higher.

Corrected for variability of scores from the first testing, the stability coefficients are somewhat less adequate with median values of .73 for the CAS subtests and .82 for the Basic and Standard Battery PASS scales. Only 16% of the subtests had corrected stability coefficients at or above .80, and no composite standard scores had corrected stability coefficients at or above .90.

CAS floors and ceilings tend to be adequate for school-age children. Test score floors extend 2 or more standard deviations (*SD*s) below the normative mean beginning with 6-year, 4-month-old children, so discrimination at the lowest processing levels is somewhat limited with 5-year-olds, particularly for simultaneous subtests. Test score ceilings extend more than 2 *SD*s above the normative mean at all age levels. Standard scores range from about 45 to 153 for the PASS scales, with a range of 40 to 160 for the Full Scale IQ (FSIQ).

Exploratory and confirmatory factor analyses of the CAS provide support for either a three- or four-factor solution (Naglieri & Das, 1997b). The four-factor solution is based on the four PASS dimensions whereas the three-factor solution combines Planning and Attention to form a single dimension. The decision to utilize the four-factor solution was based on the test’s underlying theory, meaningful discrepancies between planning and attention performance in criterion populations (e.g., attention-deficit/hyperactivity disorder [ADHD], traumatic brain injury), and differential response to treatments in intervention studies (e.g., planning-based intervention).

Critics have asserted that CAS interpretive structure (i.e., the PASS framework) does not match its factor structure, and questions have emerged about what constructs that the PASS scales actually measure. On a data set based on the tryout version of the CAS, Carroll (1995) argued that the planning scale, in which all subtests are timed, may be best conceptualized as a measure of perceptual speed. Keith, Kranzler, and Flanagan challenged the CAS factor structure based on reanalyses of the standardization sample and analyses with new samples (Keith & Kranzler, 1999; Keith et al., 2001; Kranzler & Keith, 1999; Kranzler, Keith, & Flanagan, 2000). These investigations have generally reported that the planning and attention subtests lack the specificity and factorial coherence to be interpreted separately, that they are more appropriately collapsed into a single factor strongly related to speed, and that the simultaneous and successive factors may best be reconceptualized as tapping visualization and short-term memory span. In responding to these types of criticisms, Puhan, Das, and Naglieri (2005) offered other sources of validity evidence for the factor structure of the CAS. Haddad (2004) studied the relationship between speed and planning on the CAS Planned Codes subtest and concluded that the subtest is better described as a measure of planning than speed. More recently, Deng, Liu, We, Chan, and Das (2011) found that either a four-factor PASS model or a three-factor (PA)SS model fit the data well in a confirmatory factor analysis of the Chinese-language adaptation of CAS with a Chinese sample. After conducting hierarchical exploratory factor analyses, Canivez (2011b) and the test's complex structure may have distorted the CFA results of Kranzler and Keith (1999), leading them to overestimate planning and attention factor correlations.

The CAS also deemphasizes interpretation of the overall composite *g* estimate (the Full Scale standard score) in favor of an emphasis on the four PASS scales (e.g., Naglieri, 1999). In a series of analyses, Canivez (2011a, 2011b, 2011c, 2011d) reported mixed support for this approach. Based on hierarchical exploratory factor analyses with the Schmid-Leiman procedure, Canivez (2011a, 2011b) reported that within the age 5 to 7 group, only the Number Detection subtest yields a high *g* loading, and within the ages 8 to 17 group, only Planned Connections yields a high *g* loading, seemingly supporting the interpretation of CAS as a low *g* test. However, Canivez (2011b) concluded that most of the total and common CAS variance was indeed associated with a second-order *g* factor, and interpretation of CAS at this level is supported, even if few subtests are good *g* measures. Even so, he notes that CAS yielded greater proportions of subtest variance apportioned to first order factors (i.e., PASS, or [PA]SS) than most other intelligence tests although some factors (planning and successive) appeared to explain much more variance than the simultaneous factor. Canivez (2011b) concludes, "Further research will help determine the extent to which CAS PASS scores possess acceptable incremental validity and diagnostic utility" (p. 314).

Evidence of convergent validity with other composite intelligence test scores indicates that the CAS overall composite has high correlations with other intelligence test composites—that is, with the WISC-III FSIQ ($r = .69$), the WJ III Cog Brief Intellectual Ability ($r = .70$; from McGrew & Woodcock, 2001) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-R) FSIQ ($r = .60$) (Naglieri & Das, 1997b). Evidence of convergent validity for the four individual PASS scales is not reported in the CAS *Interpretive Handbook,* and surprisingly neither the CAS nor the K-ABC or KABC-II have convergent validity studies with each other, in spite of their shared theoretical underpinnings. Some investigations suggest limitations in the validity of the PASS scales; for example, in separate investigations, the CAS Planning scale has been shown to yield low correlations with Tower of London performance (Naglieri & Das, 1997b; Ciotti, 2007), long considered a criterion measure of planning (e.g., Shallice, 1982). The CAS Planning scale also is not significantly correlated with parent and teacher reports of student planning and organizational ability (Ciotti, 2007).

CAS Full Scale standard scores also have high correlations with achievement test results. Based on a large sample ($n = 1600$) used as a basis for generating ability-achievement comparisons, the CAS Full Scale standard scores yield high correlations with broad reading and broad mathematics achievement ($r = .70$ to .72; see Naglieri & Das, 1997b). Correlations of the CAS with achievement test composites have been shown to be higher than those found for most other intelligence tests (Naglieri & Bornstein, 2003; Naglieri, DeLauder, Goldstein, & Schwebech, 2006; Naglieri & Rojahn, 2001). The comparatively high ability-achievement correlations may be interpreted as supporting a strong and possibly causal linkage between cognitive processes and the academic performances to which they may contribute.

CAS is unique among tests of cognitive abilities and processes insofar as it has been studied with several research-based programs of intervention, specifically programs of cognitive instruction that are concerned with the interface between psychology and education, particularly the cognitive processes involved in learning (e.g., Mayer, 1992). The *Process-Based Reading Enhancement Program* (PREP; see e.g., Carlson & Das, 1997; also Das & Kendrick, 1997) is a PASS theory intervention consisting of a series of exercises in global processing (to facilitate strategy development, independent from reading content) and curriculum bridging (to intervene in processing, but with similar content as required for reading) with the ultimate goal of improving word-reading and decoding skills. Another form of cognitive instruction related to PASS theory is the planning facilitation method described by Cormier, Carlson, and Das (1990); Kar, Dash, Das, and Carlson (1992); and Naglieri (1999); in it students have been shown to differentially benefit from a verbalization technique intended to facilitate planning. Participants who initially perform poorly on measures of planning earn significantly higher postintervention scores than those with good scores in planning. Extended accounts of research with CAS linking assessment to educational interventions are available (e.g., Naglieri, 1999; Naglieri & Das, 1997b; Naglieri & Otero, 2011; Naglieri & Pickering, 2010).

Interpretive Indices and Applications

The CAS yields four standard scores corresponding to the PASS processes as well as a full scale standard score. Although the subtests account for high levels of specific variance, the focus of CAS interpretation is at the PASS Scale level, not at the subtest level or full scale composite level. PASS theory guides the examination of absolute and relative cognitive strengths and weaknesses. Table 18.1 contains interpretations for each of the PASS scales.

The CAS authors suggest that the test is potentially useful in diagnosis, classification, and eligibility decisions for specific learning disabilities, attention deficit, intellectual disabilities, and giftedness as well as for planning treatment, instructional, or remedial interventions (Naglieri & Das, 1997b). Special population studies reported in validity studies include children diagnosed with ADHD, reading disabilities, intellectual disabilities/mental retardation, traumatic brain injury, serious emotional disturbance, and giftedness. Children with selected exceptionalities appear to show characteristic impairment on PASS processes or combinations of processes. Reading-disabled children tend as a group to obtain their lowest scores on measures of successive processing (Naglieri & Das, 1997b), presumably due to the slowed phonological temporal processing thresholds that have been identified as a processing deficit associated with delayed reading acquisition (e.g., Anderson, Brown, & Tallal, 1993). Children diagnosed with the hyperactive-impulsive subtype of ADHD tend to characteristically have weaknesses in planning and attention scales (Paolitto, 1999), consistent with the newest theories reconceptualizing ADHD as a disorder of executive functions (Barkley, 1997). Characteristic weaknesses in planning and attention have also been reported in samples of traumatically brain-injured children (Gutentag, Naglieri, & Yeates, 1998), consistent with the frontal-temporal cortical impairment usually associated with closed head injury.

TABLE 18.1 Cognitive Assessment System (CAS) Core Interpretive Indices

Composite Indices	Description
Full Scale	An index of complex mental activity involving the interaction of diverse cognitive processes
Planning	An index of the process by which an individual determines, selects, applies, and evaluates solutions to problems; involves generation of strategies, execution of plans, self-control, and self-monitoring
Attention	An index of the process of selectively focusing on particular stimuli while inhibiting response to competing stimuli; involves directed concentration and sustained focus on important information
Simultaneous Processing	An index of the process of integration of separate stimuli into a single perceptual or conceptual whole; applies to comprehension of relationships and concepts, understanding of inflection, and working with spatial information
Successive Processing	An index of the process of integrating stimuli into a specific, temporal order that forms a serial progression; involves sequential perception and organization of visual and auditory events and execution of motor behaviors in order

Like most of the other intelligence tests for children and adolescents, CAS is also empirically linked to achievement tests (Woodcock-Johnson–Revised and the WJ III Tests of Achievement). Through the use of simple and predicted differences between ability and achievement, children who qualify for special education services under various state guidelines for specific learning disabilities may be identified. Moreover, CAS permits the identification of selected cognitive processes (planning, attention, simultaneous, and successive processes) that if impaired may contribute to the learning problems. CAS has minimal acquired knowledge and academic skill requirements, although there are low-level requirements for fast recall of alphabetic letter sequences (Planned Connections), rapid word reading (Expressive Attention), and comprehension of language syntax (Verbal-Spatial Relations and Sentence Questions).

CAS also provides normative reference for the use of metacognitive problem-solving strategies that may be observed by the examiner or reported by the examinee on planning subtests. These strategies have been analyzed relative to developmental expectations and effectiveness in enhancing task performance (Winsler & Naglieri, 2003; Winsler, Naglieri, & Manfra, 2006). The inclusion of age-referenced norms for strategy usage provides an independent source of information about the efficiency, implementation, and maturity with which an individual approaches and performs complex tasks.

Strengths and Limitations

The CAS offers several progressive advances in intelligence testing. Its most important contribution is to include executive functions as a core element in the assessment of intelligence, a decision that preceded the inclusion of executive function subtests in the KABC-II (Kaufman & Kaufman, 2004), WISC-III as a Process Instrument (WISC-III PI; Kaplan, Fein, Kramer, Delis, & Morris, 1999) and WISC-IV Integrated (Wechsler, Kaplan, Fein, Kramer, Morris, Delis, & Maerlender, 2004), and the WJ III Cog (Woodcock, McGrew, &

Mather, 2001a). The inclusion of objective, norm-referenced problem-solving strategies also provides a potentially valuable metacognitive process-based assessment. The demonstration in CAS of higher correlations with basic academic skills (with minimal emphasis on acquired knowledge) than other intelligence tests (e.g., Naglieri & Bornstein, 2003; Naglieri et al., 2006; Naglieri & Rojahn, 2001) speaks to the merits of tapping cognitive processes over crystallized knowledge in intelligence assessment. Efforts to systematically link CAS performance with associated intervention approaches also are forward-looking and strongly needed in this field. By virtue of its theoretical underpinnings and linkages to diagnosis and treatment, the CAS builds on the earlier advances offered by the KABC (Kaufman & Kaufman, 1983a, 1983b). In an early review, Meikamp (1999) observed, "The CAS is an innovative instrument and its development meets high standards of technical adequacy. Despite interpretation cautions with exceptional populations, this instrument creatively bridges the gap between theory and applied psychology" (p. 77).

The chief limitations of the CAS stem from its ambitious yet flawed operationalization of Luria's theory of the functional systems of the brain and a mismatch between its theoretical interpretive framework (PASS) and its structure according to factor analyses. Theoretical shortcomings include a misinterpretation of Luria's first functional unit as regulating higher-order attentional processes; his first functional unit is actually associated with limbic system activation (and inhibition) of generalized arousal and alertness:

> The reticular activating formation, the most important part of the first functional unit of the brain, . . . affects all sensory and all motor [cortical] functions of the body equally, and . . . its function is merely that of regulating states of sleep and waking—the non-specific background against which different forms of activity take place. (Luria, 1973, p. 52)

The arousal and alertness mediated by the reticular system may be considered a prerequisite for attention (since a minimal level of arousal is necessary to pay attention), but not higher forms of attention per se. Higher-order attentional processes are now thought to be mediated by at least two attentional systems: a posterior cortical system associated with orienting and foveation and an anterior cortical system associated with signal detection and focal processing (e.g., Posner & Petersen, 1990). CAS attentional subtests almost certainly involve the latter system, which is associated with the same prefrontal processes that some of the planning subtests are believed to involve (e.g., Derrfuss, Brass, Neumann, & von Cramon, 2005). By including attention tasks and planning tasks, the CAS authors in effect set up two scales tapping prefrontal executive functions and Luria's third functional unit.

A second theoretical limitation was first observed in Das's early work by Paivio (1976), who asserted that the successive-simultaneous processing distinction is confounded by verbal-nonverbal sensory modality assessment methodologies. In other words, successive tasks are predominantly verbal and simultaneous tasks are predominantly visual-spatial, making it difficult to demonstrate that the two cognitive processes transcend sensory modality as Luria believed.

Problems with the CAS theoretical and interpretive structure have been documented in confirmatory factor analyses (Keith & Kranzler, 1999; Keith et al., 2001; Kranzler & Keith, 1999; Kranzler et al., 2000), providing evidence that CAS factors may more appropriately be interpreted as measuring processing speed (rather than planning and attention), memory span (rather than successive processing), and a mixture of fluid intelligence and broad visualization (rather than simultaneous processing), with a higher-order general intelligence factor (approximated by the Full Scale standard score). Hierarchical exploratory factor analyses have documented the low amounts of specificity and incremental validity offered by the PASS scales beyond the Full Scale standard score (Canivez, 2011a, 2011b, 2011c, 2011d).

Some of these challenges could be readily addressed in a revised edition, through the inclusion of untimed subtests in planning and attention domains, nonverbal or non-memory subtests in the successive processing domain, and verbal subtests in the simultaneous domains. In an explicit acknowledgment that the Lurian dimensions may be readily reinterpreted from the CHC framework, Kaufman and Kaufman (2004) offered a dual theoretical perspective in which successive (sequential) and simultaneous processing may be just as easily viewed as memory span and broad visualization. To demonstrate the structural integrity of the CAS and correspondence of the PASS processes to Luria's functional systems, the next edition of the CAS will have to redesign the content of some subtests so as to clarify the constructs they measure and to conduct the basic brain–behavior research needed to establish the subtests' neural correlates. In an examination of CAS and its underpinnings in Luria's conceptualization of the brain, McCrea (2009) recommended that CAS subtests and scales be examined through functional neuroimaging studies and brain lesion studies to more clearly establish their neural sensitivity and specificity.

Differential Ability Scales—Second Edition

The Differential Ability Scales—Second Edition (DAS-II; C. D. Elliott, 2007a, 2007b, 2007c) offer ability profiling with 20 subtests divided into two overlapping batteries standardized for ages 2½ through 17 years. The DAS-II is a U.S. adaptation, revision, and extension of the British Ability Scales (BAS; C. D. Elliott, Murray, & Pearson, 1979). Development of the BAS originally began in 1965, with plans to develop a test to measure Thurstone's (1938) seven primary mental abilities and key dimensions from Piagetian theory. Colin D. Elliott, a teacher, university faculty, and school psychologist trainer, became the director of the project in 1973. Elliott spearheaded decisions to deemphasize IQ estimation and to provide a profile of meaningful and distinct abilities as well as to support the introduction of item response theory in psychometric analyses. The first edition of the British Ability Scales was published in 1979 with an amended revised edition published in 1983 (BAS-R; C. D. Elliott et al., 1983). The development of the American edition began in 1984, and the first edition of the DAS was published in 1990. Second editions followed—the BAS-II in 1996 and the DAS-II in 2007—so that over four decades have passed since work on the BAS actually began.

The DAS-II consists of four core subtests (lower level) or six core subtests (upper level) for the Early Years Battery (ages 2:6–6:11) and six core subtests for the School-Age Battery (ages 7:0–17:11). Separate diagnostic subtests may be administered in clusters to assess working memory, phonological processing, processing speed, and foundational abilities for early school learning. The core School-Age battery typically requires about 30 to 40 minutes. Additional time is required for administration of optional diagnostic clusters in School Readiness (15–20 minutes), Working Memory (10–15 minutes), Processing Speed (10 minutes), and Phonological Processing (10 minutes). Spanish-language instructions are provided, but only for subtests that do not require a verbal response from the examinee.

Theoretical Underpinnings

The DAS-II was developed to accommodate diverse theoretical perspectives, but it now aligns most closely with the CHC framework. It is designed to yield an estimate of higher order general intelligence, the General Conceptual Ability (GCA) score, and lower order broad cognitive factors or diagnostic clusters: Verbal Ability (Gc), Nonverbal Reasoning Ability (Gf), Spatial Ability (Gv), Working Memory (Gsm), and Processing Speed (Gs). The DAS avoids use of the terms *intelligence* and *IQ,* focusing instead on profiles of cognitive abilities and processes that are either strongly related to the general factor or thought to have value for diagnostic purposes.

The GCA captures test performance on subtests that have high *g* loadings, in contrast to some intelligence tests in which all subtests (high and low *g* loading) contribute to the overall IQ composite. Through this approach, it also avoids the problems found on the Wechsler scales in which circumscribed processing deficits (such as low processing speed) depress overall ability estimates. Confirmatory factor analyses reported in the *Introductory and Technical Handbook* (C. D. Elliott, 2007b) show the *g* loadings of core subtests (i.e., those that contribute to the GCA) to range from .66 to .76 for ages 6:0 to 12:11 and from .65 to .78 for ages 6:0 to 17:11. By Kaufman's (1994) criteria, half of the DAS-II subtests are good measures of general intelligence and the remaining core subtests are fair measures of general intelligence. As expected, the diagnostic subtests do not fare as well as measures of *g*.

At a level below the GCA in hierarchical structure are cluster scores that uniformly have sufficient specific variance for interpretation (C. D. Elliott, 2007b). For children from ages 2½ years through 3½ years (Early Years Battery Lower Level), only Verbal Ability and Nonverbal Ability cluster scores may be derived. For older preschool children and school-age children, three cluster scores may be computed from core subtests: Verbal Ability, Nonverbal Reasoning Ability, and Spatial Ability, with the option to compute a Special Nonverbal Composite. From diagnostic subtests, Working Memory and Processing Speed cluster composites may be generated above age 3½, and a School Readiness composite may also be computed from diagnostic subtests in the Early Years Battery Upper Level. The increased cognitive differentiation (i.e., from two core cluster scores to three core cluster scores) from the Early Years Battery Lower Level to the Upper Level and the School-Age Battery is consistent with the fundamental developmental tenet that cognitive abilities tend to become differentiated with maturation (H. Werner, 1948).

Standardization Features and Psychometric Adequacy

Through over six years of research and development, the DAS-II underwent a national pilot, tryout, and standardization study. The standardization edition was normed on a representative U.S. sample of 3,480 children and adolescents, between ages 2:6 and 17:11. Using 2002 Current Population Survey census figures, the normative sample was stratified on the basis of race/ethnicity, parent education level, and geographic region. The sample was balanced by age and sex. Age was divided into 18 age levels, including 6-month bands for the Early Years Battery preschool ages and 12-month bands at and above age 5 years for the School-Age Battery. A total of $n = 176$ was sampled per age band for preschoolers, with $n = 200$ per age band for ages 5 through 17 years. The composition of the normative sample is detailed across stratification variables in the *DAS-II Introductory and Technical Handbook* (C. D. Elliott, 2007b) and appears to closely match its 2002 census target figures. The standardization sample inclusion criteria required English as the primary language of the examinee; prospective examinees were excluded (or referred to special population studies) when they had received diagnoses or services for any delay in cognitive, motor, language, social-emotional, or adaptive functioning.

Individual items were statistically evaluated for fit, reliability, bias, and difficulty through Rasch scaling, which was also used to divide subtests into item sets. The DAS-II reports the relative difficulty of individual items for each subtest (C. D. Elliott, 2007b, Appendix A), thereby providing item difficulty gradients. Raw score to ability score conversion tables are also reported (Appendix A), including the standard error of each ability score, from which local reliabilities may be manually calculated if desired.

The score reliabilities of the DAS-II subtests and composites were computed through calculation of coefficient alpha or item response theory (IRT) proxies for reliability (C. D. Elliott, 2007b). DAS-II subtests are administered in predetermined item sets rather than with formal basal and discontinue rules. This means that starting points and stopping decision points (as well as alternative stopping points) are designated on the Record Form according to the child's age or ability level. Within any given item set, at least three items passed and at least three items failed provide support that the appropriate item set was administered. If an examinee passes fewer than three items, a lower item set should be administered; if an examinee fails fewer than three items, a higher item set should be administered. Ideally, an examinee's performance has been optimally assessed when approximately half of the items are passed and half are failed. Through the use of item sets, examinees receive a form of tailored, adaptive testing, in which they are given items closest to their actual ability levels.

C. D. Elliott (2007b) reports that the average reliability coefficients of Early Years Battery subtests ranged from .79 (Recall of Objects-Immediate, Recognition of Pictures, Word Definitions) to .94 (Pattern Construction). A total of 86% of subtests had average reliabilities across ages at or above .80, and no subtests had substantially lower internal consistency. For cluster and composite scores, average reliabilities ranged from .89 (Nonverbal and Nonverbal Reasoning Ability and Processing Speed) to .95 (Spatial Ability, GCA, and SNC). A total of 75% of composites and cluster scores had average reliabilities at or above .90.

Average reliability coefficients of the School-Age Battery subtests ranged from .74 (Recognition of Pictures) to .96 (Pattern Construction), with 90% of subtests yielding $\alpha \geq .80$. The diagnostic subtest Recognition of Pictures is the only subtest with consistently inadequate reliability. For clusters and composites, 88% of School-Age Battery composites and cluster scores had average reliability greater than or equal to .90, with the average GCA at .96. Altogether and with the exception of the Recognition of Pictures subtest, these findings indicate that the internal consistency of DAS-II tasks and composite scores are consistently adequate.

Internal consistencies are also reported for 12 special populations, including samples of individuals diagnosed with intellectual disability, intellectual giftedness, attention deficits, learning disorders, and limited English proficiency, among others. Isolated subtests show low reliabilities, suggesting the possibility that some subtests should be interpreted cautiously in selected groups, although clearly more research on reliability generalization would be beneficial. In general, however, DAS-II subtest internal reliability coefficients appear fully adequate.

Stability coefficients were computed for 369 examinees in three age groups undergoing test and retest intervals of 1 to 9 weeks (mean interval = 23 days), with correction for restriction of range on the initial score. When these groups are combined, subtest test–retest corrected stability coefficients ranged from .63 to .91, while cluster and composite corrected stabilities ranged from .81 to .92. The stability of the GCA ranged from .91 to .94 across the three groups, indicating fully adequate test–retest stability. In general, these findings suggest that isolated subtests may not be particularly stable at specific ages, but composite score stability coefficients tend to be fairly adequate. Examinations of test–retest gain scores (i.e., practice effects) show that improvements over the span of about a month are largest for Nonverbal Reasoning (5.8 standard score points), School Readiness (5.2 points), and the GCA (5.1 points). The largest composite score practice effects are small to medium, in terms of effect sizes.

Interscorer agreement for the normative sample ranged from .98 to .99, based on double-scoring of all standardization cases. Four subtests that require scoring of drawings or verbal responses according to objective criteria were further investigated. In a stratified sample of 60 cases scored independently by four scorers, the interscorer reliability coefficients ranged from .95 (Copying) to .97 (Recall of Designs) to .99 (Word Definitions and Verbal Similarities), all of which fall within an acceptable range.

The DAS-II composites and clusters scores have considerable range, extending 9 standard deviations (*SD*) from low standard scores of about 32 (−4.5 *SD*) to ceiling scores of 169 or 170 (+4.6 *SD*). DAS-II subtest floors are sufficiently low so that its *T* scores extend 2 *SD* below the normative mean with 2½-year-old children with developmental delays. Use of Rasch scaling extrapolation also permits GCA and other composite norms to be extended downward, to standard scores as low as 32, enhancing the discriminability of the DAS-II with individuals with moderate to severe impairment including intellectual disability. DAS-II subtest ceilings are sufficiently high that they extend +4 *SD* above the normative mean for all but a single subtest (Sequential and Quantitative Reasoning, which extends +3.1 *SD*) at the highest age range served by the DAS-II. Composite and cluster score ceilings consistently extend as high as 169 or 170, to support the identification of highly gifted individuals. Subtests in the Early Years Battery Upper Level and the School-Age Battery have overlapping norms for children between the ages of 5:0 and 8:11, thereby raising the test ceiling for high-functioning younger children and lowering the test floor for low-functioning older children, permitting out-of-level assessments.

The DAS-II tends to show considerable convergence with other intelligence and achievement tests. According to analyses from C. D. Elliott (2007b), the DAS-II GCA correlates highly with composite indices from Bayley-III Cognitive Scale ($r = .59$), the WPPSI-III FSIQ ($r = .87$ for the preschool battery), the WISC-IV FSIQ ($r = .84$ for the school-age battery), the Wechsler Individual Achievement Test (WIAT-III) Total Achievement composite score ($r = .82$; see Pearson research staff, 2009), the WJ III Total Achievement Score ($r = .80$), the Kaufman Test of Educational Achievement (KTEA-II), and Comprehensive Achievement Composite ($r = .81$). Two investigations have supported the value of multiple specific DAS-II core and diagnostic scores (in lieu of the GCA composite) in predicting academic reading performance (C. D. Elliott, Hale, Fiorello, Dorvil, & Moldovan, 2010) and mathematics performance (Hale, Fiorello, Dumont, Willis, Rackley, & Elliott, 2008).

Confirmatory factor analyses reported in the *Introductory and Technical Handbook* (C. D. Elliott, 2007b) provide general support for the interpretive structure of the test. For the upper level of the Early Years Battery, a three-factor model (Verbal, Nonverbal Reasoning, and Spatial) fit the data significantly better than one- or two-factor models. When the six core subtests alone were examined for ages 7:0 to 17:11, a three-factor solution (Verbal, Nonverbal Reasoning, and Spatial) yielded optimal fit. When all core and diagnostic subtests were included in analyses for ages 6:0 to 12:11, a seven-factor model (Gc, Gf, Gv, Gsm, Glr, Gs, and Ga) provided the best fit with the data. For ages 6:0 to 17:11, a six-factor model (Gc, Gf, Gv, Gsm, Glr, and Gs) produced good fit. Independent higher-order confirmatory factor analyses by Keith and his colleagues (2010) supported the DAS-II structure, with minor exceptions, as well as supporting the factorial invariance of the DAS-II CHC model across the 4- to 17-year age span.

Interpretive Indices and Applications

The DAS-II involves some score transformations based on item response theory. Raw scores are converted first to latent trait ability scores based on tables appearing in the record form; ability scores are in turn translated into *T* scores ($M = 50$, $SD = 10$), percentile ranks, and age equivalent scores. *T* scores may be summed to produce the GCA and cluster scores ($M = 100$, $SD = 15$), percentiles, and confidence intervals. The GCA is derived only from subtests with high *g* loadings, and cluster scores consist of subtests that tend to factor together. The diagnostic subtests measure relatively independent abilities. The clusters and diagnostic subtests have adequate specific variance to support their interpretation independent from *g*. Table 18.2 contains the basic composite and cluster indices.

Dumont, Willis, and Elliott (2009) described evidence for use of DAS-II in assessment of individuals with specific learning disabilities, attention-deficit/hyperactivity disorders, intellectual disability, intellectual giftedness, language disabilities, and autism spectrum disorders, noting that the specificity of DAS-II subtests and composites enhances decision making:

> When we are asked to evaluate an individual's cognitive development or aspects of the individual's ability in

TABLE 18.2 Differential Ability Scales—Second Edition Core Interpretive Indices

Composite Indices	Description
General Conceptual Ability (GCA)	An index of ability to perform complex mental information processing, including conceptualization and transformation; derived from subtests with high *g* loadings
Special Nonverbal Composite	An alternative measure of general ability to be used when verbal subtest performances are considered invalid measures of examinee ability (e.g., when an examinee is not proficient in spoken English)
Verbal Ability	A cluster score measuring acquired verbal concepts and knowledge
Nonverbal Reasoning Ability	A cluster score measuring nonverbal mental processing, including inductive reasoning for abstract, visual problems
Spatial Ability	A cluster score measuring complex visual-spatial processing, including ability in spatial imagery and visualization, perception of spatial orientation, attention to visual details, and analytic thinking
School Readiness	A multidimensional cluster score tapping the growth of skills and abilities fundamental to early school learning, including number concepts, matching of simple graphic figures, and phonological processing (appropriate for examinees ages 5:0 to 8:11 taking the upper Early Years battery)
Working Memory	A cluster score tapping auditory short-term memory and working memory, including integration of visualization with verbal short-term memory
Processing Speed	A cluster score measuring general cognitive processing speed for simple mental operations, including visual comparison and lexical access

> information processing, we need to do this in a way that will clearly identify distinctive, interpretable cognitive factors, systems, or processes with as little overlap and ambiguity as possible. That is the reason we need subtests and composites that are not only reliable but which also have high specificity.... [T]he DAS-II has been designed for this purpose. It has between 10% and 20% more reliable specificity in its subtests and clusters than other cognitive test batteries. (p. 248)

C. D. Elliott (2007b) reported DAS-II special population studies for children identified with intellectual disability, intellectual giftedness, reading disorders, reading and written expression disorder, mathematics disorder, ADHD, ADHD and learning disorders, expressive disorders, mixed receptive-expressive language disorder, limited English proficiency, and developmental risk. Special population group mean scores are each compared with demographically matched comparison groups drawn from the standardization sample.

Strengths and Limitations

The DAS-II is widely considered to be a psychometrically superlative test battery ranking among the best preschool and school-age cognitive test batteries available. In a critical review, Davis and Finch (2008) wrote:

> The evidence around the psychometric properties of the instrument was generally well presented and of high technical quality. The psychometric properties of the DAS-II are well documented and quite stellar ... Across all age ranges, this revision continues to be child-friendly, psychometrically sound, and of high utility. There is a wealth of technical data, much of which will be excessive for the typical user, but examiners will find that virtually any imaginable psychometric study has been conducted with at least adequate results. Additionally, the development and standardization of the DAS-II set a high standard from which many other tests could benefit. Clearly, the DAS-II has been well studied and the evidence presented in the technical report supports the contention that the instrument was found to be both reliable and valid across a variety of samples.

The strengths of the DAS-II should not be wholly surprising, as it has benefited from three predecessor British editions (BAS, BAS-R, BAS-II) as well as the DAS, making it the equivalent of a fifth edition with the same lead author. It overlaps substantially in content with the Wechsler scales but is now aligned with the CHC framework and offers supplemental diagnostic subtests and scales, which will not depress the overall GCA if diagnostic impairment is present. In total, it ranks among the best intelligence tests of this era.

Hill (2005) explained the continued preference of British assessment practitioners for the Wechsler scales over the BAS-II:

> Many psychologists feel that more contemporary psychometric assessments, for example, the BAS-II, lack either the history or the extensive research profile of the Wechsler scales. In a litigious context it appears that some psychologists feel the Wechsler scales provide greater professional security. (p. 89)

The same sentiment may explain preference of American assessment practitioners for the Wechsler scales over the DAS-II (e.g., Camara, Nathan, & Puente, 2000), in spite of its technical excellence. With its assets noted, research is sorely needed to systematically link DAS-II cognitive ability profiles with academic and nonacademic interventions for children and adolescents.

Kaufman Assessment Battery for Children, Second Edition

When Alan S. Kaufman's book, *Intelligent Testing with the WISC-R*, was published in 1979, he became the leading authority after David Wechsler himself on the Wechsler intelligence scales, the most dominant tests of our era. Kaufman and his wife, Nadeen, released the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983a, 1983b) just four years later, and it constituted the boldest challenge yet encountered by the Wechsler scales. These landmark works have gone through successive editions, along with many other accomplishments, making Alan Kaufman the leading influence on the practice of applied intelligence testing in the last 30 years. In the words of his students Randy W. Kamphaus and Cecil R. Reynolds (2009), Kaufman's contribution having the greatest long-term impact was his "joining of the two disciplines of measurement science and clinical assessment practice" (p. 148).

Kaufman and Kaufman are the coauthors of the Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2004) and Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993). They have a unique training and academic lineage, and have in turn exerted strong influences on several leading test developers. Their history here is summarized from the Kaufmans' own telling (Cohen & Swerdlik, 1999). Alan Kaufman completed his doctorate from Columbia University under Robert L. Thorndike, who would head the restandardization of the Stanford-Binet L-M (Terman & Merrill, 1973) and serve as senior author of the Stanford-Binet Fourth Edition (Thorndike, Hagen, & Sattler, 1986). Kaufman was employed at the Psychological Corporation from 1968 to 1974, where he worked closely with David Wechsler on the WISC-R. Nadeen L. Kaufman completed her doctorate in special education with an emphasis in neurosciences from Columbia University, where she acquired a humanistic, intraindividual developmental approach to psychological assessment and learning disabilities that would blend uniquely with her husband's psychometric approach. Following his departure from the Psychological Corporation, Alan Kaufman joined the Educational and School Psychology Department at the University of Georgia. According to the Kaufmans, the K-ABC was conceptualized and a blueprint developed on a 2-hour car trip with their children in March of 1978. In a remarkable coincidence, they were contacted the next day by the director of test development at American Guidance Service (AGS), who asked if they were interested in developing an intelligence test to challenge the Wechsler scales. At the University of Georgia, Alan and Nadeen worked with a gifted group of graduate students on the K-ABC. Among their students were Bruce Bracken, Jack Cummings, Patti Harrison, Randy Kamphaus, Jack Naglieri, and Cecil Reynolds, all influential school psychologists and test authors.

The KABC-II measures cognitive abilities and processing in children and adolescents from age 3 years through 18 years. It consists of core and supplementary tests and may be interpreted according to a dual theoretical foundation, either the CHC psychometric model or Luria's functional systems. The KABC-II is also conormed with the Kaufman Test of Educational Achievement, Second Edition (KTEA-II), facilitating score comparisons between cognitive ability/processes and academic skills. Depending on the theoretical framework selected (the CHC model requires more tests), KABC-II batteries consist of 5 to 7 core subtests and 3 supplemental subtests at age 3; 7 to 9 core subtests and 3 supplemental subtests at age 4; 7 to 9 core subtests and 6 supplemental subtests at age 5; and 8 to 10 core subtests and 5 to 7 supplemental subtests for ages 6:0 to 17:11. The KABC-II may be administered in about 25 to 30 minutes (the core battery at youngest age) to 50 to 70 minutes (the core battery in adolescence); or 35 to 55 minutes (the expanded battery at youngest age) to 75 to 100 minutes (the expanded battery in adolescence).

The KABC-II includes Spanish-language instructions for all subtests with sample or teaching items. For items requiring verbal responses, the examiner may accept responses in any language, and the KABC-II materials list correct and incorrect verbal responses in Spanish as well as English. The Nonverbal Scale of the KABC-II is explicitly intended for use with children who are not fluent in English.

Theoretical Underpinnings

The KABC-II was developed with an unusual dual theoretical foundation, lending itself to interpretation with either the CHC framework or a Lurian neuropsychological processing framework. There was precedent for this dual foundation in the Kaufmans' body of work; after having developed the original K-ABC (Kaufman & Kaufman, 1983a) to tap Lurian processing, the Kaufmans proposed a "theoretical rerouting" based primarily on extended Gf-Gc theory when they published the KAIT (Kaufman & Kaufman, 1993). For the KABC-II, Kaufman and Kaufman (2004) added planning tasks with low speed requirements (tapping fluid reasoning, or Gf), learning tasks (thereby tapping retrieval, or Glr), while retaining measures of acquired knowledge (thereby tapping Gc). Accordingly, the KABC-II established two parallel structural frameworks. The CHC framework measures Gf, Glr, Gsm, Gv, and Gc, while the Lurian framework featured planning, learning, sequential processing, and simultaneous processing (which could be summarized by the acronym P-L-S-S and which endeavors to exclude acquired knowledge from the assessment). The KABC-II composite using the CHC framework is called the Fluid-Crystallized Index (FCI), whereas the composite using the Lurian framework is the Mental Processing Index (MPI). The authors explain that the CHC model is preferred unless the examinee comes from a background (or is diagnosed with a condition) in which verbal functioning and knowledge acquisition may be depressed, distorting the examinee's actual level of ability:

> The CHC model should generally be the model of choice, except in cases where the examiner believes that including measures of acquired knowledge/crystallized ability would 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 the authors believe that Knowledge/Gc is, in principle, an important aspect of cognitive functioning (Kaufman & Kaufman, 2004).

Standardization Features and Technical Adequacy

The KABC-II underwent national standardization from 2001 through 2003, and norms are based on a sample of 3,025 children and adolescents between the ages of 3 and 18 years. The sample was collected to be representative according to 2001 U.S. census figures, based on the stratification variables of parent education level, race/ethnicity, geographic region, and educational placement. Sample sizes were set at $n = 200$ per 12-month interval from ages 3 through 14 years; and at $n = 125$ to 150 per year for ages 15 through 18 years. Standardization sample participants were randomly selected to meet stratification targets from a large pool of potential examinees. A review of the standardization sample demographic breakdowns indicates that the KABC-II normative sample closely matches census target figures.

The score reliabilities of the K-ABC were computed with a Rasch adaptation of the split-half method, with Spearman-Brown correction (Kaufman & Kaufman, 2004). For ages 3 to 6 years, 82% of subtest score reliabilities were at or above .80; for ages 7 through 18, 73% of subtest score reliabilities were at or above .80. Reliabilities of the global scale indexes are uniformly high, averaging in the mid- to upper .90s for the FCI and MPI and in the low .90s for the Nonverbal Index (NVI). All average composite indices including global and factor scale index were at or above .90 across ages 3 to 6 years; 63% of all average composites were at or above .90 across ages 7 to 18 years.

Test–retest stability coefficients were examined for a sample of $n = 250$ children and adolescents undergoing reevaluations after an average interval of about four weeks. When correlations were adjusted for initial score, 26% of composite indices yield a correlation at or above .90 across all ages; 36% of subtest scores yield a correlation at or above .80 across all ages. These stability coefficients are somewhat low.

While exploratory factor analyses were used to guide test development, the authors reported only the results of hierarchical confirmatory factor analyses of the normative standardization sample (Kaufman & Kaufman, 2004). The core subtest configuration (a superordinate *g* and four or five CHC factors, depending on age level) fit the data well, yielding high confirmatory fit indexes (CFIs) and low root mean squared errors of approximation (RMSEAs). Independent follow-up higher-order confirmatory factor analyses by Reynolds, Keith, Fine, Fisher, and Low (2007) concluded that the "the KABC-II factor structure for school-age children is aligned closely with five broad abilities from CHC theory, although some inconsistencies were found" (p. 511). Reynolds and his colleagues (2007) also found core subtests to be age-invariant measures of their assigned constructs. More recently, Morgan, Rothlisberg, McIntosh, and Hunt (2009) found support for the KABC-II hierarchical structure (*g* plus four CHC factors) in confirmatory factor analyses with a preschool sample.

Subtest *g*-loadings were examined using the Kaufman (1994) guidelines from a single unrotated factor of the normative sample via principal axis factor analysis (Kaufman & Kaufman, 2004). For ages 3 to 4 years, only 3 subtests were good measures of *g,* four were fair measures of *g,* and five were poor measures of *g*. For ages 5 to 6 years, four subtests were good measures of *g,* 10 were fair measures of *g,* and two were poor measures of *g.* For ages 7 to 18 years, three subtests were good measures of *g,* nine subtests were fair measures of *g,* and two subtests were poor measures of *g.* The subtests tapping Knowledge (Gc) were the most consistent good measures of *g*.

KABC-II composites tend to show high levels of convergent validity with other intelligence tests (Kaufman & Kaufman, 2004). The KABC-II FCI, MPI, and

NVI respectively yield corrected correlations of .89, .88, and .79 with the WISC-IV FSIQ; .81, .76, and .81 at ages 3 to 4 and .81, .73, and .43 at ages 5 to 6 with the WPPSI-III FSIQ; .91, .85, and .77 with the KAIT Composite Intelligence Scale; and .78, .77, and .74 with the WJ III Cog GIA.

KABC-II composites also show high correlations with achievement test scores (Kaufman & Kaufman, 2004). The KABC-II FCI, MPI, and NVI respectively yield corrected correlations of .67, .69, and .66 for grades 1 to 4 and .73, .67, and .32 for grades 5 to 9 with Peabody Individual Achievement Test (PIAT-R) Total Test Achievement; .72, .65, and .52 for grades 2–5 and .87, .83, and .78 for grades 7–10 with the WIAT-II Total Achievement score; and .70, .63, and .51 for grades 2 to 5 and .79, .77, and .71 for grades 6 to 10 with WJ III Ach Total Achievement score.

Interpretive Indices and Applications

The KABC-II consists of core and supplementary subtests that vary according to the age of the examinee and the purpose of the assessment. For examinees at age 3, up to seven core subtests may be given as well as three supplementary subtests. From ages 4 to 6 years, up to 11 subtests are core with eight supplementary subtests. For ages 7 to 18 years, up to 11 subtests are core, and seven subtests are supplementary. All subtests have a normative *mean* scaled score of 10, *SD* of 3.

The KABC-II yields three global composite indices (the FCI, the MPI, and the NVI). It yields either four or five composite scales, depending on the battery given: Sequential/Gsm, Simultaneous/Gv, Learning/Glr, Planning/Gf, and Knowledge/Gc. All composite standard scores have a normative mean of 100, *SD* of 15. Table 18.3 includes basic interpretations for the KABC-II global and composite scales.

Every subtest has an optional list of Qualitative Indicators (QIs) appearing on the record form that may either disrupt or enhance task performance (e.g., "perseveres," "fails to sustain attention," "reluctant to respond when uncertain"). QIs listed with a minus sign may detract from task performance, while QIs with a plus sign may positively affect performance. The KABC-II record form has a summary page that permits QIs to be listed at a glance for all subtests. QIs can be used to facilitate checks of test validity and reliability as well as to inform observation of problem-solving processes. However, the QIs are not normed and should be compared to behaviors noted outside of the testing situation (Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005).

TABLE 18.3 Kaufman Assessment Battery for Children—Second Edition Core Interpretive Indices

Composite Indices	Description
Mental Processing Composite (MPC)	An aggregate index of information processing proficiency intended to emphasize problem-solving rather than acquired knowledge and skills
Fluid-Crystallized Index (FCI)	An index of general cognitive ability according to the CHC perspective and including measures of acquired knowledge
Nonverbal Index (NVI)	An estimate of cognitive ability based on task performances that may be administered in pantomime and responded to motorically; appropriate for students with limited English proficiency, speech or language impairments, or other language-related disabilities
Sequential Processing scale/Gsm	An index of cognitive processes that arrange input in sequential or serial order during problem solving, where each stimulus is linearly or temporally related to the previous one/Gsm = taking in and holding information and then using it within a few seconds
Simultaneous Processing scale/Gv	An index of proficiency at processing stimuli all at once, in an integrated manner interrelating each element into a perceptual whole/Gv = perceiving, storing, manipulating, and thinking with visual patterns
Learning/Glr	An index of ability to learn and retain new information with efficiency/Glr = storing and efficiently retrieving newly learned, or previously learned information
Planning/Gf	An index of proficiency at high-level, decision-making, executive processes/Gf = solving novel problems by using reasoning abilities such as induction and deduction
Knowledge/Gc	An index of breadth and depth of knowledge acquired from one's culture (included in the CHC model but not included in the Luria model).

Potential applications of the KABC-II include assessment of individuals with intellectual disability/mental retardation, ADHD, and learning disabilities as well as individuals who may be disadvantaged by a verbally loaded assessments (e.g., individuals who are deaf and hard of hearing, individuals with autism spectrum disorders, and individuals with speech and language disorders) (Kaufman et al., 2005). The KABC-II Manual reports clinical validity studies with children and adolescents diagnosed with specific learning disabilities (in reading, mathematics, written expression), intellectual disability, intellectual giftedness, autism spectrum disorder, ADHD, emotional disturbance, and hearing loss (Kaufman & Kaufman, 2004).

Strengths and Limitations

The K-ABC was introduced in 1983 with genuine innovations, including a theory-driven model of cognitive processing, careful minimization of acquired knowledge requirements, demonstration of reduced racial and ethnic group mean score differences, and conceptual links of assessment to intervention. Its subtests appeared qualitatively different from Wechsler and Binet-style procedures, it permitted examinees to be taught the task to ensure that no children do poorly because they did not understand what to do, and its easel-based test administration format quickly became the industry standard. Through various marketing campaigns the K-ABC was

> hyped by its publishers as a "revolutionary new way to define and measure intelligence!" . . . Doing away with the notion of IQ, the Kaufmans have designed the K-ABC to measure a purer form of "mental processing ability"; and in doing so, they claim, they have gone a long way toward minimizing the racial and cultural biases that plague existing tests. ("The K-ABC–Will It Be the SOMPA of the 80's?" 1984, pp. 9–10)

After the promotional efforts subsided, Kline, Snyder, and Castellanos (1996) drew some sobering lessons from the K-ABC, including (a) the need to analyze the processing demands in all tasks and scoring systems; (b) the need to critically examine test interpretive practices (i.e., subtest profile analysis) that have dubious validity; (c) the problematic position that IQ scores reflect ability (independent from achievement) even while touting unusually high correlations with achievement; and (d) the failure of efforts to match instruction to learning profiles in K-ABC's remedial model. At the very least, the advances in assessment practice made by the K-ABC may have been overshadowed by its failure to deliver on its initial promise.

The KABC-II appears to continue the advances of its predecessor edition and substantively address some limitations. Its administration remains examinee friendly and clinically flexible. It is conormed with a well-developed achievement test, the KTEA-II. Its dual theoretical model represents an unusual compromise between two opposing theoretical perspectives (CHC and Lurian) without doing damage to either. Its capacity to reduce racial/ethnic mean group score differences is reported more objectively than such findings were publicly represented on the K-ABC. (See, e.g., Kaufman et al., 2005, pp. 223–233.) Its system for identifying strengths and weaknesses is based largely on sets of subtests (scales) rather than individual subtests, thereby addressing concerns about K-ABC subtest profile analysis by Kline and his colleagues (1996) as well as criticisms of the practice by other researchers (e.g., Livingston, Jennings, Reynolds, & Gray, 2003; Macmann & Barnett, 1997; Watkins, 2000; Watkins & Canivez, 2004).

The KABC-II does have some potential weaknesses. Use of the dual theoretical model effectively causes the CHC framework to dominate the Lurian framework, and it is not clear how supportive research can be interpreted as supporting one theory or the other. The contents and processes involved in subtests are not always clear: for example, the KABC-II Manual documents instances in which subtests (e.g., the Rover subtest, see Kaufman & Kaufman, 2004) were assigned to different scales than originally intended based on factor-analytic findings. This type of discrepancy between test content and factor structure supports Kline and colleagues' (1996) call for detailed componential analyses of tasks and scoring systems, given the many cognitive processes that may be involved in cognitive test performance.

Reynolds Intellectual Assessment Scales

The Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003) offer a four- or six-subtest measure of general intelligence and two primary factors, verbal and nonverbal intelligence, normed for use with individuals between the ages of 3 years and 94 years. The RIAS is authored by Cecil R. Reynolds and Randy W. Kamphaus, both former students of Alan Kaufman at the University of Georgia and prolific authors and distinguished scholars in their own right. The RIAS is designed to be a practical and economical intelligence measure that requires no reading, motor coordination, or visual-motor speed.

The four intelligence subtests on the RIAS typically can be administered in less than a half hour through verbally administered or pictorial items appearing in a stimulus book. Instructions are brief and succinct. Following administration of one or two sample items, items are typically administered beginning at a start point, with the option to reverse, until a basal of two consecutive correct items is attained. Items are administered until a discontinue/end rule of two or three consecutive incorrect items is reached. For nonverbal subtests, the examinee is given two chances to respond to each item.

Raw subtest scores are converted to age-adjusted *T* scores ($M = 50$, $SD = 10$), which may be used to generate up to four composite index scores ($M = 100$, $SD = 15$). *T* scores for the two verbal subtests, Guess What and Verbal Reasoning, may be used to calculate a Verbal Intelligence Index (VIX). *T* scores for the two nonverbal subtests, Odd-Item Out and What's Missing, may be used to calculate a Nonverbal Intelligence Index (NIX). Results for the four intelligence subtests are used to calculate a Composite Intelligence Index (CIX). Two memory subtests, Verbal Memory and Nonverbal Memory, are available to supplement intelligence testing and generate a Composite Memory Index (CMX).

Theoretical Underpinnings

The RIAS was developed to provide efficient measurement in terms of time, cost, and yield. It emphasizes the general intelligence factor, *g,* as the most reliable factor in intelligence, measured on the RIAS with the CIX score. Drawing on the CHC framework, it also taps crystallized and fluid intelligence through its VIX and NIX scores, respectively. Reynolds and Kamphaus (2003) explained: "From our research, we have concluded that a strong measure of *g,* coupled with strong measures of verbal and nonverbal intelligence, account for nearly all of the reliable and interpretable variance in the subtests of good intelligence measures. Others have reached similar conclusions" (p. 10). The memory subtests and composite are optional but can be used to tap an additional CHC factor.

The four subtests that contribute to the CIX appear to be fair to good measures of psychometric *g*. Through exploratory principal factor analyses of the standardization sample and examination of the first unrotated factor, the *g* loadings of RIAS subtests ranged from .62 to .78 for ages 3 to 5 years, .49 to .81 for ages 6 to 11 years, .60 to .88 for ages 12 to 18 years, .66 to 87 for ages 19 to 54 years, and .69 to .85 for ages 55 to 94 years (Reynolds & Kamphaus, 2003). By Kaufman's (1994) criteria, the four RIAS subtests that contribute to CIX are good measures of general intelligence in about two-thirds of analyses conducted across five age groups. The two verbal subtests, Guess What and Verbal Reasoning, consistently have good *g* loadings ($\geq$.70), while the two nonverbal subtests do not fare quite as well. Across age groups, Odd-Item Out has fair to good *g* loadings, and What's Missing is consistently just fair (.50 to .69) in its *g* loading. Bracken (2005) concluded that the verbal subtests are the only consistently good measures of general intelligence. In an independent investigation, hierarchical exploratory factor analyses with the Schmid-Leiman procedure yielded fair *g*-loadings for the four core subtests across nearly all age ranges (Dombrowski et al., 2009).

Support for the RIAS verbal/nonverbal two-factor structure shows evidence of weakness. Exploratory factor analyses with principal factors of the four subtest core battery supports labeling the Guess What and Verbal Reasoning subtests as verbal (with factor pattern coefficients consistently greater than .65 across ages), but the Odd-Item Out subtest yields not-insubstantial loadings on the same factor from .30 to .50. A second factor is defined by factor pattern coefficients ranging from .53 to .67 on Odd-Item Out and What's Missing. However, the verbal subtests also show significant loading on this factor (.32 to .50). Nelson, Canivez, Lindstrom, and Hatt (2007) found from an independent sample that hierarchical exploratory factor analysis supported only the extraction of a general intelligence factor, accounting for the largest amount of subtest, total, and common variance. In another hierarchical exploratory factor analysis, Dombrowski et al. (2009) concluded of the RIAS factor structure: "The verbal subtests produced fair to poor factor loadings with the verbal factor, whereas the nonverbal subtests produced poor factor loadings on the nonverbal factor across all age ranges" (p. 501). Beaujean, McGlaughlin, and Margulies (2009) reanalyzed data from the standardization sample as well as from the J. M. Nelson, Canivez, Lindstrom, and Hatt (2007) sample, along with a new sample of referred cases, reporting that confirmatory factor analyses supported the two-factor structure of the RIAS although "the verbal factor showed much stronger invariance, construct reliability, and overall interpretability than did the nonverbal factor" (p. 932).

Standardization Features and Psychometric Adequacy

The RIAS was standardized from 1999 to 2002, with stratification targets based on 2001 Current Population Survey census figures. The national normative standardization sample consisted of 2,438 children, adolescents, and adults between the ages of 3 and 94 years. Sixteen age groups were established, beginning at 1-year intervals for ages 3 through 10, 2-year intervals for ages 11 to 16, and larger multiyear intervals for ages 17 through 94 years. A minimum of $n = 100$ participants were included in each interval, which may potentially be inadequate during the preschool and early school years when cognitive capacities are developing rapidly. The standardization sample exclusion criteria included color blindness, hearing or vision loss, alcohol or drug dependence, current treatment with psychotropic medication, any history of posttraumatic loss of consciousness, and any history of electroconvulsive therapy. The presence of a neuropsychiatric or psychoeducational disorder did not, however, lead to exclusion from participation in the normative standardization sample. The sample was stratified on the basis of sex, ethnicity, educational attainment, and geographic region. An examination of standardization sample demographic breakdowns suggests that while minorities as a whole were oversampled, several age groups show 3% to 4% underrepresentation for African Americans and Hispanic Americans. As a correction after stratification, standardization participants were weighted on an individual basis to precisely match census targets, and continuous norming procedures were applied to generate norms.

The internal consistency reliability of the RIAS subtest scores were computed with Cronbach's coefficient alpha. Median score reliabilities for the four RIAS intelligence subtests, and for the two memory subtests as well, range from .90 to .95 across age levels for the standardization sample. The composite scores (VIX, NIX, CIX, and CMX) show median reliabilities of .94 to .96. Reliabilities reported in the Professional Manual across gender and race are comparably high. Subtest internal consistencies are very high, sufficiently so to suggest that concern may be merited over excessively narrow, homogeneous content.

Test–retest stability coefficients for RIAS subtests were studied for $n = 86$ examinees across a broad age range from 3 through 82 years of age ($M = 11$ years, $SD = 15$) over a median time interval of 21 days (ranging from 9 to 39 days). Corrected test–retest subtest correlations ranged from .76 to .89, while corrected correlations for composite scores ranged from .83 to .91. These findings suggest that RIAS subtests appear reasonably stable over 3-week intervals. A comparison of test–retest mean composite scores suggests small practice effects, from about 2 to 4 points, over the 3-week interval.

In summary, examinations of RIAS score reliabilities indicate very high internal consistency and fairly adequate temporal stability. A small interscorer agreement study ($n = 35$) with only two raters indicated exceptionally high agreement ($r = .95$ to 1.00) between scorers of completed protocols.

In terms of floors and ceilings, the four RIAS intelligence subtests tend to have limited floors, good ceilings, and some gaps in score gradients. For preschool and school-age children, the subtests consistently extend at least two *SD*s below the normative mean of 50 *T,* down to 30 *T,* but there are some concerns with *T* score difficulty gradients. For example, answering two items on Verbal Reasoning at age 3 is enough to change resulting subtest scores by 12 *T* points, more than 1 *SD.* Across all school ages, the four intelligence subtest ceilings extend to just above +2 *SD* (e.g., What's Missing ceiling at age 17 is 71 *T*; Odd-Item Out ceiling is 73 *T*). For ages 7 to 8 years, when students are commonly selected for gifted and talented programs, RIAS subtest ceilings extend about 3 to 4 *SD*s above the normative mean, a ceiling that is certainly high enough to identify cognitively advanced students.

Correlations between RIAS scores and the WISC-III FSIQ are between .60 and .78. Correlations between RIAS scores and the WAIS-III FSIQ are above .70. Edwards and Paulin (2007) reported that, for a sample of young referred children, the correlation between the CIX and WISC-IV FSIQ was .90, while the correlation between the CIX and the WISC-IV GAI was also .90. Krach, Loe, Jones, and Farrally (2009) reported a corrected correlation of .75 between the RIAS CIX and the WJ III GIA.

Reynolds and Kamphaus (2003) report that the VIX correlates .86 and .44 respectively with the WISC-III VIQ and PIQ, while the NIX (in a pattern opposite to expectations) correlates .60 and .33 respectively with the VIQ and PIQ. The VIX correlates .71 and .61 with the WAIS-III VIQ and PIQ, while the NIX correlates .67 with the VIQ and .71 with the PIQ. These findings do not provide clear support for convergent validity of the NIX. Edwards and Paulin (2007) reported that the correlations between the RIAS VIX and WISC-IV VCI and PRI were .90 and .71, while the correlations between the NIX and the WISC-IV VCI and PRI were .53 and .72, somewhat more in line with expectations, although the magnitude of the correlations suggests that general intelligence pervades the VIX and NIX. Krach et al. (2009) reported that the VIX yields high corrected correlations of .88 and .64 with WJ III Gc and Gf, respectively, while NIX yields corrected correlations of .57 and .54 with Gc and Gf.

There are several isolated reports that RIAS yields significantly higher scores than the WISC-IV and WJ III Cog (Edwards & Paulin, 2007; Krach et al., 2009), although a validity study reported in the RIAS Professional Manual show that WISC-III yielded a significantly higher FSIQ than the RIAS CIX.

Interpretive Indices and Applications

RIAS subtest raw scores are converted to *T* scores, and age-equivalent scores corresponding to subtest raw scores are available. Subtest *T* scores may be summed to look up the composite standard scores. As reported, RIAS composites include the VIX, NIX, and CIX. The CMX may also be derived if the memory subtests are given. Composite indices are accompanied by percentile ranks and 90% and 95% confidence intervals. Table 18.4 contains descriptions of the basic composite indices.

The RIAS provides the option in Appendix I to include the memory subtests in the calculation of total composite

TABLE 18.4 Reynolds Intellectual Assessment Scales Core Interpretive Indices

Composite Indices	Description
Composite Intelligence Index (CIX)	Summary estimate of general cognitive ability
Verbal Intelligence Index (VIX)	An estimate of verbal reasoning ability and crystallized intellectual functioning
Nonverbal Intelligence Index (NIX)	An estimate of nonverbal reasoning ability and fluid intellectual functioning
Composite Memory Index (CMX)	An estimate of verbal and nonverbal memory functions for material that is meaningful, concrete, or abstract

scores. In this framework, the Total Verbal Battery (TVB) is based on the sum of *T* scores for the three verbal subtests, the Total Nonverbal Battery (TNB) is based on the sum of *T* scores for the three nonverbal subtests, and the Total Test Battery (TTB) is derived from the sum of *T* scores for all six RIAS subtests. These total battery scores are not recommended by the authors, and available research does not tend to support their use.

Subtests scores and composite scores may be compared on a pairwise basis, based on the statistical significance of differences and normative frequencies of discrepancies for a given age group in the standardization sample. For example, the statistical significance and cumulative frequency of a discrepancy between verbal and nonverbal intelligence (VIX and NIX) may be easily estimated. The discrepancy between composite intelligence and memory (CMX and CMX) is a traditional comparison used to help identify memory disorders.

Potential applications of the RIAS include identification of individuals with learning disability, intellectual disability/mental retardation, intellectual giftedness, neuropsychological impairment, memory impairment, and emotional disturbance (Reynolds & Kamphaus, 2003). The RIAS Professional Manual includes clinical samples diagnosed with intellectual disability/mental retardation, traumatic brain injury, stroke/cerebrovascular accident, seizure disorder, dementia, learning disabilities, anxiety disorders, depression, schizophrenia, bipolar disorder, and polysubstance abuse.

Strengths and Limitations

The RIAS represents a reliable, efficient, and economical measure of general intelligence. Dombrowski and Mrazik (2008) commented on its efficiency: “Although the RIAS may be expediently administered, it provides a global measure of intelligence consistent with tests more than twice its length” (p. 229). R. W. Elliott (2004) concluded that RIAS is a time and cost effective means of conducting intellectual assessments: “The RIAS will become a very popular measure of intelligence for school districts that are under serious pressure to provide measures of intelligence for determination of special education and program needs, but to do so at a lower cost and in a more rapid fashion” (p. 325).

The RIAS, however, has limitations in at least two major aspects of its validity: the degree to which its subtests all measure general intelligence and the difficulty supporting its two-factor verbal/nonverbal structure. While J. M. Nelson and his colleagues (2007) found that general intelligence accounted for a large amount of test variance, Bracken (2005) noted that the majority of RIAS subtests have loadings on the *g* factor that are considered fair or poor. The equivocal support for the factorial extraction of two factors (e.g., Dombrowski et al., 2009) is further complicated by Bracken’s (2005) observation that the NIX has correlations as high or higher with the Wechsler verbal subtests as with the Wechsler nonverbal/performance subtests. He was concerned that subtests often have a large (i.e., greater than or equal to .35) secondary loading on the opposite factor, undermining the distinctiveness of the verbal and nonverbal factors. Beaujean et al. (2009) concluded that “there appears to be a growing literature questioning the utility of interpreting the NIX score” (p. 948).

Stanford-Binet Intelligence Scales

The oldest of intelligence tests is the Stanford-Binet Intelligence Scales, now in its fifth edition (SB5; Roid, 2003a, 2003b, 2003c). The Stanford-Binet has a distinguished lineage, having been the only one of several adaptations of Binet’s 1911 scales to survive to the present time. According to Théodore Simon (cited by Wolf, 1973, p. 35), 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.” Terman (1877–1956) may arguably be considered the single person most responsible for spawning the testing industry that dominates contemporary intelligence and educational testing. The editions of the Stanford-Binet created by Terman (1916; Terman & Merrill, 1937) remain remarkable technical innovations even today. For example, the first executive function measure explicitly intended to measure planning and organization was Terman’s Ball-and-field test (Terman, 1916; see also Littman, 2004), eight decades before Naglieri and Das (1997a, 1997b) reintroduced executive functions to intelligence assessment.

A SB5 test session begins with administration of a nonverbal routing subtest (Nonverbal Fluid Reasoning) and a verbal routing subtest (Verbal Knowledge). Scores on these two subtests each provide guidance specifying which level to begin testing in the nonverbal and verbal batteries. In this manner, the SB5 lends itself to adaptive, tailored testing. It is also possible to use the routing tests as a short form, generating an Abbreviated IQ (ABIQ). The five nonverbal battery subtests are typically completed first, followed by the five verbal battery subtests, with a basal and ceiling ultimately completed for each subtest. A unique feature of the SB5 is that at different stages in development, the tasks, manipulatives, and procedures intended to tap the targeted construct (indicated by the name of the subtest) may change. The SB5 features age-appropriate tasks from 2 through 85+ years delivered in the classic spiral omnibus testing format in which a small number of items (clustered in testlets) from each of the core factors are presented at a given developmental level, before the sequence starts anew at the next developmental level.

Theoretical Underpinnings

The fourth edition of the Stanford-Binet, published in 1986, was the first intelligence test to adopt Cattell and Horn's fluid-crystallized model of cognitive abilities, with a hierarchical organization in which psychometric *g* was at the apex, and four broad group factors—crystallized ability, fluid-analytic ability, quantitative reasoning, and short-term memory—were at the second, subordinate level. The SB5 represents an ambitious effort to integrate the CHC model with the traditional verbal-nonverbal dichotomy by measuring each of five CHC factors (termed Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory, corresponding respectively to Gf, Gc, Gq, Gv, and Gsm) with separate verbal and nonverbal subtests. The SB5 deviates from the CHC approach through its inclusion of Quantitative Reasoning as a cognitive factor and its attempt to provide separate verbal and nonverbal measures of each CHC broad factor. The inclusion of Quantitative Reasoning provided continuity with the SB Fourth Edition and has a long-standing prominence among human cognitive abilities, although it was not accorded independent stratum 2 status in Carroll's (1993) three-stratum model. Still, it features prominently in other models of human cognitive abilities. A numerical facility factor appeared in Thurstone's (1938) primary mental ability structure; a quantitative reasoning factor (RQ) and broad mathematical factor (Gq) were identified by Carroll (1993, 2003); a numerical agency was specified by Cattell (1998); and a second-order quantitative knowledge (Gq) factor was identified by Horn (e.g., Horn & Blankson, 2005). There is somewhat less of a theoretical foundation for verbal and nonverbal measurement of each of the five factors, although each of the factors is thought to transcend sensory modality.

The SB5 appears to provide a sound measure of general intelligence. The *g*-loadings of individual SB5 subtests tend to be high (Roid, 2003c). Through principal axis factor analyses, the average *g* loadings of SB5 subtests across all ages are good (i.e., >.70) for nine of the 10 subtests; only Nonverbal Fluid Reasoning has a fair *g* loading (.66). Normally the Nonverbal Fluid Reasoning subtest (a matrix reasoning task) would be considered to be an optimal measure of psychometric *g*. From ages 2 to 5 years, 60% of SB5 subtests are good measures of *g* and 40% are fair measures. From ages 6 to 16 years, 60% to 70% of the subtests are good measures of *g* and the remaining subtest are fair measures of *g*. For adults, 100% of the subtests are good measures of *g* (Roid, 2003c).

Factor analyses of SB5 subtests provide strong support for interpretation of the general intelligence factor but little support for the five-factor structure and for the division of tasks into verbal and nonverbal modalities (e.g., Canivez, 2008; DiStefano & Dombrowski, 2006; Ward, Rothlisberg, McIntosh, & Bradley, 2011; Williams, McIntosh, Dixon, Newton, & Youman, 2010). An examination of test content provides possible explanations, such as the inclusion of pictorial absurdities tasks (which elicit a verbal response) among the SB5 nonverbal tasks, or the inclusion of an analogies task (dependent on word knowledge) at the highest levels of fluid reasoning. Moreover, the correlations between the VIQ and NVIQ in the entire standardization sample is unusually high ($r = .85$), providing further evidence that the verbal and nonverbal scales are not distinct. As a result, the SB5 theoretical and interpretive structure cannot be said to match its factor structure.

Standardization Features and Psychometric Adequacy

The SB5 was standardized from 2001 to 2002 on 4,800 children, adolescents, and adults in 30 age levels from 2 through 85+ years. The sampling plan provided for narrower age gradations during periods of rapid development or potential loss of function when greater cognitive change may be expected (i.e., during the preschool period and older adulthood). For the preschool ages 2 through 4 years, the age levels were divided into 6-month intervals; for child and adolescent ages 5 years through

16 years, 1-year intervals were used; for young adulthood, the intervals were 17 to 20, 21 to 25, and 26 to 29; middle adulthood intervals were by decade (30–39, 40–49, and 50–59); older adulthood (60+ years) was divided into 5-year intervals. The sample size at each age level was $n = 100$, with the exception of young adulthood, which was slightly more variable but still adequate.

The normative sample used 2001 U.S. Census figures to set sampling targets for these stratification variables: sex (balanced for most ages, except older examinees where females are more highly represented), race/ethnicity, geographic region, and socioeconomic level (level of education). Participants were excluded from the normative sample if they qualified for significant special education services, if they had any of a variety of severe medical conditions, if they had severe sensory or communication deficits, or if they were diagnosed with severe behavioral/emotional disturbance. Examinees with limited English proficiency were also excluded. Detailed breakdowns of stratification sampling sizes are reported in the SB5 Technical Manual (Roid, 2003c), and an examination of sampling proportions on the margins suggests that sampling representation was fairly accurate, compared to census proportions. There is, however, anecdotal evidence that examinees from a highly gifted special population study were added to the normative sample, with an unknown impact on mean scores (Andrew Carson, personal communication, March 9, 2010). Normative tables were produced with continuous norming methods that should correct for any vagaries at specific age levels.

The internal consistency score reliabilities of the SB5 subtests and composites appear fully adequate. Roid (2003c) reported that the average subtest score reliabilities computed with the split half method, with Spearman-Brown correction, range from .84 to .89. Average composite score reliabilities range from .90 to .98, with the Full Scale IQ yielding the highest internal consistency. Test consistency across racial/ethnic groups was compared using Feldt's (1969) F-test procedure, which yielded only a single statistically significant difference in internal consistency between Asian, Hispanic, White, and African American groups—specifically a finding that for ages 6 to 10, Hispanic groups show significantly higher reliability than Whites for the verbal subtests.

Test–retest stability coefficients were computed for four age cohorts over an interval ranging from 1 to 39 days (*Median* = 7 days). Corrected subtest test–retest correlations range from .76 to .91 for preschool children; .76 to .91 for school-age children and adolescents; .66 to .93 for adults; and .77 to .91 for older adults. The highest stability may generally be found on the Verbal Knowledge subtest, the lowest on Nonverbal Fluid Reasoning. Composite score stabilities across these cohorts range from .79 to .95, with the factor indices and ABIQ (each derived from a pair of subtests) generally being lowest, the Verbal IQ and Nonverbal IQ being higher, and the Full Scale IQ having the greatest stability. The Nonverbal IQ consistently has lower stability than the Verbal IQ, and the Fluid Reasoning Index consistently ranks near the least stable of the factor index scores.

Interscorer agreement studies were conducted by comparing three sets of ratings of polychotomously scored items (i.e., those scored as 0, 1, or 2) for clusters of items (testlets) appearing in four of the five SB5 factors. Across all polychotomous items, interscorer correlations ranged from .74 to .97, with an overall median of .90 (Roid, 2003c).

Stanford-Binet subtest floors begin to look adequate for the assessment of potentially disordered children by about ages 4 or 5 years. For 2-year-olds, subtest floors extend from −1.33 to −2.67 *SD* below the normative mean, based on the lowest scores corresponding to a raw score of 1. It is not until age 4 years, 4 months that norms for every subtest extend at least 2 *SD*s below the general population mean, to the range associated with developmental disabilities. Based on earning the lowest possible nonzero raw score on every subtest, however, composite test score floors extend down to a Full Scale IQ of 64 at age 2, fully adequate for assessing children with various impairments.

Stanford-Binet subtest ceiling scores consistently extend to scaled scores of 19 (+3 *SD*) for perfect scores for all subtests at all ages, yielding a Full Scale IQ of 160 if every item is successfully answered on the test. Roid (2003b) also offered an experimental alternative to the Full Scale IQ, called an Extended IQ (EXIQ), that is capable of describing *more* extreme scores extending up through 225. There is, however, no research as yet on this experimental index. Ceiling content may also be reason to be concerned with SB5 ceilings. For example, the Nonverbal Knowledge subtest has two items dependent on a narrow knowledge of geography at its highest level.

Cross-sectional growth curves for Rasch-based change-sensitive scores appear in the SB5 technical manual (Roid, 2003c) providing evidence of the relationship between SB5 factors and normative cognitive development. Although these scores are not directly comparable, they show that different cognitive abilities develop at different rates, that all develop rapidly through adolescence, that fluid reasoning peaks in young adulthood while acquired knowledge peaks in middle to late adulthood. Growth trends may be considered a unique way to demonstrate test construct validity.

Only confirmatory factor analyses are reported in the SB5 technical manual (Roid, 2003c), while a lengthy discussion (pp. 108–109) explains the conspicuous omission of exploratory factor analyses. Ironically, new exploratory factor analyses are actually reported for the much-older Stanford-Binet Form L (Roid, 2003c). Confirmatory factor analyses were conducted with two split-half versions of each subtest, altogether yielding 20 scores. Results were interpreted as supporting the verbal-nonverbal dichotomy as well as the five-factor structure.

Canivez (2008) conducted a hierarchical exploratory factor analysis of the SB5 standardization sample and found that large portions of total and common variance were accounted for by second-order, general intelligence, with no evidence for a five-factor (Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory) or two-factor (verbal/nonverbal) solution at any level:

> On balance, it appears that the SB-5 is a strong measure of general intelligence in children and adolescents, but little empirical evidence for additional factors was found. As such, clinicians would be wise to concentrate their interpretation on the overall global IQ score from the SB-5, even with the youngest age groups. (Canivez, 2008, pp. 539–540)

Similar concerns about SB5 factor structure have been reported by DiStefano and Dombrowski (2006), who conducted exploratory and confirmatory factor analyses with the standardization sample, concluding that the verbal/nonverbal domains were identifiable with subjects younger than 10 years of age whereas a single factor was readily identified with older age groups.

Williams and her colleagues (2010) conducted confirmatory factor analyses on a sample of 201 high-achieving third-grade students and concluded that a hierarchical, four-factor (fluid reasoning and knowledge were combined into a single factor), post hoc model provided the best fit to the data.

The SB5's overall convergence with other measures of intelligence and achievement appears fully adequate. The SB5 Full Scale IQ has an uncorrected correlation of $r = .90$ with the previous edition (SB IV) Composite Standard Age Score and .85 with the 1972 SB Form L-M IQ. The SB5 FSIQ has a corrected correlation of .83 with the WPPSI-R FSIQ, .84 with the WISC-III FSIQ, .82 with the WAIS-III FSIQ, and .78 with the WJ III Cog GIA. These correlations provide evidence of high convergent validity with other intelligence test composites. In addition, the SB5 FSIQ has corrected correlations ranging from .50 to .84 with WJ III achievement scores and a corrected correlation of $r = .80$ with the WIAT-II Total Achievement composite (Roid, 2003c).

Interpretive Indices and Applications

The Stanford-Binet yields 10 subtest scaled scores (with a mean of 10, *SD* of 3), five factor indices, a Nonverbal IQ (NVIQ) and a Verbal IQ (VIQ), and an overall composite score, the Full Scale IQ. Core composite interpretive indices, with a mean of 100 and *SD* of 15, appear in Table 18.5. The SB5 interpretive manual (Roid, 2003b) recommends a seven-step interpretive approach that first considers assumptions, purpose, and context, followed by interpretation of nonverbal versus verbal performance, interpretation of Full Scale IQ, interpretation of factor indexes, subtest profile analysis, and qualitative interpretation.

TABLE 18.5 Stanford-Binet Intelligence Scales—Fifth Edition Core Interpretive Indices

Composite/ Factor Indices	Description
Full Scale IQ	Overall cognitive functioning across a sample of verbal and nonverbal tasks
Nonverbal IQ	Estimate of functioning on problem-solving tasks with minimal or reduced language demands
Verbal IQ	Estimate of functioning on knowledge and problem-solving tasks with high language demands
Fluid Reasoning	An estimate of novel problem-solving facility, based on performance on a sample of nonverbal tasks (e.g., analysis of figural analogies and sequences) and verbal tasks (e.g., explanation of verbal absurdities, verbal analogies)
Knowledge	An estimate of previously acquired declarative and procedural learning, based on performance on a sample of nonverbal tasks (e.g., knowledge of how to perform simple tasks, identification of absurd or missing elements in pictures) and verbal tasks (ability to define words)
Quantitative Reasoning	An estimate of mathematical problem solving in words and pictures, ranging from elementary number concepts to higher order mathematical operations
Visual-Spatial Processing	An estimate of spatial cognitive ability, based on reproduction of visual-spatial patterns and capacity to verbally process spatial concepts like position and direction
Working Memory	Immediate and short-term recall/mental holding capacity, mental operating space, and the ability to mentally manipulate verbal and visual-spatial contents

Several additional SB5 scores may be of value. The change-sensitive scores (CSS) are Rasch-derived scores that measure performance on a developmental yardstick. Ironically, there is no evidence that change-sensitive scores are actually sensitive to the effects of cognitive, educational, or therapeutic interventions. Other scores, such as the EXIQ score that promises to raise composite IQ scores, should be considered experimental at this time.

Special population studies appearing in the SB5 technical manual include individuals diagnosed with ADHD, autism spectrum disorder, developmental disability, English language learners, intellectually gifted, intellectually disabled/mentally retarded, specific learning disabilities, speech and language impairment, deafness and hard of hearing, and serious emotional disturbance (Roid, 2003c).

Strengths and Limitations

The SB5 is a fast-moving, engaging, sound measure of general intelligence. In praise of its administrative qualities, Bain and Allin (2005) commented:

> The SB5 represents some departures from previous editions in organization and content while retaining some of the item variation and charm that made administration of the earlier SB L-M a pleasure, particularly with young children. We found administration time using the SB5 to be briefer than the SB-IV... Pacing of subtests is easy, as long as manipulative items are well organized. The younger children we have observed during testing have remained interested in the colorful test stimuli throughout. (p. 94)

The limitations of the SB5 stem principally from its structural difficulties and its yet-to-be-demonstrated clinical relevance. Factor analyses do not support its five-factor interpretive structure or the verbal/nonverbal interpretive dichotomy (e.g., Canivez, 2008). Consequently, the SB5 appears to represent a bold (but failed) effort to expand the CHC model by providing both verbal and nonverbal measures for each of five factors. It is not entirely clear whether the conceptualization or execution of the test was the failure, because some "nonverbal" tests actually require verbal responses, an obvious failure in developing nonverbal tests. Alternatively, CHC constructs may simply not permit equivalent verbal and nonverbal measurement, the clearest evidence of which is the prevailing tendency for all crystallized tests to be mainly verbal, all fluid ability tests to be mainly visual-spatial and nonverbal, and all short-term/working memory tests to be auditory-verbal but not spatial. If the interpretive structure of the SB5 lacks validity, as research seems to suggest, than it is invalid to interpret any indices but the FSIQ in clinical and educational decision making.

Wechsler Intelligence Scales

No brand name in psychology is better known than Wechsler, now applied to a series of intelligence scales spanning the ages 2½ through 90 years, an adult memory scale covering ages 16 through 90 years, and an achievement test covering ages 4 through 50 years as well as several ancillary tests. The remarkable success of the Wechsler measures is attributable to David Wechsler (1896–1981), a pioneering clinician and psychometrician with a well-developed sense of what was practical and clinically relevant. Decades after Wechsler's death, his tests continue to dominate intellectual assessment among psychologists (Camara et al., 2000).

Wechsler's role in the history of intelligence assessment is beginning to be critically assessed (e.g., Wasserman, 2012), but the origins of his subtests can be readily traced to testing procedures developed from the 1880s through the 1930s (e.g., Boake, 2002). Wechsler was introduced to most of the procedures that would eventually find a home in his intelligence and memory scales as a graduate student at Columbia University (with faculty including J. McKeen Cattell, Edward L. Thorndike, and Robert S. Woodworth) and as an army mental examiner in World War I. He developed his first test battery for his master's thesis completed in 1917, establishing a pattern he would later follow of appropriating practical and clinically useful procedures from other authors, making slight improvements and modifications, and synthesizing them into a streamlined battery of his own. During his military service in 1918, he learned the group-administered Army mental tests (the Army alpha and beta) as well as leading individual intelligence and performance tests, testing many recruits who suffered from limited English proficiency and illiteracy. 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). Matarazzo (1981) related that Wechsler realized the value of individual intelligence assessment after seeing recruits who functioned quite adequately in civilian life in spite of subnormal results on the group-administered tests. As part of an educational program intended for American soldiers serving overseas, Wechsler attended the University of London in 1919, where he spent some 3 months studying with Charles E. Spearman, an experience that impressed him deeply. Later Wechsler sought training from several of the leading clinicians of his day, including Augusta F. Bronner and William Healy at the Judge Baker Foundation in Boston and Anna Freud at the Vienna Psychoanalytic Institute (for 3 months in 1932). By virtue of his education and training, Wechsler should properly be remembered as one of the first scientist-clinicians in psychology.

Wechsler first introduced the Bellevue Intelligence Tests (later named the Wechsler-Bellevue; Wechsler, 1939), followed by 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). With some variation, these tests all utilize the same core set of subtests and interpretive scores. The most recent editions of Wechsler's intelligence scales are the WPPSI-III (Wechsler, 2002), the WISC-IV (Wechsler, 2003a, 2003b), the WAIS-IV (Wechsler, 2008a, 2008b), and a two- or four-subtest short form named the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2012). Since Wechsler's death, these updates have generally been developed by research and development psychologists working with expert advisory panels. The WISC-IV and WAIS-IV are emphasized in this section.

Theoretical Underpinnings

The Wechsler intelligence scales are decidedly atheoretical, beyond their emphasis on psychometric *g,* and in recent years they have appeared to be a test in search of a theory. As originally conceptualized by David Wechsler (1939), they were clearly intended to tap Spearman's general intelligence factor, *g*: "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). Wechsler purposefully included a diverse range of tasks so as to avoid placing disproportionate emphasis on any single ability: "My definition of intelligence is that it's not equivalent to any single ability, it's a global capacity . . . The tests themselves are only modes of communication" (Wechsler, 1975, p. 55). Wechsler kept in contact with Spearman long after World War I, even attempting (unsuccessfully) to identify a general emotional factor as a parallel to the general intellectual factor (Wechsler, 1925). In 1939 Wechsler wrote that Spearman's theory and its proofs constitute "one of the great discoveries of psychology" (p. 6).

Wechsler did not believe that division of his intelligence scales into verbal and performance subtests tapped separate dimensions of intelligence; rather he felt that this dichotomy was diagnostically useful (e.g., Wechsler, 1967). In essence, the verbal and performance scales constituted different means by which *g* could be assessed. Late in his life Wechsler described the verbal and performance tests merely as ways to "converse" with a person—that is, "to appraise a person in as many different modalities as possible" (Wechsler, 1975, p. 55). Wechsler's intelligence scales sought to capitalize on the established preferences of practitioners to administer both verbal and performance tests as part of a comprehensive assessment, and by packaging both sets of measurements in a single test battery, he was able to meet the needs of applied psychologists. It was not his intent to treat verbal and performance IQ as independent dimensions of intelligence:

> It was not until the publication of the Bellevue Scales that any consistent attempt was made to integrate performance and verbal tests into a single measure of intelligence. The Bellevue tests have had increasingly wider use, but I regret to report that their popularity seems to derive, not from the fact that they make possible a single global rating, but because they enable the examiner to obtain separate verbal and performance I.Q.'s with one test. (Wechsler, 1950, p. 80)

Wechsler was cognizant of multifactor models of intelligence, but he placed little emphasis on them in his writings because after the contribution of the general factor of intelligence was removed, the group factors (e.g., verbal, spatial, memory) accounted for little variance in performance (e.g., Wechsler, 1961). Wechsler also rejected the separation of abilities because he saw intelligence as resulting from the collective integration and connectivity of separate neural processes. He believed that the integrative function of intelligence would never be localized in the brain and observed, "While intellectual abilities can be shown to contain several independent factors, intelligence cannot be so broken up" (Wechsler, 1958, p. 23). Even so, Wechsler (1950) acknowledged the validity of Thurstone's (1938) primary mental abilities:

> What are the elements which factor analysis has shown our intelligence tests to measure? . . . By the use of his [Thurstone's] expanded technique, it has now been shown that intelligence tests, such as they are, contain not one but several independent factors. Some five or six have been definitely identified; they are, to repeat, induction, verbal, spatial,

> numerical, and one or two other factors. Notice, however, that these factors, like Spearman's eduction, are all cognitive. (p. 80)

Following Wechsler's death in 1981, the tests have slowly gravitated toward a multifactor interpretive model. Coverage expanded to four factors in the 1991 WISC-III (verbal-comprehension, perceptual-organization, freedom from distractibility, and processing speed) and four factors in the 1997 WAIS-III (verbal-comprehension, perceptual-organization, working memory, and processing speed). With the publication of the WISC-IV (Wechsler, 2003a, 2003b) and the WAIS-IV (Wechsler, 2008a, 2008b), both major tests now feature four identical factors (verbal comprehension, perceptual reasoning, working memory, and processing speed), and the verbal and performance IQs are no longer computed.

Standardization Features and Psychometric Adequacy

The Wechsler scales are renowned for their rigorous standardizations, with new editions being released about every 10 to 15 years. The Wechsler scales tend to utilize a demographically stratified (and quasi-random) sampling approach, collecting a sample at most age levels of about $n = 200$ usually divided equally by sex.

The WISC-IV normative sample consisted of 2,200 children and adolescents in 11 age groups between the ages of 6 years and 16 years. Unexpectedly, the Arithmetic subtest was normed on only 1,100 participants, half of the normative sample (Wechsler, 2003b). The normative sample was stratified according to race/ethnicity, parent education level, and geographic region. Some 5% to 6% of participants from special population studies were added to the normative sample. At each age, an equal number of males and females were tested. The WAIS-IV normative sample included 2,200 participants in 13 age groups, with 200 examinees in all but the four oldest age groups (70–74, 75–79, 80–84, and 85–90), which consisted of 100 examinees in each group (Wechsler, 2008b). The sample was stratified according to race/ethnicity, educational level, and geographic region. The gender composition of the five oldest age groups was based on census proportions.

The technical manuals for the Wechsler scales report the demographic breakdown of the normative samples in tables that permit several stratification variables to be examined at once in cross-tabulations (e.g., Percentages of the Standardization Sample and U.S. Population by Age, Race/Ethnicity, and Parent Education Level). These cross-tabulations make it possible to verify that stratification characteristics were accurately and proportionally distributed across groups rather than concentrated in a single group. A review of these tables for the WISC-IV and WAIS-IV indicates a fairly close correspondence to census proportions.

Internal consistency with coefficient alpha tends to be fully adequate for the Wechsler subtests and composite scales. On the WISC-IV, all subtest scaled scores had an average reliability coefficient across age groups of .80 or greater, and all composite indexes had reliability coefficients of .90 or greater. B. Thompson (2005) commented on enhanced WISC-IV reliability coefficients relative to the previous edition: "Most reliability coefficients for WISC-IV scores on the 10 retained subtests improved substantially (e.g., Arithmetic, from .78 to .88), and scores on the five new WISC-IV subtests tended to have reliability coefficients (.79 to .90) higher than those for WISC-III scores" (p. 263). For the WAIS-IV, every subtest yields an average reliability coefficient of .80 or greater, and every composite yields an average reliability coefficient of .90 or greater. Speeded subtests (Coding, Symbol Search, and Cancellation) do not lend themselves to internal consistency analyses, so test–retest stability coefficients were used to estimate reliability. Stability coefficients were slightly below .80 for Cancellation (WISC-IV and WAIS-IV) and Symbol Search (WISC-IV only), and the Processing Speed Index was slightly below .90 for the WISC-IV but not the WAIS-IV. It is noteworthy that process scores for Cancellation (Random and Structured) have stability coefficients that are significantly lower than .80 on the WISC-IV, so these scores should be interpreted with some caution.

The WAIS-IV takes a major step forward for the Wechsler scales by reporting subtest reliability coefficients for special population studies, the first step in reliability generalization. While the sample sizes used to calculate reliability coefficients in these studies are relatively small, there are several findings of interest. For example, the low coefficient alpha of .70 for WAIS-IV Letter-Number Sequencing in the small reading disordered sample ($n = 34$) may suggest that this subtest behaves less consistently for at least some learning-disabled individuals (perhaps due to reduced phonological awareness); not surprisingly, Letter-Number Sequencing shows the greatest effect size standard difference for any subtest ($ES = 1.03$) when a reading-disordered group and matched control group are compared.

Test–retest stability tends to be fairly adequate for WISC-IV and WAIS-IV subtests and composite indices, although some subtests (and nearly all process indices) have less than optimal stability. A sample of $n = 243$

participants underwent testing and retesting with the WISC-IV, with an average test–retest interval of 32 days. A sample of $n = 298$ participants underwent retesting with the WAIS-IV over a mean test–retest interval of 22 days. Across all ages for the WISC-IV, the corrected stability coefficient was greater than or equal to .80 for 12 of 15 subtests (with only Picture Concepts being substantially lower), below this benchmark for all process scores, and at or above .90 for two of five composite indices (with the PSI being substantially lower at .86). Across all ages for the WAIS-IV, the corrected stability coefficients were greater than or equal to .80 for 10 of 15 subtests (with Matrix Reasoning and Visual Puzzles substantially lower), below this benchmark for all process scores, and at or above .90 for two of five composite indices (with the remaining composite scores approaching this level).

As with previous editions of the Wechsler intelligence scales, practice effects over several weeks are reported. Over an average interval of 3 to 4 weeks, the most pronounced average practice effects on the Wechsler scales may be expected for Picture Completion (+1.8 scaled score points for the WISC-IV and +1.9 for the WAIS-IV), the Processing Speed Index (+10.9 for the WISC-IV and +4.4 for the WAIS-IV), and FSIQ (+8.3 for the WISC-IV and +4.3 for the WAIS-IV). These findings are much improved relative to older editions of the Wechsler scales and do not provide support for the commonplace recommendation that individuals undergo testing only once a year.

A final issue about the temporal stability of intelligence test results merits discussion, specifically the poor reliability of subtest profile analyses. In the last two decades, the dismal reliability of Wechsler subtest profiles (i.e., patterns of intraindividual strengths and weaknesses) has been definitively established, leading researchers to conclude that the stability of subtest profile configurations is too low for use in clinical and educational decision making (e.g., Livingston et al., 2003; Macmann & Barnett, 1997; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; Watkins, 2000; Watkins & Canivez, 2004). While intraindividual subtest profile analysis to identify relative strengths and weaknesses is still featured in most test interpretive manuals, the reliability and validity of composite scores is substantially higher and more suitable for decision making (e.g., Watkins, 2003).

In exploratory analyses documented in the WISC-IV *Technical and Interpretive Manual* (Wechsler, 2003b), the factor structure of the WISC-IV yields four clear factors (verbal comprehension, perceptual reasoning, working memory, and processing speed) that correspond to the four interpretive indices (VCI, PRI, WMI, PSI). Watkins (2006) conducted a hierarchical exploratory factor analysis with the WISC-IV standardization sample and reported a prominent general factor, *g*, that accounted for the greatest amount of common and total variance. Altogether, the general factor and four first-order factors accounted for 53.7% of the total variance. Watkins, Wilson, Kotz, Carbone, and Babula (2006) demonstrated for a clinical sample that the general factor accounts for more than 75% of the common variance, providing compelling evidence that there is considerable value in interpreting the WISC-IV at the general ability level.

Confirmatory factor analyses with multiple fit indices are reported in the technical manual for the WISC-IV and indicate that a four-factor model tends to best fit the data across multiple age groups (Wechsler, 2003b). A model with a fifth factor (arithmetic reasoning) is viable but offers little improvement over the four-factor model. Keith, Fine, Taub, Reynolds, and Kranzler (2006) conducted confirmatory factor analyses on the WISC-IV normative standardization sample and argued that a CHC-derived five-factor theoretical structure better describes the abilities underlying the WISC-IV than the four-factor model. They suggested that perceptual reasoning subtests measure a mixture of visual processing (Gv) and fluid reasoning (Gf) and that the Arithmetic subtest measures fluid reasoning more than working memory. Bodin, Pardini, Burns, and Stevens (2009) conducted confirmatory analyses with a clinical sample and reported that the second-order general intelligence factor accounted for the largest proportion of variance in the first-order latent factors and in the individual subtests. With a sample of clinically referred students, Watkins (2010) conducted confirmatory factor analyses and reported that the general factor accounted for the predominant amount of variation among the subtests, accounting for 48% of the total variance and 75% of the common variance. Watkin (2010) concluded that the general intelligence score should be the primary basis of test score interpretation.

For the WAIS-IV, the technical manual reported no exploratory factor analyses (Wechsler, 2008b). Canivez (2010) noted that fundamental psychometrics that directly affect test interpretation, such as the proportions of variance accounted for by the higher-order g-factor and the four first-order factors, as well as subtest specificity estimates, are also not reported in WAIS-IV technical materials. Accordingly, Canivez and Watkins (2010a, 2010b) conducted hierarchical exploratory factor analyses and reported that the general factor accounted for major portions of total and common variance and that all WAIS-IV subtests were associated with their theoretically proposed first-order factors. They further noted that the four factors accounted for small amounts of the total and common variance after a general factor was extracted. Confirmatory factor analyses reported in the WAIS-IV technical manual show that four-factor models fit the standardization data well, but that better fit is achieved when two split loadings are specified (i.e., Arithmetic is allowed to load on both the verbal comprehension and working memory factors, and Figure Weights is allowed to load on both the perceptual reasoning and working memory factors), and covariance of error is permitted for subtests that share method variance (i.e., Digit Span and Letter-Number Sequencing) (Wechsler, 2008b).

Convergent validity studies indicate that the Wechsler scales are strongly related to scores from intelligence tests and achievement tests. The WISC-IV FSIQ yields corrected correlations of .91 with the WAIS-IV FSIQ (Wechsler, 2008b), .86 with the WASI four-subtest FSIQ (Wechsler 2003b), and .84 with the DAS-II GCA (C. D. Elliott, 2007b). The WISC-IV FSIQ also has a corrected correlation of .82 with WIAT-III Total Achievement composite, and the FSIQ has correlations between .63 and .75 with all but one WIAT-III composite (Pearson research staff, 2009). Likewise, the WAIS-IV FSIQ correlates at .82 with WIAT-III Total Achievement composite, with correlations between .59 and .80 for all WIAT-III composites (Pearson research staff, 2009).

WISC-IV and WAIS-IV subtest floors and ceilings are improved relative to previous editions. Defined as a scaled score corresponding to a raw score of 1, subtest floors on the WISC-IV extend 2 *SD*s below the normative mean for all subtests and 3 *SD*s below the normative mean for one third of the subtests for the youngest age group (for which floor issues are particularly salient), yielding a floor FSIQ in the 40s (Wechsler, 2003a). The 2008 publication of extended WISC-IV norms raised the subtest scaled score ceiling from 19 to 28 and composite standard scores from 160 to 210 (Zhu, Cayton, Weiss, & Gabel, 2008). For the WAIS-IV oldest age group, 100% of subtests extend 2 *SD*s below the normative mean, and half of the subtests extend 3 SDs below the normative mean (Wechsler, 2008a). The WAIS-IV FSIQ ranges from 40 to 160.

Interpretive Indices and Applications

Wechsler is reported to have administered and interpreted his own tests in a clinically flexible way that would be considered unacceptable today. For example, in practice he was known to administer the Vocabulary in isolation to estimate intelligence and personality (Adam F. Wechsler, personal communication, December 3, 1993). Weider (1995) reported, “He never gave the Wechsler the same way twice” and considered the standardization of his tests to be imposed on him by the test publisher. Kaufman (1994) has described Wechsler’s clinical approach to interpreting the scales, along with his interest in qualitative aspects of examinee responses to emotionally loaded verbal and pictorial stimuli. One needs only to read Wechsler’s (1939) *Measurement of Adult Intelligence* to see that he qualitatively interpreted every test behavior, every item response, every response error, and every problem-solving strategy.

Even so, most contemporary interpretive guidelines emphasize an objective psychometric approach to interpretation of the Wechsler scales, beginning with the FSIQ, followed by the four factor scores, then profile analysis of strengths and weaknesses, and finally process analyses (see e.g., Wechsler, 2008b). The composite index scores that constitute the foundation for Wechsler scale interpretation appear in Table 18.6; these scores are the most reliable scores for interpretation.

TABLE 18.6 Wechsler Intelligence Scales (WISC-IV and WAIS-IV) Core Interpretive Indices

Composite Indices	Description
Full Scale IQ (FSIQ)	Estimate of overall cognitive functioning across a sample of verbal and nonverbal tasks; weighted to emphasize reasoning and knowledge (60%) over information processing speed and capacity (40%)
General Ability Index (GAI)	Estimate of general cognitive functioning based on performance of subtests associated with reasoning and knowledge, i.e., those strongly associated with psychometric *g*
Cognitive Proficiency Index (CPI)	Estimate of information processing efficiency, based on performance on subtests associated with processing speed and capacity (working memory); available for WISC-IV only
Verbal Comprehension Index (VCI)	Estimate of verbal reasoning and knowledge abilities, based on responses to language-based tasks requiring abstract problem solving and retrieval of previous learning
Perceptual Reasoning Index (PRI)	Performance on tasks requiring abstract visual-spatial perception, processing, and reasoning
Working Memory Index (WMI)	Auditory immediate/working memory capacity, dependent on capacity and complexity of mental operations as well as facility with number processing
Processing Speed Index (PSI)	Speed of performance on psychomotor tasks with low difficulty; nonspecifically sensitive to any source of disruptions in cognitive processing efficiency

The limitations of subtest profile analyses have already been described, although this approach remains common and recommended by some authorities in applied practice (e.g., Flanagan & Kaufman, 2009).

Two new composite indices of intelligence are available for the WISC-IV (Raiford, Weiss, Rolfhus, & Coalson, 2005; Saklofske, Weiss, Raiford, & Prifitera, 2006; Weiss, Saklofske, Schwartz, Prifitera, & Courville, 2006). The General Ability Index (GAI) provides an estimate of general intelligence or ability without penalizing for cognitive inefficiencies, such as reduced processing capacity or speed. The GAI includes subtests thought to have high *g* saturation and thereby emphasizes reasoning ability and acquired knowledge over processing capacity and speed. The GAI is also available for the WAIS-IV (Wechsler, 2008b). The Cognitive Proficiency Index (CPI) provides an estimate of mental efficiency (i.e., how much and how quickly an individual can process information). When information processing efficiency is low, the GAI may provide a more accurate estimate of reasoning and problem-solving ability than the WISC-IV FSIQ.

A few aspects of process-based test interpretation are built into the main Wechsler scales (e.g., Block Design No Time Bonus [BDN], Digit Span subtest fractionation [DSF, DSB, and DSS], and Cancellation subtest fractionation [CAS versus CAR]). Still, the process approach is exemplified in a companion battery entitled WISC-IV Integrated (Wechsler et al., 2004), which merges the WISC-IV with standardized measures of test behavior, problem-solving style, and cognitive processes, building on process-based test procedures first introduced in the WAIS-R as a Neuropsychological Instrument (WAIS-R NI; Kaplan, Fein, Morris, & Delis, 1991) and the WISC-III as a Process Instrument (WISC-III PI; Kaplan et al., 1999). The WISC-IV Integrated includes optional process-based subtests in each of the four factor-defined domains on the WISC-IV (i.e., multiple choice versions of verbal subtests; Block Design Multiple Choice, Block Design Process Approach, Elithorn Mazes; Visual Digit Span, Spatial Span, Letter Span, Letter-Number Sequencing Process Approach, Arithmetic Process Approach, Written Arithmetic; and Coding Recall and Coding Copy). The multiple-choice subtest procedures are typically administered after the standard (free-recall) verbal subtests have been administered, in an effort to discern whether difficulties are attributable to memory retrieval problems (poor performance in free recall but better performance in multiple choice recognition) or a failure to have ever encoded the knowledge in long-term memory (poor performance in both free-recall and multiple-choice recognition). The child's visual search pattern on the Cancellation subtest may be used to assign a process observation score for each item (Cancellation Random Strategy [CARS] and Cancellation Structured Strategy [CASS]), thereby facilitating comparison of planfulness and accuracy in both disorganized and organized situations. Likewise, observation of errors on Block Design permits objective norm-referenced measurement of breaks in the square 2×2 or 3×3 design configurations, which are associated with more severe visual-spatial processing deficits. WISC-IV Integrated options for the Coding subtest permit clarification as to whether difficulties were due to impaired associative learning and memory (Coding Recall) or psychomotor speed (Coding Copy). The WISC-IV Integrated also includes base rate information for the normative number of *Don't Know, No Response, Self-Correction, Repetition,* and *Prompt* response occurrences observed during specific procedures in the test session. The normative frequencies for specific test observations may potentially support various diagnostic inferences, such as an increased number of requests for item repetition in individuals diagnosed with attention deficit or language disorders. Several supplemental subtests offer additional information that extends beyond the traditional Wechsler framework for interpretation, For example, the optional Elithorn Mazes subtest is an explicit measure of planning performance and efficiency, and the Spatial Span subtest is intended as a nonverbal analogue to Digit Span. Ultimately, the WISC-IV Integrated is intended to clarify the nature of any spared and impaired cognitive processes, facilitate identification of problem-solving strategies, and enhance the interpretation of both correct and incorrect responses (e.g., Kaplan, 1988). It remains to note that the usefulness of process-based approaches has yet to be fully demonstrated in the research literature.

Supplemental tasks and procedures designed to facilitate WAIS-IV interpretation may be found in *Advanced Clinical Solutions for the WAIS-IV and WMS-IV* (ACS; Pearson, 2009a). The ACS contains tasks, procedures, and scores that yield information about cognitive processes including new subtests tapping social cognition and motivation/effort, as well as procedures for adjusting standard scores according to demographic characteristics, measuring clinically significant change across serial WAIS-IV administrations, and predicting premorbid cognitive abilities in order to quantify how much cognitive functioning may have been lost from disorders such as traumatic brain injury or Alzheimer's disease.

Strengths and Limitations

The Wechsler intelligence scales offer reliable and valid measurement of general intelligence and functioning in the areas of verbal and nonverbal intellectual ability, information processing capacity (i.e., working memory) and information processing speed. Since David Wechsler's death, his publisher continues to revise his tests according to high psychometric standards. In a review of the WISC-IV, B. Thompson (2005) commented, "Obviously, considerable resources have been invested in developing the present Wechsler revision. The marriage of resources and reflection inexorably yields impressive progeny" (p. 263). Likewise, the WAIS-IV is considered by reviewers as "one of the best measures of general intellectual functioning available . . . [I]t is extremely comprehensive and provides a reliable and valid measure of intellectual functioning relative to the demands of schooling and academic success" (Canivez, 2010, p. 688).

Yet as the industry leader and most researched of intelligence tests, the Wechsler scales have become the face of intelligence testing, and most appraisals of intelligence testing, whether positive or negative, seem to reflect the qualities (and limitations) of these scales. Among the weaknesses of the Wechsler scales are their limited coverage of the CHC broad factors (e.g., Flanagan & Kaufman, 2009), their continued support of intraindividual subtest profile analyses in spite of considerable evidence against this practice, their shrinking relevance to diagnosis (see the "Diagnostic Applications" section later in the chapter, specifically the deemphasis of intelligence tests in the identification of specific learning disabilities), and their irrelevance for intervention planning. Swanson (2010) observed that "the link between IQ and teaching is obscure (to some it may be nonexistent)" (p. 1), and it may be argued that this failure may be laid at the feet of the Wechsler scales. At the same time, efforts to keep the Wechsler scales fresh and innovative, as by extending WISC-IV norms and developing experimental products like the WISC-IV Integrated and Advanced Clinical Solutions, may help the Wechsler brand remain contemporary while advancing the science and practice of intelligence assessment.

Woodcock-Johnson Tests of Cognitive Abilities

The Woodcock-Johnson III Normative Update Tests of Cognitive Abilities (WJ III NU Cog; Woodcock, McGrew, & Mather, 2001a, 2007a) represent the most recent revision of an assessment battery with prior editions from 1977 and 1989. Normed for use from ages 2 through 90 plus years, the WJ III NU Cog is conormed with a leading achievement test, the WJ III NU Tests of Achievement (WJ III NU Ach; Woodcock, McGrew, & Mather, 2001b, 2007b). The battery's origins may be traced back to Richard W. Woodcock's early adult employment in a sawmill and a butcher shop after completion of his World War II navy military service. Upon reading Wechsler's *Measurement and Appraisal of Adult Intelligence*, Woodcock was inspired to study psychology; he quit his jobs and joined the Veteran's Testing Bureau for a wage of 55¢ per hour. Woodcock began active development of the WJ Cog in 1963 in a series of controlled learning experiments that led to development of measures of broad retrieval (now termed Glr) ability. In 1972 Woodcock formed the test development company, Measurement Learning Consultants (MLC), which developed most of his tests. During a 1974–1975 fellowship in neuropsychology at Tufts University, he created an adaptation of the Halstead Category Test that continues to be used as a measure of fluid reasoning (now termed Gf) in the WJ III NU Cog. The original Woodcock-Johnson Psycho-Educational Battery (Woodcock & Johnson, 1977) was envisioned by its primary author as being part of a comprehensive assessment system that measured cognitive abilities and aptitudes, academic achievement skills, scholastic and nonscholastic interests (the Tests of Interest Level), and adaptive functioning (separately measured by the Scales of Independent Behavior) (Woodcock, 1977). The interest inventory was dropped for subsequent editions. The 1989 Woodcock-Johnson Psycho-Educational Battery—Revised (Woodcock & Johnson, 1989, 1990) had the distinction of being the second cognitive test based on the Cattell-Horn theory of fluid and crystallized abilities (after the 1986 Stanford-Binet, Fourth Edition), not wholly surprising since consultants to its development included John L. Horn and John B. Carroll. In 2001 the third edition of the Woodcock-Johnson was published, with the WJ III Cog positioned as an intelligence test tapping general intellectual ability and specific cognitive abilities (Woodcock et al., 2001a, 2007a). A normative update (NU) edition was published in 2007.

Originally finding its primary audience with educators and best known for its companion achievement tests, the WJ III NU Cog is increasingly being utilized by psychologists in educational settings due to its CHC structural underpinnings as well as the popularity of the WJ III NU Ach. The WJ III NU Cog consists of two batteries: a 10-test standard battery and a 20-test extended battery. The Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities (Woodcock, McGrew, Mather, & Schrank, 2003, 2007) includes an additional 11 tests. All items are administered from an easel or compact disc/audiotape. The WJ III NU Cog requires computer scoring and cannot be scored by hand. A parallel Spanish-language battery of cognitive and achievement tests is available in the 2011 Batería III Woodcock-Muñoz NU.

Theoretical Underpinnings

The WJ III NU Tests of Cognitive Abilities is based on the CHC theory of cognitive abilities. The theory is a hierarchical, multiple-stratum model with *g* or general intelligence at the apex (or highest stratum), between seven to 10 broad factors of intelligence at the second stratum, and at least 69 narrow factors at the first stratum. The model has been termed an integrated or synthesized CHC framework (McGrew, 1997; McGrew & Flanagan, 1998), and it forms the basis for the cross-battery approach to cognitive assessment (e.g., Flanagan, Ortiz, & Alfonso, 2007). The WJ III NU Cog taps seven broad cognitive abilities in the CHC framework: Gc, Glr, Gv, Ga, Gf, Gs, and Gsm. When the achievement battery is also examined, the additional factors of Gq (mathematics) and language emerge (Carroll, 2003).

Individual tests are differentially weighted in the calculation of the General Intellectual Ability (GIA) score, an estimate of psychometric *g*. The WJ III NU technical manual (McGrew et al., 2007) reports the smoothed *g* weights for individual subtests, and the largest contribution to GIA consistently comes from the Verbal Comprehension test, a measure of crystallized knowledge. The weights assigned to other tests vary by age and battery. Fluid reasoning tests are never the most weighted tests, representing a major point of departure from prior investigations (e.g., Carroll, 1993; Gustafsson, 1984, 1988; Undheim, 1981) establishing Gf as the most substantial contributor to *g*. In practical terms, this finding expresses the idea that on the WJ III NU Cog, the facts you know contribute more to your intelligence than your ability to reason. It is possible that the unexpected performance of Gf in the WJ III NU Cog may be attributed to its inadequate measurement (Carroll, 2003).

A limitation of the CHC framework is that it is more a compilation of factor analytically derived cognitive abilities than an integrated, coherent theory. Its proponents assert that it has abundant empirical support, but closer examination shows that much of the validity evidence is fragmentary, applied to parts of the model rather than the model as a whole. Even Carroll's (1993) landmark study included only a handful of datasets with all of the stratum 2 broad factors. Moreover, the CHC framework was derived through factor analyses and remains somewhat method dependent. Following the lead of Thurstone, who viewed factor analysis as an early stage in theory development, Carroll (1983) himself identified a number of additional forms of validation that transcend factor analysis, including establishing the nature of a factor, 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. Unfortunately, most of these forms of validity remain to be fully explored as they apply to the factors identified in the CHC framework. The WJ III Cog's dependence on factor analysis is reminiscent of Eysenck's (1993) observation that a psychological model based primarily on factor analysis resembles a chimera ("a fabulous beast made up of parts taken from various animals" [p. 1299]).

In terms of the glue holding together the elements of a well-articulated theory, Woodcock's (1993, 1998a) cognitive performance/information processing model is a beginning to possible integration of factors in the CHC framework. It posits four higher order processes that combine to produce cognitive and academic performance: thinking abilities (Gf, Glr, Ga, Gv), stores of acquired knowledge (Gc, Grw, Gq), cognitive efficiency (Gsm, Gs), and facilitators-inhibitors (e.g., motivation, interest, attention). It also offers testable hypotheses (e.g., all performance, automatic or new learning, is constrained by the relevant stores of knowledge) that have yet to be seriously investigated.

Although McGrew and Woodcock (2001) boldly assert that "CHC taxonomy is the most comprehensive and empirically supported framework available for understanding the structure of human cognitive abilities" (p. 9), Horn and Blankson (2005) make the argument more modestly and objectively: "The extended theory of fluid and crystallized (Gf and Gc) cognitive abilities is wrong, of course, even though it may be the best account we currently have of the organization and development of abilities thought to be indicative of human intelligence" (p. 41). Arguably, scientific validation of the CHC framework is best served by objective, measured conclusions rather than dogmatic advocacy.

Standardization Features and Psychometric Adequacy

The psychometric characteristics of the WJ III NU Cog are complementarily documented in the normative update technical manual (McGrew, Schrank, & Woodcock, 2007) and the original technical manual (McGrew & Woodcock, 2001).

The WJ III NU Cog was standardized from 1996 through 1999 on 8,782 children, adolescents, and adults from ages 2 through 90+. The school-age sample consisted of 4,740 participants. Stratification targets were originally based on 1996 census projections for the year 2000, but the 2007 normative update was based on actual census 2000 findings. Sample stratification variables included sex, race, ethnicity, type of school, geographic region, community size, adult education, and adult occupation. The sample consisted of over 200 participants at each age year from 2 through 19, over 1,000 participants in their 20s, over 200 participants per decade from 30 to 59, and about 150 participants per decade after age 60. The sample was statistically weighted to correct for proportional underrepresentation of selected groups, including Hispanics and parents with education levels below high school completion. It is not possible to assess the degree to which the sample is representative of the general population, because accuracy is only reported "on the margins" without detailed reporting across stratification variables.

The need for a normative update a mere six years after the publication of the WJ III Cog has been attributed to discrepancies in the 1996 projections for the 2000 U.S. Census and actual year 2000 census statistics (McGrew, Dailey, & Schrank, 2007). The normative update resulted in as much as a 10-point standard score change, with standard score changes tending to be largest in the extreme age ranges (preschool and old age). Bootstrap sampling with replacement was used to provide stable estimates of standard error (and confidence bands). While the rationale and the methodology for the normative update is understandable, it is of concern that it was necessary and that it produced some score changes of large magnitude.

Based on examination of the two WJ III technical manuals (McGrew, Schrank, & Woodcock, 2007; McGrew & Woodcock, 2001), there appear to be some serious problems with the standardization and norming of the WJ III Cog—principally that some tests were administered to only a small number of standardization participants. For example, of the 2,216 children from ages 9 to 13 reported in the normative sample (p. 23), only 1,865 completed the Verbal Comprehension test, only 1443 took the Planning test, and only 548 obtained scores on the Pair Cancellation test (McGrew, Schrank, et al., 2007, p. 167). The small size of the Pair Cancellation sample suggests that as much as 75% of the normative sample may have not taken some tests during standardization of the WJ III Cog. It is not possible to determine whether the subsample actually given these tests is representative of the general population, although bootstrap sampling may help stabilize test parameters. It is also unclear why the sample size in this age group dropped by $n = 25$ from 2,241 (McGrew & Woodcock, 2001, p. 18) to 2,216 in the normative update (McGrew, Schrank, et al., 2007, p. 23). From the 2001 edition to the 2007 normative update, the total normative sample decreased by 36 participants.

Test internal consistency was calculated with the split-half procedure with Spearman-Brown correction and with Rasch procedures for tests that were either speeded or contained multiple point scoring. Test score reliability appears to be fully adequate, with median values across age falling below $r = .80$ for Picture Recognition and Planning only. Some 94% of WJ III Cog NU tests yield split-half or Rasch-derived consistency of .80 or higher. The standardized cluster scores also tend to be fairly reliable, with all but three having median values at or above .90 (the exceptions are Long-Term Retrieval at .88, Visual-Spatial Thinking at .81, and Short-Term Memory at .88), and across all age groups and all batteries, 84% of the clusters meet the .90 criterion. The overall composite GIA has a median reliability of .97 for the standard battery and .98 for the extended battery. Rasch scaling permits local reliabilities to be calculated:

> The Rasch procedures that underlie the W scale provide a unique estimate of the standard error of measurement for the ability score associated with each raw score for every person in the norm sample. When individual error (SEM) scores are available for all subjects who completed a test, it is possible to directly calculate test reliability. (McGrew, Schrank, et al., 2007, p. 42)

It is unfortunate that the test authors passed on the opportunity to generate automated local score reliabilities via the WJ III Cog NU's computer scoring program.

Temporal stability of WJ III Cog NU tests was not measured according to the conventional 1-month retest approach, so it is difficult to make comparisons with other intelligence tests. For the six WJ III NU Cog speeded tests administered in counterbalanced order with a test–retest interval of only a single day, 44% of the subtests yielded 1-day stability of .80 or better. When a median retest interval of over 1 year is employed, the median reliabilities for the selected tests range from .61 to .86. These long-term score stability coefficients are of considerable potential value to both practitioners and researchers, but the results are not adequately reported, missing data on most WJ III Cog NU tests, absent documentation of practice effects, and absent corrections for variability on first assessment. It also appears implausible that a test like Visual Matching should have median long-term stability of .78 to .86 over a retest interval of <1 year to 10 years, when its 1-day stability ranges from .68 to .87 across different age groups.

WJ III Cog floors and ceilings are difficult to formally evaluate because the test may only be computer-scored and no printed norms are available. The examiner's manual reports that test standard scores extend from 0 to over 200 (Mather & Woodcock, 2001), but this range seems inflated given that adequate test floors tend to be difficult to achieve with certain age groups, such as preschool children. For example, when raw scores of 1 (the best value to use to identify meaningful test floors) are entered for every test administered at the lowest level for a 6-year-old child, the resulting cluster standard scores range from 3 (Visual-Spatial Thinking) to 72 (Fluid Reasoning).

In deriving the standard scores and percentile ranks that are most commonly used for test score interpretation, the WJ III NU uses Rasch scaling to derive interval unit W-scores on a sort of yardstick of absolute performance across the life span, centered at a W of 500 for an average fifth-grade student. These scores permit developmental growth curves to be generated from the cross-sectional age samples in the normative data set, a form of evidence for test score validity that is not found in most intelligence tests. The cross-sectional growth curves suggest that fluid reasoning (Gf), processing speed (Gs), and short-term memory (Gsm) reach their highest levels at approximately age 25 to 30 before beginning a gradual decline, while comprehension-knowledge (Gc) does not reach a peak until age 50 to 60. The curves for long-term retrieval (Glr), auditory processing (Ga), and visual processing (Gv) demonstrate relatively little change with age (McGrew, Schrank, & Woodcock, 2007). These findings provide compelling evidence for the differential rates of development for the CHC broad abilities.

The WJ III GIA score tends to be highly correlated with composites from other intelligence tests, although correlations are not corrected for restricted or expanded score ranges. According to McGrew and Woodcock (2001), the standard battery WJ III Cog GIA correlates .67 to .76 with the DAS GCA; .75 with the KAIT Composite Intelligence Scale; .76 with the Stanford-Binet IV Composite SAS; .71 with the WISC-III FSIQ; and .67 with the WAIS-III FSIQ. Across all school-age groups, median correlations of the standard battery WJ III NU Cog GIA with WJ III Nu Ach scores are .76 for Total Achievement, .70 for Broad Reading, .67 for Broad Math, and .65 for Broad Written Language (McGrew, Schrank, et al., 2007). These correlations are derived from the normative update as a whole and suggest that the GIA is highly predictive of academic achievement. A number of recent investigations have examined the multivariate capacity of the WJ III Cog broad and narrow ability scores to predict academic performance in specific domains (a longtime objective of multifactor theorists; see e.g., McNemar, 1964), with results indicating that the differential prediction of academic skills varies across age and shows evidence of both direct and indirect cognitive effects (Benson, 2008; Evans, Floyd, McGrew, & Leforgee, 2002; Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007; Floyd, McGrew, & Evans, 2008). The WJ III NU software scoring programs offer three methods to predict academic performance for computation of ability-achievement discrepancies: predicted achievement derived from WJ III NU Cog ability scores regressed against achievement, predicted achievement based on GIA score, or predicted achievement based on the Oral Language Ability-Extended score. Unfortunately, the beta weights used in predicted achievement regression equations are not reported.

Factor-analytic studies of the WJ III constitute an area of concern for a test battery that has historically based its foundation on the work of Cattell, Horn, and Carroll. Exploratory factor analyses are not reported in the 2001 or 2007 technical manuals, although the addition of eight new subtests to the WJ III Cog certainly justifies these analyses. The new WJ III Cog subtests purport to measure working memory, planning, naming speed, and attention. Moreover, hierarchical exploratory factor analyses conducted by John B. Carroll (using the same approach described in his 1993 book) have been previously reported for the WJ-R (see also McGrew, Werder, et al., 1991, p. 172; reprinted in McGrew, 1997, pp. 176–177); these analyses yield findings of first-order and second-order factors that are not entirely congruent with the structure of the WJ Cog. As a basis for comparison, other tests in their third editions (e.g., WISC-III, WAIS-III) continued to report exploratory factor analyses.

Previous exploratory analyses have revealed inadequately defined factors that appear to have not been addressed in the WJ III NU Cog. For example, after conducting a hierarchical exploratory factor analyses on the WJ-R standardization data set ($n = 2{,}261$), Carroll (2003) wrote: "There is still a problem with *Gf,* namely, that it appears to be a rather weak, poorly defined factor, at least in the dataset examined here. Note the relatively small factor loadings for the two tests indicated as measuring *Gf* (p. 14)." The two tests comprising fluid reasoning (Gf) are identical from the WJ-R to the WJ III: Concept Formation and Analysis Synthesis. It may be worthwhile to consider improving these measures or substituting new procedures to tap fluid reasoning in the WJ IV, given the critical importance of this broad ability to assessment in the CHC framework.

The confirmatory factor analyses (CFAs) reported in the WJ III technical manual (McGrew & Woodcock, 2001) appear to provide marginal support for a seven-factor structure relative to two alternative models, but the RMSEA, which should ideally be less than .05 with good model fit, does not support good model fit at any age level. The CFAs involve a contrast between the seven-factor CHC structure, a WAIS-based model, and a Stanford-Binet–based model, the latter two with model specifications that Wechsler or Stanford-Binet devotees would likely argue are misrepresentations. None of the models is hierarchical; none includes a superordinate *g*; and none includes the higher-order dimensions suggested by Woodcock in his cognitive performance model. Moreover, only three goodness-of-fit indices are included whereas best practice with CFAs suggests that fit statistics should ideally include indices sensitive to model fit, model comparison, *and* model parsimony. On a model built on multifactor foundations, it may be argued that a more rigorous CFA test of alternative models is appropriate.

Interpretive Indices and Applications

Including the Diagnostic Supplement, the WJ III NU Cog consists of 31 tests purporting to measure seven broad cognitive factors and nine cluster scores. The tests are organized into a standard battery (tests 1 through 7, with three supplemental tests) and an extended battery (tests 1 through 7 and tests 11 through 17, with six supplemental tests). The Diagnostic Supplement adds more tests. The WJ III NU Cog is normed for ages 2 years through 90+ years and is conormed with 22 tests in an achievement battery, WJ III NU Tests of Achievement (Woodcock, McGrew, & Mather, 2001b, 2007b). Table 18.7 contains the fundamental interpretive indices, the most important of which are GIA and the factor-derived cluster scores: Gc, Glr, Gv, Ga, Gf, Gsm, Gs. Verbal Ability, Thinking Ability, and Cognitive Efficiency are part of Woodcock's cognitive performance model.

Cognitive cluster scores are conceptually derived and may include tests with heterogeneous content. Clusters include Phonemic Awareness, Working Memory, Cognitive Fluency, Perceptual Speed, Associative Memory, Visualization, Sound Discrimination, Auditory Memory Span, and Numerical Reasoning.

In terms of research on clinical and educational applications, the WJ III NU Technical Manual (McGrew, Schrank, et al., 2007) reports data for clinical samples totaling as much as $n = 1{,}281$, although the entire battery was not administered to every clinical sample, demographic characteristics and criteria for specific diagnoses are not fully presented, and the performance of demographically matched normative control groups are

TABLE 18.7 Woodcock-Johnson III Normative Update Tests of Cognitive Abilities Core Interpretive Indices

Composite Indices	Description
General Intellectual Ability (GIA)	A weighted estimate of general cognitive ability
Verbal Ability	Acquired knowledge in semantic and quantitative symbol systems; includes verbal conceptual knowledge (Gc), quantitative knowledge (Gq, from achievement tests), and reading-writing knowledge (Grw, from achievement tests)
Thinking Ability	Abilities that allow an individual to process information that has been placed in short-term memory but that cannot be processed automatically; consists of long-term storage and retrieval (Glr), visual processing (Gv), auditory processing (Ga), and fluid reasoning (Gf)
Cognitive Efficiency	Capacity to hold, rapidly, and automatically process information; includes short-term memory (Gsm) and processing speed (Gs)
Comprehension-Knowledge (Gc)	Breadth and depth of prior learning in culturally valued verbal areas as well as the capacity for further verbal learning
Long-Term Retrieval (Glr)	Ability to efficiently acquire and store information, measured by long-term and remote retrieval processes
Visual-Spatial Thinking (Gv)	Analysis and synthesis of spatial-visual stimuli, and the ability to hold and manipulate mental images
Auditory Processing (Ga)	Ability to discriminate, analyze, and synthesize auditory stimuli; also related to phonological awareness
Fluid Reasoning (Gf)	Ability to solve novel and abstract problems, usually of a visual and nonverbal nature
Short-Term Memory (Gsm)	Ability to hold, transform, and act on auditory information in immediate awareness; auditory-verbal mental holding capacity
Processing Speed (Gs)	Speed and efficiency in performing easy cognitive tasks

not compared with the performance of clinical samples. Among the clinical groups studied are intellectually gifted, intellectually disabled, ADHD, anxiety spectrum disorders, autism spectrum disorders, depressive spectrum disorders, head injury, language disorders, mathematics disorder, reading disorder, and written expression disorder.

Strengths and Limitations

The WJ III NU Cog represents an important step forward for intelligence assessment through its fit with the CHC model of cognitive abilities and conorming with the WJ III NU Tests of Achievement, but it largely lacks an integrated theoretical framework, established clinical correlates, and empirically demonstrated treatment utility.

The Woodcock-Johnson III Cog NU model is an elegant exemplar of the multifactor approach to cognitive abilities. Its factor-analytic lineage may be most clearly traced from the pioneering efforts in factor analysis of ability tests by Thurstone (1938) and the encyclopedic tome by Carroll (1993), along with seminal contributions by Cattell and Horn. In fact, Horn and Carroll were consultants in the development of the WJ III Cog. It is this association to a large body of factor analytic research that constitutes the WJ III Cog's main strength. It is notable, however, that as of this writing, no hierarchical exploratory factor analysis (of the type previously conducted by Carroll) has been published for the WJ III Cog or its normative update.

Unfortunately, a systematic overreliance on this same body of factor-analytic research as its primary evidence of test validity constitutes the most substantial weakness of the WJ III NU Cog. The WJ III Cog structure is a structural model missing the integrative, explanatory, and predictive glue that constitutes a scientific theory. Woodcock's (1993, 1998a) cognitive performance/information processing model represents a start toward an integrated model but requires further development and validation. To their credit, authors and advocates for the WJ Cog have acknowledged the current shortcomings of their theoretical foundation: "*Gf-Gc* provides little information on how the *Gf-Gc* abilities develop or how the cognitive processes work together. The theory is largely product oriented and provides little guidance on the dynamic interplay of variables (i.e., the processes) that occur in human cognitive processing" (Flanagan, McGrew, & Ortiz, 2000, p. 61).

The validity of the WJ III NU Cog with special populations is an emerging area of investigation (e.g., Schrank & Flanagan, 2003). The 2007 technical manual lists standard scores for a number of clinical samples, but it does not describe these samples in any detail or provide demographically matched normative comparison groups with effect size differences (McGrew, Schrank, & Woodcock, 2007). For example, the sample designated as gifted (n = 124) earned a median GIA (Standard) of 116 with median cluster standard scores ranging from 103 to 116. These scores would be unlikely to independently qualify most students for gifted and talented placements. More plausibly, the intellectually disabled/mentally retarded sample (n = 93) yielded a median GIA (Std) of 58, with cluster standard scores ranging from 56 to 77. Clearly much more research as to the clinical applications of the WJ III NU Cog is needed.

The WJ III Cog offers little in the way of empirically based assessment intervention linkages. While logical interventions are offered in Mather and Jaffe (2002; see also Wendling & Mather, 2009), there is a conspicuous absence of empirical verification for these assessment-intervention linkages directly connecting the intervention with WJ III NU Cog performance.

Finally, the claims made by WJ III NU Cog authors are frequently overstated, and research findings have a quality of being selectively reported. As an example of overstatement, the technical manual (McGrew, Schrank, et al., 2007) claims to provide "more precise measures and a wider breadth of coverage of human cognitive abilities than are found in any other system of psychological and educational assessment" (p. 3). Selective reporting extends to disproportionate emphasis on specific methodologies (e.g., CFAs, cross-battery analyses, developmental growth curves) while completely neglecting some important psychometrics (comprehensive DIF studies, hierarchical exploratory factor analyses, normative tables that can be independently reviewed, regression equations used to predict achievement, the magnitude of convergent validity correlations corrected for range restriction, or detailed descriptions of the samples used for special population studies). A more complete discussion of WJ III Cog strengths and limitations, including a dozen unanswered questions, may be found in Wasserman and Maccubbin (2004).

DIAGNOSTIC APPLICATIONS

There are numerous substantive reasons to give intelligence tests. The most common reasons among psychologists working in education and health care are to facilitate diagnosis, determine the nature of difficulty, and estimate capacity/potential (e.g., Camara et al., 2000; Harrison, Kaufman, Hickman, & Kaufman, 1988; Rabin, Barr, & Burton, 2005; Reschly, 2000). These priorities may shift as legal classification guidelines undergo change for various diagnoses. School psychologists typically spent approximately two-thirds of their time in special education eligibility determination, primarily through testing (e.g., Gresham & Witt, 1997) before the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA), which fundamentally changed procedures to assess eligibility for some exceptionalities.

In this section, a few of the most common applications of intelligence tests are described, especially in diagnostic determination. At the outset, it is noted that there is not necessarily a strong relation between the science of intelligence assessment (empirical demonstrations of value) and its various applications (i.e., those sanctioned by public institutions, based on public policy decisions). Camara (1997) effectively summarized the plethora of legal, ethical, and professional pressures impinging on assessment practices. One illustrative case showing how legislation may throw intelligence assessment practices into turmoil may be found in federal regulations for *specific learning disability.* In the Individuals with Disabilities Education Act (1997, 1999), federal legislation explicitly required a "severe discrepancy between achievement and intellectual ability" as part of the criterion for special education eligibility. Just 5 years later, the Individuals with Disabilities Education Improvement Act (variously referred to as IDEA or IDEIA, 2004) stated that educators "shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability" for specific learning disability eligibility, instead leading to the newer approach of response to intervention (RTI)." Predictably, this reversal of legal standards in a 5-year period left psychologists working in schools considerably confused about the potential value of intelligence tests in the identification of students with possible learning disability (Cangelosi, 2009). The anticipated 2013 publication of the fifth edition of *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, with its committee-based decision-making processes, will also likely be accompanied by debate for its new diagnostic methodologies. (See Widiger, 2011, for an illustration.) This section describes the role of intelligence tests in diagnostic decision making in the most recent edition, the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders, Text Revision* (*DSM-IV-TR*; American Psychiatric Association, 2000) as well as changes that appear likely for the *DSM*-5.

Beyond diagnostic decision making, several theories also posit that intelligence has a potentially far-reaching nonspecific influence on the development of psychopathology in youth (e.g., Rutter, 1987) and neuropathology in adulthood and older ages (e.g., Stern, 2002, 2009). In these models, often derived from longitudinal studies, intelligence is generally treated as a moderator variable—that is, a variable that "affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable" (Baron & Kenny, 1986, p. 1174). For Rutter (1987), intelligence is a type of protective mechanism that "is a modification of the person's response to the risk situation . . . that in ordinary circumstances leads to a maladaptive outcome" (p. 317).

In childhood psychopathology, models of risk and resilience frequently list intelligence as a protective factor or protective process facilitating an individual's ability to deal with risk-elevating factors and adverse life experiences (e.g., Rutter, 1987). Given any child's developmental history with factors that normally predict negative outcome, an average- or above-average intelligence is considered to play some role in producing developmentally appropriate and positive outcomes, while lower intelligence is associated with greater vulnerability to risk and less adaptive outcomes (e.g., Masten, 1994). E. E. Werner (2000) explained that enhanced use of coping and problem-solving strategies and better ways of responding to the risk situation are thought to account for the moderating influence of intelligence for the at-risk child across a wide range of populations:

> Youngsters who are better able to appraise stressful life events correctly are also better able to figure out effective strategies for coping with adversity, either through their own efforts or by actively reaching out to other people for help. This finding has been replicated with children from all socioeconomic groups and from diverse ethnical backgrounds, in studies of African American, Asian American, and Caucasian children who grew up under a variety of high-risk conditions, including poverty, parent mental illness, and substance abuse, as well as family discord and child abuse. (pp. 122–123)

A number of longitudinal investigations have demonstrated that childhood intelligence appears protective against later maladjustment, even into adulthood (Burt & Roisman, 2010; Egeland, Carlson, & Sroufe, 1993; Loeber, Pardini, Stouthamer-Loeber, & Raine, 2007; Luthar, D'Avanzo & Hites, 2003; Masten, Burt, Roisman, Obradovic, Long, & Tellegen, 2004; Radke-Yarrow & Sherman, 1990; E. E. Werner, 1993).

In adulthood and old age, the concept of *cognitive reserve* arises from the observation that there does not appear to be a direct relationship between the degree of brain damage and the clinical-behavioral manifestation of that damage:

> The concept of cognitive reserve provides a ready explanation for why many studies have demonstrated that higher levels of intelligence, and of educational and occupational attainment are good predictors of which individuals can sustain greater brain damage before demonstrating functional deficit. Rather than positing that these individuals' brains are grossly anatomically different than those with less reserve (e.g., they have more synapses), the cognitive reserve hypothesis posts that they process tasks in a more efficient manner. (Stern, 2002, pp. 450–451)

Individuals with high cognitive reserve are thought to better tolerate acquired brain damage, normal age-related changes in cognitive ability, and degenerative neuropathology because their cognitive processing capacities seem to compensate for (or successfully mask) the cognitive and behavioral manifestations of underlying changes in the brain. A corollary to the cognitive reserve hypothesis is that once an individual can no longer effectively compensate for underlying brain damage, any latent level of neuropathological impairment and rate of deterioration may appear more severe, no longer adequately concealed by cognitive processing efficiency. The best-known longitudinal study supporting the cognitive reserve hypothesis is known as the Nun Study (e.g., Snowdon, Greiner, Mortimer, Riley, Greiner, & Markesbery, 1997). Launched in 1986 with elderly nuns who agreed to participate in annual testing of physical and mental functions as well as to donate their brains for study upon death, the study provided compelling evidence that it was possible for individuals to present as cognitively intact even when their brains on autopsy showed prominent evidence of the structural changes associated with advanced Alzheimer's disorder. In general, the better the early language, the higher the education, and the more the positive emotional outlook that the nuns showed in early adulthood, the less prone to cognitive disability and dementia they proved to be. Scarmeas and Stern (2003) assert that through the proposed mechanism of cognitive reserve, "innate intelligence or aspects of life experience like educational or occupational attainments may supply reserve, in the form of a set of skills or repertoires that allows some people to cope with progressing Alzheimer's disease (AD) pathology better than others" (p. 625).

Dementia/Neurocognitive Disorders

> Neuropsychological evaluation and cognitive testing remain among the most effective differential diagnostic methods in discriminating pathophysiological dementia from age-related cognitive decline, cognitive difficulties that are depression-related, and other related disorders. Even after reliable biological markers have been discovered, neuropsychological evaluation and cognitive testing will still be necessary to determine the onset of dementia, the functional expression of the disease process, the rate of decline, the functional capacities of the individual, and hopefully, response to therapies. (APA Task Force to Update the Guidelines for the Evaluation of Dementia and Age-Related Cognitive Decline, 2011, p. 2)

The diagnostic term *dementia,* likely to be replaced with the term *major neurocognitive disorder* in the *DSM-5* (Ganguli et al., 2011), refers to a deterioration in (or loss of) cognitive functions, relative to a higher premorbid level of functioning. The *DSM-IV-TR* specifies that dementia is characterized by the development of multiple cognitive deficits including memory impairment and at least one of the following: aphasia, apraxia, agnosia, or a disturbance in executive functioning. Cognitive deficits must be sufficiently severe to cause impairment in occupational or social functioning and must represent a decline from a previously higher level of functioning (American Psychiatric Association, 2000). In recognition that different dementias may present with unique sequential courses and patterns of cognitive deficits (with memory loss not always the initial presenting concern), the *DSM-5* will likely deemphasize the sole criterion of memory, requiring deficits in at least one (typically two or more) of these areas: complex attention, executive functions, memory, language, visuoconstructional ability, or social cognition (American Psychiatric Association, 2011). At the time of this writing, the proposed *DSM-5* criteria require that the severity of any domain-based cognitive performance impairment must be 2 or more *SD*s below the normative mean (i.e., below the third percentile) for a diagnosis of major neurocognitive disorder, and impairment must be sufficiently severe so as to interfere with functional independence (American Psychiatric Association, 2011).

An associated diagnostic term, *mild neurocognitive disorder* (also known as *mild cognitive impairment*; see Petersen, Smith, Waring, Ivnik, Tangalos, & Kokmen, 1999; Petersen et al., 2009) has been proposed in the *DSM-5* for individuals with mild cognitive deficits in one or more of the cognitive domains but with an intact capacity for functional independence in activities of daily living. This diagnosis may serve to identify the earliest features of Alzheimer's disease and other dementias and is characterized by test performance between the 3rd and 16th percentile (i.e., 1 to 2 *SD*s below the normative mean) (American Psychiatric Association, 2011). As might be expected, individuals with this diagnosis show only slightly lower performances than control participants on intelligence and nonmemory cognitive performances, but their performance on memory tasks shows more prominent difficulties (Petersen et al., 1999). When familial Alzheimer's disease is present, prodromal/presymptomatic cognitive deficits appear prominently in the areas of general intelligence and memory (Godbolt, Cipolotti, Watt, Fox, Janssen, & Rossor, 2004).

As part of a neuropsychological assessment, intelligence testing has value in the identification of dementias, differentiation from normal age-associated cognitive changes, differential diagnosis between dementias, and monitoring of the course of cognitive decline (e.g., Petersen et al., 1999; Petersen, Stevens, Gangulli, Tangalos, Cummings, & DeKosky, 2001; Ritchie & Tuokko, 2010). However, intelligence tests by themselves are clinically considered insufficient for dementia diagnosis. Intelligence tests rank among the most frequently used measures in American neuropsychological assessments (Camara et al., 2000), but there are no surveys on American practitioner test preferences in dementia assessments. In Europe, an estimated 88% of countries use some version of the WAIS in dementia evaluations (Maruta, Guerreiro, de Mendonça, Hort, & Scheltens, 2011).

Quantification of cognitive decline, describing how much function has been lost from dementia or acquired brain injury, is an important part of monitoring the course of a disorder and estimating the magnitude of disability. There are several methods available to estimate how much cognitive function has been lost, with most methods examining the discrepancy between estimated premorbid ability and current cognitive performance. Specific ways to estimate premorbid levels of ability include: (a) review of premorbid academic or occupational test or achievement scores (e.g., Baade & Schoenberg, 2004); (b) prediction of premorbid ability based on demographic characteristics (e.g., age, education, sex, race/ethnicity; see Barona, Reynolds, & Chastain, 1984); and (c) prediction of premorbid ability based on current test performance in areas known to be resilient to brain injury (e.g., reading recognition for words with irregular phoneme-grapheme spelling; see H. E. Nelson, 1982). The WAIS-IV supplemental materials *Advanced Clinical Solutions for the WAIS-IV and WMS-IV* (Pearson, 2009a) provide estimates of premorbid intellectual and memory functioning based on methods (b) and (c), in conjunction with a new reading measure, the Test of Pre-Morbid Functioning (TOPF; Pearson, 2009b).

Intellectual Disability/Mental Retardation

> Intellectual disability is characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before age 18. (Schalock and the Ad Hoc Committee on Terminology and Classification, 2010, p. 1)

This 2010 definition from the Ad Hoc Committee on Terminology and Classification of the American Association on Intellectual and Developmental Disabilities (AAIDD; formerly American Association on Mental Retardation [AAMR]) defines *intellectual disability* (ID; formerly known as *mental retardation*; Schalock et al., 2007) in terms of intellectual functioning and functional adaptation in activities of daily living, both equal in importance and interpreted in a multidimensional, ecological context (not appearing in this main part of the definition just given). The context extends to the environmental supports needed for the individual to participate in activities linked with normative human functioning. As such, intellectual disability is considered to reside not within the individual but instead in the (mis)fit between the individual's capacities and the demands of the environmental context (microsystem, mesosystem, and macrosystem) in which the individual is expected to function (Schalock et al., 2010).

For the diagnosis of *mental retardation,* the *DSM-IV-TR* (American Psychiatric Association, 2000) requires significantly subaverage general intellectual functioning accompanied by significant limitations in adaptive functioning in at least two of these skill areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety. Onset must occur during the developmental period, and deficits are expected to adversely affect an individual's educational performance. Based on proposed changes, the *DSM-5* appears likely to rename the diagnosis *intellectual disability* or *intellectual development disorder*, aligning it with the AAIDD definition. Among the proposed criteria are a current deficit in general mental abilities approximately 2 or more *SD*s in IQ below the population mean for a person's age and cultural group; significant impairment in adaptive functioning requiring ongoing support at school, work, or independent life; and onset during the developmental period.

In both *DSM* and AAIDD diagnoses, intellectual functioning is listed as "the first, and the most salient, criterion in the definition of ID" (Borkowski, Carothers, Howard, Schatz, & Farris, 2007), although it is clear that functional adaptation, environmental context, and person–environment fit are essential considerations that have risen in relative importance compared to intelligence. Both AAIDD and *DSM-5* appear to have accorded the *general* factor of intelligence (usually represented by an overall composite score, such as the FSIQ) as the most relevant estimate of cognitive and intellectual ability. Both definitions also specify that the intellectual functioning criterion for a diagnosis of intellectual disability is approximately 2 *SD*s or more below the normative mean, but factors such as test score statistical error (standard error of measurement), test fairness, normative expectations for the population of interest, the Flynn effect, and practice effects from previous testing need to be considered before arriving at any diagnosis. In practical terms, this criterion usually corresponds to a composite IQ score below 70 or 75 (the higher number including standard error and other factors).

The severity of intellectual disability remains stratified by four levels of composite IQ scores in the *DSM-IV-TR* (American Psychiatric Association, 2000), but these levels have been omitted from the last two editions of the AAMR/AAIDD manuals and probably for the *DSM-5* to reflect an emphasis on functional adaptation and the intensity of needed environmental support rather than cognitive-intellectual ability. The levels of intellectual disability based on IQ scores, however, still appear commonly in practice and the research literature and may be summarized in this way:

- *Mild* (IQ level 50–55 to approximately 70–75). Approximately 85% of individuals with intellectual disability fall into this range, and individuals so labeled have historically been considered *educable.* These individuals may be able to acquire basic reading and mathematics literacy and are often able to successfully hold jobs and live independently, with vocational training and community and social support.
- *Moderate* (IQ level 35–40 to 50–55). Encompassing about 10% of individuals with intellectual disability, persons within this level are often considered *trainable* and may acquire academic skills at the kindergarten or first-grade level. They are likely to require regular support and supervision to function in everyday activities, and they may be able to perform repetitive employment tasks in a highly structured and sheltered environment.
- *Severe* (IQ level 20–25 to 35–40). Approximately 3% to 4% of individuals with intellectual disability fall into this group, and they will typically have low communication and social skills, with marked developmental delays. A curriculum emphasizing self-help skills (toileting, dressing with assistance) is provided with high levels of supervision and environmental support.
- *Profound* (IQ level below 20 or 25). About 1% to 2% of individuals with intellectual disabilities are in this IQ range, and they typically require lifelong care and supervision. They are typically unable to walk, talk, or carry out most basic activities of daily living, often suffering from significant physical abnormalities and sensory impairments.

The 10th edition of the AAMR definition and classification manual (Luckasson et al., 2002) enumerated four classification levels for mental retardation, based not on intelligence but rather on the level of environmental support needed: *intermittent* (need for support during stressful or transition periods but not constantly), *limited* (less intense, consistent supports needed, but needs are time-limited for changing situations), *extensive* (long-term consistent support at work and/or home), and *pervasive* (very intense, long-term, constant support needed across most or all situations). These levels do not appear in the 11th edition (Schalock & Ad Hoc Committee, 2010), which emphasizes frequency and duration of needed supports in select areas over global ratings of support needs.

Intellectual Giftedness

> 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 the 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). (Siegle & McCoach, 2010, p. 6)

In 2010 the board of directors of the National Association of Gifted Children formally approved this new definition of giftedness, based on the recommendations of a committee of 15 experts. It represented the culmination of nearly four decades of research after the first federal definition of gifted and talented, based on a 1972 report to Congress from former U.S. Commissioner of Education Sidney P. Marland:

> Gifted and talented children are those, identified by professionally qualified persons, who by virtue of outstanding abilities are capable of high performance. These children require differentiated programs and/or services beyond those normally provided by the regular school program in order to realize their contribution to self and society. Children capable of high performance include those with demonstrated high achievement and/or potential ability in any of the following areas, singly or in combination; general intellectual ability, specific academic aptitude, creative or productive thinking, leadership ability, visual and performing arts, and/or psychomotor ability. (p. 2)

These definitions share an emphasis on requiring high levels of ability or performance in either general ability or specific narrower abilities as part of a gifted eligibility determination. Federal laws do not, however, mandate educational services for gifted and talented learners, and states and individual school districts vary widely as to how they define and determine giftedness.

Norm-referenced group and/or individual intelligence tests constitute the leading criterion by which giftedness is identified, but it is considered best practice to base assessments on multiple methods providing different types of information (e.g., National Association for Gifted Children, 2008). Robertson, Pfeiffer, and Taylor (2011) reported that the most common assessment tools used by school psychologists in the identification of gifted and talented students were, in descending order, the Wechsler intelligence scales (used frequently or very frequently by 51%), Woodcock-Johnson Tests of Cognitive Abilities (24%), Stanford-Binet Intelligence Scales (17%), Differential Ability Scales (13%), and Kaufman Assessment Battery for Children (10%). Other criterion measures may include ratings of student products or portfolio, observation of in-classroom behavior, formal teacher-completed rating scales, interviews, letters of support, norm-referenced achievement test performances, curriculum-based measurement/performance assessment, and history of academic accomplishments (e.g., grades, awards), among others. Concerns about the proportional underrepresentation of minorities in gifted education programs have led to the increasing use of nonverbal intelligence tests and alternative assessment methods, but early research indicates that students identified through such methods tend to be less successful in gifted curriculums (e.g., Van Tassel-Baska, Feng, & Evans, 2007).

Characteristics associated with giftedness often include early language development, early acquisition of reading skills, high levels of memory, extended attention span, and an intense curiosity and self-motivated interest in learning and problem-solving. During their earliest school years, gifted students are typically described as active learners seeking in-depth understanding about subjects of interest and making connections between seemingly unrelated events and ideas (e.g., Damiani, 1997; Harrison, 2004; Hodge & Kemp, 2000; Jackson, 2003; Kitano, 1995; Rotigel, 2003; Sankar-DeLeeuw, 2004; Walker, Hafenstein, & Crow-Enslow, 1999). In a large sample of regular education primary school teachers for kindergarten through second grade, Moon and Brighton (2008) reported that 95% or more of primary school teacher respondents agreed with each of these characteristics of gifted learners:

- Transfers learning into other subjects or real-life situations
- Tries to understand the how and whys of things
- Has a large store of general knowledge
- Has an active imagination
- Likes to make three-dimensional structures from blocks and other manipulatives
- Completes assignments faster than same-age peers
- Can devise or adapt strategies to solve problems
- Can carry on a meaningful conversation with an adult
- Can successfully carry out multiple verbal instructions
- Demands a reason for things

While these ratings suggest that educators have meaningful insights into characteristics of giftedness, it has been known for nearly a century that teachers commonly fail to identify gifted learners who do not present as model students. Lewis M. Terman (1916) observed (as Alfred Binet had written before him) that teachers are unreliable evaluators of student cognitive ability:

> Psychological tests show that children of superior ability are very likely to be misunderstood in school. The writer has tested more than a hundred children who were as much above average intelligence as moron defectives are below. The large majority of these were found located below the school grade warranted by their intellectual level. One third had failed to reap any advantage whatever, in terms of promotion, from their very superior intelligence. Even genius languishes when kept over-long at tasks that are too easy.
>
> Our data show that teachers sometimes fail entirely to recognize exceptional superiority in a pupil, and that the degree of superiority is rarely estimated with anything like the accuracy which is possible to the psychologist after a one-hour examination. (p. 13)

In 1921–1922, Terman, the founder of the gifted child movement, launched the longest longitudinal study in the history of psychology on gifted children. With a grant from the Commonwealth Fund, Terman initiated a study entitled Genetic Studies of Genius (later renamed the Terman Study of the Gifted) that was intended to describe the characteristics of gifted children and follow their development over time. Results from this study were interpreted as dispelling misconceptions about the inadequacies of highly intelligent children, demonstrating that children with high IQ are healthier, better adjusted, better leaders, and higher academic achievers than normatively expected. Terman's findings also disproved the prevailing beliefs of the era that gifted individuals were more at risk, neurotic, or prone to mental illness as adults (e.g., "Precocity is not a menace," 1925). Neihart's (1999) review and update largely agreed but noted that the psychological well-being of gifted children is related to the type of giftedness, the quality of educational fit, and the child's personal characteristics, such as self-perceptions, temperament, and life circumstances.

In terms of intelligence tests, giftedness has traditionally been defined in terms of elevated general intelligence (e.g., Hollingworth, 1942; Terman, 1925), emphasizing the *g* factor as represented by composite intelligence test scores. Kaufman and Sternberg (2008) stated that general intelligence remains of primary importance in gifted eligibility decisions: "In the United States, a global IQ score is still the dominant criterion used for acceptance into gifted programs at the grade-school level" (p. 80).

As previously discussed, the overall composite score in intelligence tests is not always the best measure of *g*. When tests such as the WISC-IV offer an overall composite (the Full Scale IQ) and a narrower score explicitly intended to measure the *g* factor (the GAI), authorities in gifted assessment have tended to recommend use of the *g* factor score (e.g., National Association for Gifted Children, 2010). In part, this recommendation has come from findings that subtests with lower *g* loadings, such as speed measures, are quite commonly the lowest scores among gifted learners (e.g., Newman, Sparrow, & Pfeiffer, 2008), who tend to be more reflective, contemplative, or methodical in their problem-solving styles. In a consecutive series of 219 students referred for assessment for gifted program eligibility who earned a WISC-IV Full Scale IQ of 120 or higher, Wasserman (2010) reported that processing speed was the lowest index score in 59.4% of the sample; additionally, in 47.5% of the sample, processing speed fell in the average range or lower *and* was the lowest index score. Not surprisingly, gifted learners tend to earn lower composite scores on measures of cognitive processes in comparison with higher *g*-loaded tests. (See Kaufman & Kaufman, 2004, in which a gifted and talented sample earned a mean MPI of 118.7, $SD = 11.9$; see also Naglieri & Das, 1997b, in which a gifted sample earned a mean Full Scale standard score of 118.2, $SD = 10.0$).

Even as defining characteristics of psychometric *g* such as reasoning ability are usually emphasized when identifying gifted learners, there is a long-held contradictory belief that for higher levels of ability, the general factor *g* may not explain extraordinary performance as much as narrower and more independent factors. This line of research may be traced to Spearman's (1927) observation in *The Abilities of Man* that "the influence of *g* on any ability [grows] less—in just the classes of person which, on the whole, possess this *g* more abundantly" (p. 219). Now referred to as Spearman's Law of Diminishing Returns (SLODR; Jensen, 1998, pp. 585–588), this hypothesis predicts that the *g* saturation of intelligence tests, derived from test intercorrelations, declines as ability level increases. SLODR has received substantial, but somewhat equivocal, support in the research literature (e.g., Jensen, 1998; te Nijenhuis, & Hartmann, 2006). Adapting Spearman's original analogy, adding fuel (i.e., the mental energy that is g) to an automobile's engine will only increase its speed (i.e., the gifted learner's performance) up to a point before the benefits start to wane. Extending the analogy, as a test's difficulty increases, the *g* loadings are thought to decrease.

While some gifted children show fairly global and uniform elevations across all cognitive abilities, it has long been evident that subtypes of giftedness may be defined by relative superiority in narrower and specific cognitive or academic domains. The most researched domain-specific subtypes of giftedness are verbal/linguistic and logical/mathematical (e.g., Lubinski, Webb, Morelock, & Benbow, 2001; Matthews, 1997; Matthews & Keating, 1995). These subtypes not only describe cognitive profiles but also preferences in education and occupation:

> In general and irrespective of gender, students with tilted intellectual profiles tend to gravitate toward their area of strength. Those with exceptional mathematical abilities relative to verbal abilities tend to gravitate toward mathematics, engineering, and the physical sciences, while those with the inverse pattern are more attracted to the humanities, law, and social sciences. (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000, p. 474)

Another proposed subtype is visual-spatial giftedness (e.g., Humphreys, Lubinski, & Yao, 1993; Silverman, 2002b). The presence of superior visual-spatial ability has been shown to predict self-selected educational and occupational tracks, including architecture, cartography, chemistry, engineering, medical-surgery, and physics (e.g., Humphreys et al., 1993), but individuals with spatial gifts tend to be disproportionately undereducated and underemployed when compared with comparably gifted individuals with strengths in verbal and mathematical domains (Gohm, Humphreys, & Yao, 1998).

Another well-known longitudinal investigation of highly gifted children, lasting some four decades, is the Study of Mathematically Precocious Youth (SMPY), started by Julian C. Stanley in 1971 (Benbow & Stanley, 1983; Lubinski & Benbow, 2006). Its participants were originally intended to be mathematically advanced but actually included many students with greater verbal than mathematical abilities, first identified at about ages 12 or 13 through high school achievement test performances followed by out-of-level testing with the SAT in math and verbal areas while still in early middle school. Participants could be subdivided into at least three subtypes, one of which was defined by exceptional mathematical reasoning relative to verbal ability (e.g., Lubinski et al., 2001). Research findings showed that when provided with fast-paced mathematics classes, SMPY participants were twice as likely to be in math-science career tracks in their mid-20s and in their mid-30s. This study also yielded compelling evidence that gifted students benefit from being provided with educational opportunities tailored to their rates of learning:

> Intellectually able adolescents scoring 500 or more on SAT-M or SAT-V before age 13 (top 1 in 2000) can assimilate a full high school course (e.g., chemistry, English, mathematics) in 3 weeks at a summer residential program for intellectually precocious youth; yet exceptionally able adolescents, those scoring 700 or more (top 1 in 10,000), can assimilate at least twice this amount. (p. 318)

Use of the SAT by middle school students (as in the SMPY) is one way to reliably identify students with abilities in the top 0.01% (1 in 10,000) of the general population. Although descriptive classifications in the uppermost ranges of intelligence are less well known and less researched than those in the lowermost ranges, there appears to be a newfound interest in identifying and differentiating highly gifted students, considering the extended normative capacities built into the WISC-IV (Zhu et al., 2008) and the Stanford-Binet (Roid, 2003c, p. 22). While norm-referenced IQ scores have typically topped out at about 150 to 160 for decades, there is actually a long-standing tradition of identifying the highly gifted, dating back to Galton's (1869/1892) rankings of genius (eminently gifted and illustrious) and Terman's (1916) designation of near genius or genius for IQs above 140. Building on the rich qualitative descriptions generated by Leta Hollingworth (1942; see also Stanley, 1990) to describe exceptionally and profoundly gifted children, Miraca Gross (2000) identified students as highly gifted with IQs from 145 to 159, exceptionally gifted with IQs of 160 to 179, and profoundly gifted with IQs of 180 or higher (with an estimated population frequency of less than 1 per 1 million). A full range descriptive classification appears in Table 18.8 and is designed to be symmetrical around the normative mean of 100, but any descriptive system may be used to denote the highly gifted and draw attention to their unique needs. WISC-IV composite scores now extend up to 210 (Zhu et al., 2008), joining other cognitive ability tests with composite scores that also extend to 200 and beyond (e.g., Roid, 2003c; Woodcock, McGrew, & Mather, 2001a). Ironically, the Wechsler-Bellevue (Wechsler, 1939) originally yielded Full Scale IQ scores ranging up to 195, so the practice of extending norms is not particularly new.

The concept of *asynchrony* (e.g., Columbus Group, 1991) or *dyssynchrony* (Terrassier, 1985), which may be defined in gifted individuals as referring to "a lack of synchronicity in the rates of their cognitive, emotional and physical development" (Morelock, 1992, p. 11), is an apt way to close this section. With variable rates of development for the different qualities and behaviors described long ago in gifted individuals by Hollingworth (1931) and

TABLE 18.8 Descriptive Ability Levels Across an Extended IQ Range

Descriptive Level	Ability/ IQ Range	Normal Curve Cut Points
Profoundly advanced/gifted	above 176	+ 5.1 *SD* and above
Exceptionally advanced/gifted	161 to 175	+ 4.1 to +5.0 *SD*
Highly advanced/gifted	146 to 160	+ 3.1 to +4.0 *SD*
Advanced/gifted	131 to 145	+ 2.1 to +3.0 *SD*
Superior	121 to 130	+1.4 to +2.0 *SD*
High average	111 to 120	+ 0.7 to +1.3 *SD*
Average	90 to 110	+0.67 to −0.67 *SD*
Low average	80 to 89	−0.7 to −1.3 *SD*
Borderline	70 to 79	−1.4 to −2.0 *SD*
Mildly delayed/impaired	55 to 69	−2.1 *SD* to −3.0 *SD*
Moderately delayed/impaired	40 to 54	−3.1 *SD* to −4.0 *SD*
Severely delayed/impaired	25 to 39	−4.1 *SD* to −5.0 *SD*
Profoundly delayed/impaired	below 25	−5.1 *SD* and below

Note. IQ range scores are for a test with a standard score mean of 100 and *SD* of 15.

Terman (1931), the most basic and universal aspect of asynchronous development in the gifted is that cognitive development nearly always progresses at a considerably faster rate than physical development (N. M. Robinson, 2008). Asynchrony also encompasses uneven development of cognitive abilities and acquired skills, meaning that gifted individuals may commonly show a striking pattern of strengths and weaknesses, with performance discrepancies appearing more pronounced in younger students and those who are highly gifted (Gilman, 2008; Webb, Gore, Amend, & DeVries, 2007). Tolan (1994) described the complex behavioral presentation resulting from asynchrony:

> The young gifted child may appear to be many ages at once. He may be eight (his chronological age) when riding a bicycle, twelve when playing chess, fifteen when studying algebra, ten when collecting fossils and two when asked to share his chocolate chip cookie with his sister. (pp. 2–3)

Roedell (1989) described the discrepancies in intelligence, social knowledge, and actual social behaviors that are not unusual in gifted children: "It is unsettling to hold a high-level conversation with a 5-year-old who then turns around and punches a classmate who stole her pencil" (p. 22). The implications of asynchrony for understanding and counseling the gifted are quite profound (Silverman, 1993, 2002a, 2009, 2012), but for our purposes it is critical to note that uneven development of cognitive, emotional, physical, and social abilities should be normatively expected in gifted children, especially highly gifted children. An inspection of developmental growth trends suggests that cognitive abilities that are more strongly related to general intelligence tend to develop more uniformly, whereas more unique and low-*g* abilities often have different and distinctive developmental trajectories (Wasserman, 2007).

Specific Learning Disabilities

> The term "specific learning disability" means a disorder in 1 or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. . . . Such term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. . . . Such term does not include a learning problem that is primarily the result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage. (IDEA, 2004, 20 U.S.C. § 1401, 118 Stat. 2657)

While the definition of *specific learning disabilities* (SLDs) appearing in IDEA 2004 federal legislation has remained largely unchanged since 1968, the methods for SLD eligibility determination have changed significantly. In a notable scientific-legislative development, the federal government forbade the requirement for (but not the use of) the ability-achievement discrepancy method in SLD eligibility determination in favor of a Response To Intervention (RTI) methodology. The replacement of the discrepancy method, with its empirically demonstrated limitations, by the RTI method, which had little empirical support at the time of its adoption and which continues to have a weak evidence base at the time of this writing, has led some authorities to wonder if RTI is a "politically rather than scientifically motivated" model (Kavale, Kauffman, Bachmeier, & LeFever, 2008, p. 135). In this section, current federal SLD guidelines are described along with the diminished but still potentially valuable role of intelligence testing in the identification of SLDs.

From 1977 through about 2004 to 2006, the discrepancy between intellectual ability and performance on academic achievement tests was the primary federal criterion for defining specific learning disabilities in clinical and educational practice ("Procedures for evaluating specific learning disabilities," 1977). The origins of SLD discrepancy methods may be traced to Franzen's (1920) accomplishment quotient (the ratio of an educational quotient to the intelligence quotient) and Monroe's (1932) reading index (a discrepancy between actual and expected level of reading achievement). Samuel Kirk first used the term *learning disability* in print in 1962, defining it as "a retardation, disorder, or delayed development in one or more of the processes of speech, language, reading, writing, arithmetic, or other school subject" (p. 263). It was Kirk's former student, Barbara Bateman, who reintroduced the discrepancy model in 1965: "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" (p. 220). Michael L. Rutter's Isle of Wight studies (Rutter, 1978; Rutter & Yule, 1973, 1975) are generally credited with having differentiated two types of reading-impaired groups: a general reading backwardness group (an ability-achievement nondiscrepant group, with less than two standard errors of estimate from the reading achievement predicted from performance IQ) and a specific reading retardation group (an ability-achievement discrepant group, with reading more than two standard errors of the estimate below the grade level predicted from performance IQ). The two groups differed in both their cognitive characteristics and their educational prognoses.

In 1975 the Education for All Handicapped Children Act (Public Law 94–142), was signed into federal law and defined *specific learning disability* as well as promulgating the severe discrepancy approach to SLD eligibility through U.S. Office of Education regulations in 1977.

As described in the introduction to this section, federal law last reaffirmed the ability-achievement discrepancy methodology in 1999, reversing itself in IDEA (2004) by mandating that ability-achievement discrepancies may *not* be required in the determination of eligibility for specific learning disabilities, a directive that many educators and psychologists mistakenly understood as precluding the use of ability-achievement discrepancies.

Over the last two decades, a number of researchers effectively challenged the rationale, validity, reliability, and fairness of discrepancy methodologies in identification of students with SLD. For example, Stanovich (1991a, 1991b) criticized the implicit assumption in discrepancy methodologies that intelligence predicts reading potential. The concept that students must wait years until their reading achievement deficiencies have grown large enough to reach the number required for a "severe discrepancy" with intelligence was criticized as a "wait to fail" index of SLD, which unnecessarily delayed delivery of interventions (e.g., Stage, Abbott, Jenkins, & Berninger, 2003; Stuebing, Fletcher, LeDoux, Lyon, Shaywitz, & Shaywitz, 2002). Several important validity investigations, including meta-analyses, failed to support meaningful distinctions between ability-achievement discrepant and nondiscrepant groups in terms of their academic performance, cognitive/achievement characteristics, educational prognosis, and response to intervention (Fletcher et al., 2002; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Hoskyn & Swanson, 2000; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992; Stage et al., 2003; Stanovich & Siegel, 1994; Stuebing et al., 2002; Vellutino, Scanlon, & Lyon, 2000). The reliability and stability of discrepancy scores were also effectively challenged (e.g., Francis, Fletcher, Stuebing, Lyon, Shaywitz, & Shaywitz, 2005). Finally, the disproportionate identification of minority students in high-incidence special education classifications including SLD was attributed to the use of intelligence tests. The 2002 report of the President's Commission on Excellence in Special Education reported, "The Commission found that several factors were responsible for this over-representation [of minority students in special education], including the reliance on IQ tests that have known cultural bias" (U.S. Department of Education, Office of Special Education and Rehabilitative Services, 2002, p. 26). The commission left little doubt of its low regard for intelligence tests in special education:

> There is little justification for the ubiquitous use of IQ tests for children with high-incidence disabilities, except when mild mental retardation is a consideration, especially given their cost and the lack of evidence indicating that IQ test results are related meaningfully to intervention outcomes. (p. 25)

The solution in federal regulations to the discrepancy methodology was termed *Responsiveness To Intervention* (also *Response To Intervention,* or *RTI*), in which students who fail to achieve adequately are provided with a series of increasingly intensive, individualized instructional interventions across multiple conceptual stage (or tiers), coupled with continuous and systematic monitoring of student progress at each stage. Students who fail to respond positively to intervention are considered to be "at risk" for learning disabilities, potentially being referred for psychoeducational assessment in tier 3 to specify the need for special education services. Now, after several years of implementation, increasing concern about the effectiveness of RTI is being expressed (e.g., Reynolds & Shaywitz, 2009). Swanson's (2010) critical appraisal noted:

> At the present time, RTI as an assessment approach to define LD [learning disability] has a weak experimental base. There have been no controlled studies randomly assigning children seriously at risk for LD to assessment and/or delivery models [(e.g., tiered instruction versus special education (resource room placement)] that have measured outcomes on key variables (e.g., over identification, stability of classification, academic and cognitive growth in response to treatment). The few studies that compare RTI with other assessment models (e.g., discrepancy based or low achievement based models) involve post hoc assessments of children divided at post-test within the same sample. In addition, different states and school districts have variations in their interpretations on how RTI should be implemented, thereby weakening any uniformity linking the science of instruction to assessing children at risk for LD. (p. 2)

The only meta-analysis of RTI published to date (Tran, Sanchez, Arellano, & Swanson, 2011; based on 13 studies) indicated that students identified as low responders to tier 1 and tier 2 RTI reading interventions show improvement but do not reach reading levels achieved by the high responders and that students with the lowest pre-intervention reading scores consistently remain the lowest performers, even with the flexibility and breadth of RTI interventions that are delivered. This study suggested that an initial standardized assessment may be most predictive of outcome as well as a way to identify students unlikely to respond to RTI more quickly than by multiple interventions across the RTI tiers. Moreover, even with intensive intervention, significant performance weaknesses remain for some students at risk for LD when compared to students who are more responsive to instruction. According to Tran and his colleagues (2011): "Unfortunately, the validity of RTI procedures, particularly in comparisons to other assessment approaches, has not been adequately established in the present synthesis of the literature" (p. 293).

A number of researchers, awaiting evidence on the effectiveness of RTI across achievement domains, have proposed alternative multimethod assessment approaches to the identification of SLD (e.g., Kavale & Forness, 2000; Kavale, Holdnack, & Mostert, 2005; Kavale, Kaufman, Naglieri, & Hale, 2005; Mather & Gregg, 2006; C. S. Robinson, Menchetti, & Torgesen, 2002). These approaches to identification of SLD have some or all of these elements in common:

1. Document the failure to achieve adequately or make sufficient progress in one or more areas of academic achievement through examination of educational history; class grades; work samples; standardized or curriculum-based measurement; analysis of performance process and quality; teacher, student, and parent reports; and/or response to interventions.
2. Identify one or more specific cognitive abilities or processes (e.g., phonological processing for reading decoding difficulties) that plausibly explain academic performance difficulties. Any contributory impaired abilities/processes need to have a research-based association with the specific domain of academic performance that is impaired. Multiple methods can be used to identify the contributory abilities or processes including standardized cognitive-intellectual test performance and formal observation, teacher ratings, or qualitative analysis of academic performance errors.
3. Rule out other explanations for the academic difficulties including intellectual disability, sensory disability, neurological trauma or condition, emotional-psychiatric disorder, or the consequences of an impoverished, disadvantaged, or culturally/linguistically different environment.

Intelligence tests quantify important personal resources (see the section on resilience and protective factors in the introduction to "Diagnostic Applications"), measure relevant cognitive abilities and processes, and have value in the identification of exclusionary diagnoses, such as intellectual disability. Intelligence tests can also provide information about verbal cognitive abilities and word knowledge, among other predictors of successful academic performance in reading.

Finally, a meta-analytic investigation has also provided evidence that the ability-achievement discrepancy may be effective in predicting intervention outcome within circumscribed ranges. Swanson (2003) reported that studies with aggregated intelligence and reading achievement scores that are both in the low range (<25th percentile, or a standard score of 90) yield significantly higher effect sizes related to intervention outcomes than studies with reading scores in the low range (<25th percentile) but with high IQ scores (e.g., IQ > 100). Accordingly, it appears there is value in the ability-achievement discrepancy method, albeit within specific parameters, and that complete abandonment of this approach may constitute "throwing the baby out with the bathwater" (Scruggs & Mastropieri, 2002, p. 165).

One of the unintended consequences of IDEA 2004's disparagement of ability-achievement discrepancies has been its unfortunate impact on identification of gifted students with specific learning disabilities (also known as *twice-exceptional*, or *Gifted 2e* students). These students are by nature exceptionally bright and often self-motivated, using their strengths to compensate for striking academic weaknesses for as long as they are able. Even when psychometric assessments identify large ability-achievement discrepancies, federal regulations discourage their interpretation and ambiguously promote use of achievement test performance patterns as a basis for identification:

> Discrepancy models are not essential for identifying children with SLD who are gifted. However, the regulations clearly allow discrepancies in achievement domains, typical of children with SLD who are gifted, to be used to identify children with SLD. (Rules and Regulations, *71 Fed. Reg.* 46647, August 14, 2006)

No states, to our knowledge, have published specific guidelines on identification of gifted SLD students, putting these students in the tragic wait-to-fail position that IDEA 2004 was intended to solve. Nicpon, Allmon, Sieck, and Stinson (2011) reported evidence of a pervasive and harmful misconception among educators that inclusion in gifted programs and selection for special education services are mutually exclusive.

TOWARD A MATURE CLINICAL SCIENCE

Whither goest intellectual assessment?

The future is difficult to predict, but past events suggest that the measurement of intelligence will continue to be a useful professional activity, with continued gradual improvements in practice but with a shrinking number of applications depending on sociopolitical winds and scientific progress.

Actual practice has changed little since the 1960s, when the Wechsler intelligence scales began their dominance. If intelligence tests continue to be revised every 10 to 15 years, then by about 2050 it is reasonable to expect seventh or eighth edition revisions of the Wechsler and Stanford-Binet intelligence scales. There are few changes in diagnostic applications on the horizon, and recent legislation (IDEA, 2004) discouraged the use of intelligence tests in assessment of learning disabilities. The theory of general intelligence, dating back over 100 years, still tends to guide most intellectual applications.

The tenuous link between assessment and intervention continues to be an Achilles' heel for intellectual assessment. Perhaps the most telling indicator of the limited intervention utility of intelligence tests may be found in the Maruish (2004) and Antony and Barlow (2010) volumes, totaling over 1,300 pages, on the use of psychological testing for treatment planning with *no* mention of intelligence or IQ. Ironically, at the start of intelligence testing Alfred Binet (1909/1975) was unequivocal about his belief in the effectiveness of cognitive intervention, describing programs and exercises to enhance the efficiency of cognitive faculties. Efforts to systematically link intelligence assessment and intervention, such as those of Feuerstein (e.g., Feuerstein, Feuerstein, & Falik, 2010), have failed to gain traction.

Progress in reaching scientific consensus on matters related to intelligence and its assessment has been mixed, with two professional consensus statements having been published in the 1990s. The first appeared in the *Wall Street Journal* in 1994, when Linda S. Gottfredson authored a statement with 25 conclusions and 52 signatories, "Mainstream Science on Intelligence" (Gottfredson, December 13, 1994, p. A18; see also Gottfredson, 1997). In 1996 an APA task force issued an authoritative scientific consensus statement about intelligence and its assessment entitled "Intelligence: Knowns and Unknowns" (Neisser et al., 1996). While Spearman's (1904) psychometric *g* was affirmed in the 1994 statement, the1996 APA consensus 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).

There are signs of potential change and guarded optimism in intelligence assessment. The CHC model offers a potentially unifying foundational structure for thinking about human cognitive abilities and intelligence. Advances in technology would seem to make it inevitable that the tradition of one examiner testing one student with verbal inquiries, stimulus materials, and manipulables will evolve toward increased online/computerized assessment and automated scoring, reporting, and interpretation. Psychometric techniques such as Rasch scaling have had little discernible impact on the material substance of intellectual tests thus far, but they promise the potential to reduce test development time and costs, thereby offering practitioners more choices in intelligence assessment.

An appraisal of the state of the science can lead only to the conclusion that intelligence assessment has yet to achieve status as a mature clinical science. The essential requirements of a mature clinical science, according to Millon (1999; Millon & Davis, 1996), are (a) a coherent foundational theory, from which testable principles and propositions may be derived; (b) a variety of assessment instruments, operationalizing the theory and serving the needs of special populations; (c) an applied diagnostic taxonomy, derived from and consistent with the theory and its measures; and (d) a compendium of change-oriented intervention techniques, aimed at modifying specific behaviors in a manner consistent with the theory. Three of these four criteria may arguably be said to have been met: Substantial advances in theories of intelligence have been made in recent years, a variety of intelligence tests are available, and diagnostic categories related to intelligence are in widespread use. Unfortunately, the field of psychology has yet to develop a systematic model linking intelligence assessment to intervention, making this long-sought objective a sort of holy grail necessary to move forward the science and practice of intelligence assessment.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text revision). Washington, DC: Author.

American Psychiatric Association. (2011). *DSM-5 development. Major neurocognitive disorder.* Retrieved from www.dsm5.org/ProposedRevision/Pages/proposedrevision.aspx?rid=419#

Anderson, K., Brown, C., & Tallal, P. (1993). Developmental language disorders: Evidence for a basic processing deficit. *Current Opinion in Neurology and Neurosurgery, 6,* 98–106.

Antony, M. M., & Barlow, D. H. (2010). *Handbook of assessment and treatment planning for psychological disorders* (2nd ed.). New York, NY: Guilford Press.

APA Task Force to Update the Guidelines for the Evaluation of Dementia and Age-Related Cognitive Decline. (2011). *Guidelines for the evaluation of dementia and age-related cognitive change.* Washington, DC: American Psychological Association.

Ashman, A. F., & Das, J. P. (1980). Relation between planning and simultaneous-successive processing. *Perceptual and Motor Skills, 51,* 371–382.

Baade, L. L., & Schoenberg, M. R. (2004). A proposed method to estimate premorbid intelligence utilizing group achievement measures from school records. *Archives of Clinical Neuropsychology, 19,* 227–243.

Bain, S. K., & Allin, J. D. (2005). Review: Stanford-Binet Intelligence Scales, Fifth Edition. *Journal of Psychoeducational Assessment, 23,* 87–95.

Barkley, R. A. (1997). *ADHD and the nature of self-control.* New York, NY: Guilford Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52,* 885–887.

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 Publications.

Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychology in the Schools, 46,* 932–950.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science, 11,* 474–480.

Benbow, C. P., & Stanley, J. C. (Eds.). (1983). *Academic precocity: Aspects of its development.* Baltimore, MD: Johns Hopkins University Press.

Benson, N. (2008). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment, 26,* 27–41.

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

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.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417–424.

Borkowski, J. G., Carothers, S. S., Howard, K., Schatz, J., & Farris, J. R. (2007). Intellectual assessment and intellectual disability. In J. W. Jacobson, J. A. Mulick, & J. Rojahn (Eds.), *Handbook of intellectual and developmental disabilities* (pp. 261–278). New York, NY: Springer.

Bracken, B. A. (2005). Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 892–894). Lincoln, NE: Buros Institute of Mental Measurements.

Burt, K. B., & Roisman, G. I. (2010). Competence and psychopathology: Cascade effects in the NICHD Study of Early Child Care and Youth Development. *Development and Psychopathology, 22,* 557–567.

Camara, W. J. (1997). Use and consequences of assessments in the USA: Professional, ethical, and legal issues. *European Journal of Psychological Assessment, 13,* 140–152.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research & Practice, 31,* 141–154.

Cangelosi, M. D. (2009). Identification of specific learning disability: A survey of school psychologists' knowledge and current practice. *Dissertation Abstracts International: Section B: The Sciences and Engineering,* AAT 3388260.

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,* 533–541.

Canivez, G. L. (2010). Review of the Wechsler Adult Intelligence Scale—Fourth Edition. In R. A. Spies, J. F. Carlson, and K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 684–688). Lincoln, NE: Buros Institute of Mental Measurement.

Canivez, G. L. (2011a). *Cognitive Assessment System: An orthogonal hierarchical factor structure investigation.* Paper presented at the annual convention of the Midwestern Psychological Association, Chicago, IL.

Canivez, G. L. (2011b). Hierarchical factor structure of the Cognitive Assessment System: Variance partitions from the Schmid–Leiman (1957) procedure. *School Psychology Quarterly, 26,* 305–317.

Canivez, G. L. (2011c). *Incremental predictive validity of Cognitive Assessment System PASS scores.* Paper presented at the annual convention of the National Association of School Psychologists, San Francisco, CA.

Canivez, G. L. (2011d). *Interpretation of Cognitive Assessment System scores: Considering incremental validity of PASS Scores in predicting achievement.* Paper presented at the annual convention of the American Psychological Association, Washington, DC.

Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly, 25,* 223–235.

Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment, 22,* 827–836.

Carlson, J., & Das, J. P. (1997). A process approach to remediating word-decoding deficiencies in Chapter 1 children. *Learning Disability Quarterly, 20,* 93–102.

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, NY: Academic Press.

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

Carroll, J. B. (1995). Review of the book *Assessment of cognitive processing: The PASS theory of intelligence. Journal of Psychoeducational Assessment, 13,* 397–409.

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, MA: Pergamon Press.

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.

Ciotti, R. M. (2007). The relationships among executive function scores: Examining the Cognitive Assessment System (CAS), Tower of London-Drexel (TOLDX), and the Behavior Rating Inventory Executive Function (BRIEF). *ProQuest Dissertations and Theses,* AAT 3267677.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.

Columbus Group. (1991, July). Unpublished transcript of meeting. Columbus, OH.

Cormier, P., Carlson, J. S., & Das, J. P. (1990). Planning ability and cognitive performance: The compensatory effects of a dynamic assessment approach. *Learning and Individual Differences, 2,* 437–449.

Damiani, V. B. (1997). Young gifted children in research and practice: The need for early childhood programs. *Gifted Child Today, 20*(3), 18–23.

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.

Das, J. P. (2004). Theories of intelligence: Issues and applications. In G. Goldstein & S. R. Beers (Eds.), *Comprehensive handbook of psychological assessment (Vol. 1): Intellectual and neuropsychological assessment* (pp. 5–23). Hoboken, NJ: Wiley.

Das, J. P., & Kendrick, M. (1997). PASS Reading Enhancement Program: A short manual for teachers. *Journal of Cognitive Education, 5,* 193–208.

Das, J. P., Kirby, J. R., & Jarman, R. F. (1975). Simultaneous and successive synthesis: 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, NY: Academic Press.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence.* Needham Heights, MA: Allyn & Bacon.

Davis, A. S., & Finch, W. H. (2008). Review of Differential Ability Scales, Second Edition. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds.), *Eighteenth Annual Mental Measurements Yearbook.* Lincoln, NE: Buros Mental Measurements Institute.

Deng, C., Liu, M., Wei, W., Chan, R. C. K., & Das, J. P. (2011). Latent factor structure of the Das-Naglieri Cognitive Assessment System: A confirmatory factor analysis in a Chinese setting. *Research in Developmental Disabilities, 32,* 1988–1997.

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: Meta-analyses of switching and Stroop studies. *Human Brain Mapping, 25,* 22–34.

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet, Fifth Edition. *Journal of Psychoeducational Assessment, 24,* 123–136.

Dombrowski, S. C., & Mrazik, M. (2008). Test review: Reynolds Intellectual Assessment Scales. *Canadian Journal of School Psychology, 23,* 223–230.

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment, 27,* 494–507.

Dumont, R., Willis, J. O., & Elliott, C. D. (2009). *Essentials of DAS-II assessment.* Hoboken, NJ: Wiley.

Education for All Handicapped Children Act. Pub. L. No. 94–142 (1975). 20 U.S.C. § 1401 et seq.

Edwards, O. W., & Paulin, R. V. (2007). Referred students' performance on the Reynolds Intellectual Assessment Scales and the Wechsler Intelligence Scale for Children–Fourth Edition. *Journal of Psychoeducational Assessment, 25,* 334–340.

Egeland, B. R., Carlson, E., & Sroufe, L. A. (1993). Resilience as process. *Development and Psychopathology, 5,* 517–528.

Elliott, C. D. (1983). *The British Ability Scales. Manual 1: Introductory handbook.* Windsor, UK: NFER-Nelson.

Elliott, C. D. (2007a). *Differential Ability Scales—Second Edition: Administration and scoring manual.* San Antonio, TX: Pearson.

Elliott, C. D. (2007b). *Differential Ability Scales—Second Edition: Introductory and technical handbook.* San Antonio, TX: Pearson.

Elliott, C. D. (2007c). *Differential Ability Scales—Second Edition: Normative data tables manual.* San Antonio, TX: Pearson.

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., Murray, D. J., & Pearson, L. S. (1979). *British Ability Scales.* Slough, UK: NFER.

Elliott, R. W. (2004). Reynolds Intellectual Assessment Scales. *Archives of Clinical Neuropsychology, 19,* 325–328.

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.

Eysenck, H. J. (1993). Comment on Goldberg. *American Psychologist, 48,* 1299–1300.

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34,* 363–373.

Feuerstein, R., Feuerstein, R. S., & Falik, L. H. (2010). *Beyond smarter: Mediated learning and the brain's capacity for change.* New York, NY: Teachers College Press.

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

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. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., . . . Shaywitz, B. A. (2002). Classification of learning disabilities: An evidenced-based evaluation. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.). *Identification of learning disabilities: Research to practice* (pp. 185–250). Mahwah, NJ: Erlbaum.

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,* 155–171.

Floyd, R. G., Keith, T., Taub, G. McGrew, K. (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,* 200–223.

Floyd, R. G., McGrew, K. & Evans, 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.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*(2), 98–108.

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 analysis. *Journal of Educational Psychology, 88,* 3–17.

Franzen, R. (1920). The accomplishment quotient: A school mark in terms of individual capacity. *Teachers College Record, 21,* 432–440.

Ganguli, M., Blacker, D., Blazer, D. G., Grant, I., Jeste, D. V., Paulsen, J. S., . . . Sachdev, P. S. (2011). Classification of neurocognitive disorders in DSM-5: A work in progress. *American Journal of Geriatric Psychiatry, 19*(3), 205–209.

Gilman, B. J. (2008). *Academic advocacy for gifted children: A parent's complete guide.* Scottsdale, AZ: Great Potential Press.

Godbolt, A. K., Cipolotti, L., Watt, H., Fox, N. C., Janssen, J. C., & Rossor, M. N. (2004). The natural history of Alzheimer disease: A longitudinal presymptomatic and symptomatic study of a familial cohort. *Archives of Neurology, 61,* 1743–1748.

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal, 35,* 515–531.

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

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

Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12,* 249–267.

Gross, M. (2000). Exceptionally and profoundly gifted students: An underserved population. *Understanding Our Gifted, 12*(2), 3–9.

Gustafsson, J.-E. (1984). A unifying model of the structure of intellectual abilities. *Intelligence, 8,* 179–203.

Gustafsson, J.-E. (1988). Hierarchical models for individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.

Gutentag, S. S., Naglieri, J. A., & Yeates, K. O. (1998). Performance of children with traumatic brain injury on the cognitive assessment system. *Assessment, 5,* 263–272.

Haddad, F. A. (2004). Planning versus speed: An experimental examination of what Planned Codes of the Cognitive Assessment System measures. *Archives of Clinical Neuropsychology, 19,* 313–317.

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.

Harrison, C. (2004). Giftedness in early childhood: The search for complexity and connection. *Roeper Review, 26,* 78–84.

Harrison, P. L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment, 6,* 188–198.

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

Hill, V. (2005). Through the past darkly: A review of the British Ability Scales Second Edition. *Child and Adolescent Mental Health, 10,* 87–98.

Hodge, K. A., & Kemp, C. R. (2000). Exploring the nature of giftedness in preschool children. *Journal for the Education of the Gifted, 24,* 46–73.

Hollingworth, L. S. (1931). The child of very superior intelligence as a special problem in social adjustment. *Mental Hygiene, 15*(1), 3–16.

Hollingworth, L. S. (1942). *Children above 180 IQ (Stanford-Binet): Origin and development*. Yonkers-on-Hudson, NY: World Book.

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, NY: Guilford Press.

Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review, 29,* 102–119.

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78,* 250–261.

Individuals with Disabilities Education Act, Pub. L. No. 105–17 (1997, 1999). 20 U.S.C. § 1400 et seq. (Statute). 34 C.F.R. 300 (Regulations).

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108–446, 20 U.S.C. § 1401, 118 STAT. 2657 (2004).

Jackson, N. E. (2003). Young gifted children. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 470–482). Boston, MA: Allyn & Bacon.

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

The K-ABC–Will it be the SOMPA of the 80's? (1984). *Ohio School Psychologist* [Ohio School Psychologists Association], *29*(4), 9–11.

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

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 (WISC-III PI).* San Antonio, TX: Psychological Corporation.

Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *Wechsler Adult Intelligence Scale—Revised as a Neuropsychological Instrument (WAIS-R NI).* San Antonio, TX: Psychological Corporation.

Kar, B. C., Dash, U. N., Das, J. P., & Carlson, J. S. (1992). Two experiments on the dynamic assessment of planning. *Learning and Individual Differences, 5,* 13–29.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York, NY: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983a). *Kaufman Assessment Battery for Children administration and scoring manual.* Circle Pines, MN: American Guidance.

Kaufman, A. S., & Kaufman, N. L. (1983b). *Kaufman Assessment Battery for Children interpretive manual.* Circle Pines, MN: American Guidance.

Kaufman A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence Test.* Circle Pines, MN: American Guidance.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition (KABC-II): Manual.* Bloomington, MN: 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., & Sternberg, R. J. (2008). Conceptions of giftedness. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children* (pp. 71–91). New York, NY: Springer.

Kavale, K. A., & Forness, S. R. (2000). What definitions of learning disabilities say and don't say: A critical analysis. *Journal of Learning Disabilities, 33,* 239–256.

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.

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., Kaufman, A. S., Naglieri, J. A., & Hale, J. (2005). Changing procedures for identifying learning disabilities: The danger of poorly supported ideas. *School Psychologist, 59,* 16–25.

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, 35,* 108–127.

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. *School Psychology Review, 28,* 303–321.

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 (3rd ed.). *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.

Kirk, S. A. (1962). *Educating exceptional children.* Boston, MA: Houghton Mifflin.

Kitano, M. K. (1995). Language diversity and giftedness: Working with gifted English language learners. *Journal for the Education of the Gifted, 18,* 234–254.

Kline, R. B., Snyder, J., Castellanos, M. (1996). Lessons from the Kaufman Assessment Battery for Children (K-ABC): Toward a new cognitive assessment model. *Psychological Assessment, 8,* 7–17.

Krach, S. K., Loe, S. A., Jones, W. P., & Farrally, A. (2009). Convergent validity of the Reynolds Intellectual Assessment Scales (RIAS) using the Woodcock-Johnson Tests of Cognitive Ability, Third Edition (WJ-III) with university students. *Journal of Psychoeducational Assessment, 27,* 355–365.

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,* 117–144.

Kranzler, J. H., Keith, T. Z., & Flanagan, D. P. (2000). Independent examination of the factor structure of the Cognitive Assessment System (CAS): Further evidence challenging the construct validity of the CAS. *Journal of Psychoeducational Assessment, 18,* 143–159.

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

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

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

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

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

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

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

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

Livingston, R. B., Jennings, E., Reynolds, C. R., & Gray, R. M. (2003). Multivariate analyses of the profile stability of intelligence tests: High for IQs, low to very low for subtest analyses. *Archives of Clinical Neuropsychology, 18,* 487–507.

Loeber, R., Pardini, D. A., Stouthamer-Loeber, M., & Raine, A. (2007). Do cognitive, physiological, and psychosocial risk and promotive factors predict desistance from delinquency in males? *Development and Psychopathology, 19,* 867–887.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1,* 316–345.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86,* 718–720.

Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., . . . Tasse, M. J. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.

Luria, A. R. (1966). *Human brain and psychological processes.* New York, NY: Harper & Row.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222* (3), 66–78.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology.* New York, NY: Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.; B. Haigh, Trans.). New York, NY: Basic Books.

Luthar, S. S., D’Avanzo, K., & Hites, S. (2003). Maternal drug abuse versus other psychological disturbances: Risks and resilience among children. In S. S. Luthar (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. 104–129). New York, NY: Cambridge University Press.

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman’s “Intelligent Testing” approach to the WISC-III. *School Psychology Quarterly, 12,* 197–234.

Marland, S. P. (1972). *Education of the gifted and talented: Vol. 1. Report to the Congress of the United States by the U.S. Commissioner of Education.* Washington, DC: U.S. Government Printing Office.

Maruish, M. E. (Ed.). (2004). *The use of psychological testing for treatment planning and outcomes assessment. (Vol. 2) Instruments for children and adolescents* (3rd ed.). Mahwah, NJ: Erlbaum.

Maruta, C., Guerreiro, M., de Mendonça, A., Hort, J., & Scheltens, P. (2011). The use of neuropsychological tests across Europe: The need for a consensus in the use of assessment tools for dementia. *European Journal of Neurology, 18,* 279–285.

Masten, A. S. (1994). Resilience in individual development: Successful adaptation despite risk and adversity. In M. C. Wang & E. W. Gordon (Eds.), *Educational resilience in inner-city America: Challenges and prospects* (pp. 1–25). Hillsdale, NJ: Erlbaum.

Masten, A. S., Burt, K. B., Roisman, G. I., Obradovic, J., Long, J. D., & Tellegen, A. (2004). Resources and resilience in the transition to adulthood: Continuity and change. *Development and Psychopathology, 16,* 1071–1094.

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

Mather, N., & Gregg, N. (2006). Specific learning disabilities: Clarifying, not eliminating, a construct. *Professional Psychology: Research and Practice, 37,* 99–106.

Mather, N., & Jaffe, L. E. (2002). *Woodcock-Johnson III: Reports, recommendations, and strategies.* Hoboken, NJ: Wiley.

Mather, N., & Woodcock, R. W. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities examiner’s manual: Standard and extended batteries.* Itasca, IL: Riverside.

Matthews, D. J. (1997). Diversity in domains of development: Research findings and their implications for gifted identification and programming. *Roeper Review, 19* (3), 172–177.

Matthews, D, J., & Keating, D. P. (1995). Domain specificity and habits of mind: An investigation of patterns of high-level development. *Journal of Early Adolescence, 15,* 319–343.

Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology, 84,* 405–412.

McCrea, S. M. (2009). A review and empirical study of the composite scales of the Das-Naglieri Cognitive Assessment System. *Psychology Research and Behavior Management, 2,* 59–79.

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.

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. 151–180). New York, NY: Guilford Press.

McGrew, K. S., Dailey, D. E. H., & Schrank, F. A. (2007). *Woodcock-Johnson III/Woodcock-Johnson III Normative Update score differences: What the user can expect and why (Woodcock-Johnson III Assessment Service Bulletin No. 9).* Rolling Meadows, IL: Riverside.

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., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III Normative Update technical manual.* Rolling Meadows, IL: Riverside.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R) technical manual.* Allen, TX: DLM.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual. Woodcock-Johnson III.* Itasca, IL: Riverside.

McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist, 19,* 871–882.

Meikamp, J. (1999). Review of the Das-Naglieri Cognitive Assessment System. In. B. S. Plake & J. C. Impara (Eds.), *The supplement to the thirteenth mental measurements yearbook* (pp. 75–77). Lincoln, NE: Buros Institute.

Millon, T. (1999). Reflections on psychosynergy: A model for integrating science, theory, classification, assessment, and therapy. *Journal of Personality Assessment, 72,* 437–456.

Millon, T., & Davis, R. D. (1996). *Disorders of personality: DSM-IV and beyond* (2nd. ed). New York, NY: Wiley.

Monroe, M. (1932). *Children who cannot read.* Chicago, IL: University of Chicago Press.

Moon, T. R., & Brighton, C. M. (2008). Primary teachers' conceptions of giftedness. *Journal for the Education of the Gifted, 31,* 447–480.

Morelock, M. J. (1992). Giftedness: The view from within. *Understanding Our Gifted, 4*(3), 1, 11–15.

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. (1999). *Essentials of CAS assessment.* New York, NY: Wiley.

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., & Das, J. P. (1987). Construct and criterion related validity of planning, simultaneous, and successive cognitive processing tasks. *Journal of Psychoeducational Assessment, 5,* 353–363.

Naglieri, J. A., & Das, J. P. (1988). Planning-Arousal-Simultaneous-Successive (PASS): A model for assessment. *Journal of School Psychology, 26,* 35–48.

Naglieri, J. A., & Das, J. P. (1997a). *Cognitive Assessment System: Administration and scoring manual.* Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (1997b). *Cognitive Assessment System: Interpretive handbook.* Itasca, IL: Riverside.

Naglieri, J. A., DeLauder, B. Y., Goldstein, S., & Schwebech, A. (2006). WISC-III and D-N CAS: Which correlates higher with achievement for a clinical sample? *School Psychology Quarterly, 21,* 62–76.

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, NY: Springer.

Naglieri, J. A., & Pickering, E. B. (2010). *Helping children learn: Intervention handouts for use at school and home* (2nd ed.). Baltimore, MD: Brookes.

Naglieri, J. A., & Rojahn, J. (2001). Intellectual classification of Black and White children in special education programs using the WISC-III and the Cognitive Assessment System. *American Journal on Mental Retardation, 106,* 359–367.

Naglieri, J. A., & Rojahn, J. (2004). Construct validity of the PASS theory and CAS: Correlations with achievement. *Journal of Educational Psychology, 96,* 174–181.

National Association for Gifted Children. (2008). *The role of assessments in the identification of gifted students* (NAGC Position Paper). Washington, DC: Author.

National Association for Gifted Children. (2010). *Use of the WISC-IV for gifted identification* (NAGC Position Paper). Washington, DC: Author.

Neihart, M. (1999). The impact of giftedness on psychological well-being: What does the empirical literature say? *Roeper Review, 22,* 10–17.

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

Nelson, H. E. (1982). *The National Adult Reading Test (NART). Test manual.* Windsor, UK: NFER-Nelson.

Nelson, J. M., Canivez, G. L, Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology, 45,* 439–456.

Newman, T. M., Sparrow, S. S., & Pfeiffer, S. I. (2008). The use of the WISC-IV in assessment and intervention planning for children who are gifted. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 217–272). San Diego, CA: Academic Press.

Nicpon, M. F., Allmon, A., Sieck, B., & Stinson, R. D. (2011). Empirical investigation of twice-exceptionality: Where have we been and where are we going? *Gifted Child Quarterly, 55,* 3–17.

Paivio, A. (1976). Concerning dual-coding and simultaneous-successive processing. *Canadian Psychological Review, 17,* 69–72.

Paolitto, A. W. (1999). Clinical validation of the Cognitive Assessment System with children with ADHD. *ADHD Report, 7,* 1–5.

Pearson. (2009a). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV.* San Antonio, TX: Author.

Pearson. (2009b). *Test of Pre-Morbid Functioning.* San Antonio, TX: Author.

Pearson research staff. (2009). *Wechsler Individual Achievement Test–Third Edition (WIAT-III) technical manual.* San Antonio, TX: Pearson.

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R. (2009). Mild cognitive impairment: Ten years later. *Archives of Neurology, 66,* 1447–1455.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology, 56,* 303–308.

Petersen, R. C., Stevens, J. C., Gangulli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). *Neurology, 56,* 1133–1142.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Precocity is not a menace. (1925, May 31). *Los Angeles Times,* p. H9.

Procedures for evaluating specific learning disabilities. 42 *Federal Register* 65082 (Dec. 29, 1977).

Puhan, G., Das, J. P., & Naglieri, J. A. (2005). Separating planning and attention: Evidential and consequential validity. *Canadian Journal of School Psychology, 20,* 75–83.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20,* 33–65.

Radke-Yarrow, M., & Sherman, T. (1990). Hard growing: Children who survive. In R. E. Rolf, A. S. Masten, D. Cicchetti, K. H. Nuecherlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 97–119). New York, NY: Cambridge University Press.

Raiford, S. E., Weiss, L. G., Rolfhus, E., & Coalson, D. (2005/2008). *WISC-IV technical report no. 4 (2008 Update): General Ability Index.* San Antonio, TX: Pearson.

Reschly, D. J. (2000). The present and future status of school psychology in the United States. *School Psychology Review, 29,* 507–522.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales (RIAS) and the Reynolds Intellectual Screening Test (RIST): Professional manual.* Lutz, FL: PAR.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to Intervention: Ready or not? Or, from Wait-to-Fail to Watch-Them-Fail. *School Psychology Quarterly, 24,* 130–145.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (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.

Ritchie, L. J., & Tuokko, H. (2010). Patterns of cognitive decline, conversion rates, and predictive validity for 3 Models of MCI. *American Journal of Alzheimer's Disease & Other Dementias, 25,* 592–603.

Robertson, S. G., Pfeiffer, S. I., & Taylor, N. (2011). Serving the gifted: A national survey of school psychologists. *Psychology in the Schools, 48,* 786–799.

Robinson, C. S., Menchetti, B. M., & Torgesen, J. K. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learning Disabilities Research & Practice, 17,* 81–90.

Robinson, N. M. (2008). The social world of gifted children and youth. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children: Psycho-educational theory, research, and best practices* (pp. 33–51). New York, NY: Springer Science.

Roedell, W. C. (1989). Early development of gifted children. In J. VanTassel-Baska & P. Olszewski-Kubilius (Eds.), *Patterns of influence on gifted learners: The home, the self, and the school* (pp. 13–28). New York, NY: Teachers College Press.

Roid, G. H. (2003a). *Stanford-Binet Intelligence Scales, Fifth Edition: Examiner's manual.* Rolling Meadows, IL: Riverside.

Roid, G. H. (2003b). *Stanford-Binet Intelligence Scales, Fifth Edition: Interpretive manual. Expanded guide to the interpretation of SB5 test results.* Rolling Meadows, IL: Riverside.

Roid, G. H. (2003c). *Stanford-Binet Intelligence Scales, Fifth Edition: Technical manual.* Rolling Meadows, IL: Riverside.

Rotigel, J. V. (2003). Understanding the young gifted child: Guidelines for parents, families, and educators. *Early Childhood Education Journal, 30,* 209–214.

Rules and Regulations. (2006, August 14). *Federal Register, 71,* 46647.

Rutter, M. (1978). Prevalence and types of dyslexia. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 5–28). New York, NY: Oxford University Press.

Rutter, M. (1987). Psychosocial resilience and protective mechanism. *American Journal of Orthopsychiatry, 57,* 316–331.

Rutter, M., & Yule, W. (1973). Specific reading retardation. In L. Mann & D. Sabatino (Eds.), *The first review of special education* (pp. 49–62). Philadelphia, PA: JSE Press.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16,* 181–197.

Saklofske, D. H., Weiss, L. G., Raiford, S. E., & Prifitera, A. (2006). Advanced interpretive issues with the WISC-IV Full Scale IQ and General Ability Index scores. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 99–138). Burlington, MA: Academic Press.

Sankar-DeLeeuw, N. (2004). Case studies of gifted kindergarten children: Profiles of promise. *Roeper Review, 26,* 192–207.

Scarmeas, N., & Stern, Y. (2003). Cognitive reserve and lifestyle. *Journal of Clinical and Experimental Neuropsychology, 25,* 625–633.

Schalock, R. L., and the Ad Hoc Committee on Terminology and Classification. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Luckasson, R. A., Shogren, K. A., Borthwick-Duffy, S., Bradley, V., Buntinx, W. H., . . . Yeager, M. H. (2007). The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities, 45,* 116–124.

Schrank, F. A., & Flanagan, D. P. (Eds.) (2003). *WJ III clinical use and interpretation: Scientist-practitioner perspectives.* San Diego, CA: Academic Press.

Scruggs, T. E., & Mastropieri, M. A. (2002). On babies and bathwater: Addressing the problems of identification of learning disabilities. *Learning Disability Quarterly, 25,* 155–168.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 298,* 199–209.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Distribution and temporal stability of dyslexia in an epidemiological sample of 414 children followed longitudinally. *New England Journal of Medicine, 326,* 145–150.

Siegle, D., & McCoach, D. B. (2010). Redefining giftedness for a new century: Shifting the paradigm. *Journal of Advanced Academics, 22,* 6–8.

Silverman, L. K. (Ed.) (1993). *Counseling the gifted and talented*. Denver, CO: Love.

Silverman, L. K. (2002a). Asynchronous development. In M. Neihart, S. Reis, N. Robinson, & S. Moon (Eds.), *The social and emotional development of gifted children: What do we know?* National Association for Gifted Children (pp. 31–37). Waco, TX: Prufrock Press.

Silverman, L. K. (2002b). *Upside-down brilliance: The visual-spatial learner.* Denver, CO: DeLeon.

Silverman, L. K. (2009). Asynchrony. In B. Kerr (Ed.), *Encyclopedia of giftedness, creativity, and talent* (Vol. 1, pp. 67–70). Thousand Oaks, CA: Sage.

Silverman, L. K. (2012). Asynchronous development: A key to counseling the gifted. In T. L. Cross & J. R. Cross (Eds.), *Handbook for counselors serving students with gifts and talents* (pp. 261–279). Waco, TX: Prufrock.

Snowdon D. A., Greiner, L. H., Mortimer, J. A., Riley, K. P., Greiner, P. A., & Markesbery, W. R. (1997). Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. *Journal of the American Medical Association, 277,* 813–817.

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

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

Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36,* 24–33.

Stanley, J. C. (1990). Leta Hollingworth's contributions to above-level testing of the gifted. *Roeper Review, 13,* 166–171.

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.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profiles of children with reading disabilities: A regression-based test of the phonological-core variable difference model. *Journal of Educational Psychology, 86,* 24–53.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8,* 448–460.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia, 47,* 2015–2028.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39,* 469–518.

Swanson, H. L. (2003). Does IQ and reading level influence treatment outcomes? Implications for the definition of learning disabilities. *Advances in Learning and Behavioral Disabilities, 16,* 205–222.

Swanson, H. L. (2010). What about IQ again and avoiding Poison Oak. *New Times for DLD* [Newsletter for the Division for Learning Disabilities special interest group of the Council for Exceptional Children], *28*(1), 1–3.

te Nijenhuis, J., & Hartmann, P. (2006). Spearman's "law of diminishing returns" in samples of Dutch and immigrant children and adults, *Intelligence, 34,* 437–447.

Terman, L. M. (1916). *The measurement of intelligence.* Cambridge, MA: Riverside Press.

Terman, L. M. (1925). *Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children.* Stanford, CA: Stanford University Press.

Terman, L. M. (1931). The gifted child. In C. Murchison (Ed.), *A handbook of child psychology* (pp. 568–584). Worcester, MA: Clark University Press.

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

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: 1973 norms edition.* Boston, MA: Houghton Mifflin.

Terrassier, J.-C. (1985). Dyssynchrony—uneven development. In J. Freeman (Ed.), *The psychology of gifted children: Perspectives on development and education* (pp. 265–274). New York, NY: Wiley.

Thompson, B. (2005). Review of Wechsler Intelligence Scale for Children—Fourth Edition. In R. Spies & B. Plake (Eds.), *The mental measurements yearbook* (16th ed., pp. 262–263). Lincoln, NE: Buros Institute of Mental Measurements.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet intelligence scale: Fourth edition.* Itasca, IL: Riverside.

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

Tolan, S. S. (1994). Giftedness as asynchronous development. *Tip Network News, 4*(1), 1, 7. Retrieved from www.stephanietolan.com/gt_as_asynch.htm

Tran, L., Sanchez, T., Arellano, B., & Swanson, H. L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities, 44,* 283–295.

Undheim, J. O. (1981). On intelligence: II. A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology, 22,* 181–187.

U.S. Department of Education, Office of Special Education and Rehabilitative Services. (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC: U.S. Department of Education Publications Center.

Van Tassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly, 51,* 218–231.

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.

Walker, B., Hafenstein, N. L., & Crow-Enslow, L. (1999). Meeting the needs of gifted learners in the early childhood classroom. *Young Children, 54*(1), 32–36.

Ward, K. E., Rothlisberg, D. E., McIntosh, D. E., & Bradley, M. H. (2011). Assessing the SB-V factor structure in a sample of preschool children. *Psychology in the Schools, 48,* 454–463.

Wasserman, J. D. (2007). Intellectual assessment of exceptionally and profoundly gifted children. In K. Kay, D. Robson, & J. F. Brenneman (Eds.), *High IQ kids: Collected insights, information and personal stories from the experts* (pp. 48–65). Minneapolis, MN: Free Spirit.

Wasserman, J. D. (2010, March). *Recent developments in psychological testing of intelligence with gifted and talented students.* Workshop presentation to the Association of Practicing Psychologists, Montgomery-Prince George's Counties, MD.

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, NY: Guilford Press.

Wasserman, J. D., & Maccubbin, E. M. (2004). *Dark science: A critical review of the Woodcock-Johnson III Tests of Cognitive Abilities.* Unpublished test review.

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

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion? *The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 2*(2), 118–141.

Watkins, M. W. (2006). Orthogonal higher-order structure of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment, 18,* 123–125.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment, 22,* 782–787.

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment, 16,* 133–138.

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children—Fourth Edition among referred students. *Educational and Psychological Measurement, 66,* 975–983.

Webb, J. T., Gore, J. L., Amend, E. W., & DeVries, A. R. (2007). *A parent's guide to gifted children.* Scottsdale, AZ: Great Potential Press.

Wechsler, D. (1925). On the specificity of emotional reactions. *Journal of Psychology, 36,* 424–426.

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

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

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist, 5,* 78–83.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale manual.* New York, NY: Psychological Corporation.

Wechsler, D. (1958). Intelligence et fonction cérébrale. *Revue de Psychologie Appliquée, 8,* 143–147.

Wechsler, D. (1961). Intelligence, memory, and the aging process. In P. Hoch & J. Zubin (Eds.), *Psychopathology of aging* (pp. 152–159). New York, NY: Grune & Stratton.

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

Wechsler, D. (1975). A conversation with Dr. David Wechsler [Interviewers: Jerome E. Doppelt and Roger T. Lennon]. Unpublished transcript of interview. San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV): Administration and scoring manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV): Technical and interpretive manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): Administration and scoring manual.* San Antonio, TX: Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): Technical and interpretive manual.* San Antonio, TX: Pearson.

Wechsler, D. (2012). *Wechsler Abbreviated Scale of Intelligence—Second Edition (WASI-II) manual.* San Antonio, TX: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., & Maerlender, A. (2004). *Wechsler Intelligence Scale for Children Fourth Edition—Integrated (WISC-IV Integrated) administration and scoring manual.* San Antonio, TX: Harcourt Assessment.

Weider, A. (1995). An interview with Arthur Weider/Interviewer: John D. Wasserman. *Unpublished transcript of interview.* (Available from interviewer.)

Weiss, L. G., Saklofske, D. H., Schwartz, D. M., Prifitera, A., & Courville, T. (2006). Advanced clinical interpretation of WISC-IV index scores. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 139–179). Burlington, MA: Academic Press.

Wendling, B. J., & Mather, N. (2009). *Essentials of evidence-based academic interventions.* Hoboken, NJ: Wiley.

Werner, E. E. (1993). Risk, resilience, and recovery: Perspectives from the Kauai Longitudinal Study. *Development and Psychopathology, 5,* 503–515.

Werner, E. E. (2000). Protective factors and individual resilience. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 115–132). New York, NY: Cambridge University Press.

Werner, H. (1948). *Comparative psychology of mental development.* New York, NY: International Universities Press.

Widiger, T. A. (2011). A shaky future for personality disorders. *Personality Disorders: Theory, Research, and Treatment, 2,* 54–67.

Williams, T. H., McIntosh, D. E., Dixon, F., Newton, J. H., & Youman, E. (2010). A confirmatory factor analysis of the Stanford-Binet Intelligence Scales, Fifth Edition, with a high-achieving sample. *Psychology in the Schools, 47,* 1071–1083.

Winsler, A., & Naglieri, J. A. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children age 5 to 17. *Child Development, 74,* 659–678.

Winsler, A., Naglieri, J. A., & Manfra, L. (2006). Children's search strategies and accompanying verbal and motor strategic behavior: Developmental trends and relations with task performance among children age 5 to 17. *Cognitive Development, 21,* 232–248.

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

Woodcock, R. W. (1977). *Woodcock-Johnson Psycho-Educational Battery. Technical report.* Boston, MA: Teaching Resources.

Woodcock, R. W. (1993). An information processing view of Gf-Gc theory. *Journal of Psychoeducational Assessment Monograph Series: Woodcock-Johnson Psycho-Educational Assessment Battery–Revised*, pp. 80–102. Cordova, TN: Psychoeducational Corporation.

Woodcock, R. W. (1998a). Extending Gf-Gc theory into practice. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137–156). Mahwah, NJ: Erlbaum.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery.* Boston, MA: Teaching Resources Corporation.

Woodcock, R. W., & Johnson, M. B. (1989, 1990). *Woodcock-Johnson Psycho-Educational Battery—Revised.* Allen, TX: DLM.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001a, 2007a). *Woodcock-Johnson III Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b, 2007b). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003, 2007). *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Zhu, J., Cayton, T., Weiss, L., & Gabel, A. (2008). *WISC-IV extended norms (WISC-IV technical report no. 7).* San Antonio, TX: Pearson.