PEABODY JOURNAL OF EDUCATION, *77*(2), 6–22

New Looks in the Assessment of Cognitive Ability

Richard W. Woodcock
Peabody College
Vanderbilt University

The past 30 years have produced major changes in the measurement of cognitive ability and the interpretation of assessment results. Theory describing the factorial structure of cognitive ability has blossomed, and the results are visible in several recently published batteries of intellectual ability. The application of better theory to new assessment instruments has been facilitated by advances in the psychometric and statistical tools available to test developers. Attention is drawn to a concern about the capability of many clinicians to appreciate the importance of these changes and to apply them in practice without adequate continuing education.

The primary purpose for cognitive testing should be to find out more about the problem, not to obtain an IQ (Woodcock, 1997). It is essential that clinicians be aware of the major theoretical advances occurring in their field and appreciate the benefit to their assessment responsibilities accruing from the use of modern instruments reflecting those advances. Toward this end, two topics are addressed: (a) A review of the major conceptualizations of intelligence is presented, emphasizing recent advances in cognitive theory; and (b) information is provided about some recent advances in the statistical tools used by test developers that have greatly aided their work.

I wish to thank Fredrick Schrank and Kevin McGrew for their many helpful suggestions and critique of the manuscript.

Requests for reprints should be sent to Richard Woodcock, Peabody College, Box 40, 230 Appleton Place, Vanderbilt University, Nashville, TN 37203-5701.

Advances in Theory

Conceptualizations of Intelligence

There has been a clear trend in views of intelligence, from the simple to the more complex, as evidenced by the theoretical and interpretive models applied to tests of intelligence. The concept of a single, general intellectual ability dominated test development and interpretation for the first half of the 20th century. The beginning of the 21st century has witnessed the identification of a sizable set of specific, or narrow, cognitive abilities that underlie broader categories of human intellectual abilities. The history of cognitive abilities theory can be subdivided into five broad conceptualizations:

1. Intelligence as a single general ability
2. Intelligence as a pair of abilities
3. Intelligence as a limited set of multiple abilities
4. Intelligence as a complete set of multiple abilities
5. Intelligence as a hierarchy of narrow abilities underlying multiple broad abilities

Historically, the conceptualization of intelligence held by clinicians is intimately entwined with the major intelligence batteries available for their use. This is documented in Figure 1 through the portrayal of the relationship between level of conceptualization and the publication date of numerous major intelligence batteries available since 1916. Generally, a new level of conceptualization among clinicians has followed the publication of a particular intelligence battery, in turn, followed by the publication of other batteries based on a similar level of conceptualization. It is not likely that new intelligence batteries have ever grown from a recognition among clinicians that their current instruments were inadequate. Clinicians, by and large, tend to be consumers, not producers, of tests and interpretive models. Until recently, the mechanics of test administration were emphasized over theory in many professional preparation programs. Theory and research providing empirical support for test interpretation have had little effect on most clinician's evaluations of the adequacy of the instruments they use. For example, clinical lore has suggested that one of the factors identified in the Wechsler Intelligence Scale for Children (WISC–III) could be interpreted as a measure of freedom from distractibility even though no such cognitive ability has been independently verified, and many clinicians continue to interpret freedom from distractibility as a factor of intellectual ability. When advances are

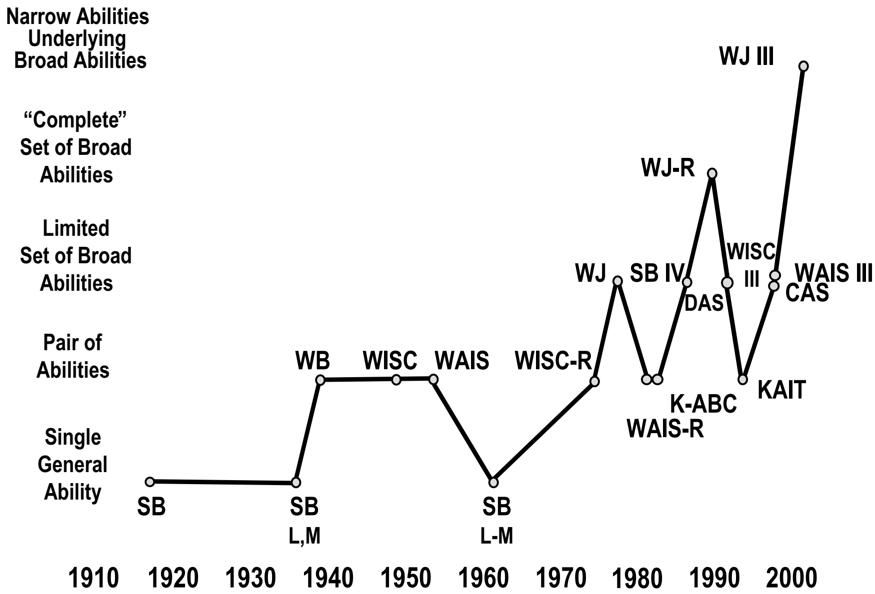

Figure 1. Interactive evolution of intelligence batteries and cognitive theory.

made practicable in the form of a measurement model, such as a theoretically based test battery, however, many clinicians can and do make changes in their conceptualizations, especially when the measurement model provides increased diagnostic capabilities and relevance to real-world interpretation.

Level 1: Intelligence as a single general ability. The earliest conceptualization of intelligence among clinicians, and still popular among some, at least operationally, is the view of intelligence as a single general ability. The Stanford–Binet (SB) test of 1916 (Terman, 1916) articulated that view of intelligence in the United States. Twenty-one years later, a revised SB was published with two alternate forms (Terman & Merrill, 1937). The third edition was published in 1960 (Terman & Merrill, 1960) as a single form that included the better test items from the previous two 1937 forms. Two scores were provided by these early SB tests: a mental age (MA) and a ratio IQ. The ratio IQ was obtained by dividing the MA by the subject's chronological age (CA). In the eyes of most clinicians using the early SB tests, that single IQ score represented all that was to be known about a person's intelligence. It should be noted, however, that the early SB tests included a rich variety of test items. Many examiners studied the pattern of correct and incorrect responses among the subject's responses to the various test items for clues about the nature of individual differences, or strengths and weaknesses in cognitive performance.

Level 2: Intelligence partitioned as a pair of abilities. At Level 2, intelligence is perceived and measured as a pair of abilities, somewhat the opposite of each other (e.g., verbal versus nonverbal). This shift in the conceptualization of intelligence among clinicians primarily occurred following publication of the Wechsler–Bellevue (WB) in 1939 (Wechsler, 1939) and several subsequent versions of the Wechsler scales. In the eyes of most users, the WB was perceived as a measure of general intelligence or full scale IQ (FSIQ), undergirded by two narrower abilities, verbal IQ (VIQ) and performance IQ (PIQ).

The WB was a scale for adults. The first WISC was published in 1949 (Wechsler, 1949) and the Wechsler Adult Intelligence Scale (WAIS), replacing the WB, was published in 1955 (Wechsler, 1955). Clinicians now had a Wechsler available for use with children and another for use with adults. Both provided VIQ and PIQ scores plus an FSIQ. These two batteries were revised and released in 1974 and 1981 as the WISC–R (Wechsler, 1974) and the WAIS–R (Wechsler, 1981) with relatively little change from their earlier editions, quite possibly because little attention was paid to cognitive ability theory, at least among clinicians and the developers of the Wechsler scales, during these decades.

Other well-known tests associated with a Level 2 conceptualization include the Kaufman Assessment Battery for Children (K–ABC; Kaufman & Kaufman, 1983) and the Kaufman Adolescent & Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993). The K–ABC provides scales of simultaneous processing and sequential processing. The KAIT provides scales of crystallized and fluid intelligence.

Level 3: Intelligence as a limited set of multiple abilities. The third level of conceptualization reflects the advent and subsequent use of intelligence batteries measuring more than two broad cognitive abilities. The first major battery to break with the pairing tradition established by the Wechsler scales was the Woodcock–Johnson Tests of Cognitive Abilities (WJ; Woodcock & Johnson, 1977). The WJ provided scores for four broad cognitive functions, identified as verbal ability, reasoning, perceptual speed, and memory.

The next Level 3 battery, published in 1986, was the Stanford–Binet Intelligence Scale, Fourth Edition (SB IV; Thorndike, Hagen, & Sattler, 1986). That battery also measured four broad categories of abilities: verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory. Other Level 3 batteries published since 1986 include the Differential Abilities Scales (DAS; Elliot, 1990) measuring three broad categories (verbal, nonverbal reasoning, and spatial) and the WISC–III (Wechsler, 1991) measuring four broad categories (verbal comprehension, perceptual organization, freedom from distraction, and processing speed). The WISC–III was the first Wechsler battery to move beyond a Level 2 conceptualization of intelligence. The Cognitive Assessment System (CAS; Naglieri & Das, 1997) measures four separate functions (planning, attention, simultaneous processing, and successive processing). The WAIS–III (Wechsler, 1997) measures four abilities (verbal comprehension, perceptual organization, working memory, and processing speed).

Level 4: Intelligence as a complete set of broad cognitive abilities. The next step advancing clinicians' conceptualizations of intelligence is associated with the availability of intelligence batteries intended to measure the complete set of broad abilities. Many contemporary scholars of intelligence would agree that the structure of cognitive ability is best portrayed by the Cattell–Horn–Carroll (CHC) theory of cognitive abilities. The CHC theory is an amalgamation of Cattell and Horn's Gf–Gc theory (Cattell, 1941; Horn, 1965, 1991; Horn & Noll, 1997) and Carroll's (1993, 1998) three-stratum theory of intelligence (Carroll & Horn, personal communication, July 1999).

The 1989 revision of the Woodcock–Johnson Tests of Cognitive Ability (WJ–R; Woodcock & Johnson, 1989b) measures seven broad abilities identified by the CHC theory. Two other broad abilities, quantitative knowledge and reading–writing, are measured as part of the companion achievement battery (Woodcock & Johnson, 1989a). The WJ–R was based on the Gf–Gc theory, now subsumed in the CHC theory.

Table 1 lists and describes nine well-defined CHC broad abilities. The acronyms presented for each broad ability are standard in the literature, though some writers may use variations. To add emphasis to the point that the purpose of testing should be to find out more about the problem, examples of implications from deficits in each of the nine abilities are included in the table.

Level 5: Cognitive ability as broad abilities undergirded by numerous narrow abilities. This level of conceptualization recognizes that 60 or more nar-

Table 1

Description of Nine Cattel–Horn–Carroll (CHC) Broad Abilities

CHC Broad Ability	*Description*	*Implications of Deficits*
Acquired knowledge:		
Comprehension–knowledge (Gc)	The breadth and depth of knowledge, including verbal communication, information, and reasoning when using previously learned procedures.	Lacks information, language skills, and knowledge of procedures.
Quantitative knowledge (Gq)	The ability to comprehend quantitative concepts and relationships; the facility to manipulate numerical symbols.	Difficulty with arithmetic and other numerical tasks; poor at handling money and making change.
Reading–writing (Grw)	An ability in areas common to both reading and writing; probably includes basic reading and writing skills, and the *skills* required for comprehension and expression.	Difficulty with word attack, reading comprehension, or other basic reading skills; writing is inconsistent and characterized by errors of spelling and usage and of poor expression.
Thinking abilities:		
Long-term retrieval (Glr)	The ability to efficiently store information and retrieve it later.	Difficulty in recalling relevant information and in learning and retrieving names; needs more practice and repetition to learn than peers; inconsistent in remembering previously learned material.
Visual–spatial thinking (Gv)	Spatial orientation, with the ability to analyze and synthesize visual stimuli and to hold and manipulate mental images.	Poor spatial orientation; misperception of object–space relationships; difficulty with art and using maps; tendency to miss subtle social and interpersonal cues.

(Continued)

Table 1 *(Continued)*

CHC Broad Ability	*Description*	*Implications of Deficits*
Auditory processing (Ga)	The ability to discriminate, analyze, and synthesize auditory stimuli.	Speech discrimination problems; poor phonological knowledge; failure in recognizing sounds; increased likelihood of misunderstanding complex verbal instructions.
Fluid reasoning (Gf)	The ability to reason and solve problems often involving unfamiliar information or procedures, which is manifested in the reorganization, transformation, and extrapolation of information.	Difficulty in grasping abstract concepts, generalizing rules, and seeing implications; has difficulty changing strategies if first approach does not work.
Cognitive efficiency:		
Processing speed (Gs)	Speed and efficiency in performing automatic or very simple cognitive tasks.	Slow in executing easy cognitive tasks; slow acquisition of new material; tendency to become overwhelmed by complex events; needs extra time in responding to well-practiced tasks.
Short-term memory (Gsm)	The ability to hold information in immediate awareness and then use it within a few seconds.	Difficulty in remembering just-imparted instructions or information; easily overwhelmed by complex or multistep verbal directions.

row abilities underlie the 9 broad abilities described in Table 1. The narrow abilities represent qualitatively different specialized abilities that have been rather well defined in the literature. Horn (1991, pp. 207–223) relates the concept of narrow abilities to the primary mental abilities concept (Thurstone, 1938) and to well-replicated cognitive factors (WERCOF) primary abilities (Ekstrom, French, & Harmon, 1979). This is followed by Horn's presentation of several kinds of measures that are associated with each of 9 broad Gf–Gc abilities. Carroll (1993) identifies narrow abilities as in the first stratum of his three-stratum theory. The WJ–III (Woodcock, McGrew, & Mather, 2001a, 2001b) is a Level 5 battery. Each of the broad CHC abilities is measured by at least two qualitatively different narrow ability tests. Some 21 narrow abilities are documented as measured in the WJ–III Tests of Cognitive Ability and 19 other narrow abilities are measured in the WJ–III Tests of Achievement.

Table 2 provides several examples of narrow abilities associated with each of the broad CHC abilities. The acronyms listed for the narrow abilities in Table 2 are rather standard in the literature. Comprehension–knowledge (Gc) is an example of a broad ability with a list of underlying narrow abilities (e.g., language development, listening ability, and general information). Each narrow ability is a verifiably separate and measurable aspect of broad comprehension–knowledge and provides qualitatively different information. An examinee may demonstrate a significant strength or weakness on one of the measures but not on the others. A parallel can be drawn with the assessment of reading, a broad area of achievement. A variety of reading tests may be administered, each assessing a different narrow reading ability (e.g., word attack, word identification, or reading comprehension). To find out more about a reading problem, it may be necessary to measure several narrow aspects of reading so that the nature of the problem can be determined and appropriate instruction planned. The same strategy applies to the assessment of a problem in one of the broad cognitive abilities. A thorough discussion of narrow abilities and their measurement is presented in the *Intelligence Test Desk Reference* (ITDR; McGrew & Flanagan, 1998). Other useful references include Flanagan and Ortiz (2001) and Flanagan, McGrew, and Ortiz (2000).

The Cognitive and Academic Performance Model

An individual's observed cognitive and academic performance results from a complex interaction of many components. These components may be assigned to four broad categories differentiated by function: stores of acquired knowledge, thinking abilities, cognitive efficiency,

Table 2

Examples of Cattell–Horn–Carroll (CHC) Narrow Abilities

CHC Broad Ability	*CHC Narrow Abilities*	
Acquired knowledge:		
Comprehension–knowledge (Gc)	Language development (LD)	Oral production and fluency (OP)
	Lexical knowledge (VL)	General information (K0)
	Listening ability (LS)	Information about culture (K2)
Quantitative knowledge (Gq)	Mathematical achievement (A3)	Mathematical knowledge (KM)
Reading–writing (Grw)	Reading decoding (RD)	Writing ability (WA)
	Reading comprehension (RC)	English usage knowledge (EU)
	Spelling ability (SG)	
Thinking abilities:		
Long-term retrieval (Glr)	Associative memory (MA)	Ideational fluency (FI)
	Meaningful memory (MM)	Naming facility (NA)
	Figural fluency (FF)	
Visual–spatial thinking (Gv)	Visualization (Vz)	Length estimation (LE)
	Spatial relations (SR)	Visual memory (MV)
	Flexibility of closure (CF)	Spatial scanning (SS)
Auditory processing (Ga)	Phonetic coding (PC)	Resistance to auditory stimulus distortion (UR)
	Speech-sound discrimination (US)	
	General sound discrimination (U3)	Memory for sound patterns (UM)
Fluid feasoning (Gf)	Induction (I)	Quantitative reasoning (RQ)
	General sequential reasoning (RG)	
Cognitive efficiency:		
Processing speed (Gs)	Perceptual speed (P)	Rate of test taking (R9)
	Semantic processing speed (R4)	Number facility (N)
Short-term memory (Gsm)	Working memory (WM)	Memory span (MS)

and facilitator–inhibitors. These categories are logically derived, based on similar and dissimilar characteristics. Each of these four categories includes components that contribute in a common way to cognitive performance but also contribute differently from the common contribution of the other three categories. Figure 2 illustrates the relationship between the four functional categories and cognitive/academic performance.

In the cognitive and academic performance model, the fourth oval represents the influence of facilitator–inhibitors. These are the noncognitive factors that modify cognitive and academic performance for better or worse, often overriding the effects of strengths and weaknesses in the individual's cognitive and achievement profiles. The source of some facilitator–inhibitors is internal (e.g., health, emotional state, and motivation/volition), while the source of other facilitator–inhibitors is situational and environmental (e.g., the presence of visual and auditory distractions, the response format, or the types of tests selected for a cognitive examination).

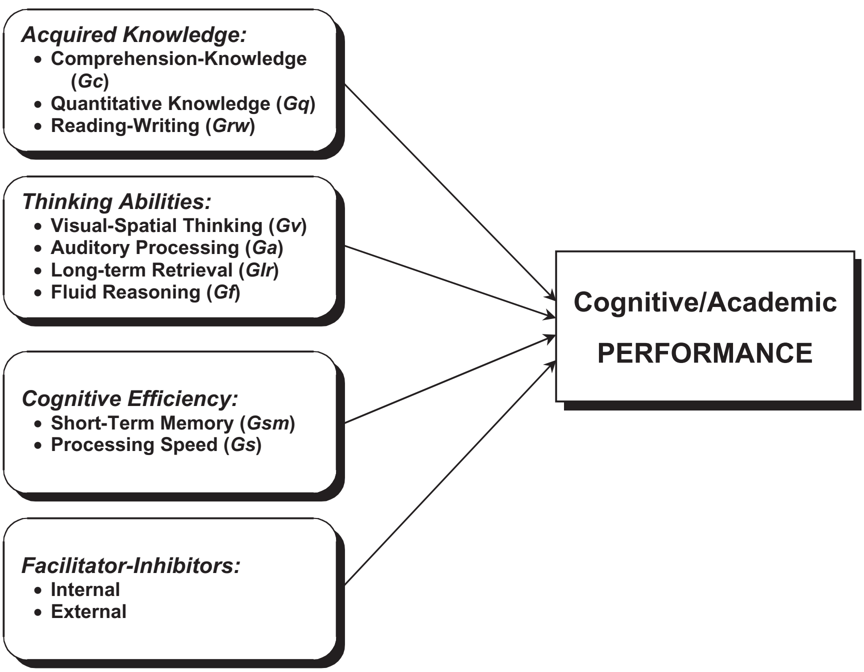

Figure 2. The cognitive and academic performance model.

There are several ways in which the cognitive performance model can be helpful in interpretation of the CHC broad and narrow abilities. Grouping the broad abilities by similarities can help practitioners understand some practical implications of test performance. For example, the stores of acquired knowledge include Gc, quantitative knowledge (Gq), and reading–writing ability (Grw). All of these abilities are learned. Once a piece of information is learned, it can become a building block for new learning. Similarly, if a piece of information is not learned, it can become an impediment to future learning. Additionally, the stores of acquired knowledge are mutable; that is, teaching strategies and enrichment opportunities can impact an individual's performance levels.

The thinking abilities are the processes through which new learning occurs. These processes include visual–spatial thinking (Gv), auditory processing (Ga), long-term retrieval (Glr), and fluid reasoning (Gf). Limitations in one or more of these thinking abilities will likely constrain new learning, possibly requiring alternative forms of instruction. Also, new learning will be constrained by any limitations in the relevant stores of acquired knowledge.

Cognitive efficiency includes short-term memory (Gsm) and processing speed (Gs). Automatic cognitive performance will be constrained by any limitations in these broad abilities. Lowered levels of cognitive efficiency frequently require accommodations in instruction and in activities such as group testing situations.

All performance, especially new learning, is constrained by any limitations among the facilitator–inhibitors. For example, uncorrected poor visual acuity will likely result in missed opportunities for learning. Any significant health problems that result in poor school attendance can disrupt learning opportunities. Poor motivation to learn, or low interest in academics, will likely affect academic task engagement. Certain cognitive style or temperament characteristics, such as impulsivity, can negatively affect the quality of one's work. Other factors, such as emotional stability, organization, and ability to concentrate, can positively or negatively affect learning opportunities. Consequently, the depiction of this broad class of facilitator–inhibitors as a component of cognitive performance can help practitioners to pay greater attention to these variables when evaluating an individual's cognitive and/or academic performance.

Other, more complex and informative, models of cognitive performance based on the CHC theory include the Gf–Gc Information Processing Model (Woodcock, 1993) and the Cognitive Neuropsychology Model (Dean & Woodcock, 1999, 2003; Woodcock, 1998).

Advances in Statistical and Test Development Tools

Advances in their statistical tools of trade have impacted the work of cognitive theorists and intelligence test developers. For the theorist, the most notable among these advances has likely been the development of powerful confirmatory factor analysis programs. For the test developer, item response theory (IRT), better procedures for the imputation of missing data, and the availability of complex curve-fitting programs have transformed their approach to the task.

Tools Applied to Cognitive Theory

Toward the advancement of cognitive theory, among the most useful tools are confirmatory factor analysis (CFA) programs (e.g., LISREL, AMOS, M-PLUS, EQS). These computer programs, which are largely based on the iterative power of maximum-likelihood analysis, have enhanced efforts toward describing the factorial structure of human cognitive abilities. Included in these programs are procedures for evaluating the comparative fit of competing measurement models, a feature that provides for a degree of objectivity in the hands of sophisticated users.

CFA methods are used during various stages of test development. First, theory-driven CFA methods have been applied to published editions of intelligence and achievement batteries to ascertain how an existing battery of tests measures, or does not measure, certain abilities according to a specific cognitive theory. The results are then used to specify a revision plan that will result in a battery of tests better representing the major domains specified by the theory. Once data gathering has begun, CFAs performed on early processed data can assist test developers in determining if new or revised tests are behaving as expected (i.e., loading on the factor constructs they were designed to measure). Revisions to tests, or the development of new tests, can then occur prior to gathering the majority of the norm data. Finally, CFA analysis at the end of a test norming project provides important structural validity information in support of the organizational structure of a battery of tests.

Though not a statistical program in itself, *joint factor analysis* can assist in identifying the subset of CHC broad and narrow abilities measured by an intelligence battery that does not contain at least two or three separate measures for each factor implicit in the battery. In an early joint factor analysis study that employed both exploratory and confirmatory factor analysis techniques, Woodcock (1990, 1994) investigated the factorial structure of six widely used, individually administered intelligence batteries. Two observations from that study are of interest here. First, it was demonstrated that the breadth of factor coverage in certain batteries was greater than reflected in the interpretation systems provided by the publishers. For example, the WISC–R provided verbal and performance scores, but Woodcock reported that the WISC–R tests measured five distinct CHC factors (Gc, Gq, Gv, Gs, and Gsm) lumped together into the two interpretive scores. Newer versions of the Wechslers provide interpretive schemes that more closely reflect CHC theory.

Second, the results of this study drew attention to the fact that different labeled abilities (and composite scores) across different intelligence batteries were often factorially equivalent even though their labels differed and clinicians perceived them as measuring different traits. For example, both the Wechsler perceptual organization and the K–ABC simultaneous processing scales were demonstrated to be measures of the same broad CHC ability, visual–spatial thinking (Gv). The K–ABC sequential processing scale was demonstrated to be a measure of short-term memory (Gsm). Clinicians need to be cognizant that tests with unique names that purport to measure unique traits may simply be new tests measuring well-established CHC factors. The most famous description of this property of names is perhaps Juliet's line to Romeo:

> What's in a name? that which we call a rose
> By any other name would smell as sweet.
> (Shakespeare, *Romeo and Juliet*)

Woodcock's early study led to a series of subsequent joint factor analysis studies, often called cross-battery studies. As a result, there now exists a classification of the individual tests from all major intelligence and achievement batteries within a common CHC nomenclature (see Flanagan et al., 2000; Flanagan & Ortiz, 2001; Flanagan, Ortiz, Alfonso, & Mascolo, 2002; McGrew, 1997; McGrew & Flanagan, 1998).

Tools Applied to Test Development

Arguably, the single most important advancement in psychometrics for the test developer has been the introduction of IRT (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985). Particularly influential has been the Rasch model, a single-parameter logistic test model used to analyze item response data (Embretson, 1996; Rasch, 1960; Wright, 1968; Wright & Stone, 1979). Several benefits realized from the application of the Rasch model to test development include the following:

• Sample-free item calibration: Difficulty levels assigned to items are independent of the level and distribution of ability in the sample chosen to calibrate the items.

• Item-free measurement: After a bank of items fitting the model is calibrated, any subset of items may be used to construct a test and the ability scores from this new test will be on the same scale as the bank of items. As an example of this benefit, multiple forms of a test can be constructed easily on the same underlying measurement scale.

• Item difficulties and ability scores are on the same scale (Woodcock & Dahl, 1971). The distance on the scale between task difficulty and a person's ability provides a direct and quantifiable implication for performance on that task. This information can be generalized to the prediction of performance with similar tasks.

• The equating of tests is simplified, even across languages (Woodcock & Muñoz-Sandoval, 1993).

• Options for interpreting test performance are extended (Woodcock, 1999). For example, test performance can be reported as a criterion-referenced score that describes quality of performance or proficiency compared to others of the same age or grade, or as a developmental zone extending from a fixed level of easiness to a fixed level of difficulty for a given subject.

Missing data are the bane of existence for test developers and many applied researchers. By definition, all standard multivariate statistical methods (e.g., multiple regression or factor analysis) require a complete set of data on all variables in order for a subject's data to be included in the analyses. The absence of a score for a single variable can render 100% of that subject's available data useless. As a result, less accurate estimates of statistical parameters are obtained due to the reduced size of the database. Due, in many respects, to the increased availability of high-speed and low-cost computers, archaic improvisations such as replacing a missing score by the mean (mean substitution) or by the value of an adjacent subject are rarely used today. They have been replaced by procedures such as regression substitution, stochastic regression imputation, hot-deck imputation, multiple imputation, and maximum-likelihood methods (including the expectation maximization or EM algorithm).

A third significant advancement in test development tools is the emergence of better programs that facilitate the discovery of equations for complex curves that characterize large data sets. For example, the TableCurve 2D (Systat, 1997a) and TableCurve 3D (Systat, 1997b) programs are two- and three-dimensional curve- and surface-fitting programs. These programs readily accommodate missing data and generate sets of possible complex equations that provide the best possible fit of a curve or a surface to a large set of data. The use of such programs has had a major impact on the computation of norms for a test vis-à-vis (a) the reduction of human error present in earlier hand curve-fitting procedures, (b) the generation of sets of viable equations/curves from which to select, (c) the generation of a variety of fit statistics by which to evaluate solutions, (d) the ability to superimpose and visually compare multiple curve solutions, (e) the ability to apply *pre-smoothers* that reduce wild data points resulting from sampling error, and (f) the quick and efficient transfer of the final smoothed values and/or computer code to other software.

Conclusion

At the beginning of this article, it was stated that the primary purpose of a comprehensive cognitive assessment should be to gain information about the problem, not to obtain an IQ. Advances in cognitive theory, and the subsequent availability of intelligence batteries based on these advances, provide clinicians with knowledge and tools that are more informative and that allow for finer diagnostic interpretations. Tests of the future will become more informative but not necessarily more complex. As a consequence, it is incumbent on clinicians and the trainers of clinicians to stay abreast with advances in cognitive theory and measurement and with their corresponding implications for practice. Attention to these advances should not be limited to the primary training programs for clinicians but needs to be a major component of continuing professional education. All professionals—whether they are tax accountants, physicians, or educational/psychological clinicians—have an ethical responsibility to their clients to stay abreast of changes in their field.

References

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

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Dean, R. S., & Woodcock, R. W. (1999). *The WJ–R and Batería–R in neuropsychological assessment: Research report number 3.* Itasca, IL: Riverside.

Dean, R. S., & Woodcock, R. W. (2003). *Dean–Woodcock Neuropsychological Assessment System.* Itasca, IL: Riverside.

Ekstrom, R. B., French, J. W., & Harmon, M. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*(2).

Elliot, C. D. (1990). *Differential Abilities Scales.* San Antonio, TX: Psychological Corporation.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8,* 341–349.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and G–Gc theory.* Boston: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. W. (2001). *Essentials of cross-battery assessment.* New York: Wiley.

Flanagan, D. P., Ortiz, S. W., Alfonso, V., & Mascolo, J. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities.* Boston: Allyn & Bacon.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Norwell, MA: Kluwer Academic.

Horn, J. L. (1965). Fluid and crystallized intelligence. Unpublished doctoral dissertation, University of Illinois, Urbana–Champaign.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *WJ–R technical manual* (pp. 197–232). Chicago: Riverside.

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

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

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

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. 151–189). New York: Guilford.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): A cross-battery approach to intelligence test interpretation.* Boston: Allyn & Bacon.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system.* Itasca, IL: Riverside.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish Institute for Educational Research.

Systat (1997a). TableCurve 2D [Computer Software]. Richmond, CA: Systat Software.

Systat (1997b). TableCurve 3D [Computer Software]. Richmond, CA: Systat Software.

Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.

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

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

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

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1.*

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

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–Third Edition.* San Antonio, TX: Psychological Corporation.

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

Woodcock, R. W. (1993). An information processing view of Gf–Gc theory. *Journal of Psychoeducational Assessment Monograph Series:* Advances in Psychoeducational Assessment, *Woodcock–Johnson Psycho-Educational Battery–Revised.* Brandon, VT: Clinical Psychology Publishing Company.

Woodcock, R. W. (1994). Measures of fluid and crystallized theory of intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 452–456). New York: Macmillan.

Woodcock, R. W. (1997). The Woodcock-Johnson Tests of Cognitive Ability—Revised. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 230–245). New York: Guilford.

Woodcock, R. W. (1998). *The WJ–R and Batería–R in neuropsychological assessment: Research report number 1.* Itasca, IL: Riverside.

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

Woodcock, R. W., & Dahl, M. N. (1971). *A common scale for the measurement of person ability and test item difficulty* (AGS Paper No. 10). Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson Psycho-Educational Battery.* Itasca, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989a). *Woodcock–Johnson Tests of Achievement–Revised.* Itasca, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989b). *Woodcock–Johnson Tests of Cognitive Ability–Revised.* Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001a). *Woodcock–Johnson III Tests of Achievement.* Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1993). An IRT approach to cross-language test equating and interpretation. *European Journal of Psychological Assessment, 9,* 233–241.

Wright, B. D. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on Testing Problems* (pp. 85–101). Princeton, NJ: Educational Testing Service.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.