

CHAPTER 1

Understanding the Strengths and Weaknesses of Intelligence and Achievement Tests

Jack Naglieri, Sam Goldstein

It is essential that any study and measurement of human intelligence and academic achievement recognize the importance of the brain. The human brain is an amazing organ—a product of an ongoing, six-billion-year construction project. In its physical form and function, the human brain represents millions upon millions of trial-and-error adaptive adjustments. Comprised of an estimated 100 billion neurons and many more glial cells, it is organized into thousands of regions. The human brain, in a seamlessly integrated manner, governs body functions and movement, but more important, it is the seat of intelligence and regulates cognition and achievement. Not surprisingly, although the brains of different animals may not look exactly alike, they all work according to the same principles and mechanisms. These neurons and glial cells communicate using a nearly infinite number of synaptic connections, yet the entire organ in humans weighs only about three pounds. Consider also that an infant is born with a brain of 300 to 400 gm, tripling in size by the adult years. Yet, between birth and the conclusion of the first two decades of life, a nearly infinite acquisition

of knowledge and behaviors characterizes human development. Gram for gram, the human brain delivers an almost-dazzling array of motoric, behavioral, cognitive, and emotional capacities nearly impossible to fathom in light of its size.

In her extremely cogent and interesting book, *Brain Dance* (2004), Dean Falk, a professor of anthropology at Florida State University, describes the conditions and circumstances that allowed a group of ape-like individuals to evolve over a period of at least five million years into *Homo sapiens*. During this process, the brain became increasingly more specialized, evolving a broad range of abilities as well as right-brain/left-brain and male/female differences. As Falk notes, in less than two million years, brain size doubled in the *Homo* species, from around 650 gm to 1350 gm. Only a small portion of this newly evolved, larger brain was tied to increasing body size. As Falk points out, this process was unprecedented in the evolutionary histories of other mammals. As brain size increased, neurons enlarged and became more widely spaced and the cerebral cortex became more convoluted. No new structures were found in these larger human brains.

However, these larger brains set the foundation for an accelerated evolutionary process never before witnessed in any earthbound, mammalian species. In this process, the prefrontal cortex and the posterior areas of the brain associated with sensory processing in particular became especially convoluted. The shift in neurochemicals, anatomy of neurons, and brain function provided the underlying mechanics of our rapid evolutionary progression, a pattern that was most certainly driven by natural selection.

It is a fascinating historical phenomenon that although scientists and philosophers have written about the intellectual and academic capabilities of the human brain over the last 2,000 years, it is just within the past 100 years, but especially the past 50 years, that we have witnessed an explosion of theories and tests to measure these qualities. This growth has created an often-confusing menu of clinical and educational tools from which to choose. The goal of this book is to help the reader contextualize the strengths and weaknesses of the tests presented and to apply a balanced perspective on their critique. To do so, we asked authors to cover the following 15 essential aspects:

Theory Underlying the Test

1. Historical information, definition of the constructs, and development of the subtests

Description of the Test

2. Subtest background
3. Scales the test yields
4. Structure of the test

Administration and Scoring

5. Tips on administration
6. How to score the test

Standardization, Norms, and Reliability

7. Characteristics of the standardization sample
8. Reliability of the scales

Use of the Test (Including Validity Research on Each Topic)

9. Interpretation methods
10. Identification of special populations
11. Interventions based on test results

Validity

12. Relationships between ability and achievement
13. Fairness, sex, race, and ethnic differences
14. Profiles of abilities and their relationship to diagnosis and treatment planning
15. Factor structure

In this chapter, we review critical points and issues the reader should consider when evaluating the various instruments within the context of the field today and in consideration of what can be reasonably expected for the future. We further urge the reader to carefully examine each author's contribution with attention to the specific details presented and the extent to which each test addresses the issues that were covered.

ASSESSING STRENGTHS AND LIMITATIONS OF ABILITY TESTS

Traditional tests of intelligence were initiated in 1905 with the publication of the Stanford-Binet and further solidified in 1939 with the publication of the Wechsler-Bellevue Scales. These tests made a substantial contribution to our society, shaped how we define intelligence, and influenced the lives of countless children and adults in the United States and around the world (Anastasi & Urbina, 1997). Even though high-quality intelligence and achievement tests are among the most influential contributions made by psychology to society in general, they have also been the subject of considerable criticism. The value of these tests has been demonstrated by a substantial literature (see Graham & Naglieri, 2002; Jensen, 1998; and Ramsey & Reynolds, 2004)

even though there are criticisms that range from emotionally to scientifically based. The strengths and weaknesses of the tests presented in this book should be apparent by the content of each chapter, and especially what information was and was not provided.

In order to best determine the utility of each test, readers should consider the strengths and limitations as reported by the authors or their representatives. One point that needs to be stressed is that sometimes a “limitation” of a test may not be evidenced by lack of empirical evidence or poor quality, but rather by misapplication or misinterpretation of scores. For example, sometimes tests are criticized when they are used for a reason for which they were not developed or intended (e.g., using verbal and performance IQ scores for determining which instructional method to employ). Similarly, some tests are criticized when used in ways for which research suggests the test is ill-suited (e.g., subtest profile analysis). Readers should also consider the source of the criticisms, and the extent to which charges are valid and supported by research findings and not based solely on a critic’s opinion or straw-man arguments.

The information provided in these chapters helps us understand the authors’ position on the most appropriate use of their test. For example, a test such as the Wechsler Nonverbal Scale of Ability (Wechsler & Naglieri, 2006) will not be able to inform us about the examinee’s level of verbal expression and verbal comprehension; it was not developed for that purpose. That is not a limitation of the test itself, but rather an indication of how it should and should not be used. Similarly, using the WISC-IV with an examinee who has limited English-language skills will yield an estimate of ability that, according to Yoakum and Yerkes (1920), will likely be inaccurate because it has been known for nearly 100 years that a person could do poorly on verbal and quantitative tests because of limited skills in English. To avoid “injustice by reason of relative unfamiliarity with English” (p. 19), these persons should be tested with

nonverbal measures of general ability (see the tests included in the nonverbal test section of this book). It is important to understand that using a verbal test to measure general ability for someone who does not have an adequate knowledge of English does not mean that the test is invalid. Instead, the score should be interpreted differently (as a measure of English-language comprehension and expression) because of the context within which it was used. The point is, however, that a test of general ability that contains verbal examinations is not inherently faulty, but rather its application to a person who does not speak English undermines the effort to estimate intelligence accurately. Criticisms of the test, therefore, must be viewed within the context of the test’s use.

The limitations of traditional IQ tests containing verbal, quantitative, and performance scales have been recognized by practitioners and test developers in the field. This has led to an effort to provide so-called *nonverbal tests of ability*, which comprise a substantial portion of this book. These nonverbal tests of general ability represent an effort on the part of test developers to meet the need for a way to measure ability for diverse populations. By measuring ability using tests of general ability that do not involve verbal or quantitative content, the tests offer a specific methodology designed explicitly for assessing diverse populations. That is, nonverbal tests of ability meet a particular need in the field that traditional multicontent tests like the Wechsler and Binet could not adequately address. Just as nonverbal tests of general ability have changed the field, the availability of cognitive processing-based measures of ability has also altered the landscape considerably.

Tests built to measure cognitive processes meet the need for a new approach to assessment that reconceptualizes ability within the context of basic psychological processes. The KABC-II and the CAS stand out as methods of conceptualizing and measuring ability very differently from the verbal/quantitative/nonverbal test content perspective utilized since the early 1900s.

These tests are also distinguished from more traditional tests because they were explicitly developed to measure basic psychological processes while avoiding achievement-laden test content. The availability of these tests has also raised the awareness of users and test developers of the need for theoretically-based measures of ability that are appropriate for assessment of culturally and linguistically diverse populations, strongly correlated with achievement, and are linked to intervention. The KABC-II and CAS chapters should be viewed within this larger context and with recognition that theory-based instruments offer considerable advantages.

INTELLIGENCE TESTS AND THEORIES

The field of intelligence testing has evolved considerably over the past 20 years and, most important, the value of having a test that is based in a theory has become much more salient. Although we may discuss the theory behind a test, we often do not reflect carefully upon precisely what this means. According to the United States National Academy of Sciences, in the general population the word *theory* means a hunch or speculation and is often used as an antonym for a hypothesis or opinion. In contrast, a scientific definition of the word *theory* “refers to a comprehensive explanation of an important feature of nature that is supported by many facts gathered over time. Theories also allow scientists to make predictions about as yet unobserved phenomena” (Wikipedia, 2008). In the science of assessment psychology, we should use the word *theory* only to describe a model of a construct that has been tested using scientific experimentation, has demonstrated validity, and can be used to predict future behavior. It is quite important to ask the question, “What theory underlies each of the various ability tests included in this book?”

There is considerable variability in the answers to the question, “What theory was used to

describe this test of ability?” For example, the term *general intelligence* is used to describe tests like the WISC-IV as well as the UNIT, CTONI, and WNV. Some authors merge different views to describe their approach to explaining what the test measures (e.g., KABC-II) and the theory behind the test. Other authors take a much more liberal approach and apply a model to a previously developed test, and still others take a strict approach of building a test explicitly according to one theory (e.g., CAS). Ultimately, it is the responsibility of the test authors to precisely define the idea, model, or theory behind the test, demonstrate the extent to which the test represents their view, and provide evidence for both the approach and its operationalization in that particular test.

Readers may wonder about the relevance of the underlying conceptualization of a test. How important is a theory for application in the real world? Does a test built on a theory offer advantages over one built on an atheoretical basis? Does this really matter? These are the kinds of questions that must be understood in order to effectively use tests of intelligence, ability, or cognitive processing. In our view, understanding the authors’ definition of what their intelligence, ability, or cognitive processing test measures is critical to understanding what the test measures, the relevance that information has for diagnosis and treatment, and how the test will be applied in practice, and especially, explained to teachers and parents.

Readers may wonder what importance a test’s theory holds for the practicing psychologist or the educator. The answer is that theory holds considerable importance to understanding and interpreting the scores from a test. We suggest that (1) the user must understand where each test falls on the continuum between an idea and a theory; (2) if the test was built on a theory, then that theory must be clearly articulated and well understood so that it can be used for interpretation of the test; (3) the validity for the organization of the test must be carefully scrutinized by each and every user; and

(4) communication of the findings to consumers must be based on the theory. Why? Because in order to interpret the results from any test, the user must know what the test was designed to measure, whether it does so adequately, and what validity evidence there is for the utility of the constructs. Readers should carefully consider the extent to which this information is provided in each of the chapters.

STANDARDIZATION, NORMS, AND PSYCHOMETRICS

The field of assessment, and especially intelligence and achievement testing, has advanced considerably in the past 25 years to a point where we assume that any test sold by the major companies is well developed, standardized, and normed. The quality of all intelligence and achievement tests is amply described in the chapters contained in this book. Readers should not, however, assume that every test was normed with the same level of sophistication and adequately documented. For example, whereas some of the tests are exemplifiers of excellence, some are less well developed; for example, documentation of the representative nature of the standardization group is inadequate (e.g., DIBELS). Readers should also carefully examine the types of test scores obtained and be cautious of tests that yield only raw scores, which can be misleading (Naglieri & Crocket, 2005). For a review of important psychometric issues, the value of a representative sample, and the statistical characteristics of a test with which the practitioner should be provided, see Naglieri and Chambers (2008) and Urbina (2004).

DIAGNOSIS AND INTERVENTION

Educational diagnosis and eligibility determination are being substantially influenced by the

Individuals with Disabilities Education Improvement Act (IDEA) of 2004. This law is having considerable influence on the way school psychologists use intelligence and achievement tests. This law has also influenced community-based decision making in mental health and clinical settings. One of the most important changes particularly relevant to the tests included in this book involves the identification of children with specific learning disabilities (SLD) (see Hale, Flanagan, & Naglieri, 2008; and Hale, Kaufman, Naglieri, & Kavale, 2006), but other issues, such as bias in testing, relevance to intervention, and ongoing treatment monitoring, are also implicated. The issues emphasized in IDEA are included in the chapters contained in this book. For authors of ability tests, this includes evidence of test bias, assessment of basic psychological processes, relevance to identification of SLD, and instructional implications (see KABC-II and CAS). For authors of achievement tests, this also includes relevance to ongoing progress monitoring (see DIBELS and WRAT-PM).

The IDEA and federal regulations have created an opportunity for change in the way school psychologists use ability and achievement tests that will be clear throughout the chapters in this book. Readers should pay particular attention to the similarities and differences among the authors' methods for applying their tests to the question of SLD diagnosis. For example, the deemphasis of the IQ-Ability test difference (e.g., WISC-IV versus WIAT-II) as a criterion for SLD has led to greater emphasis on tests of basic psychological processes (e.g., K-ABC and CAS) and the need for screening of academic skills (e.g., BASI) and ongoing progress monitoring (WRAT-PM). Test authors are developing methods of assessment that can address the demands of today's practitioners. For example, the presence of a processing disorder along with learning failure represents the essence of SLD (Kavale, Kaufman, Naglieri, & Hale, 2005). The cognitive processing measures included in this book enable practitioners to document the essential operational marker for SLD—consistency

between cognitive deficits and academic deficits coupled with a significant discrepancy between cognitive assets and cognitive deficits (e.g., Hale & Fiorello, 2004; Kavale et al., 2005; Naglieri, 1999, 2000, 2008). One of the most important advantages of this approach is that it unites an identification method with the definition included in IDEA as suggested by Hale, Kaufman, Naglieri, and Kavale (2006). The next step, of course, is the determination of appropriate interventions.

There is considerable controversy about the connection between tests of intelligence and cognitive processing with instruction, perhaps best illustrated by the view proposed by Vellutino, Scanlon, Small, and Fanuele (2003) that tests of ability are irrelevant because they do not predict response to treatment. Some of the chapters in this book provide a view of how instruction can be linked to assessment of cognitive processes and how instruction can be guided by monitoring the academic progress of students on a frequent basis (e.g., DIBELS and WRAT-PM). The extension of assessment information to instruction will likely remain a contentious issue for some time.

INTELLIGENCE TEST BIAS

The characteristics of the U.S. population continue to change and the need for fair assessment of culturally and linguistically diverse populations, especially the Hispanic population, has become increasingly important. IDEA 2004 clearly states that assessments must be selected and administered so as to be nondiscriminatory on a racial or cultural basis. It is critical, therefore, that any measure of ability that is used should have been evaluated for test bias (see Reynolds & Ramsay, 2003, for a summary of test bias issues).

The ability tests presented in the book differ considerably in their conceptualizations and operationalizations of the construct. The two types of tests of general ability (those that

include verbal, quantitative, and nonverbal content and those that use only nonverbal test questions) are distinct from the cognitive processing-based measures. Readers should consider the psychometric and impact-related issues related to the scores that these three types of tests yield for various race and ethnic groups. What will become apparent is that while researchers have recognized the value of general intelligence tests with diverse content, these tests yield the largest mean-score race and ethnic differences. When ability is measured using nonverbal tests or conceptualized according to basic psychological processes, both of which avoid the knowledge required to answer verbal and quantitative questions, race and ethnic differences are reduced (Fagan, 2000; Suzuki & Valencia, 1997). Nonverbal and cognitive processing-based tests that do not rely on questions that contain language and quantitative content are, therefore, deemed more appropriate for assessment of culturally and linguistically diverse populations (Fagan, 2000; Suzuki & Valencia, 1997). Importantly, the chapters on nonverbal and cognitive processing tests provide two important approaches that measure ability without the loss of predictive validity and at the same time result in a more equitable system for evaluating culturally and linguistically diverse populations.

THE CHANGING FACE OF ACHIEVEMENT TESTING

The chapters included in this book that cover evaluation of achievement illustrate important changes in the way academic skills are conceptualized and measured. The tests vary in their scope. For example, tests such as the K-TEA and WIAT-II offer measurement of a broad spectrum of skills that are standardized on large samples that represent the U.S. population. These achievement tests also vary in scope. For example, the WJ-III offers a large number of

subtests, whereas the WRAT-4 measures relatively fewer areas. The WRAT-4 also provides a way to measure progress over time, as does the BASI, but the BASI provides for both individual as well as group testing options. Tests like the CTOPP and GORT provide specialized measures of achievement. These tests differ in their psychometric characteristics and normative samples, which have important implications for usefulness and defensibility.

SUMMARY

We anticipate that this book will greatly aid users' understanding of the strengths and weaknesses of the various tests presented. Ultimately, it is the responsibility of the test authors to provide evidence of the quality of their tests and the contexts within which they are most defensibly used. It is the responsibility of the test user to be informed of the various quality issues associated with each test and the interpretive methods the authors recommend. Our goal is, of course, to provide the very best information to the consumers of this information for the benefit of the client.

REFERENCES

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Fagan, J. R. (2000). A theory of intelligence as processing: Implications for society. *Psychology, Public Policy, and Law*, 6, 168–179.
- Falk, D. (2004). *Brain dance—revised and expanded edition*. Gainesville, FL: University Press.
- Graham, J. R., & Naglieri, J. A. (Eds.) (2002). *Handbook of assessment psychology*. New York: Wiley.
- Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.
- Hale, J. N., Flanagan, D. P., & Naglieri, J. A. (2008). Alternative research-based methods for IDEA 2004 identification of children with specific learning disabilities. *Communiqué*, 36, 14–15.
- Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools*, 43, 753–770.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kavale, K. A., Kaufman, A. S., Naglieri, J. A., & Hale, J. B. (2005). Changing procedures for identifying learning disabilities: The danger of poorly supported ideas. *The School Psychologist*, 59, 16–25.
- Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.
- Naglieri, J. A. (2000). Can profile analysis of ability test scores work? An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly*, 15, 419–433.
- Naglieri, J. A. (2008). Best practices in linking cognitive assessment of students with learning disabilities to interventions. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 679–696). Bethesda, MD: NASP.
- Naglieri, J. A., & Chambers, K. (2008). Psychometric issues and current scales for assessing autism spectrum disorders. In S. Goldstein, J. A. Naglieri, & S. Ozonoff (Eds.), *Assessment of autism spectrum disorders* (pp. 55–90). New York: Springer.
- Naglieri, J. A., & Crockett, D. (2005). Response to Intervention (RTI): Is it a scientifically proven method? *NASP Communiqué*, 34, 38–39.
- Ramsay, M. C., & Reynolds, C. R. (2004). Relations between intelligence and achievement tests. In G. Goldstein and S. Beers (Eds.), *Comprehensive handbook of psychological assessment* (pp. 25–50). New York: Wiley.
- Reynolds, C. R., & Ramsay, M. C. (2003). Bias in psychological assessment: An empirical review and recommendations. In J. Graham and J. A. Naglieri (Eds.), *Assessment psychology* (pp. 67–93). New York: Wiley.
- Suzuki, L. A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence. *American Psychologist*, 52, 1103–1114.
- Urbina, S. (2004). *Essentials of psychological testing*. New York: Wiley.
- Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. (2003, December). Response to intervention as a vehicle for distinguishing between reading disabled and non-reading disabled

children: Evidence for the role of kindergarten and first grade intervention. Paper presented at the *National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium*, Kansas City, MO.

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Pearson.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York: Henry Holt & Company.