CHAPTER
4

Traditional IQ: 100 Years of Misconception and Its Relationship to Minority Representation in Gifted Programs

Jack A. Naglieri

Introduction

The underrepresentation of minority children in classes for the gifted has been and continues to be one of the most important problems facing educators of gifted students (Ford, 1998; Naglieri & Ford, 2005). The severity of the problem was made obvious in the United States Department of Education's recent report that Black, Hispanic, and Native American students are underrepresented by 50–70% in gifted education programs (Naglieri & Ford, 2003). Efforts to address this situation include, for example, use of multiple criteria for inclusion, refinement of the referral procedures, and reexamination of the very definition of the term *gifted*. Some have argued that the content of the ability tests used and procedures followed fail to take into consideration the characteristics of culturally, ethnically, and linguistically diverse populations (Frazier et al., 1995; Naglieri & Ford, 2005).

The concept of intelligence has been defined by the tests used to measure this construct since the early 1900s. Traditional intelligence tests have had the now familiar verbal, quantitative, and nonverbal format since Binet and Simon (1905) and Wechsler (1939) published their influential tests. The division

of items by content was not based on a theory of verbal, quantitative, and nonverbal intelligences. In fact, the division was a practical one as noted by Yoakum and Yerkes when they wrote that the Army Beta (nonverbal) tests were used because it was known that a person could fail the Army Alpha (verbal and quantitative) tests because of limited skills in English. To avoid "injustice by reason of relative unfamiliarity with English" (1920, p. 19), these persons were then tested with the nonverbal tests. It is important to note that there is no mention of the need to measure different types of intelligence even though verbal, nonverbal, and quantitative tests were all used to measure general ability.

Content of Traditional IQ Tests

Traditional IQ tests measure *general ability* through questions that are verbal (e.g., vocabulary or word analogies), spatial (e.g., arranging blocks to match a simple design or assembling puzzles to make a common object), or quantitative (e.g., math word problems or math calculation). The spatial tests have been described as *nonverbal*, because it is an easier concept to understand, not because of any intention to measure nonverbal ability. In fact, this lack of theoretical basis was noted by Pintner (1923) when he wrote "we did not start with a clear definition of general intelligence . . . [but] borrowed from every-day life a vague term implying all-round ability and . . . we [are] still attempting to define it more sharply and endow it with a stricter scientific connotation" (p. 53). The use of a vague definition of intelligence leaves unspecified the differences between a test of intelligence and a test of achievement. The result has been that our tests have been used to define the theory of intelligence the test is intended to measure.

Traditionally, IQ has been measured using verbal, quantitative, and nonverbal tests since the tests were initially formulated in 1905 with the publication of the Stanford-Binet (Binet & Simon, 1905) and in 1939 with the publication of the Wechsler-Bellevue Scales (Wechsler, 1939). These tests made a significant and long-lasting contribution to our understanding of how to

measure and conceptualize intelligence. The results obtained from these tests have influenced the lives of countless children and adults in the United States and around the world. Although intelligence tests represent one of the most influential contributions made by psychology to society in general (Anastasi & Urbina, 1997), they also have become engrained in our culture as *the* way to measure ability.

There is considerable experimental support for the concept of general intelligence as measured by tests such as the Wechsler and Binet (see Jensen, 1998, for a review). Among the most important sources of validity evidence for IQ tests is the fact that the scores the tests yield are a good prediction of school achievement (Naglieri & Bornstein, 2003; Ramsey & Reynolds, 2004). It made sense in the early 1900s, as it does today, however, that limited academic skills interfere with the measurement of general ability when verbal tests are used. For example, in the immigration museum at Ellis Island there is a story about a young woman whose verbal skills suggested she may have been mentally retarded. Once a nonverbal test was administered, which she completed easily, it became clear that it was a mistake to think she was not smart, even though she did not know English. The issue is no different today, but compounded when verbal and quantitative tests are more closely examined and their similarity to tests used to measure achievement become more apparent.

If a student has not had the chance to acquire verbal and quantitative skills due to limited opportunity to learn, or a disability, verbal and quantitative tests designed to measure general ability may be a good predictor of current academic performance but not a good reflection of his or her ability to learn after having had ample instruction. For example, typical Native American Navajo children living on a reservation in northern Arizona earn low scores on the Verbal but average scores on the Performance scale of the Wechsler (Naglieri & Yazzie, 1983) because they speak English as a second language and have had insufficient exposure to the language of a typical American child. Suzuki and Valencia (1997) argued that verbal and quantitative questions found on most traditional IQ tests interfere

with accurate assessment of minority children. Importantly, the similarity in knowledge and skills required to complete IQ and achievement tests is becoming more apparent. To illustrate, the similarities of the verbal and quantitative questions included in tests of intelligence that include verbal, nonverbal, and quantitative components and administered tests of achievement will be explored.

The oldest intelligence test in use today is the Stanford-Binet 5 (SB-5; Roid, 2003). This test has Quantitative Reasoning items that, for example, require the student to calculate the total number of circles on a page (e.g., two circles in one box plus three in a second box plus one in a third box). The same type of question appears on the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) Arithmetic subtest (now a supplemental subtest), which requires the child to count, for example, the number of birds pictured on a page. Very similar items appear on the Wechsler Individual Achievement Test (WIAT-II, Wechsler, 2001). On that test of knowledge, for example, a Numerical Operations subtest item requires the child to determine the total number of balls shown (e.g., 3 plus 5). Similarly, a Woodcock-Johnson Tests of Achievement (WJ-III ACH; Woodcock, McGrew, & Mather, 2001a) Applied Problems subtest item asks the child to count the number of pencils pictured (e.g., 4). Moreover, a SB-5 Quantitative Reasoning item requires the child to complete a simple math problem (e.g., 4 + 2 = ?) just as the WJ-III ACH Math Fluency (e.g., 7 + 2 = ?) and the WIAT-II Numerical Operations (e.g., 3 + 2 = ?) tests do. This also is found on the Cognitive Abilities Test's (CogAT; Lohman & Hagen, 2001) Quantitative battery and the Iowa Tests of Basic Skills' (ITBS; Hoover, Dunbar, & Frisbie, 2001) mathematics tests. The CogAT Equation Building Test, for example, demands basic math skills to determine how numbers and symbols can be combined to yield a specific numerical value (e.g., 8 x 3 = ? and 12 + 4 - 6 + 2 = ?) and the ITBS Mathematics tests include one test in particular (Math Concepts) that also involves understanding equations. In that test, the student is shown a math problem and asked to select which of four possible equations answers the question. Thus, knowledge of equations is used to test *ability* in

the CogAT and *achievement* in the ITBS. Although it seems reasonable that math skills should be part of a test of achievement, it does not seem reasonable that math skills should be used to measure ability because acquired skills are influenced by both instruction and ability. The same issue applies to verbal tests.

Verbal questions are found on both traditional IQ tests and measures of achievement. For example, all traditional IQ tests include a measure of word knowledge and, amazingly sometimes use the same words on both types of tests. For example, students are required to define a word like *bat* on subtests included in the SB-5 or WISC-IV intelligence tests and the WJ-III ACH. The WJ-III Tests of Cognitive Abilities (WJ-III COG; Woodcock, McGrew, & Mather, 2001b) battery contains a Verbal Comprehension subtest that has an item similar to "Tell me another word for *small*," and the WJ-III ACH contains a Reading Vocabulary question like "Tell me another word for *little*." In addition, an item on the WJ-III ACH Reading Vocabulary test is something like "Tell me another word for (examiner points to the word *big*)," and in the WJ-III COG, the examiner asks something like "Tell me another word for *tiny*." Additionally, the WJ-III COG Verbal Comprehension test contains 23 Picture Vocabulary items and the WJ-III ACH includes 44 Picture Vocabulary items, some of which are the same between the tests. The CogAT Verbal battery also contains tests that demand knowledge of words. The Verbal Classification items require the child to determine how words such as "red, green, and yellow" are alike by choosing from options such as "color, crayon, blue, and marker." Similarly, the reading portion of the ITBS includes a Vocabulary test. A word is presented in a short phrase or sentence, and the student is required to select the answer that has the same meaning as the target word. For example, the child reads the phrase "To *look* in the room" and chooses a corresponding word from among this list: push, sit, fix, peek. These items also require reading skills that sometimes exceed the reading level of those students for whom the test is intended (Naglieri & Ford, 2005).

Naglieri and Ford (2005) evaluated the reading levels required for the items for the CogAT Form 6 Level D Sentence

Completion test. This test is intended for children of average ability in grades 5 and 6. They calculated the readability of the items using the Flesch-Kincaid Grade Level method (Flesch, 1948), which is among the most widely used methods of evaluating reading requirements of text (Chall & Dale, 1995). The Sentence Completion test readability grade level was 6.1 and the readabilities of the individual items ranged from grade 3.7 to 10.4. These findings indicate that children with poor reading skills potentially due to a learning disability, language difference, or limited exposure to English will be at a disadvantage when tested with the CogAT Sentence Completion test because of the achievement demands of this measure of ability.

Acquisition of reading, math, and language skills is a fundamental goal of any formal educational system and in addition, often encouraged, if not explicitly taught, in the home environment. The quality of the educational system and the level of enrichment at home play an important role in the knowledge and skills the child attains. For some children this means that there may be more or less opportunity. For example, Hispanics ages 25 and older are less likely to have a high school diploma (57%) than Whites (88.7%). Importantly, 27% of Hispanics have less than a ninth-grade education compared with only 4% of Whites and only 14.2% of Hispanics are in managerial or professional occupations compared with 35.1% of Whites (Ramirez & de la Cruz, 2002). In order to equitably evaluate the level of ability for a population such as this, or any others with limited opportunity to learn, tests that do not gauge intelligence on the basis of verbal and quantitative skills are necessary.

Practitioners need to understand that the conceptualization of general intelligence that has dominated the field for more than 100 years and which most professionals in education and psychology readily accept as what intelligence is needs to be reexamined. The notion that verbal, quantitative, and nonverbal intelligences are real must be understood within a more accurate historical perspective. The methods used by the U.S. military in the early 1900s (Yoakum & Yerkes, 1920) had utility, but the results must be interpreted differently when applied to

those who have limited English skills and learning experiences. Perhaps most importantly, practitioners need to understand that the originators of these tests did *not* think that the content of the tests represented separate constructs of intelligence; but rather, the different content was used to more fairly assess a wide variety of individuals, many of whom did not have requisite language and math skills.

Wechsler's Influence

Wechsler's view of intelligence was that verbal and nonverbal were *not* two different types of intelligence; despite the fact that for years his tests yielded both Verbal and Performance (nonverbal) IQ scores. He argued that nonverbal tests help to

> minimize the over-diagnosing of feeble-mindness that was, he believed, caused by intelligence tests that were too verbal in content . . . and he viewed verbal and performance tests as equally valid measures of intelligence and criticized the labeling of performance [nonverbal] tests as measures of special abilities (Boake, 2002, p. 396).

There has been widespread acceptance of the inclusion of verbal, quantitative, and nonverbal tests in both ability and achievement tests even though their similarity is obvious. There has been obvious failure to differentiate these constructs, apparent in the descriptions of the tests themselves provided by the authors. Interestingly, the ITBS Vocabulary test is described as "a useful indicator of overall verbal *ability* [emphasis added]" on the publisher's Web page (http://riverpub.com/products/itbs/details.html). Identification of a score from a test of *achievement* as a measure of *ability* does not seem defensible. Nonverbal measures of general ability can be used as a way to circumvent this problem. Using a score from a test of achievement as a measure of ability is illogical because achievement and ability tests should be measuring different constructs—acquired knowledge and skills in contrast to intelligence, respectively. Nonverbal mea-

sures of general ability can be used as a way to circumvent this problem.

General Ability Measured Using Nonverbal Tests

The essence of a nonverbal test of general ability is that it does not contain verbal and quantitative test questions, although it may involve verbal solutions to the problem. For example, Figure 4.1 shows a simple nonverbal test question that could be included in a test described as a progressive matrix. The matrix varies across the horizontal and vertical dimensions. The difference between the top and bottom rows is that the shape inside that square changes (a circle appears on the top row and a triangle on the bottom row). The difference between the first and second column included in the top row is the addition of shading to the circle. The child needs to understand the interrelationships among these variables (shape and shading across the columns and rows) to arrive at the correct answer (option 3). The child may, or may not, use a verbal description (in any language) of the matrix as just described or the child may simply look at the shapes and understand which option is the answer with minimal verbal analysis.

Tests that measure general ability nonverbally may have different types of nonverbal questions, but the essential aspect of these tests is measuring ability nonverbally. Although there is consistency across nonverbal tests in terms of the content of the questions, there are some differences in views about the directions. For example, nonverbal test directions for administration may be spoken as in the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997). Another method is to use pictorial directions as found in the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006), and some authors argue that the entire test must be administered using pantomime, which is perhaps best illustrated by the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998). The slight variation in administration format notwithstanding, the goal is the same: to measure general ability nonverbally. Two examples of tests are provided in the section that follows.

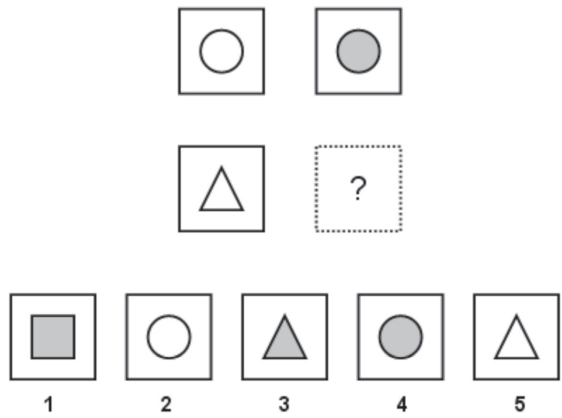

Figure 4.1. Illustration of a nonverbal test (NNAT) item.

Group and Individual Nonverbal Measures of Ability

The NNAT

The Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997) is a group-administered nonverbal test of general ability organized into multiple levels of items composed of diagrams as illustrated in Figure 4.1.

The NNAT consists of seven separate booklets organized into levels, each of which is comprised of 38 items presented in the colors blue, white, and yellow, because these colors are least influenced by color-impaired vision. The seven levels and corresponding grades for which they are intended are as follows: Level A: Kindergarten; Level B: Grade 1; Level C: Grade 2; Level D: Grades 3–4; Level E: Grades 5–6; Level F: Grades 7–9; Level G: Grades 10–12. Each level contains items shared from both the adjacent higher and lower levels, as well as exclusive items. The shared items were used to develop a continuous scaled score across the entire standardization sample. These items yield a total raw score that is converted to a Nonverbal Ability Index standard score set at a mean of 100 with a standard

deviation of 15 through an intermediate Rasch value called a scaled score. Thus, each child's raw score is converted to a scaled score (Rasch value) based upon the NNAT level administered, and then the scaled score is converted to a standard score based upon the age of the child. For more information see Naglieri (1997).

The NNAT was standardized on a nationally representative sample of 89,600 children in grades K–12 (ages 5–18 years). The sample included 22,600 children tested in the fall of 1995 and 67,000 children tested in the spring of 1996. The final complete sample used to create the NNAT norms tables closely matches the U.S. population on the basis of geographic region, socioeconomic status, urbanicity, ethnicity, and school setting (private and parochial). The sample included children in special educational settings such as those with emotional disturbance, learning disabilities, hearing and visual impairment, and those who were mentally handicapped. Children with limited English proficiency also were included in the standardization sample. This standardization procedure also involved concurrent administration of the Stanford Achievement Test–Ninth Edition (SAT-9; 1995). More details may be obtained from the NNAT Technical Manual (Naglieri, 1997). It is important to note that as of the time of this writing, the NNAT is being revised and a second edition of the test, as well as an online version, is expected to be published at the end of 2007.

Summary of NNAT Research

The validity of the NNAT that has particular relevance to the assessment of gifted minority children has been examined in a series of published research papers and will be briefly described here. This will include the examination of White and minority populations, bilingual children, gender differences, and relationships to achievement.

Naglieri and Ronning (2000a, 2000b) studied mean score differences and correlations to achievement for matched samples of White (n = 2,306) and Black (n = 2,306); White (n = 1,176) and Hispanic (n = 1,176); and White (n = 466) and Asian American (n = 466) students in grades K–12. The three

pairs of groups were carefully selected from a larger sample included in the NNAT standardization sample and matched on the demographic characteristics of the U.S. population, including geographical region, socioeconomic status, ethnicity, and type of school setting (public or private). Only small differences were found between the NNAT scores for the White and Black samples (Cohen's *d*-ratio = .25 or about 4 standard score points). Minimal differences between the White and Hispanic (*d*-ratio = .17 or about 3 standard score points), as well as White and Asian American (*d*-ratio = .02 or less than one standard score point) groups also were reported. Additionally, the correlations between NNAT and academic achievement were strong and consistent across grades K–12. Importantly, the NNAT correlated similarly for the White, Black, and Hispanic samples. The small mean score differences and the strong correlations strongly suggest that the NNAT has utility for fair assessment of White and minority children and that the scores the test yields are good for statistical prediction of academic achievement.

Naglieri, Booth, and Winsler (2004) examined the performance of Hispanic children with limited English-language skills. They studied the differences between Hispanic children with (*n* = 148) and without (*n* = 148) limited English proficiency who were administered the NNAT (Naglieri, 1997) and the SAT-9 (1995). The two groups of Hispanic children were selected from 22,620 children included in the NNAT standardization sample and matched on geographical region, gender, socioeconomic status, urbanicity, and ethnicity. The results showed that there was only a small difference (*d*-ratio = 0.1) between the NNAT standard scores for the Hispanic children with limited English proficiency (mean = 98.0) and those without limited English proficiency (mean = 96.7). In addition, the NNAT correlated similarly with achievement for the Hispanic children with and without limited English proficiency. The results suggested that the NNAT scores have use for assessment of Hispanic children with and without limited English proficiency and that these children earned scores that were close to average.

Perhaps one of the most important studies of the NNAT and racial/ethnic differences that suggested that the NNAT may be particularly useful as a fair measure of general ability for gifted minority children was reported by Naglieri and Ford (2003). They studied the practical question—if the NNAT yields small mean score differences between minority and majority groups, would it identify similar percentages of White, Black, and Hispanic children as gifted? Of course, the NNAT would be one part of the larger assessment process used to determine placement in classes for the gifted, albeit an important part. If children would be similarly identified as gifted using the NNAT scores, then the numbers of minority children who may have the opportunity to be selected for gifted programs might increase. To study this question, Naglieri and Ford (2003) used a sample of 20,270 children from the NNAT standardization sample tested during the fall of 1995. These students were representative of the national school population according to socioeconomic status, urban background, and ethnicity and the characteristics of the separate Black, Hispanic, and White groups were also similar in composition. The question addressed by Naglieri and Ford (2003) was: Are the percentages of children who earned NNAT standard scores from 120 to 140 comparable by racial and ethnic groups? To answer this question, standard score frequency distributions were compared to obtain the percentage of each group that would meet the intellectual ability criteria based upon a standard score of 120, as well as 125, 130, 135, and 140 or above (corresponding to the 91st, 95th, 98th, 99th, and 99.6th percentile ranks).

Naglieri and Ford (2003) found that 5.6% of the White (n = 14,141), 5.1% of the Black (n = 2,863), and 4.4% of the Hispanic (n = 1,991) children earned an NNAT standard score of 125 (95th percentile rank) or higher and 2.5% of White, 2.6% of Black, and 2.3% of Hispanic children earned NNAT standard scores of 130 or higher (98th percentile). The identification rates at each 5-point interval from 120 to 140 are shown in Figure 4.2.

These data suggest that the percentages of children that would be identified if the NNAT was used are similar across race

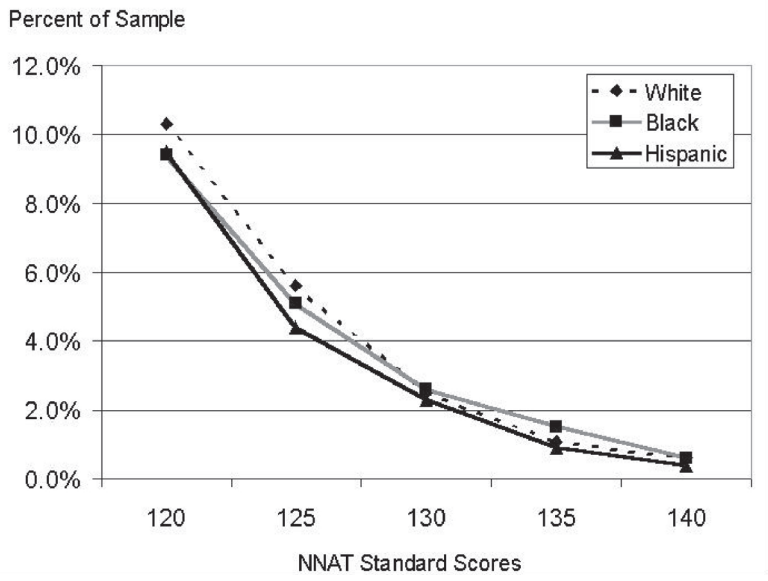

Figure 4.2. Percentage of White, Black, and Hispanic students earning various NNAT standard scores.

and ethnic groups and that the NNAT was effective at identifying diverse students at levels appropriate for gifted education services. The results also suggest that the use of this instrument may help address the persistent problem of the underrepresentation of diverse students in gifted education.

Rojahn and Naglieri (2006) examined gender differences on the NNAT for the entire standardization sample. They found that the NNAT scores for children ages 6–9 (14,468 males and 14,668 females) did not differ (100.2 for both genders). Males (n = 14,273) and females (n = 14,443) ages 10–13 scored the same on the NNAT (100.0 and 100.2, respectively). Finally, males (n = 5,681) and females (n = 5,940) ages 15–17 also scored the same on the NNAT (99.1 for both genders). Scores for this sample by NNAT level then yielded the same results, indicating that on average, males and females earn the same scores on this nonverbal measure of ability.

Wechsler Nonverbal Scale of Ability

Sometimes an individually administered nonverbal measure of general ability is desired and in this case a test like the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006) could be considered. The test yields a full-scale standard score (mean of 100 and *SD* of 15) based on the combination of either two or four subtests scaled using a T-score metric (mean of 50 and *SD* of 10). The WNV is comprised of subtests that were either adapted from other Wechsler tests, are new, or are modeled after the Naglieri Nonverbal Ability Tests (Naglieri, 1997, 2003). The WNV consists of six subtests—Matrices, Coding, Object Assembly, Recognition, Spatial Span, and Picture Arrangement—carefully selected to take into consideration developmental differences between the ages of 4 years, 0 months, and 21 years, 11 months (4:0–21:11). For this reason, the age range was divided into two age bands, ages 4:0–7:11, and ages 8:0–21:11, with each age band having different combinations of subtests comprising both a two- and four-subtest battery. This test was standardized on a large representative sample of children ages 4–21 who closely represented the U.S. population on a number of important demographic variables. The WNV was also standardized on a large representative sample of Canadian children ages 4–21 who closely represented the characteristics of that country (for more details see Wechsler & Naglieri, 2006).

The WNV uses a new method for informing the examinee of the demands of the subtest called Pictorial Directions, designed to provide a nonverbal and engaging method of communicating the task requirements to the examinee. Students are shown a series of pictures that illustrate what is required along with gestures by the examiner that draw attention to the correspondence between the pictured directions and the stimuli in front of the subject.

Pictorial directions are supplemented by simple verbal directions in English, French, Spanish, Chinese, German, and Dutch. The translated verbal directions are used only as needed and by a professional who is able to perform the testing in the

examinee's preferred language. If the use of the pictorial directions and supplemental verbal directions proves ineffective for explaining the demands of the subtest, examiners are instructed to provide additional help as needed. That is, the examiner may interact with the examinee (e.g., refer back to pictorial directions, gesture, demonstrate, and say or sign additional words) to ensure that he or she understands the task requirements. The amount of assistance offered is based on professional judgment, reactions of the examinee, and the particular subtest.

The composition of the WNV reflects the authors' recognition of the value of measuring general ability and the particular advantage of using nonverbal tests to do so. The WNV is like other Wechsler tests in that it uses subtests that vary in content and specific requirements, but different from other Wechsler tests because it was designed to measure general ability using tests that do not have verbal content. The advantage of using nonverbal tasks to measure general ability is that the need for language skills is minimized, and requirements that the examinee have spoken or written language, as well as mathematical, skills are greatly reduced. Although the nonverbal tests on the WNV are all alike in that they do not require language or arithmetic skills, they are diverse in their specific requirements. For example, some of the subtests have a strong visual-spatial requirement, others demand paper-and-pencil skills, and others require the recall of the sequence of information. This multidimensionality of task requirements distinguishes the WNV from tests that use one type of task requirement, such as the NNAT (Naglieri, 1997). Despite the variability of subtest content and task demands, the WNV, like other nonverbal tests have essentially the same goal of measuring general ability nonverbally.

Summary of WNV Research

Due to the recent publication of the WNV, there are comparatively fewer studies on it than on the NNAT, but there are important preliminary findings that bear on the assessment of gifted children that will be briefly described here (see the test manual for more details).

The WNV is strongly correlated with other Wechsler tests (see Wechsler & Naglieri, 2006) but more importantly it is an effective tool for measuring general ability for diverse populations. Gifted children earn high scores on the WNV and the test yields Full Scale scores as high as 170 for both the two- and four-subtest versions. Importantly, the WNV Manual provides a study of English Language Learners. The sample included students whose native language was not English, the primary language they spoke was not English, a language other than English was spoken at home, and/or their parents had resided in the United States fewer than 6 years. The 55 students ages 8–21 were administered the WNV and compared to a group matched on basic demographics. The results showed that the students learning English earned essentially the same score (mean = 101.7) as the matched control of English speaking students group (mean = 102.1). These results indicate that the WNV measures general ability effectively and fairly for those with limited English language skills.

General Ability and Gifted Children

There is considerable need to carefully examine the tests used to help identify children who are gifted and to select those tests that provide all children an equal opportunity to perform. Bracken and Naglieri (2003) argue that traditional tests of intelligence with their verbal, nonverbal, and quantitative tests are best described as measures of general ability. They go on to state that "general intelligence tests with verbal content and nonverbal content measure essentially the same construct as general ability tests that are entirely nonverbal" (p. 247). Both types of tests measure general ability, but one test measures general ability with varying content (verbal, quantitative, and nonverbal); and the other takes an exclusively nonverbal approach. It is important to recognize that the term *nonverbal assessment* describes the methods used to measure the construct of general intelligence, not a theoretical construct of nonverbal ability (Bracken & McCallum, 1998). That is, there is no assumption that nonverbal, as opposed to verbal or quantitative, *abilities* are

being measured. Instead, general ability is measured using nonverbal tests so that a wide variety of individuals may be assessed, using the same set of questions.

The importance of excluding tests with obvious achievement content from a test of ability is particularly salient for children with limited English language skills or those from lower socioeconomic levels where enrichment in the home is limited. It is well-known that high poverty is correlated with low test scores because of issues associated with educational enrichment at home and at school. Many students who live in poverty receive low test scores because of limited opportunity to learn. These students, who may be from all racial and cultural backgrounds, are sometimes penalized on traditional tests of intelligence and subsequently denied access to gifted education programs and services.

Nonverbal measures of general ability are less influenced by limited English language and quantitative skills, making them more appropriate for assessment of culturally and linguistically diverse children (Hayes, 1999; Naglieri & Ford, 2005; Naglieri & Yazzie, 1983; Suzuki & Valencia, 1997). For this and other reasons, nonverbal tests of ability are considered appropriate for a wide variety of persons, especially those with limited English language skills and academic failure (Bracken & McCallum, 1998; Zurcher, 1998). Nonverbal tests can help identify children with high ability who may lack verbal and quantitative skills. The identification method, therefore, has considerable influence on who is served.

There is no consensus about how gifted children should be identified. Although standardized tests often are used as part of the identification process, there is considerable variability as to which tests should be used and what other information should be gathered. Some (e.g., Lohman, 2005) argue that verbal, quantitative, and nonverbal tests are absolutely necessary to identify academically talented students but others (e.g., Naglieri & Ford, 2003, 2005) argue that limiting the definition of gifted to those who demonstrate high achievement and excluding children with high nonverbal scores but lower academic scores perpetuates the problem of underrepresentation of minority

children in gifted programs. They suggest that nonverbal tests are advantageous to provide a more equitable way of evaluating a wide variety of children and give greater opportunity for those from culturally and linguistically diverse populations to participate in gifted programs.

More Inclusive Screening

The identification of gifted children who may not excel in academic skills despite high ability poses an important dilemma to those working in the field of gifted education. The implications for our understanding of what a gifted child is, as well as how he or she should be instructed, are considerable. What the use of a nonverbal test of general ability provides is a way to identify children who might otherwise have been excluded, including children with high ability but lower academic skills. The high nonverbal score suggests that the child can acquire the information, and when combined with an understanding of the child's background, provides a more complete picture of what could be expected. Current academic achievement does predict later academic achievement in most children, but not everyone. Using a nonverbal test of general ability allows us to identify those children who have great potential for academic attainment because of high ability and to give more children the opportunity to get additional educational services.

It is important to stress that a high score on a nonverbal test of ability does not mean that instruction should be nonverbal. The term *nonverbal* describes the method of testing, not the type of ability or thinking. Children identified as gifted on a nonverbal test have the intellectual ability to succeed, and they have the ability to understand and learn at a fast pace. They see the big picture and can understand the detail, but their communication and knowledge base may limit the extent to which they can demonstrate their ability. Their curriculum should provide delivery of academic skills at their level and at a pace that is consistent with their fast rate of learning. Performance is, of course, the desired outcome of many variables, not just ability. Importantly, the interaction of ability with knowledge,

motivation, emotional status, and the classroom, school, and home environments plays an important role. Smart children who earn very high scores on a nonverbal test of ability but whose achievement skills are not at the same high level should be viable candidates for gifted programming and provided the opportunity to raise their academic skills to a level commensurate with their ability.

Conclusions

The need to identify more minority children who are gifted is clear. How to achieve this goal is complicated. In this chapter, I have outlined one dimension of the problem. That is, ability tests that are achievement-laden can become a barrier to smart children who do not have adequate academic skills. Nonverbal measures of ability are, therefore, more appropriate for identification of gifted minority children, especially those who come from disadvantaged homes (Naglieri & Ford, 2003, 2005). Any apparent psychometric advantage verbal and quantitative tests have over nonverbal tests for prediction of achievement is due to the similar skills needed to solve the items included in the verbal and quantitative portions of ability tests and verbal and quantitative portions of achievement tests. The disadvantage of such tests outweighs any advantages, and the failure to include diverse populations because of limited academic skills can be described as a social injustice.

There is a well-documented achievement gap between minority students and those from low-income homes, in contrast to middle and upper socioeconomic and White students, and documented underrepresentation of minority children in programs for the gifted (Bemak & Chung, 2005; Naglieri & Ford, 2003). The methods that have been used and the assumptions about who is gifted have influenced who has been selected to receive additional academic instruction. There is a need for administrators of gifted programs and teachers of the gifted to reduce the achievement gap and foster social justice and equity for minority students who have high ability, yet lower academic skills. We must, therefore, carefully consider the implications of

the test choices we make in addition to the other methods we use for the identification of gifted children.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing.* Upper Saddle River, NJ: Prentice Hall.

Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, 11, 191–244.

Bemak, F., & Chung, R. (2005). Advocacy as a critical role for urban school counselors: Working toward equity and social justice. *Professional School Counseling, 8*, 196.

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

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test.* Itasca, IL: Riverside.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 243–273). New York: Guilford.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education, 32*, 4–14.

Frazier, M. M., Martin, D., Garcia, J., Finley, V. S., Frank, E., Krisel, T, et al. (1995). *A new window for looking at gifted children.* Storrs: National Research Center on the Gifted and Talented, University of Connecticut.

Hayes, S. C. (1999). Comparison of the Kaufman Brief Intelligence Test and the Matrix Analogies Test-Short Form in an adolescent forensic population. *Psychological Assessment, 11*, 108–110.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Tests of Basic Skills.* Itasca, IL: Riverside.

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

Lohman, D. F. (2005). Review of Naglieri and Ford (2003): Does the Naglieri Nonverbal Ability Test identify equal proportions of high-scoring White, Black, and Hispanic students? *Gifted Child Quarterly, 49*, 19–28.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test.* Itasca, IL: Riverside.

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test multilevel form.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (2003). *Naglieri Nonverbal Ability Test individual form.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A., Booth, A., & Winsler, A. (2004). Comparison of Hispanic children with and without limited English proficiency on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 16*, 81–84.

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., & Ford, D. Y. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47*, 155–160.

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly*, *49*, 29–36.

Naglieri, J. A., & Ronning, M. E. (2000a). Comparison of White, African-American, Hispanic, and Asian Children on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 12,* 328–334.

Naglieri, J. A., & Ronning, M. E. (2000b). The relationships between general ability using the NNAT and SAT reading achievement. *Journal of Psychoeducational Assessment, 18,* 230–239.

Naglieri, J. A., & Yazzie, C. (1983). Comparison of the WISC-R and PPVT-R with Navajo children. *Journal of Clinical Psychology, 39*, 598–600.

Pintner, R. (1923). *Intelligence testing.* New York: Henry Holt.

Ramirez, R. R., & de la Cruz, G. (2002). *The Hispanic population in the United States: March 2002* (Current Population Reports, 20-545). Washington, DC: U.S. Census Bureau.

Ramsey, M. C., & Reynolds, C. R. (2004). Relations between intelligence and achievement tests. In G. Goldstein & S. Beers (Eds.), *Comprehensive handbook of psychological assessment* (pp. 25–50). New York: John Wiley.

Rojahn, J., & Naglieri, J. A. (2006). Developmental gender differences on the Naglieri Nonverbal Ability Test in a nationally normed sample of 5–17 year olds. *Intelligence, 34*, 253–260.

Roid, G. (2003). *Stanford-Binet Intelligence Scale* (5th ed.). Itasca, IL: Riverside.

Stanford Achievement Test (9th ed.). (1995). San Antonio, TX: The Psychological Corporation.

Suzuki, L. A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence. *American Psychologist*, *52*, 1103–1114.

Wechsler, D. (1939) *Wechsler-Bellevue Intelligence Scale.* New York: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Individual Achievement Test* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: The Psychological Corporation.

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

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.

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

Zurcher, R. (1998). Issues and trends in culture-fair assessment. *Intervention in School and Clinic, 34*, 103–106.

About the Editor

Joyce L. VanTassel-Baska is the Jody and Layton Smith Professor of Education and Executive Director of the Center for Gifted Education at The College of William and Mary in Virginia, where she has developed a graduate program and a research and development center in gifted education. Formerly, she initiated and directed the Center for Talent Development at Northwestern University. She has also served as the state director of gifted programs for Illinois, as a regional director of a gifted service center in the Chicago area, as coordinator of gifted programs for the Toledo, OH, public school system, and as a teacher of gifted high school students in English and Latin. She has worked as a consultant on gifted education in all 50 states and for key national groups, including the U.S. Department of Education, National Association of Secondary School Principals, and American Association of School Administrators. She has consulted internationally in Australia, New Zealand, Hungary, Jordan, Singapore, Korea, England, Germany, The Netherlands, and the United Arab Emirates. She is past-president of The Association for the Gifted of the Council for Exceptional Children, and the Northwestern University Chapter of Phi Delta Kappa.

Dr. VanTassel-Baska has published widely, including 22 books and more than 500 refereed journal articles, book chapters, and scholarly reports. Recent books include: *Comprehensive Curriculum for Gifted Education* (3rd ed.; 2006) with Tamra Stambaugh; *Curriculum for Gifted*

Students (2004); *Designing and Utilizing Evaluation for Gifted Program Improvement* (2004) with Annie Feng; *Content-Based Curriculum for Gifted Learners* (2003) with Catherine Little; and *Curriculum Planning and Instructional Design for Gifted Learners* (2003). She also served as the editor of *Gifted and Talented International*, a publication of the World Council on Gifted and Talented, from 1998–2005.

Dr. VanTassel-Baska has received numerous awards for her work, including the National Association for Gifted Children's Early Leader Award in 1986, the State Council of Higher Education in Virginia Outstanding Faculty Award in 1993, the Phi Beta Kappa faculty award in 1995, the National Association of Gifted Children Distinguished Scholar Award in 1997, and the President's Award, World Council on Gifted and Talented Education in 2005. She has received awards from Ohio, Virginia, Colorado, South Carolina, and Illinois for her contribution to the field of gifted education in those states. She was selected as a Fulbright Scholar to New Zealand in 2000 and a visiting scholar to Cambridge University in England in 1993. Her major research interests are on the talent development process and effective curricular interventions with the gifted. She holds her bachelor's, master's, and Ed.D. degrees from the University of Toledo.

are formative and summative assessment, standardized testing, and school improvement.

Joni Lakin is a doctoral student in educational psychology at the University of Iowa. Her research interests include cognitive and noncognitive predictors of academic achievement and the use of assessments to identify gifted students, especially gifted students who are English language learners.

David F. Lohman is professor of educational psychology at the University of Iowa. His primary research interests are the nature and measurement of reasoning abilities, the identification of children with extraordinary academic aptitude, and adapting instruction to the needs and proclivities of learners. He also coauthors the *Cognitive Abilities Test* with Elizabeth Hagen.

Jack A. Naglieri is professor of psychology and director of the school psychology program at George Mason University. He is a Fellow of APA Division 16, recipient of APA Division 16 2001 Senior Scientist Award, and holds an appointment as a senior Research Scientist at the Devereux Foundation's Institute for Clinical Training and Research. Dr. Naglieri worked as a school psychologist in the New York area, obtained his Ph.D. in Educational Psychology from the University of Georgia, and taught school psychology at Northern Arizona University and The Ohio State University. The author of more than 250 publications, he has emphasized nonverbal assessment, cognitive processing, cognitive interventions, and measurement of psychopathology and resilience. He has published a number of books including *Essentials of CAS Assessment* and *Helping Children Learn* and tests such as the *Cognitive Assessment System, Wechsler Nonverbal Scale of Ability*, and the *Naglieri Nonverbal Ability Tests*.

Paula Olszewski-Kubilius is currently the director of the Center for Talent Development at Northwestern University and a professor in the School of Education and Social Policy. She received both her master's degree and Ph.D. from Northwestern University in educational psychology. She has worked at the Center for almost 25 years during which she has conducted educational programs of all types for learners of all ages including programs for underrepresented gifted students. She has conducted research and published more than 80 articles or book chapters on issues of talent development, particularly the effects of accelerated educational programs and the needs of special popu-