

Addressing Underrepresentation of Gifted Minority Children Using the Naglieri Nonverbal Ability Test (NNAT)

Jack A. Naglieri
George Mason University

Donna Y. Ford
The Ohio State University

ABSTRACT

A persistent problem in education is the underrepresentation of diverse students in gifted education programs. Many educators attribute the poor participation of diverse students in gifted programs to the ineffectiveness of standardized tests in capturing the ability of these students. Thus, a primary agenda of school selection committees is to find more culturally sensitive measures. This study examined the effectiveness of the Naglieri Nonverbal Ability Test (NNAT) in identifying gifted Black and Hispanic students in comparison to White students. The sample was comprised of 20,270 students in grades K–12 who were similar to the U.S. population on several demographic variables. The distributions of NNAT standard scores were studied separately for White, Black, and Hispanic groups. Results indicate that similar percentages of White (5.6%), Black (5.1%), and Hispanic (4.4%) children earned an NNAT standard score of 125 (95th percentile rank). These findings suggest that the NNAT may be useful as part of a procedure to identify diverse students for gifted education services.

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). As of 1993, the U.S. Department of Education reported that Black, Hispanic, and Native American students were underrepresented by 50–70% in gifted education programs. Ford reported that, despite recent efforts to redress this problem, the underrepresentation of minority students in gifted programs has been persistent and, for some groups, has increased. School personnel and researchers have sought to resolve this problem by examining the ability tests used and procedures followed (de Bernard, 1985; Sternberg, 1985). Many reports

attribute the problem to standardized tests, contending that these tests fail to assess the strengths and abilities of culturally, ethnically, and linguistically diverse populations (e.g., Frazier et al., 1995). Support for this assertion comes from reports showing that Black, Hispanic, and Native American students consistently score lower than White students on traditional standardized tests (Brody, 1992; Sattler, 1988).

Despite the fact that intelligence tests such as the Wechsler Intelligence Scale for Children—Third Edition

PUTTING THE RESEARCH TO USE

There are many smart minority children in our country who may not be considered gifted because they lack the reading, writing, and arithmetic skills typically seen in gifted children and they are identified, in part, by tests of ability that demand school-related knowledge and skills. School ability tests that have verbal and quantitative sections put at a disadvantage minority children with limited educational skills, and, therefore, these children are more likely to earn lower IQ scores. This problem has led some educators to suggest the use of alternative means of assessment that may have limited validity or reliability. In this study, we have shown that a nonverbal test of ability that does not require children to answer verbal and quantitative questions to earn a high IQ can help find more gifted children from minority groups. Those children who demonstrate high performance on a nonverbal test of general ability, but lack academic skills should be served in classes for the gifted. Because there has historically been a reliance on traditional intelligence test scores, rather than nonverbal tests, we have little data on how children who are identified with nonverbal tests might perform in gifted programs. Exactly how they can be served is a question for researchers and policy makers to ponder.

(Wechsler, 1991) and the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986) yield lower scores for minority children (see Kaufman, 1994, for a discussion of the WISC-III), they have been widely used for gifted identification. Wasserman and Becker (2000) have provided a summary of recent research on the WISC-III (Wechsler, 1991), Stanford-Binet Intelligence Scale, Fourth Edition (Thorndike, Hagen, & Sattler), and Woodcock-Johnson Tests of Cognitive Ability (WJ-R; Woodcock & Johnson, 1989) that used samples matched on key demographic variables. They found that the average *differences* in favor of Whites between standard scores for matched samples of Black and White groups were as follows: WISC-III = 11.0; Stanford-Binet IV = 8.1; and Woodcock-Johnson Tests of Cognitive Ability = 11.7. These sizable mean score differences suggest that fewer minority children might be identified when such tests are used for determination of giftedness.

Other reports contend that policies and procedures have a disparate impact on the participation of diverse students in gifted programs, especially the common procedure used by schools: teacher referral. Some researchers (e.g., Ford, 1998) have suggested that teachers often underrefer diverse students for gifted education screening and placement. An additional policy used in some school systems is that students must be assessed in English, which has a profound impact on linguistically diverse or limited English proficient students (de Bernard, 1985).

Given the widespread concerns about testing and assessing diverse students with traditional measures, it is important that school administrators closely examine tests considered culturally fair. Many of these tests fall under the label of "nonverbal tests." Nonverbal tests like the Raven's Progressive Matrices (Raven, 1947) and the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) have been used to evaluate diverse populations of children. Raven's Progressive Matrices and the NNAT are comprised of nonverbal, geometric designs arranged in a 2 x 2 or 3 x 3 matrix. These items can be considered to have content that is culturally reduced because they do not contain items that require the child, for example, to define words or solve oral (English) arithmetic problems. The tests seem especially useful for identification of gifted minority children because the nonverbal content is more appropriate for a wide variety of children (Jensen, 1980; Naglieri & Prewett, 1990; Sattler, 1988).

Researchers have found that the nonverbal measures are less influenced by limited English language skills and, therefore, are more appropriate for bilingual children (Hayes, 1999; Naglieri & Yazzie, 1983). Verbal test scores can be adversely influenced when children have poor lan-

guage skills and live in poverty (Kaufman, 1994; Naglieri, 1999). The use of nonverbal tests helps reduce problems associated with measuring ability through the use of language tests like vocabulary, for example. For these 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).

Raven's Progressive Matrices (Raven, 1947) is the oldest and most widely used nonverbal test. This test has been studied in many countries around the world and with a substantial variety of individuals. Despite its widespread use in the United States, the test has been consistently criticized for its poor psychometric qualities, including the lack of a well-constructed norm group, uneven gradients of item difficulty, inadequate numbers of items, and the need for better documentation of psychometric qualities in the test manual (Jensen, 1980; Nicholson, 1989). Most importantly, however, the difficulty with Raven's Progressive Matrices most relevant to this discussion is findings of higher mean score differences between White and minority children (see Mills & Tissot, 1995; Vincent, 1991). The purpose of this study, therefore, was to examine the question of identification of minority children as gifted using a different nonverbal test: the NNAT.

The Naglieri Nonverbal Ability Test uses the same progressive matrix format as Raven's tests, but there are some important differences. First, the NNAT, like its original versions, the Matrix Analogies Test-Short Form (MAT-SF; Naglieri, 1985a) and Matrix Analogies Test-Expanded Form (MAT-EF; Naglieri, 1985b), was constructed using items that are least influenced by color-impaired vision (only the colors white, black, blue, and yellow are used). Second the NNAT is well standardized on a sample of more than 89,000 students in grades K-12. Third, the psychometric properties of the test are amply documented (Naglieri, 1997b). Finally, there is a research base on the NNAT and its earlier versions (the MAT-EF and MAT-SF) that support its use for diverse populations of children.

Naglieri's progressive matrices tests have a history of yielding small differences between White and minority groups. Naglieri (1985b) summarized the results of two studies involving minority children conducted using the original versions of the NNAT, the MAT-SF and MAT-EF standardization sample. White ($n = 336$) and Black ($n = 336$) children matched on school, gender, and age in years performed similarly (effect size = 0.17 or about 2.6 standard score points) on the MAT-SF. Results for the MAT-EF were similar: Matched samples of White ($n = 55$) and Black ($n = 55$) children earned standard scores (mean of 100, *SD*

of 15) of 90.6 and 90.0, respectively. In other research, the MAT correlated significantly with the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974) Performance IQ Scale ($r = .43, p < .001$) and Raven's Progressive Matrices ($r = .64, p < .001$) for a sample of 114 Native American students (Naglieri, 1985b).

In addition to these initial studies conducted on the first editions of progressive matrices tests by Naglieri (1985a, 1985b), there has been one published study that examined differences between matched samples of White with Black, Hispanic, and Asian American children on the second edition (NNAT; Naglieri, 1997a). In this study, Naglieri and Ronning (2000a) examined differences between three 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 ($n = 466$) children on the NNAT. They found only small differences between the NNAT mean scores for the White and Black samples (d -ratio = .25 or about 4 standard score points) and minimal differences between the White and Hispanic (d -ratio = .17 or about 2.5 standard score points), as well as White and Asian groups (d -ratio = .02 or less than one standard score point). Additionally, Naglieri and Ronning (2000b) found that the NNAT correlated similarly with achievement as measured by the Stanford Achievement Test (Ninth Edition) for the White and minority groups. This implies that children's performance on the NNAT is predictive of their scores on a test of academic achievement (SAT). The results also suggested that the NNAT scores had utility for assessment of White and minority children and that, should the NNAT be used for identification of gifted children, similar numbers of each population might be identified. The present study was conducted to examine the question of identification based upon an ability score and to meet the need for more research on the second edition of this nonverbal test of ability.

Method

Participants

The sample included 20,270 children from the NNAT standardization sample tested during the fall of 1995. These students are representative of the national school population according to socioeconomic status (SES), urbanicity, and ethnicity (see Table 1). The data provided in Table 1 show that the characteristics of the separate Black, Hispanic, and White groups are similar in composition. There were comparable percentages of chil-

dren from the four regions except for West, which had more Black and Hispanic children. The groups differed slightly on urban, suburban, rural community settings. Most of the Black and White children were from suburban and rural settings, while the Hispanic children were fairly evenly dispersed from each setting. Socioeconomic data for the groups showed that the White and Black samples were similar except that there were more Whites at the middle SES level. The Hispanic sample had a large percentage at the low and low middle levels of SES.

Instrument

The Naglieri Nonverbal Ability Test (Naglieri, 1997a) is a brief nonverbal measure of ability that does not require the child to read, write, or speak (Naglieri, 1997b). The test is a nonverbal measure of general ability comprised of progressive matrix items that utilize shapes and geometric designs interrelated through spatial or logical organization. All of the NNAT items require the child to examine the relationships among the parts of the matrix and determine which response is the correct one based only on the information provided in the matrix. The NNAT items are organized into seven levels, each containing 38 items selected to be most appropriate for children at the grade or grades for which that level is intended. Each level contains items that overlap with adjacent higher and lower levels, as well as unique items. Shared items were used to develop a continuous scaled score across the entire standardization sample. The seven levels and corresponding grades for which they are appropriate are as follows: A / K; B / 1; C / 2; D / 3 & 4; E / 5 & 6; F / 7–9; G / 10–12.

NNAT raw scores are converted to Nonverbal Ability Index (NAI) standard scores set at a mean of 100 and *SD* of 15 through an intermediate Rasch value called a scaled score. Level D of the NNAT was used as the base level to which all other levels were equated. The appropriate equating constant was then added to the spring standardization Rasch item difficulties of each level to produce a continuous Rasch ability scale across all levels of the tests (for more information, see Naglieri, 1997b). The internal reliability coefficients for the NNAT by grade range from .83 to .93 with a median internal reliability across all levels of .87 (Naglieri, 1997b).

Data Analyses

The following question was addressed in this study: Are the percentages of children who earned NNAT standard scores from 120 to 140 comparable by racial and ethnic groups? This question is essential given the underrepresent-

Table 1
Demographic Characteristics of the NNAT Samples

Variable	White		Black		Hispanic	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Total <i>n</i>	14,141		2,863		1,991	
Gender						
Male	7,090	50.0	1,519	53.0	1,058	53.0
Female	7,088	50.0	1,346	47.0	939	47.0
Region						
Northeast	2,220	15.7	678	23.7	192	9.6
Midwest	4,629	32.6	484	16.9	137	6.8
Southeast	3,459	24.4	556	19.4	229	11.4
West	3,872	27.3	1,147	40.0	1,444	72.1
Urbanicity						
Urban	411	3.3	302	11.0	604	31.2
Suburban	5,476	44.6	1,536	56.1	827	42.8
Rural	6,392	52.1	899	32.8	503	26.0
SES status						
Low	2,353	19.2	568	20.8	813	42.0
Low middle	2,464	20.1	716	26.2	567	29.3
Middle	2,510	20.4	231	8.4	58	3.0
High-middle	2,910	23.7	533	19.5	119	6.2
High	2,042	16.6	689	25.2	377	19.5

tation of minority children in gifted education programs. Frequency distributions of standard scores were computed by race or ethnic group to answer this question. Comparison of these frequency distributions allowed for determination of the percentage of each group that would meet the intellectual ability criteria based upon a standard score and corresponding percentile cut-off point. In summary, the numbers and percentages of children who earned standard scores of 120, as well as 125, 130, 135, and 140 or above (corresponding to the 91st, 95th, 98th, 99th, and 99.6th percentile ranks) on the NNAT, were computed for each group.

Results

The sample of 14,141 White children earned a mean NNAT score of 99.3 ($SD = 16.7$), which was similar to the mean scores earned by the Black ($n = 2,863$; mean = 96.1; $SD = 17.5$) and Hispanic ($n = 1,991$; mean = 97.3; $SD = 16.8$) children. The percentages of children who earned NNAT standard scores of 120 or higher, 125 or higher,

through 140 or higher are provided in Table 2. These values show that there was similarity in the relative proportions of students from the three groups. That is, 2.5% of White, 2.6 % of Black, and 2.3% of Hispanic children earned NNAT standard scores at the 98th percentile (a standard score of 130). The NNAT standard score of 125 (95th percentile) resulted in 5.6, 5.1, and 4.4% of the White, Black, and Hispanic samples, respectively. These data imply that, if the NNAT were used as one of the criteria in a system of identification of gifted children, similar percentages of White, Black, and Hispanic children would be selected using the cut-offs of 125 or 130.

Discussion

The results of this investigation are consistent with previous research on the NNAT, which has shown that samples of White and minority children perform similarly on this nonverbal measure of ability. The findings, however, go beyond the examination of mean score differences and cor-

Table 2
NNAT Scores

	White		Black		Hispanic		Expected %
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
120 & above	1,571	10.3	269	9.4	190	9.5	9.0
125 & above	906	5.6	145	5.1	88	4.4	5.0
130 & above	467	2.5	75	2.6	46	2.3	2.0
135 & above	190	1.1	42	1.5	18	0.9	1.0
140 & above	90	0.6	19	0.6	9	0.4	0.4
Total Sample <i>n</i>	14,141		2,863		1,991		

Note. Expected percentage values are those associated with normal curve probabilities.

relations to achievement provided by Naglieri and Ronning (2000a, 2000b) to include an important examination of the differential rates of identification for diverse groups. These results are similar to previous studies of the NNAT and its earlier version, the MAT (Naglieri, 1985a, 1985b), which demonstrated that the instrument yielded small differences between majority and minority groups (Naglieri, 1985b; Naglieri & Ronning, 2000a). More importantly, however, the similar percentages of White, Black, and Hispanic children who earned NNAT standard scores of 125, for example, illustrated the potential utility of this instrument for the identification of gifted minority children. With the exception of the Cognitive Assessment System (Naglieri & Das, 1997) and Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) other measures of intelligence have not produced similar findings.

The underrepresentation of Black and Hispanic children in classes for the gifted has been and continues to be of interest to many educators and psychologists. While there is ample evidence that traditional intelligence tests yield differences between groups that do not favor minority populations (e.g., Brody, 1992), they continue to be used. The small mean score differences for the NNAT previously reported by Naglieri and Ronning (2000a), in combination with the data presented in this study, imply that ability can be assessed for these minority populations in a manner that may not lead to underrepresentation. Additionally, because the NNAT can be administered in a group setting (or individually using the NNAT-Individual [Naglieri, 2002]), the information can be obtained in considerably less time per student.

The importance of this study and those that preceded it illustrate how a nonverbal test can be used to evaluate fairly minority children's cognitive ability and, subse-

quently, provide access to gifted education services. The primary difference between the NNAT and other group ability tests is that the latter typically include verbal, quantitative, as well as nonverbal tests. Some researchers have argued that a general ability test with verbal and quantitative items is limited in utility because it demands English language skills and knowledge directly taught in school (Naglieri, 1999; Naglieri & Prewett, 1990). This study showed that the NNAT found similar percentages of White, Black, and Hispanic children to be intellectually gifted. These results further support Naglieri and Ronning's (2000a) suggestion that a nonverbal measure can be a more appropriate measure of general ability for minority children than a measure of general ability that contains both verbal and nonverbal content.

The most important finding of this study (and previous ones) is that, when the NNAT was used, the mean score differences and percentages of children with high standard scores between White and minority groups were small. These results suggest that, when this approach is used as part of the identification process, it could help diverse students gain access to gifted education services.

The next step in the study of the utility of this approach is to examine what classroom modifications and interventions, if any, are necessary when children are identified and placed in gifted programs partially on the basis of nonverbal measures. When children with high NNAT scores and low achievement (because of language differences or limited exposure to academic content) are identified, a curriculum that meets their particular educational needs will be necessary. These children will be different from those who have high NNAT and high verbal and quantitative scores (e.g., they may have poor basic skills), and some type of differentiated instruction will be needed.

Researchers should carefully study the implications of such a potential change in the results of identification and how to teach children who score very well on nonverbal tests of ability, but have lower levels of achievement. They should also examine issues like how long it will take for these students to obtain high levels of achievement. These and other issues should be examined so that we can more fully address the problem of underrepresentation of minority children in classes for the gifted. In the mean time, we must remember that a significant segment of our student population—culturally and linguistically diverse students—have been consistently and significantly underrepresented in our gifted programs; thus, new answers to this persistent problem must be pursued. Nonverbal intelligence tests show promise for increasing the opportunity for diverse students to participate in gifted education programs.

References

- Bracken, B. A., & McCallum, R. S. (1998). *Universal nonverbal intelligence test*. Itasca, IL: Riverside.
- Brody, N. (1992). *Intelligence*. San Diego: Academic Press.
- de Bernard, A. E. (1985). Why Jose can't get into the gifted class: The bilingual child and standardized reading tests. *Roeper Review*, 8, 80–82.
- 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, S., 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.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Kaufman, A. S. (1994) *Intelligent testing with the WISC-III*. New York: Wiley.
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman assessment battery for children*. Circle Pines, MN: American Guidance Service.
- Mills, C.J., & Tissot, S. L. (1995). Identifying academic potential in students from under-represented populations: Is using the Ravens Progressive Matrices a good idea? *Gifted Child Quarterly*, 39, 209–217.
- Naglieri, J. A. (1985a). *Matrix analogies test—Short form*. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A. (1985b). *Matrix analogies test—Expanded form*. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A. (1997a). *Naglieri nonverbal ability test*. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A. (1997b). *NNAT multilevel technical manual*. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Itasca, IL: Riverside.
- Naglieri, J. A., & Prewett, P. N. (1990). Nonverbal intelligence: A selected review of instruments and their use. In R. W. Kamphaus & C. R. Reynolds (Eds.), *Handbook of psychological and educational assessment: Volume I: Intelligence and achievement* (pp. 348–370). New York: Guilford Press.
- 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.
- Nicholson, C. L. (1989). Matrix Analogies Test (MAT). *Diagnostique*, 15, 115–123.
- Raven, J. C. (1947). *Coloured progressive matrices*. London: H. K. Lewis.
- Sattler, J. M. (1988). *Assessment of children* (3rd ed). San Diego, CA: Sattler.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Thorndike, R. I., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet intelligence scale: Fourth edition*. Itasca, IL: Riverside.
- Vincent, K. R. (1991). Black/White IQ differences: Does age make the difference? *Journal of Clinical Psychology*, 47, 266–270.
- Wasserman, J. D., & Becker, K. A. (2000). *Racial and ethnic group mean score differences on intelligence tests*. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.
- Wechsler, D. (1974). *Wechsler intelligence scale for children—Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler intelligence scale for children—Third edition*. San Antonio, TX: The Psychological Corporation.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson revised tests of achievement: Standard and supplemental batteries*. Itasca, IL: Riverside.
- Zurcher, R. (1998). Issues and trends in culture-fair assessment. *Intervention in School and Clinic*, 34, 103–106.

Author Note

This research was supported in part by grant R215K010121 from the U.S. Department of Education.