BRIEF REPORTS

Comparison of Hispanic Children With and Without Limited English Proficiency on the Naglieri Nonverbal Ability Test

Jack A. Naglieri
George Mason University

Ashley L. Booth
University of Virginia

Adam Winsler
George Mason University

Hispanic children with ($n = 148$) and without ($n = 148$) limited English proficiency were given the Naglieri Nonverbal Ability Test (NNAT; J. A. Naglieri, 1997a) and the Stanford Achievement Test—9th edition (SAT–9; 1995). The groups were selected from the NNAT standardization sample ($N = 22,620$) and matched on geographic region, gender, socioeconomic status, urbanicity, and ethnicity. There was a very small difference (d ratio = 0.1) between the NNAT standard scores for the children with limited English proficiency ($M = 98.0$) and those without limited English proficiency ($M = 96.7$). The NNAT correlated moderately and similarly with achievement for the 2 groups. The sample of children with limited English proficiency earned considerably lower scores on SAT–9 Reading and Verbal subtests. Results suggest that the NNAT may be useful for the assessment of Hispanic children with and without limited English proficiency.

Assessment of intelligence for persons with limited English language skills has been an important issue since the familiar verbal–nonverbal organization of tests was initially made popular in the Army Alpha and Beta tests (Yoakum & Yerkes, 1920). The value of a nonverbal test for evaluation of diverse populations was noted by Yoakum and Yerkes more than 80 years ago: "Men who fail in alpha [the verbal tests] are sent to beta [the nonverbal tests] in order that injustice by reason of relative unfamiliarity with English may be avoided" (p. 19). The Beta tests and other similar nonverbal tests have, therefore, served an important role in effective assessment of diverse populations because their content is more culturally neutral (Jensen, 1980; McCallum, Bracken, & Wasserman, 2001; Naglieri & Prewett, 1990; Sattler, 1988). It is logical and empirically supported that (a) bilingual persons will likely perform poorly on English measures of general intelligence that contain verbal tests because of limited English language skills and (b) nonverbal measures of general ability are, therefore, more useful across cultures (Bracken & McCallum, 1998; Hayes, 1999; Kaufman, 1994; Naglieri & Yazzie, 1983; McCallum et al., 2001; Zurcher, 1998).

Recent research on the nonverbal approach to measuring general ability has shown that the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) can be an effective way to assess general ability, yields small race and ethnic group differences, and shows good prediction of achievement. Naglieri and Ronning (2000a) provided a detailed study of mean score differences between 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. Only small differences were found between the NNAT scores for the White and Black samples (Cohen's d ratio, 1988, $d = 0.25$, or about 4 standard score points), White and Hispanic groups ($d = 0.17$, or about 3 standard score points), and White and Asian groups ($d = 0.02$, or less than 1 standard score point). Additionally, the correlations between NNAT and academic achievement were moderate (overall median $r = .59$), which suggested that the nonverbal measure of general ability had utility for assessment of White and minority children. No researchers, however, have compared samples of children with different primary languages. Such research is critical to avoid underrepresentation of language-minority children in gifted education programs and overrepresentation of minority children in special education programs. The purpose of this study was to examine matched samples of Hispanic children with and without limited English proficiency (LEP) on the NNAT. This could provide important practical information to guide a psychologist's selection of instruments when assessing a child or adolescent with limited English language skills. An additional aim of the study was to examine the relationship between the NNAT and achievement for each sample.

Method

This study involved 296 Hispanic children that were divided into matched samples of Hispanic children with and without limited English proficiency as designated during standardization data collection. These children were selected from a larger sample of 22,620 children included in the fall 1995 data collection phase of the NNAT standardization group because of the administration of additional achievement tests. The sample of 22,620 children in kindergarten through Grade 12 closely matches the demographic characteristics of the U.S. national school population in terms of geographic region, ethnicity, and type of school setting (public or private; Naglieri & Ronning, 2000a). Included in this sample were 2,002 Hispanic children who are similar to the U.S. population in terms of gender and urbanicity. Most of this sample came from the western region of the United States, and more than 60% had parents with low and low-middle education levels.

The Hispanic sample was further divided into two groups: those who were designated during the testing phase by school administrators or the students themselves as having LEP and those who did not designate themselves as LEP (non-LEP). Only students who were administered both the NNAT and the Stanford Achievement Test—Ninth edition (SAT–9; 1995) were included in this study. Each LEP child was matched to a non-LEP child on the basis of geographic region (Northeast, Midwest, Southeast, and West), gender, level of NNAT taken (which corresponds to grade), socioeconomic status (low, low-middle, middle, high-middle, and high), and urbanicity (urban, suburban, and rural). Table 1 shows that the two matched groups were nearly identical with regard to their respective demographic characteristics. The two groups were also very similar on the basis of age; t test for age in months showed no significant difference between the LEP ($M = 113.3$, $SD = 44.9$) and the non-LEP ($M = 115.7$, $SD = 38.1$) groups, $t(294) = 0.50$, $p > .05$. However, much like the total Hispanic population, the subgroups displayed the differences from the overall U.S. population in terms of geographic region and socioeconomic status. For example, a lower concentration of suburban cases was found within the two subgroups as opposed to either the Hispanic sample overall or the U.S. population.

NNAT

The NNAT is a general ability test designed so that it does not require the child to read, write, or speak (Naglieri, 1997b). The NNAT measures the nonverbal general ability of its participants through a series of complex matrix items that involve shapes and geometric designs interrelated through spatial or logical organization. Each item within the NNAT is similar in that the child must realize the relationship between the parts of the matrix to successfully choose the correct response. The NNAT consists of seven levels, each containing 38 dichotomously scored items. The level of NNAT indicates the appropriate grade level and age of the intended child. 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.

A Nonverbal Ability Index standard score was converted from the child's NNAT raw score. The standard score is set at a mean of 100 ($SD = 15$) through an intermediate Rasch value called a *scaled score.* Level D, assessing Grades 3–4, was used as the initial level on which all other levels were based. 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 test. Thus, each child's raw score was converted to a scaled score (Rasch value) based on the NNAT level administered. Then the scaled score was converted to a standard score with a mean of 100 ($SD = 15$) based on the age of the child. For more information, see Naglieri (1997b).

The NNAT was standardized on 89,600 children, kindergarten to Grade 12, during the 1995–1996 school year. Of these, 22,600 were tested in fall

Table 1

Similarity of Hispanic LEP and Non-LEP Matched Samples on Demographic Variables

Variable	U.S. (%)	LEP (N = 148)		Non-LEP (N = 148)		Hispanics (N = 2,002)	
		n	%	n	%	n	%
Region							
Northeast	19.6	2	1.4	2	1.4	192	9.6
Midwest	23.8	0	0	0	0	137	6.8
Southeast	24.1	24	16.2	24	16.2	229	11.4
West	32.4	122	82.4	122	82.4	1,444	72.1
Gender							
Male	51.4	79	53.4	79	53.4	1,058	52.8
Female	48.6	69	46.6	69	46.6	939	46.9
SES status							
Low	19.6	15	10.1	15	10.1	813	40.6
Low-middle	21.4	107	72.3	107	72.3	567	28.3
Middle	20.8	0	0	0	0	58	2.9
High-middle	17.8	0	0	0	0	119	5.9
High	20.3	26	17.6	26	17.6	377	18.8
Urbanicity							
Urban	26.8	60	40.5	60	40.5	604	30.2
Suburban	48.0	30	20.3	30	20.3	827	41.3
Rural	25.2	58	39.2	58	39.2	503	25.1

Note. U.S. percentages are from the National Center for Education Statistics, U.S. Department of Education 1993–1994 (see Naglieri, 1997b). Because percentages were rounded, they may not sum to 100. Subsamples (n) may not equal sample size (N) because of missing data. Naglieri Nonverbal Ability Test standardization data Copyright 1995 by The Psychological Corporation. Used by permission. All rights reserved. LEP = limited English proficiency; Non-LEP = nonlimited English proficiency.

1995; the remaining 67,000 were tested in spring 1996. The complete sample of 89,600 children used to create the NNAT norms was representative of the U.S. population in terms of geographic region, socioeconomic status, urbanicity, ethnicity, and school setting (public or private; see Naglieri, 1997b). The current study involved the children included in the fall 1995 sample when the SAT–9 was also administered to the same children. The KR-20 internal reliability coefficients for the NNAT by grade found in Naglieri (1997b) range from .83 to .93. The median internal reliability across all levels is .87.

SAT–9

The SAT–9 is composed of multiple-choice and open-ended assessment questions that measure Total Reading, Vocabulary, and Reading Comprehension subtests. Word reading and sentence reading are also evaluated within the Vocabulary and Reading Comprehension subtests, respectively. The Reading Comprehension subtest, administered to all grades except kindergarten, is composed of a scale of questions that range from interpreting simple sentences to understanding more complex paragraphs. The complex paragraphs ask the child to recognize directly stated details or relationships as well as implicit information and relationships that demand integration of what is provided in the text. The Vocabulary subtest involves identification of a range of words—some simple, some complex—as well as antonyms and synonyms. The Listening subtest assesses students' listening comprehension of dictated selections followed by questions. Standard scores ($M = 100$, $SD = 15$) for each of these tests were obtained by converting national percentile scores to a standard score using a normal cumulative distribution function (e.g., a Z score).

Results and Discussion

Descriptive statistics were calculated for the LEP and non-LEP groups based on the five categories used for matching: geographical region, gender, NNAT level, parent's socioeconomic status, and urbanicity. Standard scores ($M = 100$, $SD = 15$) were used in all analyses, and the differences between the two matched groups were described using d ratios (Becker, 1991). The d ratio is an expression of the difference between the means in units of standard deviation based on the standard deviations of the samples within the two subgroups being compared. Pearson correlations were computed between the NNAT and achievement test standard scores, and moderated regression equations were used to assess group differences between NNAT scores and SAT–9 scores.

Means, standard deviations, and d ratios are presented in Table 2 for NNAT and the achievement variables for the two matched

samples of children. The mean NNAT standard score for the LEP group of 98.0 ($SD = 19.9$) differed minimally from the mean of 96.7 ($SD = 17.6$) for the non-LEP group. The d ratio for the difference was 0.1, which is considered small according to Cohen's (1988) suggestion to interpret d ratios less than 0.5 as small. A t test for independent samples showed no significant difference between the NNAT means for the LEP and non-LEP groups. The means earned by the LEP children and non-LEP children on the tests of achievement suggest that these children as a whole scored in the lower half of what is often considered the average range of performance (90–110). The LEP and non-LEP groups scored similarly on measures of Math Problem Solving and Procedures (all d ratios < 0.2). The LEP children earned means scores on the reading tests that were somewhat lower than those for the non-LEP children (d ratios were small, 0.2–0.3). The largest d ratio was for the Listening standard score ($d = 0.5$), which demands comprehension of English language and written communication in English. These scores suggest that the language-based tests posed challenges for the LEP children, but their math test performance was more similar to the non-LEP sample. Thus, tests with minimal English language demands (math as well as NNAT) yielded smaller differences than those with verbal content.

Simple correlations between the NNAT and achievement test standard scores are presented in Table 3. These illustrate that the test scores for the LEP sample correlated as well as or better than the scores for the non-LEP sample. The correlations ranged from .49 (NNAT with Vocabulary) to .83 (NNAT with Math Problem Solving) for the LEP sample and from .39 (NNAT with Vocabulary) to .77 (NNAT with Math Procedures) for the non-LEP samples. The somewhat higher correlation between NNAT and achievement for the LEP samples is consistent with previous research (e.g., Naglieri & Ronning, 2000a).

To test whether the NNAT scores predict achievement better for LEP or non-LEP students (i.e., to determine whether the slopes or correlations between NNAT and achievement scores were significantly different for the two groups), a series of moderated regression models was conducted. Each of the SAT–9 achievement scores was, in turn, regressed simultaneously on the children's NNAT scores, LEP status, and the interaction term created by the cross product of LEP and NNAT (Aiken & West, 1991). If the interaction term is a statistically significant predictor in the model, it means that the degree of association–prediction between the

Table 2

Means, Standard Deviations, and Sample Sizes for Hispanic LEP and Non-LEP Matched Samples on Ability and Achievement Measures

Measure	LEP			Non-LEP			d ratio	F
	M	SD	n	M	SD	n		
NNAT	98.0	19.8	148	96.7	17.6	148	0.1	0.4
Total Reading	91.7	14.7	148	95.4	12.8	144	0.3	5.3*
Vocabulary	90.1	17.0	133	94.7	13.0	143	0.3	6.5*
Reading Comprehension	93.6	14.7	130	97.1	13.0	130	0.2	4.1*
Listening	88.9	14.5	137	96.2	12.4	137	0.5	19.8**
Total Math	98.5	15.7	148	96.7	14.0	144	0.1	1.1
Problem Solving	97.3	14.8	113	97.2	13.6	110	0.0	0.0
Procedures	102.3	18.2	113	100.4	17.6	113	0.1	0.7

Note. LEP = limited English proficiency.
* $p < .05$.   ** $p < .01$.

Table 3

Correlations Between Naglieri Nonverbal Ability Test and Stanford Achievement Test Standard Scores for Hispanic LEP and Non-LEP Matched Samples

Measure	LEP	n	Non-LEP	n
Total Reading	.70	148	.59	144
Vocabulary	.49	133	.39	143
Reading Comprehension	.69	130	.62	130
Listening	.78	137	.41	137
Total Math	.80	148	.73	144
Problem Solving	.83	113	.73	110
Procedures	.77	113	.77	113

Note. All correlations are significant at $p < .01$. LEP = limited English proficiency.

NNAT score and achievement differs as a function of student LEP status. This was the case only once: for the model predicting children's listening achievement ($\beta = -.284$, $t = -4.24$, $p < .001$), indicating that the association between children's NNAT scores and SAT–9 listening achievement is significantly stronger for LEP students ($r = .78$) than for non-LEP students ($r = .41$).

The results of this study suggest that carefully matched groups of Hispanic children with different levels of English proficiency showed small differences on the NNAT and that NNAT–achievement correlations were at least as strong for the LEP as non-LEP samples. These results provide support for the validity of the NNAT when used with Hispanic children but, more importantly, suggest that these groups can be assessed with this instrument with the expectation of small mean differences and good prediction to achievement.

The current results suggest that nonverbal assessment of bilingual children is an important alternative to tests of general intelligence with verbal content on which bilingual children consistently perform poorly (Kaufman, 1994). Intelligence tests that show sizable mean score differences can lead to overrepresentation of Hispanic children in classes for the mentally impaired (Naglieri & Rojahn, 2001) and fewer minority children in classes for the gifted (Ford, 1998; Naglieri & Ford, 2003). Hispanic children with limited English language skills who are assessed with verbal tests are less likely to qualify, for example, for gifted education programs. Research has shown (Naglieri & Ford, 2003) that the NNAT identifies similar percentages of White and Hispanic children as gifted. Thus, instrument selection is an important factor when making important high-stakes decisions about children.

This study is important because it suggests that practitioners can use this nonverbal test of general ability with the expectation that the scores will be minimally influenced by primary language. The advantage of nonverbal tests like NNAT is that general ability can be assessed using the same content for all children. Given that the size of the correlation between the NNAT and achievement found in this study and in previous research (Naglieri & Ronning, 2000a, b) is similar to that found using ability tests that have verbal and nonverbal content (Naglieri, 1999), it can be concluded that tests of general ability similarly predict achievement despite the differences in their content. The advantage of a nonverbal test, as noted by Yoakum and Yerkes more than 80 years ago, is that nonverbal tests play an important role in assessment of diverse populations. Practitioners, therefore, can use NNAT with Hispanic children and

expect (a) minimal differences between those with and without limited English language skills and (b) good correlation with achievement. Researchers should study children from other linguistic groups who live within the United States and extend this research to international settings.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Thousand Oaks, CA: Sage.

Becker, G. (1991). Alternative methods of reporting research results. *American Psychologist, 46,* 654–655.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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.

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.

McCallum, R. S., Bracken, B. A., & Wasserman, J. D. (2001). *Essentials of nonverbal assessment.* New York: Wiley.

Naglieri, J. A. (1997a). *Naglieri Nonverbal Ability Test.* San Antonio, TX: 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., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47,* 155–160.

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., & Rojahn, J. (2001). Evaluation of African-American and white children in special education programs for children with mental retardation using the WISC-III and Cognitive Assessment System. *American Journal of Mental Retardation, 106,* 359–367.

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.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Sattler.

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

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

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