Journal of Research Methodology, Volume 20, Number 3 (September-December 2007)

วารสารวิธีวิทยาการวิจัย ปีที่ 20 ฉบับที่ 3 (กันยายน-ธันวาคม 2550)

Race Differences on the UNIT: Evidence from Multi-Sample Confirmatory Factor Analysis

Harrison Kane

Kamonwan Tangdhanakanond

ABSTRACT

The Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) is a standardized, norm referenced measure of intelligence in which administration and response formats are entirely nonverbal, requiring only universal hand gestures from the examiner and examinee. The UNIT was constructed in order to provide a more appropriate and valid assessment for children who are not fluent in a community's dominant language and culture. Using multi-sample confirmatory factor analysis, the present study examines standardization for four different racial/ethnic groups (i.e., Whites, Black, Asians, and Hispanics), concluding that the observed differences in subtest performance are likely attributable to group differences in factor means. Suggestions for future study are offered.

Introduction

The demographics of the United States have shifted dramatically over the past decade. According to both the U.S. Bureau of the Census (2006) and the National Center for Culturally Responsive Education Systems (NCCRES, 2006), nearly 30% the nation's school age children belong to a minority group (i.e., African American, Hispanic, Asian, or Native American). Already, these demographic shifts have a marked impact in many of the major cities in the U.S.A., with minorities constituting an overwhelming percentage of the school population in Miami (84%), Chicago (89%) and Houston (88%). Concomitant with the growth of minority students in the general student population is the longstanding concern regarding the discrepancies in the referral and placement of minority children in special education. For example, although Black children comprise 16% of the student population, they constitute nearly 35% of the enrollment in special education classes for the mentally handicapped (National Association of School Psychologists, 2005; Office of Special Education and Rehabilitative Services, 2004). Overall, minorities are nearly twice as likely as Whites to be placed in special education classes (NCCRES, 2006). These group differences in treatments and outcomes warrant more than a passing theoretical interest: academic success translates to real world occupational success (Jencks & Phillips, 1998). These troubling statistics prompted the Individuals With Disabilities Education Act (IDEA) of 1997 (P.L. 105-17) and subsequent revisions (P.L. 108-446) to state explicitly that "greater efforts are needed to prevent the intensification of problems connected with mislabeling minority children with disabilities" (p. 4). Concomitant with this admonition, both the American Psychological Association (APA; 1999) and the National Association of School Psychologists (NASP; 2002, 2005) counsel psychologists to select, and administer psychometrically sound tests that do not discriminate on the basis of race or culture.

Central to the issue of psychoeducational assessment of the nonwhite population in the US is the stubbornly inaccurate conception, by the public and professionals alike, that IQ tests are biased (Meile, 2002; Seligman, 1994). Historically, adherents of this notion point to the oft-observed IQ differences between Blacks and Whites. These group differences are substantial, usually on the order of 15 to 18 IQ points. The judgment that average group differences indicate bias is a seriously flawed simplistic standard and has been abandoned in favor of more sophisticated techniques that scrutinize potential sources of variance. Consequently, average group differences do not necessarily imply test bias. Rather, bias in assessment occurs "when deficiencies in the test itself or the manner in which it is used result in different meanings for scores earned by members of different identifiable subgroups" (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1999, p. 74). To this end, an essential step in ensuring that a test is a fair and appropriate instrument for its intended audience is to examine the factorial invariance across racial/ ethnic groups.

One of the more recent entrants into the field of nonverbal assessment is the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998). The UNIT is a standardized, norm referenced measure of intelligence. The salient feature of the UNIT is that administration and response formats are entirely nonverbal, requiring only universal hand gestures from the examiner and examinee. With particular relevance to the intellectual assessment of nonwhites, the UNIT was constructed in order to provide a more appropriate and valid assessment for children who are not fluent in a community's dominant language and culture. Since its publication, the UNIT has garnered considerable praise from practitioners and researchers (Hooper & Bell, 2006; Sattler, 2001), describing it as "much needed means of obtaining reliable and valid assessments of intelligence for children with a wide array of disabilities who cannot be tested accurately with existing instrument. It is a carefully developed instrument with excellent reliability and impressive evidence of validity for use as supplement to or substitute for more traditional measures such as the WISC-III" (Bandalos, 2001, p. 1298). Despite its widespread use and acceptance, no studies have investigated the structural fidelity of the UNIT's theoretical models across the racial/ethnic groups that comprise its target audience.

The aim of the present study is to ascertain the invariance of the UNIT constructs across groups of Whites, Blacks, Asians, and Hispanics. Evidence of factorial invariance would suggest that the UNIT subtests are valid measures of the same intellectual constructs regardless of race/ethnicity.

Method

Instrument

The UNIT is an individually administered test of intelligence that requires only nonverbal gestures on the part of examiners and examinees. Brief descriptions of the UNIT subtest and scales are offered in Table 1. Across the age levels represented in the standardization sample, average reliabilities (i.e., Cronbach's Alpha) for the various UNIT subtests range from .64 (Mazes) to .91 (Cube Design). For the UNIT Scales, reliabilities are naturally higher, all on the order of .90. Full Scale reliability for the standardization sample is .93. Thus, the UNIT meets and exceeds commonly accepted standards for measures of cognitive and intellectual ability (e.g., Anastasi & Urbina, 1997).

Table 1

Descriptions of the UNIT Subtests and Scales

UNIT	Description
Subtest	
Symbolic Memory	The examinee is presented with universal symbolic stimuli (e.g., baby, girl, boy, woman, and man) of selected colors. The examinee is then required to recall and then reproduce each presented sequence.
Cube Design	The examinee must reproduce abstract, three-dimensional geometric designs constructed of one-inch green and white cubes. One of only two timed subtests, Cube Design is akin to the Block Design subtest common to the familiar Wechsler Scales.
Spatial Memory	The examinee must recall and recreate a pattern of black and green chips presented on a response grid.
Analogic Reasoning	The examinee completes geometric and symbolic analogies and patterns.
Object Memory	The examinee is visually presented with an array of common objects under timed conditions. Following a delay, the examinee must then recall and identify the objects from a larger array.
Mazes	The examinee traces a path through each maze, beginning at the center and ending at an exit. Mazes is one of two timed subtests.
Quotient	
Memory Quotient	MQ measures short-term memory functioning, according to content, location, and sequence of visually presented stimuli.
Reasoning Quotient	RQ measures general problem solving ability, using familiar and novel stimuli.
Symbolic Quotient	SQ measures the ability to problem solve with familiar and meaningful stimuli.
NonSymbolic Quotient	NSQ measures the ability to problem solve with novel and abstract stimuli.
Full Scale Intelligence Quotient	FSIQ provides an index of overall intellectual functioning.

The UNIT subtests can be organized into two measurement models. The primary Theoretical Model arranges subtests according to a two-tiered hierarchical model of intelligence consistent with Jensen's (1980) conceptualization of cognitive ability. Jensen proposed that intelligence consists of Level I (Memory) and Level II (Reasoning) abilities. Nested within the Theoretical Model, the UNIT incorporates two organizational strategies (Symbolic and NonSymbolic categories). Therefore, the theoretical foundation of the UNIT bears some resemblance to existing hierarchical models of intelligence. The Memory/Reasoning abilities correspond roughly to short-term memory (gsm) and fluid reasoning *(gf)* factors characteristic of a *gc-gf* model of fluid and crystallized abilities (e.g., Woodcock, 1998), while the Symbolic/NonSymbolic organizational strategies are akin to the verbal-performance dichotomy often associated with the Wechsler scales (Wechsler, 1997). The Theoretical and Interpretive models of the UNIT are presented in Figure 1.

	Memory	**Reasoning**
Symbolic	Symbolic Memory Object Memory	Analogic Reasoning
NonSymbolic	Spatial Memory	Cube Design Mazes

Figure 1. Conceptual Model of the UNIT.

Participants

The technical manual for the UNIT reports that the normative sample was comprised of 2,100 children and adolescents ranging from 5 to 17 years of age. The standardization sample was distributed evenly, with 175 individuals at each of the twelve age levels. The sample matched the US population on a number of variables, including sex, race, ethnicity, educational placement, geographic region, and parental educational attainment. For the present study, the standardization sample was augmented by additional data collected in a series of validity studies. Participants receiving selected special education services (e.g., Sensory Impairment, Specific Learning Disability, ADHD, Behavioral/Emotional Disorders, and Language Impairment) were excluded from the analyses, as were individuals who indicated a primary language other than English. Using the smallest sample (i.e., Asians) as a basis for matching, racial/ethnic groups were matched additionally according to parental education. Ultimately, the immediate sample consisted of 77 examinees in each group of Whites (M=10.7, SD=3.6 years), Blacks (M=10.2, SD=3.5 years), Asians (M=10.2, SD=3.4 years), and Hispanics (M=10.3, SD=3.3 years).

Analysis

The UNIT scores of students in four different racial/ethnic groups were analyzed using MANOVA to see if there were significant differences. A multi-sample confirmatory factor analysis (MCFA) was also employed to detect test construct invariance among four different racial/ethnic groups of examinee. In this research, the Theoretical and Interpretive models of the UNIT were tested separately by conducting several different sets of MCFA, with the following indices gauging the goodness of fit: Chi-square statistics (χ^2), goodness of fit index (GFI), the Tucker-Lewis index (TLI), and root mean squared error of approximation (RMSEA).

The variance-covariance matrices for all four race/ethnic groups served as input data. The first set of analyses examined the configural invariance of the Theoretical and Interpretive models, specifying the same factor structure and pattern of path coefficients across groups. The purpose of these initial analyses was to determine if the UNIT subtests measure the same latent constructs across race groups. Each of the four subsequent analyses imposed additional constraints on specific parameters of the factor structure (e.g., pattern coefficients, residuals, latent means), with the intent of testing a particular aspect of the factor model that may account for observed group differences. In each stepwise comparison of models, if the more restrictive model results in a significant diminishment in fit, as evaluated by the change in chi-square (subtracting chi-square of the less restrictive from the more restrictive model, which is equal to chi-square with associated degrees of freedom), the factor model is found to differ between groups. The second set of analyses assessed the metric or factorial invariance of the models by requiring the factor loadings to be constrained across groups. Invariance of loadings indicates that both models share a common factor structure. The third set of analyses imposes the additional constraint of equal residual variances for the subtests. If comparison with the previous models reveals no appreciable change in chi-square, observed group differences are not a function of factors not included in the model. Providing the measurement models are deemed invariant, a fourth set of analyses focuses on the latent mean structure by constraining the intercepts of each subtest (e.g., Symbolic Memory) and its associated factor (e.g., Memory). Conventionally, the common factor analysis model makes no assumptions about the means of the common factors. However, Sörbom (1974) showed that it is possible to make inferences about estimated *differences* in factor means under reasonable assumptions. If these additional constraints lack invariance (i.e., comparisons result in a significant change in chi-square), then group differences in subtest performance may be attributed to the respective underlying ability.

Results

Descriptive statistics for the UNIT subtests are offered in Table 2. Predictably, MANOVA reveals significant group differences across the six UNIT subtests, $F(18, 818) = 6.7$, $p < .05$. These mean differences are better appreciated when expressed in reference to the White sample, by dividing the differences by the standard deviation of the White sample. The standardized differences were then corrected for attenuation by dividing by the square root of the reliability for the respective subtest and scale. The ranking of Full Scale IQs for the groups were, in order: Asians (*M*=113.22, *SD*=11.80), Whites (*M*=106.33, *SD*=14.92), Hispanics (*M*=99.58, *SD*=15.48), and Blacks (*M*=92.25, *SD*=16.42). Note in Table 2 that among the UNIT subtests, the most pronounced differences occurred in the Black-White comparisons, where post-hoc t-tests confirmed all comparisons were statistically significant ($p < .05$). Similarly, statistically significant differences were found in the Asian-White comparisons, with Asians uniformly outperforming Whites across the various subtests and quotients. Among the subtests, the largest differences were noted on Black-White comparisons of Cube Design and Analogic Reasoning, at .89 and .98 sigma units, respectively.

Table 2

Mean, and Standard Deviations for Race/Ethnic Groups on the UNIT Subtests and Quotients

UNIT	White		Black		Asian		Hispanic	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Subtest								
Symbolic Memory	11.13	2.65	8.98*	2.89	12.18*	2.84	9.71*	3.25
Cube Design	11.12	3.10	8.50*	2.81	11.46	3,23	10.00*	3.37
Spatial Memory	11.05	2.92	8.79*	3.04	11.88	2.53	10.57	2.78
Analogic Reasoning	11.10	2.64	8.81*	3.20	11.86	2.81	9.01*	3.08
Object Memory	10.80	3.22	9.59*	3.58	11.94*	2.53	10.31	2.98
Mazes	10.47	3.15	9.20*	2.95	11.04	2.95	10.38	2.65
Quotient								
Memory	106.00	15.56	94.14*	17.16	112.63*	12.35	100.92*	15.78
Reasoning	105.80	15.20	92.01*	15.62	109.67	12.98	98.53*	15.63
Symbolic	106.00	14.58	94.22*	17.42	113.33*	13.09	97.68*	15.53
NonSymbolic	105.40	14.97	91.88*	14.49	109.82	12.02	101.76	15.01
Full Scale	106.33	14.92	92.25*	16.42	113.22*	11.80	99.58*	15.48

*$p < 0.5$, when compared to Whites.

MCFA results are summarized in Table 3. Taking each model in turn, the first set of analyses upheld the configural invariance of the Theoretical Model across racial/ethnic groups. The fit statistics universally affirmed an excellent fit of the data to the UNIT factor structure ($\chi^2(32) = 38$, $p = .23$, TLI =.96, GFI = .92, RMSEA = .02). Contrast of this initial measurement model to the next in the progression of constraints yielded a nonsignificant change in the chi-square statistic ($\Delta\chi^2 = 8$ (12), $p > .05$). Consequently, loadings from the common factors are considered equivalent across racial/ethnic groups. Hierarchically extending these constraints to include the covariance between factors further affirmed the invariance of the Theoretical Model ($\Delta\chi^2 = 2$ (3), $p > .05$). Requiring the subtest residuals to be invariant ruled out the influence of factors not included in the model and established the invariance of the measurement model ($\Delta\chi^2 = 19$ (18), $p > .05$). Subtest intercepts were found to be equivalent, regardless of race ($\Delta\chi^2 = 14$ (12), $p > .05$). However, subsequent examination reveals statistically appreciable group differences on the latent means of the Reasoning and Memory factors ($\Delta\chi^2 = 105$ (4), $p < .05$). Regardless of factor, differences in latent means mirrored the ranking of observed performance (see Table 4). Overall, the progression of constraints imposed on the Interpretive model replicated the outcomes from analyses of the Theoretical model. When all parameters are free, the Interpretive model provided a fair fit across the groups ($\chi^2 = 57$ (32), $p < .05$, TLI = .82, GFI = .87, RMSEA = .05). Comparison of this initial measurement model to a model in which the loadings from the common factors to subtests are constrained across groups yielded a nonsignificant change in the chi-square statistic ($\Delta\chi^2 = 8$ (12), $p > .05$). As in the case of the Theoretical model, extending these constraints to include the covariance between the Symbolic and NonSymbolic factors offered no significant change in chi-square ($\Delta\chi^2 = 6$ (3), $p > .05$). Thus, the metric model is invariant across groups. Invariance was further determined when the subtest residuals were constrained ($\Delta\chi^2 = 13$ (18), $p > .05$). Correspondent with the findings from the previous analyses, while subtest intercepts were equivalent ($\Delta\chi^2 = 16$ (12), $p > .05$), while the Interpretive model demonstrated significant group differences in the means of the Symbolic and NonSymbolic ($\Delta\chi^2 = 106$ (4), $p < .05$). Rankings of latent means were identical across factors, with Asians followed by Whites, Hispanics, and Blacks, respectively.

Table 3

Summary of Multi-Sample Confirmatory Factor Analysis for the UNIT Subtests

Model	TLI	GFI	RMSEA	df	X^2	p	$\Delta\ X^2$
Theoretical Model							
All Paramters Free	.96	.92	.02	38	32	.23	--
Common Loading Fixed	.99	.90	.01	46	44	.41	8
Covariance Fixed	.99	.90	.01	48	47	.43	2
Subtest Residuals Fixed	.99	.85	.01	67	65	.42	19
Subtest Intercepts	.98	.82	.01	81	77	.34	14
Factor Means	.71	.58	.06	186	81	.00	105*
Interpretive Modle							
All Paramters Free	.82	.87	.05	57	32	.00	--
Common Loading Fixed	.89	.86	.02	65	44	.02	8
Covariance Fixed	.88	.84	.02	71	47	.02	6
Subtest Residuals Fixed	.93	.81	.03	84	65	.05	13
Subtest Intercepts	.93	.78	.02	100	77	.03	16
Factor Means	.65	.55	.07	206	81	.00	106*

*p < 0.5

Table 4

Means of UNIT Factors for Each Racial/Ethnic Group, Expressed as a Standardized Difference from the White Sample.

Racial/Ethnic Group	Blacks	Hispanics	Asians	*F (3, 304)*
Memory	-1.82+	0.10	1.76+	77.83*
Reasoning	-2.37+	-0.28	1.42+	92.33*
Symbolic	-1.95+	-0.17	1.72+	79.65*
NonSymbolic	-2.17+	0.07	1.40+	86.27*

* $p < .05$

+ $p < .05$, when compared to Whites.

Discussion

Despite completely nonverbal administration and response formats, group differences emerged across the UNIT subtests. Surprisingly, the range of IQ is substantial, spanning 24 IQ points. Therefore, the hope that the UNIT would diminish group differences in IQ is unsatisfied. As with more traditional IQ tests, Blacks anchor the low end of the intellectual continuum (FSIQ = 92.25), while Asians outperform all other groups (FSIQ = 113.22). The cognitive abilities of Whites and Hispanics bore quite a bit of similarity, with FSIQs of 106.33 and 99.58, respectively. Therefore, group differences in IQ remain a stubborn and well-established phenomenon. Importantly, these findings contravene the common criticism that IQ tests are "Eurocentric" and cater to culturally privileged knowledge of the dominant (i.e., White) majority. Apparently, even in the absence of symbolic content, language demands, and academic skills, group differences are manifested in the same fashion commonly observed with more traditional IQ tests that emphasize verbal ability and acquired knowledge. Therefore, group differences in IQ should be accompanied by the established and empirical societal outcomes that are usually acquainted with individual differences in intelligence (e.g., employment status & parenting).

Regarding the factor structure of the UNIT, MSCFA affirmed the general forms of the Theoretical (Memory/Reasoning) and Interpretive (Symbolic/NonSymbolic) models. However, a direct comparison of these models advances the Theoretical model as the better representation of the UNIT data ($\Delta\chi^2 = 19$). Therefore, practitioners should prefer the Theoretical over the Interpretive model, and rely primarily on the former to guide their clinical interpretations. The progression of MSCFA corroborated the invariance of the Theoretical and Interpretive measurement models as a function of group membership. Essentially, the patterns of subtest and factor coefficients were equivalent across these different groups, indicating scores may be interpreted as having comparable meanings regardless of race/ethnicity. Thus, comparison of scores across race groups is a tenable practice. These results also document that the UNIT meets the Standards on Educational and Psychological Testing (AERA et al., 1999).

While the configural and metric invariance of the Theoretical and Interpretive models of UNIT are tenable, several important parameters of these factor models vary across race groups. Notably, groups differ markedly in the distribution of the second-order common factors. Consequently, these particularities represent plausible and authentic disparities in identifiable cognitive abilities, and serve as a credible source of group differences in subtest performance. While this finding may be controversial, it is not new. The causative association between cognitive abilities and observed group differences is well documented (Jensen, 1998, 2001; Lynn, 1996. 2001; Rushton, 1998, 2000, 2002; Spearman, 1927). In each investigation of group differences in intelligence, Blacks occupy the lower end of the spectrum of the cognitive ability in question, while Asians anchor the upper ranges. Certainly, that is the case in the four groups examined in the present study. Findings implicate each of the factors as a source of both within population (individual) and between population (group) differences in nonverbal intelligence.

Of course, some researchers (Gould, 1996; Gustafsson, 1992; Kamin, 1974) are strident in their opposition to investigations of race differences in intelligence, offering that such investigations are morally suspect "distortions and misrepresentations of the data which constitute a truly venomous racism, combined with scandalous disregard for scientific objectivity." (Kamin, 1995, p. 87). Others (Helms-Lorenz, Van de Vijver,

& Pooringa, 2003) are more measured in their criticisms of such research, stating that any findings are necessarily confounded by the entangling of cognitive complexity with culture and linguistic complexity. The present study directly addresses this concern, and finds that group differences in observed performance are intact, even in the absence of culturally laden item content.

There are obvious study limitations that suggest room for additional research. First and foremost, although adequate for the purposes of the present study, sample sizes need to be increased. Efforts to match the samples on parental education drastically reduced the number of examinees. Second, in light of the limited number of subtests and factors, the structure of the UNIT did not permit identification of a factor model that incorporated a hierarchical second order general factor of intelligence (i.e., Spearman's g). Stated differently, the imposition of constraints that were required to answer the questions posed by the presented study prevented the identification of a factor model that incorporated Spearman's g. Most research investigating race differences in intelligence has examined group test performance exclusively as a function of Spearman's g (e.g., Rushton & Jensen, 2005). Thus, future studies should repeat these analyses with data drawn from intelligence tests that include an adequate and necessary number of subtests and factors that would permit the identification of Spearman's g. Finally, the inclusion of data obtained from international sources would determine whether these findings are limited to the particularities of USA. If similar findings were obtained in societies in which the dominant group is a minority in the USA (e.g., Asians tested in Asia or Blacks tested in Africa), such evidence would lend tentative support to a genetic explanation for group differences in IQ.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychological Association. (1999). *The Standards for educational and psychological testing.* Washington, DC: Author

Anastasi, A, & Urbina, S. (1997). *Psychological testing* (7th ed.). New York, NY: Prentice Hall.

Bandalos, D. L. (2001). Review of the Universal Nonverbal Intelligence Test. In B. S. Plake & J. C. Impara (eds.), *Fourteenth mental measurements yearbook* (pp.1296-1298). Lincoln, NE: Buros Institute.

Bracken, B. A, & McCallum, R. S. (1998). *The universal nonverbal intelligence test.* Itasca, IL: Riverside.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatures, history and bibliography, *Intelligence, 24(1),* 13-23.

Gottfredson, L. S. (2002). g: Highly general and highly practical. In R. J. Sternberg & E.L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp.331-380). Mahwah, NJ: Erlbaum.

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311-319.

Gould, S. J. (1996). *The mismeasure of man (2nd ed.).* New York: Norton.

Gustafsson, J. E. (1992). The Spearman Hypothesis is false. *Multivariate behavioral research, 27,* 265-267.

Helms-Lorenz, M., Van de Vijver, F. J. R., & Pooringa, Y. H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: g or c? *Intelligence, 31,* 9-29.

Hooper, S. V., & Bell, S. (2006). Concurrent validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale-Revised. *Psychology in the Schools, 43,* 143-148.

Jencks, C. & Phillips, M. (1998). The black-white test score gap: An introduction. In C. Jencks & M. Phillips (Eds.). *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: Free Press.

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

Jensen, A. R. (2001). Spearman's hypothesis. In S. Messick & Collis, J. M. (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement, (pp. 3-24).* Mahwah, NJ: Lawrence Erlbaum Associates.

Kamin, L. J. (1974). *The science and politics of IQ.* Hillsdale, NJ: Erlbaum.

Kamin, L. J. (1995). Lies, Damned Lies and Statistics. In R. Jacoby & N. Glauberman (Eds.), *The bell curve debate: History, documents, opinions.* (pp.81-105). New York, NY: NY Times Books.

Lynn, R. (1996). Racial and ethnic differences in intelligence in the United States on the Differential Ability Scale. *Personality and Individual Differences, 20,* 271-273.

Lynn, R. (2001). *The science of human diversity: A history of the Pioneer Fund.* Lanham, MD: University Press of America.

McCallum, R. S., Bracken, B. A., & Wasserman, J. D. (2001). *Essentials of nonverbal assessment.* Hoboken, NJ: John Wiley & Sons.

Miele, F. (2002). *Intelligence, race, and genetics: Conversations with Arthur R. Jensen.* Boulder, CO: Westview Press.

National Association of School Psychologists. (2002). *Professional conduct manual.* Retrieved from http://www.nasponline.org/html

National Association of School Psychologists. (2005). *Culturally competent practice.* Retrieved from http://www.nasponline.org/culturalcompetence/index.html

National Center for Culturally Responsive Education Systems. (2006). *Race and risk.* Retrieved from http://nccrest.eddata.net/maps/index.php=RACE_RRW

National Center for Education Statistics.(2001). *Educational achievement and black-white inequality.* U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001061

Office of Special Education and Rehabilitative Services. (2004). *Annual Report To Congress.* Retreived from http://www.ed.gov/about/offices/list/osers/osep/research.html

Office of Special Education and Rehabilitative Services. *Number of people served under Part B of IDEA*. Retrieved from http://www.ed.gov/about/reports/annual/osep/

Rushton, J. P. (1998). The 'Jensen Effect' and the 'Spearman-Jensen hypothesis' of Black-White IQ differences, *Intelligence, 26*, 217-225.

Rushton, J. P. (2000). Performance on Raven's matrices by African and White university students in South Africa, *Intelligence, 28*, 251-265.

Rushton, J. P. (2002). Jensen effects and African/Coloured/Indian/White differences on Raven's Standard Progressive matrices in South Africa. *Personality and Individual Differences, 33*, 1279-1284.

Rushton, J. P., Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11*, 235-294.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Author.

Seligman, D. (1994) *A Question of intelligence*. New York: Birch Lane Press.

Spearman, C.E. (1927). *The abilities of man: Their nature and measurement*. New York, NY: Macmillan.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.

U.S. Bureau of the Census. (2006). *Nation's population one third minority*. Retrieved from http://www.census.gov/Press-Release/www/releases/archives/index.html

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale*. (3rd ed.). San Antonio, TX: The Psychological Corporation.

Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In R. W. Woodcock & J. J. McArdle (Eds.). *Human cognitive abilities in theory and practice* (pp. 137-156). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Woodcock, R.W., McGrew, K.S., Mather, N. (2001). *Woodcock-Johnson Tests of Cognitive Ability*. (3rd ed.). Itasca, IL: Riverside Publishing.