Canadian Journal of
School Psychology
Volume 22 Number 2
December 2007 235-248

10.1177/0829573507303056
http://cjsp.sagepub.com
hosted at
http://online.sagepub.com

Structural Validity of the WISC-III for a National Sample of Native American Students

Joseph C. Kush
Duquesne University

Marley W. Watkins
Pennsylvania State University

Abstract: Test bias research with Native American participants is uncommon, although individual tests of intelligence are often used with Native American students to determine eligibility for special education services. Only two studies with minimally adequate sample sizes have addressed the structural validity of major tests of intelligence in Native American populations. It is unfortunate that both used an obsolete test and included students from only two tribes. This study used confirmatory factor analyses to examine the structure of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) among 344 Native American students representing 12 Bureau of Indian Affairs (BIA) Nations attending BIA schools in 11 states. Results indicated that Wechsler's four-factor oblique model exhibited the best overall statistical fit. Thus, the underlying factor structure of the WISC-III with a national sample of Native Americans was similar to that found in the normative sample. Implications for school psychologists are presented and recommendations for further research are provided.

Résumé: La recherche concernant le biais des tests utilisés avec les participants américains indigènes a été rare, bien que différentes épreuves d'intelligence soient souvent employées avec les étudiants américains indigènes pour déterminer l'acceptabilité pour des services d'adaptation scolaire. Seulement deux études avec des échantillons minimales ont adressé la validité structurale des épreuves principaux de l'intelligence dans les populations américaines indigènes. Malheureusement, ils ont utilisé un test hors d'usage et des étudiants inclus de seulement deux nations. L'étude courante a employé des analyses factorielles confirmatoires pour examiner la structure de la échelle d'intelligence de Wechsler pour enfants—troisième édition (WISC-III) parmi 344 étudiants américains indigènes représentant douze nations du bureau des affaires indiennes (BIA) fréquentant des écoles du BIA dans onze états. Les résultats ont indiqué que le modèle oblique de quatre facteurs proposé par Wechsler a montré la meilleure représentation statistique globale. Ainsi, la structure fondamentale de facteur du WISC-III avec un groupe national des étudiants américains indigènes était semblable à cela trouvée dans

Authors' Note: Please address correspondence to Joseph C. Kush, School of Education, 327C Fisher Hall, Duquesne University, Pittsburgh, PA 15282; kush@duq.edu.

l'échantillon normatif de WISC-III. Des implications pour des psychologues scolaires sont présentées et des recommandations pour la future recherche sont fournies.

Keywords: *intelligence; IQ; factor analysis; Native American; structural validity*

Individual tests of intelligence remain among the most commonly used measures by school psychologists (Stinnett, Havey, & Oehler-Stinnett, 1994) and, as part of the special education eligibility process, are administered to more than one million students in the United States each year (Gresham & Witt, 1997). Given that almost 40% of the total U.S. public school population is linked to minority status (National Center of Educational Statistics, 2000), thousands of African American, Hispanic, Asian American, and Native American children are administered IQ tests each year to determine their eligibility for special education services. Among minorities, Native American students have been found to be more likely to be referred and are overrepresented in special education classrooms (Nielson Salois, 1999; Sparks, 1999).

The use of IQ tests with minority populations has been controversial (Dauphinais & King, 1992; Samuda, 1975) and the concept of test bias has been vigorously debated over the past 30 years (Bond, 1981; Jensen, 1980; Oakland, 1977; Reynolds, 2000a, 2000b). Initially, test bias was thought to be reflective of commonly found mean score differences between majority and minority students, but Thorndike (1971) and others offered more sophisticated alternatives for identifying test bias. Specifically, psychometric test bias is thought to exist when the measure is found to exhibit differential validity across subgroups for which the test will be used (Cole & Moss, 1993; Reynolds, 1983).

Three types of validity evidence have traditionally been identified in test bias research: content, predictive, and construct (Reynolds & Kaiser, 1990). Content bias exists when test items exhibit different statistical properties from group to group for people who have the same underlying skills, and it is typically assessed with differential item functioning (DIF) techniques. Predictive bias is defined by an error in prediction from test scores that is a function of membership in a particular group. Construct validity is generally thought of as the most important from a scientific standpoint (Jensen, 1980) and is often established through factor analysis. A test is considered to be free of construct bias when comparable factor structures, or latent traits, are shown across majority and minority groups. When a test fails to assess the same underlying constructs across ethnic or cultural groups, then that test is not measuring the same constructs for each group and the appropriateness of using the scores for diagnostic and placement purposes is called into question (Kush & Watkins, 1997).

Although there have been hundreds of publications on ability testing with Native Americans (Vraniak, 1994), test bias research with Native American participants has been rare (Suzuki & Valencia, 1997). One comprehensive review of the published literature on IQ test bias found only six studies that involved Native Americans

(Valencia, Suzuki, & Salinas, 2001). In contrast, much of the extant research focused on majority-minority group mean differences and subtest score patterns. For example, Native American examinees have been found to evidence lower verbal IQ scores on the Wechsler Intelligence Scale for Children (St. John, Krichev, & Bauman, 1976); the Wechsler Intelligence Scale for Children–Revised (WISC-R; Beiser & Gotowiec, 2000; Browne, 1984; Hynd, Quackenbush, Kramer, Conner, & Weed, 1979; McShane, 1980; Naglieri & Yazzie, 1983; Taylor, Ziegler, & Partenio, 1984; Teeter, Moore, & Petersen, 1982; Tempest, 1987; Tempest & Skipper, 1988; Whorton & Morgan, 1990; Wilgosh, Mulcahy, & Watters, 1986); the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Nielson Salois, 1999; Tempest, 1998); and the Wechsler Adult Intelligence Scale–Revised (WAIS-R; McCullough, Walker, & Diessner, 1985). In addition, several studies have found that Native American children exhibited a pattern of subtest scores that favored nonverbal and spatial subtests over verbal subtests (Ducheneaux, 2002; Hynd et al., 1979; Naglieri, 1984). However, mean IQ and subtest levels are not informative absent evidence of construct equivalence across groups. That is, tests must be shown to measure the same construct(s) across ethnic groups before the levels of performance across ethnic groups can be compared.

Valencia et al. (2001) found only two studies that assessed the content validity of IQ tests among Native Americans. After conducting an item analysis, Mishra (1982) concluded that approximately 20% of the items examined on the WISC-R were potentially biased against a sample of Navajo children. Similar results were obtained by Ross-Reynolds and Reschly (1983) with a sample of Papago students. Studies completing item analyses of these instruments were an initial step in the examination of test bias, but item bias in the absence of predictive bias has been attributed to methodological error (Hunter & Schmidt, 2000).

Likewise, only two predictive bias studies have been conducted with Native American children (Valencia et al., 2001). In the first study, the WISC-R was the predictor, and group achievement test scores in reading and mathematics were the criteria among White, Black, Chicano, and Native American (Papago) groups (Reschly & Reschly, 1979). Correlations between the WISC-R and academic achievement were lower for the Native American students than for the other groups. These data were subjected to a more sophisticated regression analysis by Reschly and Sabers (1979), who found that WISC-R scores, although lower for the Native American group, were not biased in terms of predictive accuracy. Based on these results, Reschly and Sabers concluded that "the WISC-R appears to be equally valid for different groups as a measure of academic aptitude" (p. 7).

Construct validity has several aspects (Messick, 1995). Of these, the structural aspect is most pertinent for test bias research because it analyzes how the items of a test relate to each other and to the underlying construct. The construct, or structural, validity of IQ tests among Native American children has been addressed by five studies. Reschly (1978) compared the factor structure of the WISC-R for 950

White, African American, Mexican American, and Native American (Papago) students. The broad Verbal and Performance factors were recovered for all four groups, but the Freedom from Distractibility factor did not emerge among the Black and Native American groups. Similar results were reported by Zarske, Moore, and Peterson (1981), who investigated the factor structure of the WISC-R with a sample of 192 Navajo and 50 Papago children with learning disabilities. The expected Verbal and Performance dimensions were similar across groups, but the Freedom from Distractibility factor failed to emerge for either group. It is interesting that studies with the WISC-R among majority students, especially clinical samples, also failed to find the Freedom from Distractibility factor (O'Grady, 1989; Petersen & Hart, 1979), so there was no clear evidence of construct bias. It is unfortunate that several factor analytic studies with the WISC-R and WISC-III among Native Americans were methodologically inadequate, including too few participants for stable estimates (McShane & Plas, 1982; Mishra, Lord, & Sabers, 1989; Wiseley, 2001).

Native Americans have been "vastly under-studied in test bias research investigations" (Suzuki & Valencia, 1997, p. 1109). Specifically, very little research has addressed the factor structure of major tests of intelligence in Native American populations. Factor analytic studies necessitate relatively large numbers of participants, typically a minimum of 200 to 300 (Comrey, 1988; Guadagnoli & Velicer, 1988). Thus, only two studies with minimally adequate sample sizes have been published (Reschly, 1978; Zarske et al., 1981). Both studies generally supported factor congruity (Reynolds, 1982) but used the WISC-R, a test that is now obsolete, and included students from only two tribes (Navajo and Papago).

Given this dearth of construct validity evidence for IQ tests among Native American children, coupled with the fact that no study to date has used confirmatory factor analytic techniques with this population, this study will use confirmatory factor analytic methods to discern the underlying factor structure of the WISC-III with a national sample of Native American students. Empirical support for similar factor structures as found in majority populations would confirm that similar latent traits are being measured among Native American students.

Method

Participants

The sample included 344 Native American students attending Bureau of Indian Affairs (BIA) schools who received comprehensive psychological evaluations in 11 states: Arizona, California, Maine, Michigan, North Carolina, North Dakota, New Mexico, South Dakota, Utah, Washington, and Wyoming. Twelve BIA Nations were represented, including Apache, Arapaho, Cherokee, Chippewa, Navajo, Ojibiwa, Penobscot, Potawatomi, Puyallup, Siboba, Sioux, and Tohono O'odham (i.e., Papago). All students were selected from archival records contributed from recent psychological evaluations and reevaluations. The sample included 227 boys (66%) and 117 girls (33%) in kindergarten through 11th grade (median age = 10; median grade = 4) with a relatively equal distribution across grades 1 through 8. Subsequent to these evaluations, special education status was determined to include 220 students with learning disabilities, 30 students with emotional disabilities, 21 students with mild mental retardation, 6 students with moderate mental retardation, 3 students with multiple handicaps, 5 students with speech-language disabilities, 2 students with hearing impairments, and 2 students categorized as other health impaired. Fifty-five of the students were determined to be ineligible for special education services.

Measures

The WISC-III is an individually administered test of intellectual ability for children aged 6 years to 16 years, 11 months (Wechsler, 1991). It was standardized on a nationally representative sample of 2,200 children, with 100 boys and 100 girls included at each of 11 age levels. The WISC-III includes 13 subtests ($M = 10$; $SD = 3$), which combine to yield Verbal, Performance, and Full Scale IQs ($M = 100$; $SD = 15$). Because the Mazes subtest is not included in the calculation of any IQ scores, it was excluded from all subsequent analyses.

Procedure

Letters requesting participation were mailed to the building principals or special education directors at each of 117 BIA schools in the United States. School administrators were asked to provide anonymous WISC-III data and demographic information. Following a 1-month interval, follow-up postcards were sent requesting participation; following a 2-month interval, individual phone calls were made. This resulted in the receipt of data for 2,301 Native American students; however, Digit Span and Symbol Search subtests were not administered to the vast majority of these students who were consequently excluded from this study. The elimination of a large part of the dataset is unfortunate, as all 12 subtests are required to examine for full factor structure, but is explained by previous research that has demonstrated that many psychologists do not administer the optional WISC subtests (Glutting, Konold, McDermott, Kush, & Watkins, 1997; Ward, Ward, Hatt, Young, & Mollner, 1995) either due to time constraints or because they believe they offer no incremental utility. Special education placements were independently determined by multidisciplinary teams based on federal and state special education rules and regulations. All tests were administered by school psychologists using standard test administration procedures; all administrations were completed in English with no translations.

Table 1
Factor Models Tested in Confirmatory Factor Analyses of the Wechsler Intelligence Scale for Children–Third Edition

	Factor Model				
Subtest	One	Two	Three	Four	Five
Information	1	1	1	1	1
Similarities	1	1	1	1	1
Vocabulary	1	1	1	1	1
Comprehension	1	1	1	1	1
Picture completion	1	2	2	2	2
Picture arrangement	1	2	2	2	2
Block design	1	2	2	2	2
Object assembly	1	2	2	2	2
Arithmetic	1	1	1	3	3
Digit span	1	1	1	3	4
Coding	1	2	3	4	5
Symbol search	1	2	3	4	5

Data Analyses

Confirmatory factor analyses (CFAs) using maximum likelihood estimation methods were conducted on covariance matrices with EQS 6.1. The traditional chi-square statistic was retained to allow a test of exact fit between the model and observed covariances. The comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) fit statistics were also selected a priori for coverage of multiple dimensions of model fit (Bentler, 2007; Hu & Bentler, 1999). The CFI represents the proportion of improvement in fit relative to a null model, RMSEA reflects the covariance residuals adjusted for degrees of freedom, and SRMR indicates how well, on average, the correlation matrix is reproduced. High values of CFI (near 1.0) and low values of SRMR and RMSEA (near 0.0) indicate good model fit. Hu and Bentler (1999) recommended a combination rule that requires a CFI cutoff value close to .95 and SRMR or RMSEA values near .06 to minimize both Type I and Type II error rates.

As noted by MacCallum, Wegener, Uchino, and Fabrigar (1993), "Without adequate consideration of alternative equivalent models, support for one model from a class of equivalent models is suspect at best and potentially groundless and misleading" (p. 196). As a consequence, the five alternative models specified by Wechsler (1991) were tested. As illustrated in Table 1, alternatives included the normative four-factor oblique structure as well as one-, two-, three-, and five-factor models. Within the five-factor model, the error term was set to a maximum of 1 because there were two single-indicator factors (Digit Span and Arithmetic).

Table 2
Descriptive Statistics for Wechsler Intelligence Scale for Children–Third Edition Indices and Subtests for 344 Native American Students

Variable	*M*	*SD*	Range	Skew	Kurtosis
Verbal IQ	77.03	12.80	46–122	.095	–.083
Performance IQ	89.28	14.20	46–140	.017	.964
Full scale IQ	81.36	12.42	40–123	–.137	.718
Verbal comprehension index	77.57	13.28	50–128	.251	.021
Perceptual organization index	90.65	14.34	50–136	–.087	.572
Freedom from distractibility index	81.67	11.52	50–115	–.134	–.107
Perceptual speed index	90.94	14.94	50–146	.227	.384
Picture completion	8.93	3.35	1–18	–.006	.198
Information	5.53	2.63	1–15	.335	–.047
Coding	7.97	3.34	1–19	.554	.743
Similarities	5.97	3.39	1–17	.296	–.318
Picture arrangement	7.25	3.16	1–17	.170	–.107
Arithmetic	6.44	2.43	1–14	.140	.156
Block design	8.37	3.08	1–19	–.106	.227
Vocabulary	5.39	2.88	1–16	.385	.020
Object assembly	8.67	3.29	1–17	–.401	–.309
Comprehension	6.30	3.37	1–19	.398	–.019
Symbol search	8.12	3.52	1–19	.163	.290
Digit span	6.76	2.53	1–15	.312	.568

Table 3
Wechsler Intelligence Scale for Children–Third Edition Confirmatory Factor Analysis Results for Five Alternative Models

Model	X^2	*df*	CFI	RMSEA	90% RMSEA	SRMR	ΔX^2
One	315.61	54	.783	.119	.106—.131	.085	—
Two	145.59	53	.923	.071	.058—.085	.056	170.02*
Three	123.36	51	.940	.064	.050—.079	.050	22.23*
Four	103.29	48	.954	.058	.042—.073	.041	20.07*
Five	102.89	46	.984	.060	.044—.075	.041	0.40

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
$*p < .001$.

Results

Descriptive statistics for the WISC-III among this sample are presented in Table 2. As has been previously found with exceptional samples (Kavale & Nye, 1985), overall scores were lower than found in the normative sample. Verbal scores were

Table 4
Factor Loadings and Communalities for the Wechsler Intelligence Scale for Children–Third Edition Four-Factor Model

	Factor				
Subtest	I	II	III	IV	R^2
Information	.75				.56
Similarities	.74				.54
Vocabulary	.71				.50
Comprehension	.71				.50
Picture completion		.62			.38
Picture arrangement		.68			.46
Block design		.68			.46
Object assembly		.68			.46
Arithmetic			.60		.36
Digit span			.50		.25
Coding				.54	.29
Symbol search				.72	.52

particularly depressed in this sample. However, variability was near normal; thus, restriction of range did not appear to affect the score distribution. Similarly, univariate skewness and kurtosis indices appeared to reflect expected variability.

The results in Table 3 indicate that the normative four-factor oblique model was the best fit for this sample. That model met the combinatorial rule of Hu and Bentler (1999), and chi-square difference tests demonstrated that it was a significantly better fit than models with fewer factors. Although the five-factor model demonstrated a higher CFI index, its RMSEA value deteriorated and it was not significantly different from the four-factor model.

Factor loadings and communalities from the preferred four-factor model are presented in Table 4. The first two factors mirrored the Verbal Comprehension (VC) and Perceptual Organization (PO) dimensions of the WISC-III normative sample, but the communality of the Picture Completion subtest was low. The Freedom from Distractibility (FD) factor was clearly identified by the Arithmetic and Digit Span subtests, but unique variance disproportionally exceeded common variance. The fourth factor, Processing Speed (PS), was loaded by the Symbol Search and Coding subtests, but Coding also displayed a large amount of unique variance.

Discussion

Test bias research with Native American participants has been rare, although individual tests of intelligence are often used with Native American students in the special education eligibility process. Specifically, very little research has addressed the factor structure of major tests of intelligence in Native American populations. Only two studies with minimally adequate sample sizes have been published (Reschly, 1978; Zarske et al., 1981). Both studies generally supported factor congruity (Reynolds, 1982) but used the WISC-R, a test that is now obsolete, and included students from only two tribes (Navajo and Papago). This study examined the factor structure of the WISC-III among Native American students of 12 separate BIA nations. The WISC-III normative oblique four-factor structure was found to be the best fit to these data. Thus, the underlying factor structure of the WISC-III with a national sample of Native American students was configurally similar to that found in the WISC-III normative sample.

Keith and Witta (1997) also found support for this four-factor configuration in a reanalysis of the WISC-III standardization data. While performing a hierarchical CFA, they identified four first-order factors as well as a second-order factor reflecting general intellectual ability. Support has been shown for the four-factor solution identified in the WISC-III standardization sample in an independent nationally representative sample of American children (Roid, Prifitera, & Weiss, 1993) as well as with the normative sample of Canadian children (Roid & Worrall, 1997).

The picture has been more confusing when data derived from special populations of children have been examined (Watkins & Kush, 2002). For example, Konold, Kush, and Canivez (1997) found support for all four WISC-III factors in three independent samples of children receiving special education. Similarly, when examining a sample of Mexican American students with learning disabilities, Kush and Watkins (1994) found support for the four factors, although only partial support could be found for the FD factor. However, other research that has examined children with learning disabilities has shown a two-factor (Kush, 1996), three-factor (VC, PO, and PS; Logerquist-Hansen & Barona, 1994), or four-factor solution (Bell, 1994). Finally, when examining White and Black students from the WISC-III standardization sample and a group of Black students referred for psychological evaluation, Kush et al. (2001) evidenced full support for the VC and PO factors, mixed support for the FD factor, and very limited evidence of the PS factor. Thus, the VC and PO factors have been robust, but the FD and PS factors have received inconsistent support in clinical samples.

Kamphaus (2001) pointed out that concurrent and predictive validity evidence for the WISC-III FD and PS factors remains unconvincing. Similarly, these factors have failed to demonstrate validity for prediction of behavior problems (Riccio, Cohen, Hall, & Ross, 1997) and have failed to display diagnostic precision in the identification of exceptional students (Watkins, Kush, & Glutting, 1997). Relatedly, FD and PS reliability coefficients are relatively weak (Salvia & Ysseldyke, 1998), as are their short- and long-term stability (Canivez & Watkins, 1998). The large amounts of unique variance found in the FD and PS factors within this sample suggest that similar weaknesses may obtain in this sample of Native American students.

IQ scores remain powerful predictors of academic and economic success (Gottfredson, 2005; Schmidt & Hunter, 1998) and, as a result, a scientific examination of possible group IQ differences continues to be of significant societal importance (Gordon, 1997; Gottfredson, 1997). Although these findings are consistent with recent research demonstrating the factor structure of measures of intelligence to be similar for Blacks and Whites (Owen, 1992; Rushton & Skuy, 2000) as well as for Africans, East Indians, and Whites (Rushton, Skuy, & Fridjhon, 2002, 2003), the finding of similar configurations of factor loadings on the WISC-III is an important advancement for psychologists and educators who consider the Wechsler Scales to be the most appropriate measures of intelligence for use with North American Indian children (Browne, 1984; Mishra et al., 1989; Mueller, Mulcahy, Wilgosh, Watters, & Mancini, 1986; Teeter et al., 1982). Establishment of configural invariance indicates that the WISC-III exhibits "similar, but not identical, latent variables" (Chen, Sousa, & West, 2005, p. 474) when used with Native Americans as it does with the normative population.

Nevertheless, several limitations of the study must also be addressed. First, there was no attempt to separate students who had been referred for initial evaluation from those who were receiving a periodic reevaluation. Relatedly, there was no separation by special education classification, by grade level, or by state or region in which the student resided. The rationale behind these decisions was to produce a sample that was nationally representative and large enough to be appropriate for the factor analytic procedures. Further research might consider examining independent subgroups, such as only children with learning disabilities. In addition, although all tests were administered in English, no measure of English proficiency was collected. Finally, because data from the referred sample were gathered from archival records, competence in administration by the examiner can only be assumed.

It is clear that these results cannot be directly extended to the newly created Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003). No studies are reported in the technical manual concerning the psychometric properties of the WISC-IV across diverse populations, and it remains unknown whether the factor structure is invariant across major subgroups or whether the test differentially predicts important outcomes such as academic achievement. Research examining the construct validity of the WISC-IV, including strong indices of factorial invariance, with minority populations must be conducted, but it will take considerable time for these data to be generated. Until then, the importance of the current WISC-III study cannot be minimized, as it remains the single published study to use CFA techniques with a medium- to large-sized population of Native American students.

References

Beiser, M., & Gotowiec, A. (2000). Accounting for native/non-native difference in IQ scores. *Psychology in the Schools*, *37*, 237-252.

Bell, W. M. (1994). *A confirmatory factor analysis of the Wechsler Intelligence Scale for Children–Third Edition on children with learning disabilities*. Unpublished doctoral dissertation, Mississippi State University.

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*, 825-829.

Bond, L. (1981). Bias in mental tests. In B. F. Green (Ed.), *New directions for testing and measurement: Issues in testing—coaching, disclosure, and ethnic bias* (No. 11, pp. 55-77). San Francisco: Jossey-Bass.

Browne, D. B. (1984). WISC-R scoring patterns among Native Americans of the Northern Plains. *White Cloud Journal*, *3*, 3-16.

Canivez, G. L., & Watkins, M. W. (1998). Long term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment*, *10*, 285-291.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*, 471-492.

Cole, N. S., & Moss, P. A. (1993). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201-220). Phoenix, AZ: Oryx Press.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*, 754-761.

Dauphinais, P., & King, J. (1992). Psychological assessment with American Indian children. *Applied & Preventive Psychology*, *1*, 97-110.

Ducheneaux, T. W. (2002). *WISC-III performance patterning differences between Native American and Caucasian children*. Unpublished doctoral dissertation, University of North Dakota.

Glutting, J. J., Konold, T. R., McDermott, P. A., Kush, J. C., & Watkins, M. W. (1997). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *School Psychology Review*, *26*, 176-188.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, *24*, 203-320.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79-132.

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

Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly*, *12*, 249-267.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265-275.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hunter, J. E., & Schmidt, F. L. (2000). Racial and gender bias in ability and achievement tests: Resolving the apparent paradox. *Psychology, Public Policy, and Law*, *6*, 151-158.

Hynd, G. W., Quackenbush, R., Kramer, R., Conner, R., & Weed, W. (1979). Clinical utility of the WISC-R and the French pictorial test of intelligence with Native American primary grade children. *Perceptual and Motor Skills*, *49*, 480-482.

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

Kamphaus, R. W. (2001). *Clinical assessment of children's intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kavale, K. A., & Nye, C. (1985). Parameters of learning disabilities in achievement, linguistic, neuropsychological, and social/behavioral domains. *Journal of Special Education*, *19*, 443-458.

Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly*, *12*, 89-107.

Konold, T. R., Kush, J. C., & Canivez, G. L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. *Journal of Psychoeducational Assessment*, *15*, 123-137.

Kush, J. C. (1996). Factor structure of the WISC-III for students with learning disabilities. *Journal of Psychoeducational Assessment*, *14*, 32-40.

Kush, J. C., & Watkins, M. W. (1994, March). *Factor structure of the WISC-III for Mexican-American, learning disabled students*. Paper presented at the annual meeting of the National Association of School Psychologists, Seattle, WA.

Kush, J. C., & Watkins, M. W. (1997). Construct validity of the WISC-III verbal and performance factors for Black special education students. *Assessment, 4*, 297-304.

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review, 30*, 70-88.

Logerquist-Hansen, S., & Barona, A. (1994, August). *Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and non-Hispanic White children with learning disabilities*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

McCullough, C. S., Walker, J. L., & Diessner, R. (1985). The use of Wechsler scales in the assessment of Native Americans of the Columbia River basin. *Psychology in the Schools, 22*, 23-28.

McShane, D. (1980). A review of the scores of American Indian children on the Wechsler Intelligence Scales. *White Cloud Journal, 1*, 3-10.

McShane, D., & Plas, J. M. (1982). Wechsler Scale performance patterns of American Indian children. *Psychology in the Schools, 19*, 8-17.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Mishra, S. P. (1982). The WISC-R and evidence of item bias for Native American Navajos. *Psychology in the Schools, 19*, 458-464.

Mishra, S. P., Lord, J., & Sabers, D. L. (1989). Cognitive processes underlying WISC-R performance of gifted and learning disabled Navajos. *Psychology in the Schools, 26*, 31-36.

Mueller, H. H., Mulcahy, R. F., Wilgosh, L., Watters, B., & Mancini, G. J. (1986). An analysis of WISC-R item responses with Canadian, Inuit children. *Alberta Journal of Educational Research, 32*, 12-36.

Naglieri, J. A. (1984). Concurrent and predictive validity of the Kaufman assessment battery for children with a Navajo sample. *Journal of School Psychology*, 22, 373-380.

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

National Center of Educational Statistics. (2000). *The condition of education*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Nielson Salois, K. A. (1999). *A comparative study of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) test performance: Northern Cheyenne and Blackfeet reservation Indian children with the standardization sample*. Unpublished doctoral dissertation, Pace University.

Oakland, T. (Ed.). (1977). *Psychological and educational assessment of minority children*. New York: Brunner/Mazel.

O'Grady, K. E. (1989). Factor structure of the WISC-R. *Multivariate Behavioral Research, 24*, 177-193.

Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences, 13*, 149-159.

Petersen, C. R., & Hart, D. H. (1979). Factor structure of the WISC-R for a clinic-referred population and specific subgroups. *Journal of Consulting and Clinical Psychology, 47*, 643-645.

Reschly, D. J. (1978). WISC-R factor structure among Anglos, Blacks, Chicanos, and Native-American Papagos. *Journal of Consulting and Clinical Psychology, 46*, 417-422.

Reschly, D. J., & Reschly, J. E. (1979). Validity of WISC-R factor scores in predicting achievement and attention for four sociocultural groups. *Journal of School Psychology, 17*, 355-361.

Reschly, D. J., & Sabers, D. (1979). Analysis of test bias in four groups with the regression definition. *Journal of Educational Measurement, 16*, 1-9.

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Wiley.

Reynolds, C. R. (1983). Test bias: In God we trust; all others must have data. *Journal of Special Education, 17*, 241-260.

Reynolds, C. R. (2000a). Methods for detecting and evaluating cultural bias in neuropsychological tests. In F. Strickland & C. R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 249-285). New York: Plenum.

Reynolds, C. R. (2000b). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, *6*, 144-150.

Reynolds, C. R., & Kaiser, S. M. (1990). Bias in assessment of aptitude. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 611-653). New York: Guilford.

Riccio, C. A., Cohen, M. J., Hall, J., & Ross, C. M. (1997). The third and fourth factors of the WISC-III: What they don't measure. *Journal of Psychoeducational Assessment*, *15*, 27-39.

Roid, G. A., Prifitera, A., & Weiss, L. G. (1993). Replication of the WISC-III factor structure in an independent sample. *Journal of Psychoeducational Assessment*, *11*, 6-21.

Roid, G. H., & Worrall, W. (1997). Replication of the Wechsler Intelligence Scale for Children–Third Edition four-factor model in the Canadian normative sample. *Psychological Assessment*, *9*, 512-515.

Ross-Reynolds, J., & Reschly, D. J. (1983). An investigation of item bias on the WISC-R with four sociocultural groups. *Journal of Consulting and Clinical Psychology*, *51*, 144-146.

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

Rushton, J. P., Skuy, M., & Fridjhon, P. (2002). Jensen effects among African, Indian, and White engineering students in South Africa on Raven's Standard Progressive Matrices. *Intelligence*, *30*, 409-423.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence*, *31*, 123-137.

Salvia, J., & Ysseldyke, J. E. (1998). *Assessment* (5th ed.). Boston: Houghton Mifflin.

Samuda, A. J. (1975). *Psychological testing of American minorities: Issues and consequences*. New York: Dodd.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Sparks, S. (1999). Educating the Native-American exceptional learner. *Advances in Special Education*, *12*, 73-89.

St. John, J., Krichev, A., & Bauman, E. (1976). Northwestern Ontario Indian children and the WISC. *Psychology in the Schools*, *13*, 407-411.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment*, *12*, 331-350.

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

Taylor, R. L., Ziegler, E. W., & Partenio, I. (1984). An investigation of WISC-R verbal-performance differences as a function of ethnic status. *Psychology in the Schools*, *21*, 437-441.

Teeter, A., Moore, C., & Petersen, J. (1982). WISC-R verbal and performance abilities of Native American students referred for school learning problems. *Psychology in the Schools*, *19*, 39-44.

Tempest, P. (1987). The physical, environmental, and intellectual profile of the fifth grade Navajo. *Journal of American Indian Education*, *26*(3), 29-40.

Tempest, P. (1998). Local Navajo norms for the Wechsler Intelligence Scale for Children–Third Edition. *Journal of American Indian Education*, *37*(3), 18-30.

Tempest, P., & Skipper, B. (1988). Norms for the Wechsler Intelligence Scale for Children–Revised for Navajo Indians. *Diagnostique*, *13*, 123-129.

Thorndike, R. L. (1971). Concepts of culture fairness. *Journal of Educational Measurement*, *8*, 63-70.

Valencia, R. R., Suzuki, L. A., & Salinas, M. F. (2001). Test bias. In R. R. Valencia & L. A. Suzuki (Eds.), *Intelligence testing and minority students: Foundations, performance factors, and assessment issues* (pp. 111-150). Thousand Oaks, CA: Sage.

Vraniak, D. (1994). Native Americans. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 747-754). New York: Macmillan.

Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools, 32,* 267-276.

Watkins, M. W., & Kush, J. C. (2002). Confirmatory factor analysis of the WISC-III for students with learning disabilities. *Journal of Psychoeducational Assessment, 20,* 408-423.

Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997). Prevalence and diagnostic utility of the WISC-III SCAD profile among children with disabilities. *School Psychology Quarterly, 12,* 235-248.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition.* San Antonio, TX: Psychological Corporation.

Whorton, J. E., & Morgan, R. L. (1990). Comparison of the Test of Nonverbal Intelligence and Wechsler Intelligence Scale for Children–Revised in rural Native American and White children. *Perceptual and Motor Skills, 70,* 12-14.

Wilgosh, L., Mulcahy, R., & Watters, B. (1986). Assessing intellectual performance of culturally different Inuit children with the WISC-R. *Canadian Journal of Behavioural Science, 18,* 270-277.

Wiseley, M. C. (2001). *Non-verbal intelligence and Native-American Navajo children: A comparison between the CTONI and the WISC-III.* Unpublished doctoral dissertation, University of Arizona.

Zarske, J. A., Moore, C. L., & Peterson, J. D. (1981). WISC-R factor structures for diagnosed learning disabled Navajo and Papago children. *Psychology in the Schools, 18,* 402-407.

Joseph C. Kush is an associate professor in the School of Education at Duquesne University. He received his PhD in school psychology from Arizona State University. His research interests include issues related to test bias and fairness and instructional leadership and technology.

Marley W. Watkins is a professor in the school psychology program at The Pennsylvania State University and a Diplomat of the American Board of Professional Psychology. He received his PhD in school psychology from the University of Nebraska–Lincoln.