Psychology in the Schools, Vol. 50(10), 2013
View this article online at wileyonlinelibrary.com/journal/pits

DOI: 10.1002/pits.21724

FACTOR STRUCTURE OF THE WECHSLER INTELLIGENCE SCALES FOR CHILDREN–FOURTH EDITION AMONG REFERRED NATIVE AMERICAN STUDENTS

SELENA NAKANO

Arizona State University

MARLEY W. WATKINS

Baylor University

The Native American population is severely underrepresented in empirical test validity research despite being overrepresented in special education programs and at increased risk for psychoeducational evaluation. The structural validity of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) was investigated with a sample of 176, six-to-sixteen-year-old Native American children referred for a psychoeducational evaluation. Confirmatory factor analysis procedures replicated the normative first-order factor structure and a higher-order general ability factor that accounted for the greatest amount of common (69%) and total (33%) variance. These results support the structural validity of the WISC-IV with a referred Native American sample and suggest that interpretation of the WISC-IV scores should not neglect the strong general ability factor. © 2013 Wiley Periodicals, Inc.

Popular perceptions of bias in intelligence testing are ubiquitous, resulting in serious concern about the use of intelligence tests with some ethnocultural minority groups in the United States (Suzuki & Valencia, 1997). These perceptions of bias are frequently based on the observation that Hispanics, African Americans, and Native Americans have, as groups, historically scored lower on intelligence tests than have majority Whites (Coutinho & Oswald, 2000). This underperformance of some ethnic minority groups has been singled out as evidence of the cultural bias of intelligence tests (Dent, 1996; Gould, 1995; Helms, 1992).

Schools are a leading consumer of intelligence tests, using them to determine eligibility for special education services (Suzuki & Valencia, 1997). For example, it has been estimated that 1.0 to 1.8 million individual intelligence tests are administered to American students each year (Gresham & Witt, 1997). Given the life-altering decisions made based on intelligence tests and their widespread use within schools, it is imperative that psychologists be knowledgeable of the reliability and validity of intelligence test scores and, of special import, not rely on biased tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999).

In accord with the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), empirical studies of test bias have utilized a variety of statistical methods for identifying bias (Jensen, 1980; Reynolds, 1983; Reynolds, Lowe, & Saenz, 1999), but have focused on evidence of validity across test items (content validity); evidence that the measure is appropriately related to measures of alternative constructs (predictive validity); and evidence for the measure's internal structure (structural validity). Empirical studies of these aspects of test bias have occurred most often with the Wechsler scales of intelligence (Suzuki & Valencia, 1997), as they are the most widely used cognitive tests with school-aged children (Flanagan & Genshaft, 1997). A meta-analysis of test bias research found that studies investigating structural validity were conducted more frequently than were studies of content and predictive validity (Valencia & Suzuki, 2001).

This study is based on the dissertation of the first author. The contributions of Dr. Linda Caterino and Dr. Sylvia Cohen are gratefully acknowledged.

Correspondence to: Marley W. Watkins, Baylor University, Department of Educational Psychology, One Bear Place #97301, Waco, TX 76798-7301. E-mail: Marley_Watkins@baylor.edu

Structural validity is established when the internal structure of a scale is consistent with what is known about the structure of the construct being measured (Messick, 1995). Factor analysis is a primary method of investigating the internal structure of a cognitive measure (Carroll, 1995). Studies using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have evaluated the consistency of the factor structure of the Wechsler scales across different groups of test-takers to investigate structural validity. Comparability in the constructs measured for different groups and the degree to which the underlying factor structure of a measure is consistent with the major research findings and common interpretations of the test are requisite conditions for the validity of the test and support for its use across diverse groups (Kaufman & DiCuio, 1975). If a test does not measure equivalent constructs across groups, then scores for the groups do not have comparable meaning and the test fails at being useful or appropriate for use (Sandoval, 1982); that is, it is biased.

The Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) is the most recent version of the Wechsler intelligence scales for children. A representative percentage of the ethnic groups found in the March 2000 census were included in the standardization sample for the WISC-IV. Thus, the standardization sample included White ($n = 1,402$), African American ($n = 343$), Hispanic ($n = 335$), and Asian American ($n = 92$) children and a group of children described as "Other" ($n = 28$). The Other category included Native American Indians, Alaskan Natives, and Pacific Islanders (Wechsler, 2003b) and comprised 1.2% of the total standardization sample, whereas the Native American and Alaska Native population comprised 1.5% of the total U.S. population at that time (U.S. Census Bureau, 2000).

With the addition of new subtests and the retention and deletion of several Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991) subtests, the WISC-IV factor structure continued to include four first-order factors and a higher-order general ability factor. The WISC-IV technical manual (Wechsler, 2003b) reported that its 10 core and 5 supplemental subtests are organized in a first-order structure of four oblique factors. The results reported for the normative exploratory and confirmatory factor analyses indicated the clear division of the subtests across each of the four factors but did not include information on the higher-order general ability factor.

Neglect of the general intelligence factor by Wechsler (2003b) prompted independent researchers to investigate the higher-order factor structure and to examine the structural validity of the WISC-IV across gender and age groups. Exploration of the WISC-IV higher-order factor structure was first conducted by Watkins (2006) using EFA and the Schmid–Leiman (1957) orthogonalization procedure with the WISC-IV standardization sample. Watkins (2006) found that the general ability factor accounted for the majority of variance at the subtest level compared with the four first-order oblique factors. Given the relative variance accounted for by the general and first-order factors, Watkins (2006) recommended that WISC-IV interpretation not discount the strong general factor. Another investigation of the WISC-IV factor structure was conducted across male and female subsamples of the standardization sample. Chen and Zhu (2008) used multi-group CFA to verify that the normative four first-order oblique factor structure of the WISC-IV was invariant across gender. Further structural validity studies found that the normative factors were also invariant across age groups (Keith, Goldenring, Taub, Reynolds, & Kranzler, 2006; Sattler & Dumont, 2008; Wechsler, 2003b).

Only four studies have investigated the structure of the WISC-IV in clinical samples. The first study applied EFA methods to the WISC-IV core subtest scores of 432 students who were referred for a psychoeducational evaluation in Pennsylvania schools (Watkins, Wilson, Kotz, Carbone, & Babula, 2006). The second study applied CFA methods to WISC-IV scores from a national sample of 355 students who were referred for evaluation to determine special education eligibility (Watkins, 2010). The third study assessed the WISC-IV factor structure by applying CFA methods to a sample of 344 children who received neuropsychological examinations in the southeastern United States (Bodin, Pardini, Burns, & Stevens, 2009). The final study applied CFA methods to the WISC-IV scores of the 550 children included in the heterogeneous clinical samples collected during standardization research on the WISC-IV (Chen & Zhu, 2012). All four studies supported a hierarchical structure with four first-order factors plus a general intelligence factor. Each study found that the general factor accounted for the largest proportion of common and total variance.

Although structural bias has been empirically rejected in previous versions of the Wechsler scales (Kush et al., 2001; Reschly, 1978), cross-cultural research has focused primarily on African American, Hispanic, and White groups. Of the groups represented in the 2000 Census, Native Americans have been severely underrepresented in structural bias research but are overrepresented in special education (Dauphinais & King, 1992; Hibel, Faircloth, & Farkas, 2008; Marks, Lemley, & Wood, 2010). This group also suffers from environmental deprivation (Vraniak, 1994), high rates of suicide (Centers for Disease Control and Prevention, 2007), and high rates of school dropout (Sparks, 2000). Over the past 50 years, only five structural validity studies of the core subtests of the Wechsler scales have focused on Native American children. Three structural bias studies of the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974a) found that the normative two-factor structure emerged with Native American samples (McShane & Plas, 1982; Reschly, 1978; Zarske, Moore, & Peterson, 1981). The remaining two studies investigated the WISC-III and found the normative factor structure to be consistent with the respective Native American samples (Kush & Watkins, 2007; Wiseley, 2001). Structural bias research on the WISC-IV has yet to be conducted with a Native American sample. Accordingly, the current study employed CFA with a sample of referred Native American students attending public schools in the American Southwest.

Method

Participants

The sample included 176 Native American students (115 boys) attending three school districts in central Arizona and three school districts in northern Arizona who received comprehensive psychoeducational evaluations to determine their eligibility for special education services. Students were enrolled in kindergarten through Grade 12 and were between the ages of 6 and 16 years ($M = 10.6$; $SD = 2.74$). Navajo tribal affiliation was reported in 40% of student files. Multi-tribal affiliation was reported in two instances, wherein Hopi/Navajo and Sioux/Navajo were indicated. The demographic data of the remaining 105 students in the sample indicated Native American without identification of a specific tribal affiliation.

The special education eligibility classifications represented in the sample included learning disability (77%), cognitive impairment (5%), other health impairment (4.5%), emotional disturbance (4%), autism (1%), and traumatic brain injury (1%). Hearing impairment and orthopedic impairment were each reported once (0.6%), and 5% of the sample was not eligible for special education. The primary language of each child was reported to be English, but other languages were spoken in the homes of 8 children. Per the policies of the respective school districts, no further identifying data could be collected.

Instrument

The WISC-IV is an individually administered test of intellectual ability for children ages 6 years, 0 months to 16 years, 11 months. The WISC-IV was standardized using a nationally representative sample of the U.S. population that was stratified according to the population demographics of the U.S. Census Bureau data collected in March, 2000. This measure includes 15 subtests ($M = 10$; $SD = 3$), 10 of which are mandatory for calculation of a Full-Scale IQ (FSIQ). Each subtest contributes to one of the four cognitive domains: Verbal Comprehension Index, Perceptual Reasoning

Index, Processing Speed Index (PSI), and Working Memory Index. Each index score has a mean of 100 and a standard deviation of 15. These four indices are summed to compute the FSIQ ($M = 100$; $SD = 15$).

The WISC-IV technical manual (Wechsler, 2003b) reported the internal consistency reliability of the WISC-IV across standardization and special education samples. Internal consistency coefficients for the test's four indices ranged from .88 (PSI) to .94 (FSIQ) with the standardization sample. For individual subtests, internal consistency coefficients ranged from .85 (Word Reasoning) to .93 (Letter–Number Sequencing). For the special education groups, internal consistency coefficients of the subtests ranged from .72 (Coding) to .94 (Vocabulary).

Wechsler (2003b) also reported evidence on the concurrent validity of the WISC-IV. High correlation coefficients between the WISC-IV and other Wechsler intelligence scales were supportive of its external validity. For example, the WISC-IV FSIQ was correlated with the WISC-III FSIQ ($r = .89$), Wechsler Preschool and Primary Scale of Intelligence–Third Edition (Wechsler, 1974b) FSIQ ($r = .89$), Wechsler Adult Intelligence Scale–Third Edition (Wechsler, 1997) FSIQ ($r = .89$), and the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) FSIQ-4 ($r = .86$). The WISC-IV FSIQ was also correlated with the Wechsler Individual Achievement Test (Wechsler, 1993) Total Achievement composite ($r = .87$).

Procedure

The sample for this study was taken from three school districts in central Arizona and three school districts in northern Arizona. These school districts were asked to participate based on their proximity and diversity of student populations. Following institutional review board approval, a data collection team of eight school psychology graduate students reviewed all psychoeducational files in these school districts and extracted relevant WISC-IV test scores. Participants were included if (a) WISC-IV scores for the 10 core subtests were available, (b) the students were Native American, and (c) English was the child's primary language. Of the 3,297 student files reviewed, 176 students of Native American ancestry met these criteria and were included in the current study.

Analyses

CFA is an appropriate method for testing the structural validity of a theoretically derived and empirically validated construct such as intelligence (Brown, 2006). Additionally, CFA is "most appropriately applied to measures" such as the WISC-IV "that have been fully developed and their factor structures validated" (Byrne, 2012, p. 95). Accordingly, CFA was conducted using maximum likelihood estimation in Mplus 6.1 (Muthén & Muthén, 2010).

Although debatable (Rouquette & Falissard, 2011), some experts have suggested that there be 10 times the number of participants as the number of variables to be analyzed (Hair, Black, Babin, & Anderson, 2010). Thus, a minimum of 100 participants were recruited for this study. Tabachnick and Fidell (2001) suggested that about 150 participants should be sufficient for factor solutions that contain several high loading variables. For similar situations, Guadagnoli and Velicer (1988) also obtained good results when sample size was ≥ 150. Tompson (2000) opined that 100 to 200 subjects were sufficient in straightforward CFA analyses. Given that the tested WISC-IV structures were relatively simple (i.e., no complex loadings, no correlated residuals, etc.), communalities tended to be strong, and measured variables were normally distributed and exhibited relatively high reliabilities, a sample size of 176 was deemed to be adequate (Boomsma, 1987).

Following previous WISC-IV structural analyses, four first-order models and two hierarchical models were specified: (a) one factor; (b) two oblique verbal and nonverbal factors; (c) three

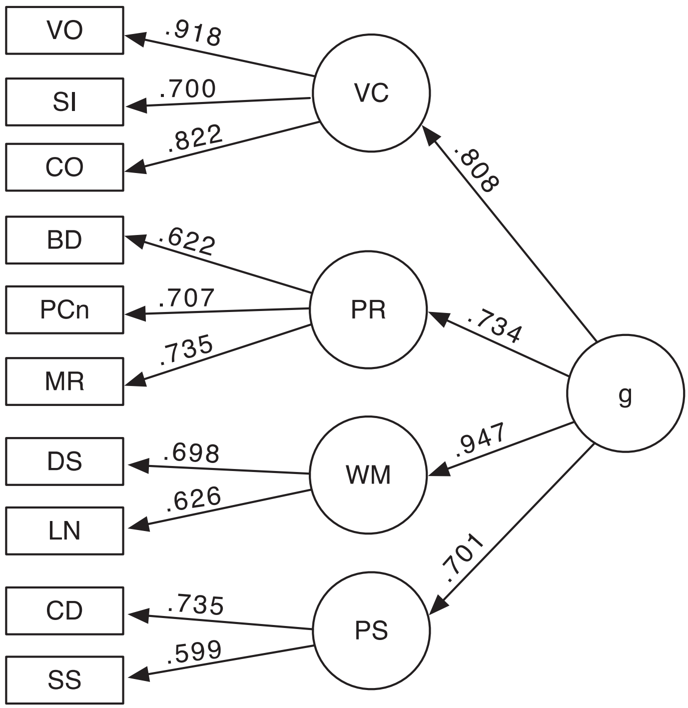

FIGURE 1. Indirect hierarchical structure with standardized coefficients for the Wechsler Intelligence Scale for Children–Fourth Edition in a referral sample of 176 Native American students. g = general intelligence; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; VO = Vocabulary; SI = Similarities; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search.

oblique verbal, perceptual, and working memory/processing speed factors; (d) four oblique verbal, perceptual, working memory, and processing speed factors; (e) an indirect hierarchical model (as per Bodin et al., 2009) with four first-order factors; and (f) a direct hierarchical model (as per Watkins, 2010) with four first-order factors. The final two models conceptualized general intelligence in disparate ways: as a superordinate factor or as a breadth factor, respectively. In the indirect hierarchical model (Gignac, 2008), general intelligence has a direct influence on the first-order factors but only influences the subtests indirectly through the first-order factors (see Figure 1). In contrast, general intelligence has a direct effect on the 10 subtests in the direct hierarchical model (see Figure 2), and each of the first-order factors also has a direct effect on specific subtests, but the general factor has no effect on the first-order factors.

Although contentious (Marsh, Hau, & Wen, 2004), Hu and Bentler (1998, 1999) recommended a dual cut-off value of .95 for the comparative fit index (CFI) *and* .06 for the root mean square error of approximation (RMSEA) to ensure against both type I and type II errors. Higher CFI values and lower RMSEA values indicate better fit. These two indices were supplemented with chi-square and Bayesian information criterion (BIC) values. Nonsignificant chi-square values tend to indicate relatively good model fit. Smaller BIC values indicate better fit after accounting for model complexity. Not all models were nested, so meaningful differences between good models were evaluated using ΔCFI $> +.01$ (Cheung & Rensvold, 2002), ΔRMSEA $> -.015$ (Chen, 2007), and ΔBIC $> +2$ (Raftery, 1995) as standards.

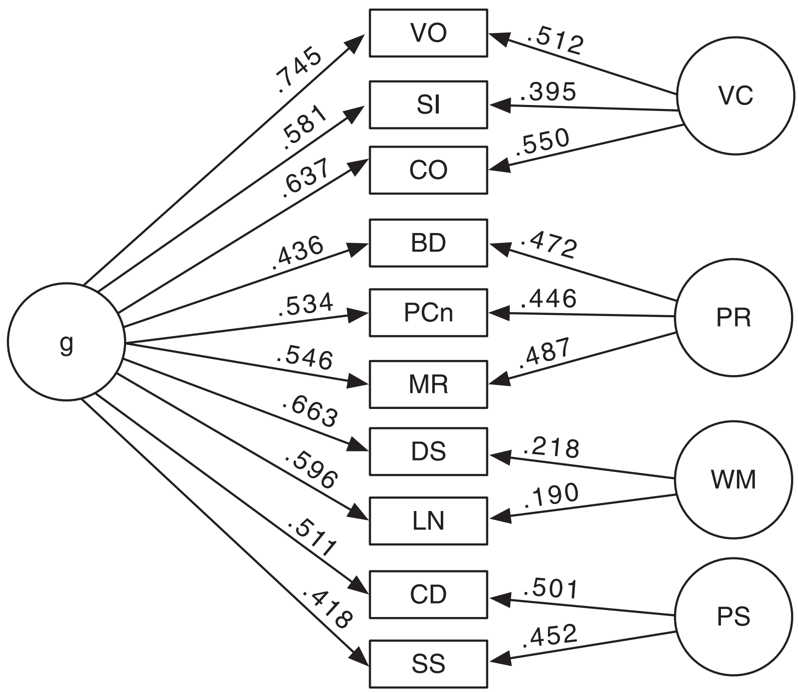

FIGURE 2. Direct hierarchical structure with standardized coefficients for the Wechsler Intelligence Scale for Children–Fourth Edition in a referral sample of 176 Native American students. g = general intelligence; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; VO = Vocabulary; SI = Similarities; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search.

RESULTS

The WISC-IV subtest, factor, and IQ scores of the referred Native American sample were lower and somewhat less variable than were the normative sample (see Table 1). Univariate score distributions from the current sample appeared to be relatively normal, with .66 the largest skew and 1.80 the largest kurtosis. Mardia's (1970) multivariate kurtosis was 0.20, indicating no significant deviation from multivariate normality.

CFA fit statistics are reported in Table 2. As with other structural analyses of the WISC-IV, the first-order models with one to three oblique factors were inferior to the four-factor oblique model and the two hierarchical models. According to chi-square, CFI, and RMSEA values, the final three models all exhibited good fit. Ideally, one of those three would demonstrate superior ΔCFI, ΔRMSEA, and ΔBIC values. However, the ΔBIC favored the indirect hierarchical model, the ΔCFI was neutral, and the ΔRMSEA favored the first-order oblique and indirect hierarchical models. Given its support by two (ΔBIC and ΔRMSEA) of the three indices, the indirect hierarchical model was tentatively accepted as the superior fit to this data and is presented in Figure 1.

When the sources of variance were decomposed (as in Table 3), the general intelligence factor accounted for the majority of variance in the first-order factors (49% for PS to 90% for WM factor) and explained the largest amount of total variance (33.1%) compared with the first-order factors (0.7% to 5.5%). Only one subtest had more variance explained by its first-order factor than by the general factor (17.8% explained by the PS factor versus 17.6% explained by the general factor). For

Table 1
Descriptive Statistics for 176 Native American Children Tested on the Wechsler Intelligence Scale for Children–Fourth Edition

Score	Mean	*SD*	Skewness	Kurtosis
Block Design	9.8	2.9	0.11	– 0.06
Similarities	6.7	2.7	0.21	– 0.62
Digit Span	6.3	2.5	0.25	0.27
Picture Concepts	9.1	2.9	– 0.32	0.31
Coding	8.0	2.8	0.66	1.80
Vocabulary	5.7	2.3	– 0.12	– 0.49
Letter–Number Sequencing	7.2	2.9	– 0.38	– 0.63
Matrix Reasoning	8.7	2.8	0.11	– 0.42
Comprehension	6.6	3.0	– 0.09	– 0.45
Symbol Search	8.2	3.1	– 0.14	0.01
Verbal Comprehension Index	78.7	13.8	– 0.12	– 0.47
Perceptual Reasoning Index	95.2	14.2	– 0.06	– 0.22
Working Memory Index	80.1	13.0	– 0.30	– 0.61
Perceptual Speed Index	89.3	14.1	0.26	0.06
Full-Scale IQ	82.4	12.9	– 0.25	0.01

Table 2
Fit Statistics for Alternative Models for the Wechsler Intelligence Scale for Children–Fourth Edition in a Referral Sample of 176 Native American Students

Model	χ^2	*df*	CFI	ΔCFI	RMSEA (90% CI)	ΔRMSEA	BIC	ΔBIC
One Factor	130.19***	35	.842	–	.124 (.101–.147)	–	8331.6	–
Two Factors	64.55**	34	.949	+.107	.071 (.044–.097)	– .053	8271.1	– 60.5
Three Factors	44.82*	32	.979	+.030	.047 (.000–.078)	– .024	8261.8	– 9.3
Four Factors	26.40	29	.999	+.020	.000 (.000–.051)	– .047	8258.9	– 2.9
Indirect Hierarchical	31.13	31	.999	±.000	.005 (.000–.056)	+.005	8253.3	– 5.6
Direct Hierarchical[a]	29.71	27	.995	– .005	.024 (.000–.065)	+.019	8272.6	+19.3

Note. *df* = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; BIC = Bayesian information criteria. [a]Loading of two subtests on third and fourth first-order factors constrained to be equal to allow identification of model.
*$p < .05$. **$p < .01$. ***$p < .001$.

each of the remaining nine subtests, the general factor explained more variance than its corresponding first-order factor.

Discussion

The preferred factor structure for the WISC-IV with a referred sample of 176 Native American students was an indirect hierarchical model with general intelligence at the apex and four first-order factors. This structure was configurally similar to that found with the WISC-IV normative sample (Keith, 2005) and appears to measure the same underlying constructs as those hypothesized by Wechsler (2003b). Thus, there was no evidence of structural bias with this sample of Native American students.

This obtained structure is also consistent with previous research with clinical samples of non-Native American children. However, the general ability factor only accounted for 33% of the total

Table 3
Decomposition of the Indirect Hierarchical Four-Factor Model of the Wechsler Intelligence Scale for Children–Fourth Edition in a Referral Sample of 176 Native American Students

	General		VC		PR		WM		PS	
Subtest	*b*	Var%	*b*	Var%	*b*	Var%	*b*	Var%	*b*	Var%
SI	.566	32.0	.366	13.4						
VO	.742	55.1	.420	17.6						
CO	.664	44.1	.397	15.8						
BD	.457	20.9			.406	16.5				
PCn	.519	26.9			.432	18.7				
MR	.539	29.1			.443	19.6				
DS	.661	43.7					.192	3.7		
LN	.593	35.2					.182	3.3		
CD	.515	26.5							.469	22.0
SS	.420	17.6							.422	17.8
Total		33.1		4.7		5.5		0.7		4.0
Common		69.1		9.8		11.4		1.5		8.3

Note. *b* = loading of subtest on factor; Var% = percent of variance explained; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; and SS = Symbol Search.

variance and 69% of the common variance of the WISC-IV in the current study. In contrast, the general ability factor accounted for, on average, 47% of total variance and 76% of common variance in three other clinical samples (Bodin et al., 2009; Watkins, 2006, 2010).

Similar patterns of observed scores have been found with other samples of referred students (Canivez & Watkins, 1998), including referred Native American students (Dolan, 1999; Ducheneaux, 2002; Kush & Watkins, 2007; Reschly, 1978; Tanner-Halverson, Burden, & Sabers, 1993), namely, the current sample presented with mean verbal comprehension performance approximately 1.5 standard deviations below the normative mean of 100, perceptual reasoning ability .33 of a standard deviation below the normative mean, processing speed ability .66 of a standard deviation below the normative mean, working memory performance 1.33 standard deviations below the normative mean, and overall ability that was approximately 1 standard deviation below the normative mean.

Limitations

There are several limitations to the current study. First, the data were collected from archival special education records, and the accuracy of the professionals who initially gathered and recorded the data was assumed. Second, the current sample was derived from a small number of school districts in northern and central Arizona. Third, only Navajo (of the 21 federally recognized tribes in Arizona) was indicated in nearly all of those cases that included a specific tribal affiliation. Fourth, some children lived primarily on a reservation, whereas others lived primarily in rural or urban environments. Previous research has identified differences in performance on cognitive measures between children who live in rural and urban environments (Jensen, 1984; Tanner-Halverson et al., 1993; Tempest, 1998). Finally, there was no quantification of participants' English-language proficiency. A substantial body of research has implicated English-language proficiency as a primary source of variance in Native American performance on cognitive measures (Beiser & Gotowiec, 2000; Tanner-Halverson et al., 1993; Tsethlikai, 2011). Given these limitations, the generalizability of these results to other samples of Native American children from different tribes and who reside in different regions of the United States is uncertain.

Future Research

Future research on the cognitive assessment of Native American children is tasked with answering the decades-old call to produce a systematic examination of the cognitive assessment practices used with this population. Beyond the current study's verification of the structural validity of the WISC-IV with a referred Native American sample, there are no empirical investigations of the content or predictive validity of the WISC-IV with Native American children. It is crucial that empirical research be used to bring insight to the cognitive assessment of Native American children, especially considering the frequent use of cognitive tests in placing these children into special education programs.

Additional information that explores the possible relationship between background experiences distinctive of some Native American children and their experience as learners is also needed to inform best practices with this population. Thus far, efforts to understand the respective roles of social, cultural, and linguistic factors in the normative development of Native American children have identified English language skills (Beiser & Gotowiec, 2000; Dauphinais & King, 1992; Tsethlikai, 2011), cultural practices (Dauphinais & King, 1992; Tsethlikai, 2011), and school readiness (Hibel et al., 2008) as factors contributing to the educational experiences of Native American children.

Conclusion

The Wechsler scales have been the most frequently employed intelligence measure among Native American students (McShane & Plas, 1982); yet, studies of Wechsler scales' psychometric characteristics with Native American populations have been rare. The current study is the sole investigation of the latest Wechsler child scale with a Native American sample. Although the current study offers evidence-based support for the structural validity of the WISC-IV with a referred Native American sample, this finding is but one strand of information pulled from the tapestry of unknowns that currently shroud the cognitive assessment practices of Native American children, leaving implications for best practices yet to be uncovered. Until additional research is available, these results support preliminary recommendations (Bodin et al., 2009; Watkins, 2006; Watkins et al., 2006) that practitioners using the WISC-IV with Native American students should not neglect the global ability factor.

References

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

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

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. Child Neuropsychology, 15, 417 – 424.

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), Structural modeling by example: Applications in educational, sociological, and behavioral research (pp. 160 – 188). New York, NY: Cambridge University Press.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York, NY: Guilford.

Byrne, B. M. (2012). Structural equation modeling with Mplus: Basic concepts, applications, and programming. New York, NY: Routledge.

Canivez, G.L., & Watkins, M.W. (1998). Long term stability of the WISC-III. Psychological Assessment, 10, 285 – 291.

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. Multivariate Behavioral Research, 30, 429 – 452.

Centers for Disease Control and Prevention. (2007, May 29). Injuries among Native Americans: Fact sheet. Retrieved from http://www.cdc.gov/ncipc/factsheets/nativeamericans.html

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling, 14, 464 – 504.

Chen, H., & Zhu, J. (2008). Factor invariance between genders of the Wechsler Intelligence Scales for Children – Fourth Edition. Personality and Individual Difference, 45, 260 – 266.

Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. Personality and Individual Differences, 52, 161 – 166.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233 – 255.

Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. Journal of Child and Family Studies, 9, 135 – 156.

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

Dent, H. E. (1996). Non-biased assessment or realistic assessment? In R. L. Jones (Ed.), Handbook of tests and measurement of Black populations, (Vol. 1, pp. 103 – 122). Hampton, VA: Cobb & Henry.

Dolan, B. (1999). From the field: Cognitive profiles of First Nations and Caucasian children referred for psychoeducational assessment. Canadian Journal of School Psychology, 15, 63 – 71.

Ducheneaux, T. W. (2002). WISC-III performance patterning differences between Native American and Caucasian children (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3088043)

Flanagan, D. P., & Genshaft, J. L. (1997). Issues in the use and interpretation of intelligence tests in the schools: Guest editors' comments. School Psychology Review, 26, 146 – 149.

Gignac, G. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? Psychology Science Quarterly, 50, 21 – 43.

Gould, S. J. (1995). Curveball. In S. Fraser (Ed.), The bell curve wars: Race, intelligence, and the future of America (pp. 11 – 22). New York, NY: BasicBooks.

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.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? American Psychologist, 47, 1083 – 1101.

Hibel, J., Faircloth, S. C., & Farkas, G. (2008). Unpacking the placement of American Indian and Alaska Native students in special education programs and services in the early grades: School readiness as a predictive variable. Harvard Educational Review, 78, 498 – 528.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. Psychological Methods, 3, 424 – 453.

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.

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

Jensen, A. R. (1984). The Black-White differences on the K-ABC: Implications for future tests. Journal of Special Education, 18, 377 – 408.

Kaufman, A. S., & DiCuio, F. F. (1975). Separate factor analyses of the McCarthy scales for groups of Black and White children. Journal of School Psychology, 13, 10 – 18.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 581 – 614). New York, NY: Guilford.

Keith, T. Z., Goldenring, F., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? School Psychology Review, 35, 108 – 127.

Kush, J. C., & Watkins, M. W. (2007). Structural validity of the WISC-III for a national sample of Native American students. Canadian Journal of School Psychology, 22, 235 – 248.

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.

Mardia, K. V. (1970), Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519 – 530.

Marks, S.U., Lemley, C.K., & Wood, G.K. (2010). The persistent issue of disproportionality in special education and why it still hasn't gone away. PowerPlay: A Journal of Educational Justice, 1, 4 – 21.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural Equation Modeling, 11, 320 – 341.

McShane, D. A., & Plas, J. M. (1982). WISC-R factor structure of Ojibwa Indian children. White Cloud Journal, 2, 18 – 22.

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.

Muthén, L. K., & Muthén, B. O. (1998–2010). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

Raftery, A. E. (1995). Bayesian model selection in social research. Sociological Methodology, 25, 111 – 163.

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

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

Reynolds, C. R., Lowe, P. A., & Saenz, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (3rd ed., pp. 549 – 596). New York, NY: Wiley & Sons.

Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. International Journal of Methods in Psychiatric Research, 20, 235 – 249.

Sandoval, J. (1982). The WISC-R factorial validity for minority groups and Spearman's hypothesis. Journal of School Psychology, 20, 198 – 204.

Sattler, J. M., & Dumont, R. (2008). Wechsler Intelligence Scales for Children – Fourth Edition (WISC-IV): Description. In J. M. Sattler (Ed.), Assessment of children: Cognitive foundations (5th ed., pp. 265 – 315). San Diego, CA: Jerome M. Sattler, Publisher.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53 – 61.

Sparks, S. (2000). Classroom and curriculum accommodations for Native American students. Intervention in School & Clinic, 35, 259 – 264.

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

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Boston, MA: Allyn and Bacon.

Tanner-Halverson, P., Burden, T., & Sabers, D. (1993). WISC-III normative data for Tohono O'odham Native-American children. In B. A. Bracken, & R. S. McCallum (Eds.), Wechsler intelligence scale for children: Third edition. (pp. 125 – 133). Brandon, VT: Clinical Psychology Publishing.

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

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 261 – 284). Washington, DC: American Psychological Association.

Tsethlikai, M. (2011). An exploratory analysis of American Indian children's cultural engagement, fluid cognitive skills, and standardized verbal IQ scores. Developmental Psychology, 47, 192 – 202.

U. S. Census Bureau. (2000). American FactFinder fact sheet: Allegany County, N.Y. Retrieved from http://factfinder.census.gov.

Valencia, R. R., & Suzuki, L. A. (2001). Intelligence testing and minority students: Foundations, performance factors, and assessments issues. Thousand Oaks, CA: Sage Publications.

Vraniak, D. A. (1994). Native Americans. In R. J. Sternberg (Ed.), Encyclopedia of human intelligence. New York, NY: MacMillan Reference Books.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children-Fourth Edition. Psychological Assessment, 18, 123 – 125.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children – Fourth Edition among a national sample of referred students. Psychological Assessment, 22, 782 – 787.

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scales-Fourth Edition among referred students. Educational and Psychological Measurement, 66, 975 – 983.

Wechsler, D. (1974a). Wechsler Intelligence Scale for Children-Revised. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1974b). Wechsler Preschool and Primary Scale of Intelligence-Third Edition. San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (1993). Wechsler Individual Achievement Test. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale-Third Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003a). Wechsler Intelligence Scale for Children-Fourth Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003b). Wechsler Intelligence Scale for Children-Fourth Edition: Technical and interpretive manual. San Antonio, TX: The Psychological Corporation.

Wiseley, M. C. (2001). Non-verbal intelligence and Native-American Navajo children: A comparison between the CTONI and the WISC-III. Dissertation Abstract International: Section A. Humanities and Social Sciences, 62, 2030.

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.