

Factor Structure of the Wechsler Intelligence Scales for Children—Fourth Edition

Among Referred Native American Students

By

Selena Nakano

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Approved August 2011 by the  
Graduate Supervisory Committee:

Marley Watkins, Co-Chair  
Linda Caterino, Co-Chair  
Sylvia Cohen

ARIZONA STATE UNIVERSITY

December 2011

## ABSTRACT

The Native American population is severely underrepresented in empirical test validity research despite being overrepresented in special education programs and at an increased risk for special educational evaluation. This study is the first to investigate the structural validity of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) with a Native American sample. The structural validity of the WISC-IV was investigated using the core subtest scores of 176, six-to-sixteen-year-old Native American children referred for a psychoeducational evaluation. The exploratory factor analysis procedures reported in the WISC-IV technical manual were replicated with the current sample. Congruence coefficients were used to measure the similarity between the derived factor structure and the normative factor structure. The Schmid-Leiman orthogonalization procedure was used to study the role of the higher-order general ability factor. Results support the structural validity of the first-order and higher-order factors of the WISC-IV within this sample. The normative first-order factor structure was replicated in this sample, and the Schmid-Leiman procedure identified a higher-order general ability factor that accounted for the greatest amount of common variance (70%) and total variance (37%). The results support the structural validity of the WISC-IV within a referred Native American sample. The outcome also suggests that interpretation of the WISC-IV scores should focus on the global ability factor.

## ACKNOWLEDGEMENTS

This dissertation project started with the contagious passion Dr. James Cox shared about his work with Native American children in Arizona. I began eagerly searching for empirical studies with Native American children and found that Dr. Marley Watkins, the training director and my soon-to-be-advisor, had published one of the few empirical studies with this population. Being an advisee of Dr. Watkins has been the most challenging and fruitful roles I have taken on in my life as a student. His expectations for my work were always high and unwavering yet consistently matched by his expert ability to develop in me the skills to rise to meet his expectations. I thank Dr. Watkins. I benefitted from my committee members Dr. Linda Caterino, a faithful servant to students and defender of our program; and Dr. Cohen, a natural choice for a committee member given her reputation as a passionate and pioneering school psychologist. My colleagues Paula McCall encouraged me through a mock proposal presentation, Kerry Lawton peered over the "Green book" as I picked his statistically-saavy brain, and Tracy Strickland (who became Tracy Allen along the way) acted as my empathetic sounding-board and confidant while she endured her own long nights of "dissertating." My family kept their eyes fixed on the end of the project knowing it meant me going home to be with them. Monica & Denise stayed excited & optimistic for me, and Sheryl never let me forget why I needed to finish strong. My loving, patient grandfather made it impossible to give up. Grandpa, I just finished my last big paper. We did it!! Where to now, Lord? You lead.

## TABLE OF CONTENTS

	Page
LIST OF TABLES.....	v
CHAPTER	
1 Introduction.....	1
Mean Group Differences.....	3
Content Bias .....	4
Predictive Bias .....	5
Structural Bias.....	8
Structural Bias Research Among Asian American and Native American Children.....	27
Current Study.....	28
2 Methods.....	30
Participants.....	30
Measure.....	33
Procedure.....	35
Analyses.....	35
3 Results.....	39
Replicatory Factor Analysis.....	40
Congruence Coefficient of Obtained Factor Structure.....	41
Schmid-Leiman Orthogonalization Procedure.....	42

CHAPTER	Page
4 Discussion.....	44
Limitations.....	47
Future Research.....	49
Conclusion.....	50
REFERENCES.....	52
APPENDIX	
A Institutional Review Board Approval of Data Collection.....	64

## LIST OF TABLES

Table	Page
1. Changes in Subests Across WISC Revisions.....	10
2. Characteristics of Students and Academic Achievement in the Sampled School Districts.....	32
3. Mean and Standard Deviations on Wechsler Intelligence Scale for Children-Forth Edition (WISC-IV) Subtest, Factor, and IQ Scores of 176 Native American Students Tested for Special Education Eligibility.....	39
4. Structure of WISC-IV with Principal Axis Extraction and Promax Rotation of Four Factors Among 176 Native American Students Tested for Special Education Eligibility .....	41
5. Percent of Variance Accounted for in the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) for 176 Native American Students Referred for Special Education Testing According to an Orthogonalized Higher Order Factor Model.....	43

## **Introduction**

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 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). Students who are placed in special education generally have limited access to higher education and are, subsequently, less qualified for a variety of higher income jobs (Green, 2007; Hocutt, 1996). Given the life-altering decisions made with 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 (AERA, APA, & NCME, 1999; Reynolds, Lowe, & Saenz, 1999).

The American Psychological Association Committee on Psychological Tests first presented what now comprises the foundation of test validation (AERA, APA, & NCME, 1954). Specifically, a test must have evidence of content, predictive, and construct validity. Messick (1995) further expanded the concept of validity when he discussed validity, not as a property of a test, but rather as being based on an empirical evaluation of the meaning and consequences of the measurement.

Following Messick's conceptualization of validity, the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) specified that sources of validity evidence include evidence based on test content, response processes, internal structure, relations to other variables, and consequences of testing. In contrast, test bias can arise when "deficiencies in a test itself or the manner in which it is used result in different meanings for scores earned by members of different identifiable subgroups" (AERA, APA, & NCME, 1999, p. 74). Consequently, evidence of test bias may be "sought in the content of the tests, in comparisons of the internal structure of test responses for different groups, and in comparisons of relationships of test scores to other measures" (AERA, APA, & NCME, 1999, p. 77). 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, et al., 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). Decades of research on the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974), and Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) have been conducted (Sattler, 2008). Several recent empirical studies of the current Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003a) continue to contribute to this body of test bias research.

### **Mean Group Differences**

The popular perception of test bias has been almost isomorphic with differences in mean scores across diverse ethnocultural groups (Brown, Reynolds, & Whitaker, 1999). Mean score differences in performance across groups are well-documented and acknowledged by researchers (Jensen, 1980; Sackett, Borneman, & Connelly, 2008). For example, African Americans typically score around one standard deviation below Whites and Hispanics score about one-half standard deviation below Whites on IQ tests. Native Americans score approximately one standard deviation lower than Whites on verbal measures but

score equal to or higher than Whites on nonverbal or performance tests.

Conversely, Asians typically score slightly higher than Whites on IQ tests (Lynn, 2006; Mackintosh, 1998; McShane, 1980; McShane & Plas, 1984; Tanner-Halverson, Burden, & Sabers, 1993).

However, psychometricians ultimately rejected mean score difference as evidence of test bias on the grounds that there is no a priori evidence that there *should* or *should not* be differences between groups on intelligence measures (Jensen, 1980; Reynolds, Lowe, et al., 1999). This *egalitarian fallacy* regarding inter-group differences was described by Jensen (1980) as “the gratuitous assumption that all human populations are essential identical or equal in whatever trait or ability the tests purport to measure” (p. 370). Given the educational and social disparities among groups, Sattler (2008) claimed that “mean group differences are to be expected among groups that live in different environments” (p. 162). Additionally, the variability *within* groups is far greater than the variability *between* groups. For example, 30% of the total variance of IQ scores of a random sample of Black and White children was related to race and social class whereas 65% of the IQ score variance was due to family differences, completely unrelated to race and social class (Jensen, 1998).

### **Content Bias**

Beginning with the WISC, the earliest test bias studies focused on content validity (Reynolds, Lowe, et al., 1999). The first federal court decisions regarding test bias (*Diana v. Board of Education*, 1970; *Guadalupe v. Tempe*

*Elementary School District*, 1978; *Larry P. v. Riles*, 1984; *Marshall v. Georgia*, 1984) were based on claims of differential content validity. In these cases, notions of face validity were used to make subjective evaluations of the presence of bias in the test items. For example, the judge in the *Larry P. v. Riles* case personally reviewed intelligence test items and found them biased while another federal judge in the *Marshall v. Georgia* case read the same items and concluded that they were unbiased (Kaufman, 1990; Sandoval, 1982). The first empirical study of item bias was conducted by Sandoval (1979), who used items from the WISC-R with samples of White, African American, and Mexican American children. Sandoval found that, although the three samples of children performed differently on the test, the reliabilities across the three groups were large and generally consistent. Thus, there was negligible evidence of item bias on the WISC-R across the three minority groups. A host of studies on the content validity of later versions of the WISC have been consistent with Sandoval's (1979) conclusion that item difficulties generally did not differ between majority and minority samples (Hunter & Schmidt, 2000; Reynolds & Ramsay, 2003; Ross-Reynolds & Reschly, 1983). Given this evidence, it appears that WISC items are not biased against minority students. This supports the content validity of the Wechsler scales for use across culturally diverse samples.

### **Predictive Bias**

A test's predictive validity lies in its ability to predict a specific outcome or behavior (Reynolds, Lowe, et al., 1999). In psychoeducational assessment, IQ

scores are often used to predict an individual's academic achievement scores. Consequently, IQ and achievement tests are often used in conjunction to make decisions regarding special education eligibility. Prediction of academic achievement scores using an IQ score is possible because IQ and achievement test scores are moderately correlated. Predictive validity is frequently quantified using correlation coefficients (Cleary, Humphreys, Kendrick, & Wesman, 1975; Sattler, 2008) and regression models are often used to identify bias in test use across groups (Cleary et al., 1975). When a test predicts performance on a related measure as a function of group membership, the test is considered to have predictive bias.

In cross-cultural predictive validity studies a common regression line is computed and comparisons of the slopes and intercepts are made across groups (Cleary et al., 1975). In an early study comparing the WISC and WISC-R regression lines of Black and White students, Reynolds and Hartlage (1979) found results that supported the use of a common regression line when predicting achievement scores. In contrast, use of the common regression line for prediction across a diverse sample was challenged by Reschly and Sabers (1979) in their study of the WISC-R regression lines of four ethnic groups: White, African American, Mexican American, and Native American Papago. They found significant differences in both the slopes and intercepts across groups and concluded that predictive bias resulted in the *overprediction* of academic

achievement scores for the ethnic minorities in the sample, and *underprediction* for the Whites in the sample.

Subsequent WISC-III predictive bias research studies have also found a pattern of overprediction of academic achievement scores for minority groups (Glutting, Oh, Ward, & Ward, 2000; Weiss & Prifitera, 1995). For instance, Weiss and Prifitera (1995) studied the WISC-III for evidence of predictive bias using those White, African American, and Hispanic children from the Wechsler Individual Achievement Test (WIAT; Wechsler, 1993) standardization sample who had WISC-III scores. Academic achievement scores were predicted and the slopes and intercepts of the observed regression lines for each ethnic group were compared. In this study, the WISC-III overestimated the reading abilities of Hispanic children by 2.0 points. Another study of the WISC-III and WIAT with a sample of White and African American students also found that when a combined regression line was applied to the distribution of WISC-III Verbal IQ scores and predicted WIAT Reading scores, African American students' reading scores were overestimated and White students' reading scores were underestimated (Glutting et al., 2000). These results converge on the general findings of predictive validity research with IQ tests among racially diverse samples in finding that when the IQ assessments exhibit biased predictive validity, the bias favors ethnic minorities by way of overestimation of academic achievement (Glutting et al, 2000; Reschly & Reschly, 1979; Reschly & Sabers, 1979; Reynolds & Gutkin, 1980; Reynolds & Hartlage, 1979; Saccuzzo & Johnson, 1995; Weiss & Prifitera, 1995).

## **Structural Bias**

Construct validity is concerned with whether the test is measuring the latent construct that it intends to measure. Central to construct validity is structural validity. 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). A meta-analysis of test bias research found that studies investigating structural validity were conducted more frequently than studies of content and predictive validity (Valencia & Suzuki, 2001).

Factor analysis is a primary method of investigating the internal structure of a measure (Carroll, 1966). Exploratory (EFA) and confirmatory (CFA) factor analytic studies 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.

Historically, there has been much contention over the empirical factor structure of the Wechsler intelligence scales. The addition and removal of

subtests across the WISC revisions often led to changes in the normative factor structure from one version of the WISC to the next. Table 1 depicts the changes in subtests across the four versions of the WISC. Compounding this was the critique that the Wechsler intelligence scales were not based on a theory of intelligence (Jensen, 1987; Witt & Gresham, 1985), but rather Wechsler's conception that "intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (1944, p. 3). The lack of a clear theoretical framework to drive test construction contributed to the debates over the clinical utility and structural validity of the WISC in measuring intelligence (Bodin, Pardini, Burns, & Stevens, 2009; Kamphaus, Benson, Hutchinson, & Platt, 1994; Keith, Goldenring, Taub, Reynolds, & Kranzler, 2006).

Table 1

*Changes in Subtests Across WISC Revisions*

Subtest	WISC	WISC-R	WISC-III	WISC-IV
Information	New <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>s</sup>
Comprehension	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>
Similarities	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>
Vocabulary	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>
Arithmetic	New <sup>c</sup>	Unchanged <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>s</sup>
Digit Span	New <sup>c</sup>	Unchanged <sup>s</sup>	Unchanged <sup>s</sup>	Revised <sup>c</sup>
Picture Completion	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>s</sup>
Picture Arrangement	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	n/a
Block Design	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>
Object Assembly	New <sup>c</sup>	Unchanged <sup>c</sup>	Revised <sup>c</sup>	n/a
Coding	New <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>	Revised <sup>c</sup>
Mazes	New <sup>s</sup>	Revised <sup>s</sup>	Unchanged <sup>s</sup>	n/a
Symbol Search	n/a	n/a	New <sup>c</sup>	Revised <sup>c</sup>
Word Reasoning	n/a	n/a	n/a	New <sup>s</sup>
Picture Concepts	n/a	n/a	n/a	New <sup>c</sup>
Matrix Reasoning	n/a	n/a	n/a	New <sup>c</sup>
Letter-Number	n/a	n/a	n/a	New <sup>c</sup>
Sequencing				
Cancellation	n/a	n/a	n/a	New <sup>s</sup>

*Note.* WISC is the Wechsler Intelligence Scale for Children, WISC-R is the Wechsler Intelligence Scale for Children-Revised, WISC-III is the Wechsler Intelligence Scale for Children-Third Edition, and WISC-IV is the Wechsler Intelligence Scale for Children-Fourth Edition. New = new subtest in this version of the WISC. Revised = subtest was revised in this version of the WISC.

Unchanged = subtest was not changed from previous version of the WISC.

<sup>c</sup>Subtest in core battery. <sup>s</sup>Subtest in supplemental battery. n/a = no applicable change.



*WISC*. The structure of the original *WISC* included two factors: Verbal Comprehension (VC) and Perceptual Organization (PO). The verbal factor included the Information (IN), Similarities (SI), Arithmetic (A), Vocabulary (V), Comprehension (CO), and Digit Span (DS) subtests. The Perceptual Organization factor included the Picture Completion (PC), Picture Arrangement (PA), Block Design (BD), Object Assembly (OA), and Coding (CD) subtests. Structural validity studies were uncommon during the first half of the twentieth century, so, not surprisingly, when Silverstein (1973) searched over 800 references on the original *WISC* in the *Mental Measurement Yearbook* contemporary to that time he found only two that had investigated the structural validity of the *WISC*.

Silverstein (1973) echoed the argument for the importance of evidence of structural validity in saying, “Yet before comparing the scores of various groups, it seems critical to demonstrate that the test is measuring the same abilities in each group to preclude a comparison of ‘apples and oranges’” (p. 408). Using a random sample of 1,310 White, African American, and Mexican American public school children, Silverstein (1973) found the normative VC-PO two-factor structure was consistent across the three ethnic groups. Silverstein concluded that although his results verified that the *WISC* measured the same constructs across the three groups included in the sample, his results *did not* say it is “fair or proper to use the *WISC* indiscriminately with other ethnic groups since it was standardized and normed using ONLY whites” (p. 410).

*WISC-R*. The next development in the Wechsler scales came in 1974 with the *WISC-R*. A surge of studies investigating statistical bias in the *WISC-R* followed the numerous civil rights cases challenging the use of IQ tests with racially diverse populations (*Diana v. Board of Education*, 1970; *Guadalupe v. Tempe Elementary School District*, 1984; *Larry P. v. Riles*; *Marshall v. Georgia*, 1984). As a result, The *WISC-R* factor structure was studied across several independent samples of clinical and culturally diverse groups.

The *WISC-R* retained the VC-PO factor structure and subtests of the original *WISC* (see Table 1 for details). Factor analytic studies across diverse groups detected a third, smaller factor that Kaufman (1975) identified as the Freedom from Distractibility (FD) factor originally discussed by Cohen (1952a, 1952b). The new FD factor included the Arithmetic, Digit Span, and Coding subtests. Dean (1980) studied the structure of the *WISC-R* with White and Mexican American children and found a three-factor structure for both groups of children, whereas Gutkin and Reynolds (1980) and Sandoval (1982) found only the VC-PO two-factor structure with their samples of African American, Mexican American, and White children.

Studies that included Native American children also resulted in factor configurations that varied from the normative structure. For example, a sample of 1,040 randomly selected Native American Papago, African American, Mexican American, and White children were included in a structural validity study of the *WISC-R* conducted by Reschly (1978). Whereas a two-factor structure

consistently emerged across all groups, a three-factor structure emerged only for the White and Mexican American children. When the third factor did emerge for the Black and Native American children, the mix of subtests that represented the factor made it uninterpretable. Overall, the two-factor solution for all groups of children mimicked the VC and PO factors of the standardization sample, and when the third factor emerged for the White and Mexican American children it mimicked the FD factor found by Kaufman (1975). Similarly, in another study including 192 Navajo children and 50 Papago children with learning disabilities, Zarske, Moore, and Peterson (1981) found that the normative Verbal Comprehension and Perceptual Organization factors emerged across each subsample of children, but the FD factor failed to emerge for either group.

In a smaller scale study, McShane and Plas (1982) used a sample of 77 Ojibwa students who attended a reservation school. The researchers utilized empirical methods consistent with the majority of the WISC-R factor analytic studies at that time; however, their results varied significantly from other studies. Three factors emerged that were quite discrepant from those found in previous samples. For example, the Information, Similarities, and Vocabulary subtests loaded on both factors 1 and 2, whereas Comprehension loaded uniquely on factor 1. Factor 2 also included Arithmetic, Digit Span, and Block Design. Factor 3 had loadings between .50 to .73 from Mazes, Block Design, and Object Assembly. The factor structure found in this small sample was significantly different from

that of the normative sample, and from other studies using diverse samples including Reschly's (1978) study with a Papago sample.

A possible explanation for the unusual factor structure found by McShane and Plas (1982) among Ojibwa students could be attributed to the small size of the sample. Decades of research (Comrey & Lee, 1992; Guadagnoli & Velicer, 1988; MacCallum, Widaman, Zhang, & Hong, 1999; Stevens, 1996; Tabachnick & Fidell, 2001) have been devoted to determination of the optimal sample size for factor analytic studies. The results have been inconsistent, with recommendations ranging from a 2:1 participant to variable ratio (Stevens, 1996) to guidelines recommending criteria for the desirability of total sample size (i.e., 100 is a poor sample size, 300 is good, and 1,000 is excellent; Comrey & Lee, 1992). What is generally agreed upon, however, is that small sample sizes can negatively affect the factor analysis by making the factor solution unstable (Guadagnoli & Velicer, 1988).

Additionally, McShane & Plas (1982) used a Varimax rotation, forcing the variables to be orthogonal. Restricting variables that are known to be correlated, such as cognitive ability tests, can misrepresent the structure of the data and lead to unexpected results (Gorsuch, 2003). In addition to the small sample size used in McShane and Plas (1982) study, the use of Varimax rotation with the Ojibwa data could have also accounted for the odd factor structure. It is plausible that the structural bias found in the McShane and Plas (1982) study was the result of methodological flaws as opposed to the validity of the test.

With the exception of McShane and Plas (1982), cross-cultural research of the WISC-R generally converged on the interpretation of the non-biased structural validity of the verbal and performance WISC-R factors across various ethnic groups, whereas the Freedom From Distractibility factor was found to vary across groups. Unfortunately, there were only two empirical structural validity studies of the WISC-R with a Native American sample (McShane & Plas, 1982; Reschly, 1978).

*WISC-III.* In a departure from the traditional VC-PO two-factor structure of previous versions of the WISC, the reported factor structure for the WISC-III normative sample included a second-order general ability, or *g*, factor and four, first-order factors named Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and the newly created Processing Speed. The updated VC factor included the WISC and WISC-R Information, Comprehension, Similarities, and Vocabulary subtests. All but Coding were retained for the PO factor and the updated FD factor now included Arithmetic and Digit Span subtests. The new Processing Speed factor was comprised of the retained Coding subtest and the one new Symbol Search subtest. See Table 1 for additional details.

A host of studies investigated the factor structure of the WISC-III in clinical and non-clinical samples across cultures. In a large scale study, Georgas, Van de Vijver, Weiss, & Saklofske (2003) tested the internal structure of the WISC-III normative samples from twelve countries. The four first-order factors

reported for the American WISC-III standardization sample was also generally observed in these 12 countries. An exception was found with the FD factor, which was unstable in some countries due to the Arithmetic subtest that loaded on the VC factor rather than the FD factor. Missing from Georgas et al. (2003) was a discussion of the second-order general ability factor in accounting for the constructs being measured by the WISC-III. Nevertheless, the main finding supported the normative first-order factor structure consisting of four oblique factors.

Similar results were found by Reynolds and Ford (1994) who compared the 3-factor solution in the WISC-R and WISC-III. Using the WISC-III standardization sample, they found that the 3-factor structure (VC, PO, and FD) was most interpretable when the Symbol Search subtest was not used. Reynolds and Ford (1994) cautioned against use of Symbol Search, but said that the four-factor structure included in the WISC-III technical manual (Wechsler, 1991) was appropriate if Symbol Search was used.

Further research with special education samples continued to yield structures with abnormalities in the FD and PS factors. In a sample of 1,201 students with learning disabilities, Watkins and Kush (2002) tested twelve possible structures based on previous empirical research and a priori theories of intelligence to identify which latent constructs were being measured by the WISC-III. Their findings concurred with previous research (Kush et al., 2001) that found the WISC-III to primarily be a measure of the general ability factor that

accounted for 37.8% of the total score variance compared to the low percentage of variance accounted for by the four first-order factors: VC, 3.7%; PO, 3.0%; FD, 0.5%; and PS, 1.5%. Given the small amount of variance accounted for by the FD and PS factors, Watkins and Kush (2002) suggested that these two factors be used with extreme caution, “if at all” (p. 16).

Other studies with racially diverse samples also cautioned against use of the FD and PS indices (Kush et al., 2001; Logerquist-Hansen & Barona, 1994). In a study comparing the factor structure for White and African American subsamples of children of the WISC-III standardization sample and a sample of Black students referred for a psychoeducational evaluation, Kush et al. (2001) conducted an exploratory and confirmatory factor analysis to investigate the factor structure across these groups. The exploratory factor analysis indicated that the Verbal Comprehension, Perceptual Organization, and Processing Speed factors best accounted for the pattern of correlations found between the WISC-III subtests. Inconsistencies arose with the two FD subtests Arithmetic and Digit Span. Arithmetic consistently loaded on the Verbal Comprehension factor and Digit Span failed to load substantially on any one factor. Additionally, the PS factor only accounted for 6-8% of total variance.

Results of the confirmatory factory analysis found the normative four-factor structure for the African American and White standardization subsamples, but results for the referred sample were unclear (Kush et al., 2001). With this referred subsample, both a three and four-factor structure emerged. In the four-

factor model, anomalies in the FD and PS factors were found. FD only accounted for 10% of the variance in the Digit Span subtest and the PS factor accounted for 25% of the variance in the Coding subtest. Although Kush et al.'s (2001) study supported the normative factor structure across the African American and White groups, the authors stated that the "lack of theoretical support, weak factorial invariance, inadequate long-term stability, and trivial incremental validity" (p. 83) of the FD and PS factors prompted them to discourage clinical interpretation of the WISC-III beyond the verbal and performance indices and the Full Scale Intelligence Quotient (FSIQ).

Not all studies with ethnically diverse samples found abnormalities in the WISC-III factor structure. For example, Wiseley (2001) used principal components analysis (PCA) to study the structural validity of the Verbal Comprehension and Perceptual Organization factors of WISC-III with a small sample ( $n = 50$ ) of Navajo children with a learning disability. The expected subtests had high loadings on each of the factors. Wiseley also calculated coefficients of congruence ( $r_c$ ) to measure the agreement of the derived VC and PO factors with the normative VC and PO factors. The coefficients of congruence indicated similarity between the derived and normative VC ( $r_c = .96$ ) and PO ( $r_c = .87$ ) factors. The overall results of this PCA supported the normative structure for the VC and PO factors in the sample of Navajo children.

Likewise, the WISC-III normative structure was supported in a referred Native American sample of 344 students (Kush & Watkins, 2007). The authors



used CFA to test factor models ranging from one to five factors. The root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and change in chi-square were reported for each of the five models tested. The normative four-factor oblique model had the best fit to the data, yielding good fit statistics (RMSEA = .058, SRMR = .041, CFI = .954) and a significant chi-square difference test that indicated its superiority over the other models. The results of their confirmatory factor analysis supported the normative four-factor structure with a sample of Native American students, demonstrating a lack of cross-cultural structural bias of the WISC-III.

In summary, factor analytic studies with the WISC-III standardization sample and independent samples demonstrated the general utility of the Verbal Comprehension and Perceptual Organization factors of the WISC-III as non-biased and valid measures of intelligence across diverse samples. In contrast, the FD and PS factors were unstable across many of the WISC-III structural validity studies regardless of eligibility status or racial background of the sample. Structural bias studies of the WISC-III failed to support the validity of four-factor normative structure across all groups.

*WISC-IV.* The WISC-IV 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 ( $n = 335$ ); Asian American ( $n = 92$ ); and a group of children described as “Other” ( $n = 28$ ). The “Other” category included Native American Indians, Alaskan Natives, and Pacific Islanders (Weiss, Saklofske, Prifitera, & Holdnack, 2006). The “Other” category comprised 1.2% of the total standardization sample, while the Native American and Alaska Native population comprised 1.5% of the total U.S. population during that time (U.S. Census, 2000).

Differing from previous Wechsler intelligence scales, the WISC-IV was developed to align with a current theory of intelligence (Wechsler, 2003b). Specifically, the Cattell-Horn-Carroll *Gf-Gc* (CHC) theory of intelligence (McGrew & Flanagan, 1998) was claimed as the theoretical model that guided the development of the WISC-IV. The current Cattell-Horn-Carroll theory is an integration of Carroll’s (1997) three-stratum theory and Horn-Cattell’s *Gf-Gc* theory of intelligence (Cattell, 1941, 1957; Horn, 1965).

The CHC model posits a hierarchical structure with three strata. Stratum III consists of general intelligence, or *g*, as conceptualized by Spearman (1904). Strata II is comprised of the ten broad *Gf-Gc* cognitive abilities: Fluid Intelligence (*Gf*), Crystallized Intelligence (*Gc*), Quantitative Knowledge (*Gq*), Reading and Writing (*Grw*), Visual Processing (*Gv*), Auditory Processing (*Ga*), Short-term Memory (*Gsm*), Long-term Storage and Retrieval (*Glr*), Processing Speed (*Gs*), and Decision/Reaction Time/Speed (*Gt*). Strata I includes around 70 narrow

abilities that are encompassed by the broad abilities. Based on this theory, five new subtests were included in the WISC-IV, including three for measurement of the broad ability *Gf* (i.e., Matrix Reasoning, Picture Concepts, and Word Reasoning); one to measure the short-term memory broad ability *Gsm* (i.e., Letter-Number Sequencing), and one for measurement of the broad ability *Gs* or processing speed (i.e., Cancellation).

With the addition of these new subtests and the retention and deletion of several WISC-III subtests, the WISC-IV factor structure continued to include four first-order factors and a higher-order general ability factor. The Verbal Comprehension factor was renamed the Verbal Comprehension Index (VCI); retained four core verbal subtests from the WISC-III, and made the Information subtest and the new Word Reasoning subtest supplemental. The Perceptual Organization factor of prior WISC versions was updated and renamed the Perceptual Reasoning Index (PRI). The updated PRI eliminated the Picture Arrangement and Object Assembly subtests from the WISC-III, included the new Picture Concepts and Matrix Reasoning subtests, retained Block Design and made Picture Completion a supplemental subtest. The Freedom From Distractibility factor of the previous WISC versions was updated and renamed the Working Memory Index (WMI) and included the new Letter-Number Sequencing subtest along with the retained Digit Span subtest. Arithmetic was retained as a supplemental subtest for this factor. The Processing

Speed factor was updated and retained with the addition of the new supplemental Cancellation subtest.

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. For higher order analysis of cognitive measures, Carroll (1993, 1995) recommended use of the orthogonalization procedure originally described by Schmid and Leiman (1957). The Schmid-Leiman procedure explicates “the independent influence of the first-order and higher-order factors on a set of primary variables” (Wolff & Preising, 2005, p. 48) by transforming the first-order factors to be orthogonal to the second-order factors and extracting the variance accounted for at each factor level. In this procedure, the “variance accounted for by the higher-order factor is extracted first. The first-order factors are then residualized of all variance present in the second-order factors” (McClain, 1996, p. 10). Use of the Schmid-Leiman (1957) procedure provides information about the direct relationships between the WISC-IV subtest variables and the higher-order general ability factor, thereby demonstrating the role of the WISC-IV subtests and factors in measuring children’s intelligence.

The limitations of the data reported in the WISC-IV technical manual (Wechsler, 2003b) prompted independent researchers to investigate the higher-

order factor structure and 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 replicatory exploratory factor analysis and the Schmid-Leiman (1957) orthogonalization procedure. Using the WISC-IV standardization sample, Watkins found that the general ability factor accounted for the majority of variance at the subtest level compared to the four, first-order oblique factors. For example, among the other first-order factors, the VC factor accounted for the highest percentage of total variance, 6.5%, and common variance, 12.1%, whereas the general ability factor accounted for 71.3% of the common variance and 38.3% of the total variance. Given the low percentage of variance accounted for by the first-order factors, Watkins (2006) recommended WISC-IV interpretation be limited to the FSIQ.

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 confirmatory factor analysis to test four levels of measurement invariance of the WISC-IV in the male and female participants of the standardization sample. The four levels of invariance tested were: configural invariance, factorial (metric) invariance, unique variance invariance, and factor covariance invariance. Successive tests of each type of invariance imposed more constraints than the previous one, as suggested by Meredith (1993). Fit statistics between models were compared and suggested that each model fit equally well across the male and female subsamples. The authors concluded 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 et al., 2006; Sattler & Dumont, 2008; Wechsler, 2003b).

Keith and colleagues (2006) also identified the alignment of the current WISC-IV scoring structure with the CHC theory of broad cognitive abilities. Results of their multisample confirmatory factor analysis across the 11 age-groups supported the normative four-factor structure; however, Keith et al. claimed that the CHC-derived model was a better fit to the WISC-IV data than the normative structure. When the CHC theory was applied, the WISC-IV appeared to measure five factors: crystallized ability (*Gc*), visual processing (*Gv*), fluid reasoning (*Gf*), short-term memory (*Gsm*), and processing speed (*Gs*). Keith and colleagues asserted that the results of their higher-order CFA revealed that a “CHC-derived theoretical structure describes the abilities underlying the WISC-IV better than the VCI, PRI, WMI and PSI scores that define the actual organization of the scale” (p. 122). However, a comparison of the fit statistics for each model shows relatively comparable fits. For example, the normative WISC-IV model had the following fit statistics: RMSEA = .041; SRMR = .038; and CFI = .979, compared to the CHC-derived model: RMSEA = .040; SRMR = .037, and CFI = .981. RMSEA values below .06, SRMR values below .08, and CFI values near 1.0 indicate that both models were good fits. Additionally, in the CHC-derived model the *Gf* factor loaded on the second-order *g* factor at unity level, suggesting that *Gf*

did not account for any unique variance in the constituent subtests. The CHC model also exhibited *g* loadings near unity for the first-order factors Working Memory (.94) and Perceptual Reasoning (.91), indicating that these factors, like the *Gf* factor, were not accounting for much unique variance. Given the comparable fit statistics of the normative model and CHC-derived models for the WISC-IV, the most parsimonious of the competing models was a model that included the second-order general intelligence factor and four, first-order factors as described by Wechsler (2003b).

All 15 core and supplemental subtests were included in the Keith et al. (2006) study; however, only the 10 core subtests are needed to derive a FSIQ. Using the 10 core subtests of the WISC-IV, Watkins, Wilson, Kotz, Carbone, and Babula (2006) were the first to evaluate the normative factor structure in a clinical sample. Sixty-five percent of this referred sample was ultimately identified as eligible for special education services under the categories of learning disability (37%), mental retardation (5%), emotional disability (7%), gifted (8%), speech disability (2%), and multiple disabilities (6%). With only the 10 core subtests, the first-order four-factor oblique structure reported for the normative sample was replicated. When the first-order oblique structure was transformed into an orthogonal higher-order model using the Schmid-Leiman (1957) procedure, the general factor accounted for the largest proportion of common (75.7%) and total variance (46.7%).

Structural bias of the ten core subtests was also studied by Bodin et al (2009). Bodin and colleagues applied confirmatory factor analysis to a predominantly male (60%) sample of 344 children with neurological disorders. The children ranged in age from 6 to 16 years and had diagnoses of attention deficit/hyperactivity disorder, epilepsy, learning disability, traumatic brain injury, cerebral palsy, meningitis/encephalitis, spina bifida, in-utero perinatal conditions, and other medical conditions. Four alternative, nested models were tested; each subsequent model included an additional factor, beginning with a one-factor baseline model. The model preferred by Bodin et al. (2009) revealed that the WM factor accounted for only 1% of the variance in the WM subtests, and loaded at near unity (.99) on the second-level *g* factor. PR also accounted for a low percentage of total variance (2.5%). These results were consistent with previous research (Keith et al., 2006; Watkins, 2006; Watkins et al., 2006) regarding the saliency of the higher-order *g* factor in accounting for variance in the first-order factors and individual subtests.

The higher-order structure of the WISC-IV factors has been empirically validated in factor analytic studies, in lieu of such an analysis in the WISC-IV technical manual (Wechsler, 2003b). The structural bias studies of the WISC-IV have consistently reported evidence supporting the normative factor across various samples of children. In summary, the WISC-IV has been shown to have comparable structural validity for clinical (Bodin et al., 2009; Watkins et al., 2006) and non-clinical groups (Chen & Zhu, 2008; Keith et al., 2006; Sattler &



Dumont, 2008; Watkins, 2006), to have similar age-raw score correlations across groups (Keith et al., 2006; Sattler & Dumont, 2008), and to have comparable structural validity across gender (Chen & Zhu, 2008).

### **Structural Bias Research Among Asian American and Native American Children**

Although structural bias has been empirically rejected in previous versions of the Wechsler scales, cross-cultural research has focused primarily on African American, Hispanic, and White groups. Of the groups represented in the 2000 Census, Asian American and Native American groups have been severely underrepresented in structural bias research. Historically, Asian American children have also been underrepresented in special education (Suzuki & Valencia, 1997). The research that has been published with Asian American groups has found that they perform as well as or better than their White counterparts on cognitive assessments (Jensen, 1980; Sue & Okazaki, 2009; Suzuki & Valencia, 1997). Structural bias research on the Wechsler scales has yet to be conducted with an Asian American sample.

Unlike the Asian American population, Native American students are *overrepresented* in special education nationally (Dauphinais & King, 1992; Hibell, Faircloth, & Farkas, 2008; Marks, Lemley, & Wood, 2010). This group also suffers from environmental deprivation (Vraniak, 1994), high rates of suicide (CDC, 2007), and high rates of school dropout (Sparks, 2000). Over the past five decades, 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 WISC-R found that the normative two-factor structure emerged with Native American samples (McShane & Plas, 1982; Reschly, 1978; Zarske et al., 1981). The remaining two studies investigated the WISC-III and both found the normative factor structure to be consistent with the respective Native American samples (Kush & Watkins, 2007; Wiseley, 2001). Unfortunately, the quality of the methodology and sample size varied widely in these studies and few studies included samples of adequate size (Kush & Watkins, 2007; Reschly, 1978; Zarske et al., 1981). Structural bias research on the WISC-IV has yet to be conducted with a Native American sample.

### **Current Study**

Test bias is a paramount concern for practitioners when considering the usefulness of tests across different groups of test-takers. Researchers have used both exploratory and confirmatory factor analysis to study psychometric test validity across various samples of children composing both non-clinical and clinical groups across a range of disabilities and cultural backgrounds. Lacking in the current body of structural validity research is an understanding of the structural composition of the WISC-IV with a Native American sample.

Accordingly, the current study will be a replicatory exploratory factor analysis (Ben-Porath, 1990) of the analysis reported in the WISC-IV technical manual (Wechsler, 2003b) but with a sample of referred Native American students attending public schools in the American Southwest. The current study will also

apply the Schmid-Leiman (1957) orthogonalization procedure to investigate the role of the higher-order general ability factor with this sample.

## Methods

### Participants

The sample included 176 Native American students (115 boys and 61 girls) 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 grades kindergarten through 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 a specific tribal affiliation. The special education eligibility classifications represented in the sample included: learning disability (LD, 77%), Cognitive Impairment (Mental Retardation) (CI, 5%), Other Health Impairment (OHI, 4.5%), emotional disturbance (ED, 4%), Autism (1%), and Traumatic Brain Injury (TBI, 1%). Hearing Impairment and Orthopedic Impairment were each reported once (0.6%), and 5% of the sample were not eligible for special education. Per the policies of the respective school districts, no further identifying data could be collected on the student sample. However, Table 2 contains information on achievement test scores as well as the ethnic background of student populations of the school districts included in the sample as reported by

the Arizona Department of Education (AZDE) and the National Center for Educational Statistics (NCES).

Table 2

*Characteristics of Students and Academic Achievement in the Sampled School Districts*

	District A	District B	District C	District D	District E	District F
Total Students	26,604	34,083	11,162	1,500	1,820	18,281
Students in Sample	11	22	99	5	15	24
Student Characteristics						
Students with IEP	11.1%	11.4%	16.2%	14.8%	10.0%	8.7%
English Language						
Learners (ELL)	6.6%	8.9%	12.6%	5.7%	45.8%	1.8%
Free/Reduced Lunch <sup>b</sup>	24.5%	32.6%	40.6%	61.3%	76.2%	23.8%
White	75%	69%	50%	63%	0%	61%
Hispanic	16%	23%	21%	21%	0%	17%
Asian	4%	4%	1%	0%	0%	9%
Black	3%	3%	2%	1%	0%	9%
Native American	2%	1%	25%	14%	99%	3%
Unspecified Ethnicity	0%	0%	1%	1%	1%	0%
Student Academic Performance						
AIMS <sup>a</sup>						
Elementary Math	85	81	74	55	68	87
Elementary Reading	84	79	69	56	72	87
High School Math	86	80	73	53	30	n/a
High School Reading	89	83	78	69	34	n/a
Terra Nova <sup>b</sup>						
Elementary Math	61	59	48	36	35	n/a
Elementary Reading	72	65	53	32	41	n/a
High School Math	70	66	56	44	37	n/a
High School Reading	68	62	52	50	36	n/a

*Note.* Data from the National Center for Educational Statistics and the Arizona Department of Education. AIMS = Arizona's Instrument to Measure Standards.

<sup>a</sup>Elementary performance expressed as percent passing scores attained by third grade students, and high school performance as percent passing scores attained by tenth grade students. <sup>b</sup>Elementary performance expressed in national percentile scores attained by third grade students and high school performance as national percentile scores attained by ninth grade students. n/a = scores not available.

## **Measure**

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. Each subtest contributes to one of the four cognitive domains: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Processing Speed Index (PSI), and Working Memory Index (WMI). 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 (Processing Speed) 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 to support the predictive validity of the WISC-IV. Reports of high correlation coefficients between the WISC-IV and other Wechsler intelligence scales support 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 (WPPSI-III; Wechsler, 1974) FSIQ ( $r = .89$ ), Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997) FSIQ ( $r = .89$ ), and the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) FSIQ-4 ( $r = .86$ ). The WISC-IV FSIQ was also correlated with the WIAT-II Total Achievement composite ( $r = .87$ ).

Additionally, the WISC-IV technical manual (Wechsler, 2003b) reported evidence of structural validity. Specifically, an exploratory factor analysis (EFA) indicated high factor loadings of the core subtests on the predicted factors. The technical manual also reported results of a confirmatory factor analysis (CFA) with the core subtests. The CFA tested factor models with one, two, three, and four factors and indicated the normative four-factor oblique structure was the best fit for the 10 core subtests. Independent research on the WISC-IV has reported the presence of a higher-order structure including four, first-order oblique factors and a second-order general ability “g” factor (Bodin et al., 2009; Keith et al., 2006; Watkins, 2006; Watkins et al., 2006).



## **Procedure**

The sample for this study was taken from a larger database of student psychoeducational data from three school districts in central Arizona and three school districts in northern Arizona (see Table 1). These school districts were asked to participate based on their proximity to the researcher and diversity of student population. Following IRB approval, a data collection team of five school psychology graduate students reviewed all psychoeducational files in these school districts and extracted relevant WISC-IV test scores. Criteria for inclusion in the sample included: (a) WISC-IV scores for the ten core subtests used to compute a FSIQ score; and (b) Native American racial background reported in the demographic data file. Of the 3, 297 student files reviewed, 176 students of Native American ancestry were identified and included in the current study.

## **Analyses**

Ordinarily, confirmatory factor analysis would be an appropriate method for testing the structural validity of a theoretically derived and empirically validated construct such as intelligence (Costello & Osborne, 2005; Fabrigar, Wegener, MacCallm, & Strahan, 1999). Alternatively, Carroll (1993, 1995) and others (Browne, 2001; Dolan, Oort, Stoel, & Wicherts, 2009; Goldberg & Velicer, 2006; Gorsuch, 2003) have supported the use of exploratory factor analysis for studying structural validity. EFA functions to describe the observed associations among variables in the underlying factor structure and is not restricted by a priori hypotheses of factor structure or other model specifications (Gorsuch, 2003);

whereas, CFA tests an a priori hypothesis about the underlying structure and is guided by an assumption that each variable is a pure measure of one factor only (Brown, 2006; Lee & Ashton, 2007; Sass & Schmitt, 2010). The requirement of zero cross-loading in CFA may lead to distorted factors and inflated factor intercorrelations if variables are not pure measures of each factor (Asparouhov & Muthen, 2009). Additionally, Goldberg and Velicer (2006) noted that “repeated discoveries of the same factor structure derived from exploratory techniques [across independent samples] provide stronger evidence for that structure than would be provided by the same number of confirmatory factor analyses” (p. 233). Replication of factor structure using multiple EFAs was also recommended by several other researchers (Dolan et al., 2009; Gorsuch, 2003; Lee & Ashton, 2007). Given these considerations and recommendations, exploratory factor analysis was applied in the current study.

Carroll (1993, 1995) also recommended the use of the Schmid-Leiman (Schmid & Leiman, 1957) orthogonalization procedure when investigating a higher-order factor structure. The Schmid-Leiman (1957) procedure transforms first-order factors to be orthogonal to second-order factors. This procedure partitions the variance of each individual subtest and extracts the variance accounted for by the higher-order factor (i.e., general ability or *g*) followed by variance attributable to the group factor (e.g., Verbal Comprehension). In the present analysis, applying the Schmid-Leiman (1957) procedure will provide

information regarding the proportion of WISC-IV subtest variance accounted for by the second-order factor (i.e., *g*) independent of the first-order factors.

The exploratory factor analysis procedures conducted in the current study were identical to those reported in the WISC-IV technical manual (Wechsler, 2003b) with the addition of the Schmid-Leiman procedure. Essentially, this analysis was considered a replicatory factor analysis (Ben-Porath, 1990), a cross-validation technique to investigate the cross-cultural validity of an instrument (Butcher, 1985; Geisinger, 2003). Ben-Porath (1990) specified methods for investigating an instrument's factor structure across groups using replicatory factor analysis. In this procedure,

a representative sample of the group with whom the instrument is to be adopted completes the assessment instrument; the data is then factor analyzed using the same EFA techniques for extraction, estimation of communalities, and rotation, as were used in the original development and validation of the instrument. In this new analysis, the number of factors extracted is constrained to the number of factors identified in the research with the instrument in its culture of origin (Allen & Walsh, 2000, p. 70).

Accordingly, the current study used the principal axis method for factor extraction with two iterations, constrained the number of factors to retain at four, and performed a Promax rotation, as was conducted with the standardization sample (Wechsler, 2003b). The factor solution derived from the principal axis factoring method was then orthogonalized using the Schmid-Leiman (1957) procedure.

In replication studies aimed at investigating the consistency of factor structures across independent samples, coefficients of factor similarity are used to compare a derived factor structure to a hypothesized factor structure (Davenport, 1997; Guadagnoli & Velicer, 1991; Reise, Waller, & Comrey, 2000). Calculation of a coefficient of congruence ( $r_c$ ) is one method of identifying consistency of factor structure (Dolan et al., 2009; Lee & Ashton, 2007; Lorenzo-Seva & ten Berge, 2006). Empirical studies aimed at identifying critical values for the coefficient of congruence have generally interpreted values of  $r_c > .95$  to indicate good factor similarity,  $r_c$  values between .85-.94 to indicate fair congruence, and values of  $r_c$  less than .85 to indicate the factor structure is not similar (Lorenzo-Seva & ten Berge, 2006). In the current study, a coefficient of congruence was calculated to measure similarity of the derived WISC-IV factor model of the referred Native American sample compared to the normative factor model reported in the technical manual (Wechsler, 2003b). The current study used the  $r_c$  critical values suggested by Lorenzo-Seva and ten Berge (2006) to determine factor congruence.

## Results

The WISC-IV subtest, factor, and IQ scores of the referred Native American sample are reported in Table 3. These results indicate participants' mean scores were slightly lower and somewhat less variable than the normative sample. Similar patterns of depressed scores have been found with other samples of referred students (Canivez & Watkins, 1998; Watkins et. al., 2006), including referred Native American students (Dolan, 1999; Ducheneaux, 2002; Kush & Watkins, 2007). Score distributions from the current sample appear to be relatively normal, with .66 the largest skew and 1.80 the largest kurtosis (Fabrigar et al., 1999).

Table 3

*Mean and Standard Deviations on Wechsler Intelligence Scale for Children-Forth Edition (WISC-IV) Subtest, Factor, and IQ Scores of 176 Native American Students Tested for Special Education Eligibility*

Variable	Mean	SD	Skewness	Kurtosis
BD	9.8	2.9	0.11	-0.06
SI	6.7	2.7	0.21	-0.62
DS	6.3	2.5	0.25	0.27
PCn	9.1	2.9	-0.32	0.31
CD	8.0	2.8	0.66	1.80
VC	5.7	2.3	-0.12	-0.49
LN	7.2	2.9	-0.38	-0.63
MR	8.7	2.8	0.11	-0.42
CO	6.6	3.0	-0.09	-0.45
SS	8.2	3.1	-0.14	0.01
VCI	78.7	13.8	-0.12	-0.47
PRI	95.2	14.2	-0.06	-0.22
WMI	80.1	13.0	-0.30	-0.61
PSI	89.3	14.1	0.26	0.06
FSIQ	82.4	12.9	-0.25	0.01

*Note.* BD =Block Design; SI =Similarities; DS =Digit Span; PCn =Picture Concepts; CD =Coding; VC= Vocabulary; LN =Letter-Number Sequencing; MR

=Matrix Reasoning; CO =Comprehension; SS =Symbol Search; VCI =Verbal Comprehension Index; PRI =Perceptual Reasoning Index; WMI = Working Memory Index; PSI =Processing Speed Index; FSIQ =Full-Scale IQ.

### **Replicatory Factor Analysis**

Results from Bartlett's Test of Sphericity (Bartlett, 1954) indicated that the correlation matrix was not random ( $\chi^2 = 635.43$ ;  $df = 45$ ;  $p < .001$ ). The Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) statistic was .86, which exceeds the minimum standard suggested by Tabachnick and Fidell (2007). These results suggested the matrix was appropriate for factor analysis.

Factor extraction and rotation methods for the current study included the principal axis method for factor extraction constrained to retain four factors after two iterations and an oblique Promax rotation, as was performed in the WISC-IV standardization sample (Wechsler, 2003a). Simple structure was evidenced by first-order pattern coefficients meeting at least the minimum criteria of .30 (Comrey & Lee, 1992). Pattern coefficients ranged from .49 to .84 with all 10 subtests loading on expected factors. Factor intercorrelations ranged from .48 between PSI and VCI to .65 between WMI and VCI (Table 4). The intercorrelation between factors suggested the presence of a second-order factor.

Table 4

*Structure of WISC-IV with Principal Axis Extraction and Promax Rotation of Four Factors Among 176 Native American Students Tested for Special Education Eligibility*

Subtest/Factor	Pattern Coefficients			
	VCI	PRI	WMI	PSI
BD	.13	<b>.61</b>	-.24	.15
SI	<b>.73</b>	.16	-.02	-.15
DS	.19	-.01	<b>.57</b>	.03
PCn	.04	<b>.58</b>	.12	-.06
CD	.08	-.02	.01	<b>.64</b>
VC	<b>.84</b>	-.03	.09	.03
LN	.10	.04	<b>.50</b>	.04
MR	-.10	<b>.61</b>	.24	.01
CO	<b>.80</b>	-.07	.06	.08
SS	-.12	.11	.06	<b>.60</b>
PRI	.51			
WMI	.65	.58		
PSI	.48	.61	.53	

*Note.* BD =Block Design; SI =Similarities; DS =Digit Span; PCn =Picture Concepts; CD = Coding; VC= Vocabulary; LN =Letter-Number Sequencing; MR =Matrix Reasoning; CO = Comprehension; SS =Symbol Search; VCI =Verbal Comprehension factor; PRI =Perceptual Reasoning factor; WMI = Working Memory factor; PSI =Processing Speed factor; FSIQ = Full Scale IQ. Salient pattern coefficients ( $\geq .30$ ) are indicated in bold.

### **Congruence Coefficient of Obtained Factor Structure**

To test the obtained Native American factor structure against the WISC-IV normative sample, the congruence coefficient ( $r_c$ ) was calculated for each factor. The resulting  $r_c$  values were .978, .983, .973, and .981 for the VCI, PRI, WMI, and PSI factors, respectively. Lorenzo-Seva and ten Berge (2006) suggested that values of  $r_c > .95$  indicate good factor similarity,  $r_c$  values between .85-.94 indicate fair congruence, and values of  $r_c$  less than .85 indicate the factor structure is not similar. Based upon these guidelines, the obtained congruence coefficients

indicate all four first-order factors in this sample have good factor similarity with the normative factor structure. Similar  $r_c$  levels were called “excellent” by MacCallum et al., (1999) and “practical identity of the factors” by Jensen (1998, p. 99).

### **Schmid-Leiman Orthogonalization Procedure**

The Schmid-Leiman (1957) transformation was used to decompose the variance of the first-order, four-factor oblique structure of the WISC-IV into several orthogonal components. The results presented in Table 5 indicate that the second-order general ability factor ( $g$ ) accounted for more variance in each of the 10 core WISC-IV subtests than any orthogonal first-order factor. The WISC-IV general factor accounted for between 21% and 55% of the variance in the core subtests. The VC factor accounted for an additional 18.7% to 24.7% of the variance in the three VC subtests. Beyond  $g$ , the PR factor accounted for an additional 7% of the variance in the three PR subtests, the WM factor contributed 8.3% to 11% of the variance in the two WM subtests, and the PS factor provided 18.2% to 20.8% of the variance of its two subtests. The first-order factors accounted for 4.6% (WM) to 13.1% (VC) of common variance and 2.4% (WM) to 6.9% (VC) of total variance. In contrast, the higher-order general ability factor accounted for approximately 70% of common variance and 37% of the total variance. An analysis without the 19 students who did not indicate English as a primary language produced almost identical results.



Table 5

*Percent of Variance Accounted for in the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) for 176 Native American Students Referred for Special Education Testing According to an Orthogonalized Higher Order Factor Model*

Subtest	General		VCI		PRI		WMI		PSI	
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var
BD	.564	31.8	.079	0.6	.280	7.8	-.136	1.8		1.1
SI	.613	37.6	.432	18.7	.075	0.6	-.013	00	-.104	1.1
DS	.628	39.4	.114	1.3	-.006	00	.331	11.0	.021	00
PCn	.605	36.6	.024	0.1	.265	7.0	.068	0.5	-.040	0.2
CD	.501	25.1	.045	0.2	-.010	00	.008	00	.456	20.8
VC	.744	55.4	.497	24.7	-.014	00	.050	0.3	.024	0.1
LN	.547	29.9	.062	0.4	.017	00	.288	8.3	.026	0.1
MR	.665	44.2	-.059	0.3	.281	7.9	.137	1.9	.008	00
CO	.677	45.8	.468	21.9	-.032	0.1	.033	0.1	.055	0.3
SS	.463	21.4	-.071	0.5	.049	0.2	.033	0.1	.427	18.2
% Total Variance		36.7		6.87		2.37		2.39		4.19
% Common Variance		69.9		13.1		4.51		4.55		7.97

*Note.* BD =Block Design; SI =Similarities; DS =Digit Span; PCn =Picture Concepts; CD = Coding; VC= Vocabulary; LN =Letter-Number Sequencing; MR =Matrix Reasoning; CO =Comprehension; SS = Symbol Search; *b* = loading of the subtest on the factor; Var = percent variance explained in the subtest.

## Discussion

The present study investigated the factor structure of the WISC-IV and found evidence to support its structural validity with a referred sample of 176 Native American children. Results of the first-order factor analysis revealed that the normative WISC-IV factor structure was replicated with a referred Native American sample. Obtained congruence coefficients supported the similarity of the factor structure of the current Native American sample to the normative factor structure. As expected, the Schmid-Leiman orthogonalization procedure identified a higher-order factor structure with a second-order, general ability factor. When used to measure global ability of a referred sample of Native American students, the WISC-IV appears to measure the same underlying constructs as those hypothesized by Wechsler (Wechsler, 2003b) and does not exhibit evidence of structural bias.

In the current Native American sample, the general ability factor accounted for approximately 70% of common variance and 37% of the total variance among the four, first-order factors. The present results are consistent with previous research with non-referred (Watkins, 2006) and referred (Bodin, et al., 2009; Watkins et al, 2006) non-Native American samples. In a higher-order factor analysis of the WISC-IV normative sample, Watkins (2006) found the general ability factor to account for approximately 71% of common variance and 38% of total variance in the four, first-order factors. Similarly, studies with referred samples also identified a second-order general ability factor as the primary source of variance in the first-order factors. In a sample of children

referred for neuropsychological testing, Bodin, et al. (2009) found the general ability factor to account for 77% of common variance and 48% of total variance. Additionally, Watkins et al. (2006) found *g* to account for approximately 75% of common variance and 46% of total variance in a sample of students referred for psychoeducational testing. These results were essentially replicated in another study with a referred sample by Watkins (2010) that found the higher order general ability factor to account for 75% of common variance and 48% of total variance in the four, first-order factors. The current study contributes to the existing research that has found the normative WISC-IV factor structure across groups and the primacy of a second-order general ability factor in accounting for variance in the four, first-order factors. These results support recommendations that practitioners using the WISC-IV should favor interpretation of the global ability factor over the four, first-order factors (Bodin et al., 2009; Watkins, 2006; Watkins et al., 2006).

The performance pattern of the current referred Native American sample is representative of patterns observed among samples of referred students and samples of Native American students. Overall WISC-IV performance that is below the normative mean is characteristic of referred samples (Canivez & Watkins, 1998; Watkins et. al., 2006), whereas verbal comprehension performance that is more than one standard deviation below the normative mean and accompanied with a significant discrepancy with the perceptual reasoning factor is a trademark of the Native American samples studied to date (Kush & Watkins, 2007; McCullough, Walker, & Diessner, 1985; McShane & Plas, 1984;

Tanner-Halvorsen et al., 1993; Wiseley, 2001). Research with Native Americans and the earlier versions of the WISC found a pattern of performance coined by McShane and Plas (1984) as the “Native American pattern” of cognitive performance (McCullough et al., 1985; McShane & Plas, 1984; Tanner-Halvorsen et al., 1993). This pattern is characterized by an eight to nineteen point discrepancy favoring the performance scales and “unique performance profiles characterized by low verbal subtest scores, selected elevated nonverbal subtest scores, [and] typical and large performance IQ-verbal IQ discrepancies” (McShane & Plas, 1984, p. 61). Recent studies of the WISC-III factor structure and Native American samples (Kush & Watkins, 2007; Wiseley, 2001) also reported evidence of this unique pattern.

The pattern of performance observed in the current sample of referred Native American children is congruous with previous research. 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, and overall ability that was approximately one standard deviation below the normative mean. Native American performance on the processing speed factor of previous Wechsler scales has not has been the focus of empirical investigation and the WISC-IV working memory factor was not represented on previous versions. The current sample presented with disparate performance in these areas as well. That is, processing speed ability .66 of a standard deviation below the normative mean and working memory performance 1.33 standard deviations

below the normative mean. Although an exploratory finding, evidence of this unique performance pattern in the current sample is noteworthy considering it is consistent with the extant research of Native American performance on the Wechsler scales. Further empirical investigation aimed at explaining this pattern is needed.

### **Limitations**

There are apparent limitations in the current study that can be improved upon in future replications. Foremost, the current sample is relatively small. However, the number of subjects was appropriate for the analyses used for investigation (MacCallum et al., 1999). The method of data collection was also a limiting factor. The data were collected from archival special education records and the accuracy of the professionals who initially gathered and recorded the data was assumed. The current sample consisted of exceptional students who were referred for a special education evaluation. It is uncertain if the present results generalize to samples of non-referred students.

The current sample was derived from a small sample of school districts in northern and central Arizona. Consequently, of the 21 federally recognized tribes that exist in Arizona (U.S. Census, 2000), Navajo was indicated in nearly all of those cases that included a specific tribal affiliation. Additionally, the living situations represented in the sample were not accounted for as potential sources of variance. Specifically, some children have lived primarily on a reservation, whereas others have lived primarily in rural or urban environments. Previous research 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), making the child's living situation a possible source of variance that should be accounted for. Given the limited range of sampling, 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.

Lastly, the sample included students with varying levels of English language proficiency. The collected data only indicated whether a second language was noted in the education records reviewed and does not identify the student's 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).

Empirical examination of the cognitive assessment of Native American children includes an exiguous body of research and a poorly understood profile of mental abilities. Consequently, when conducting intelligence testing to make special education eligibility decisions for Native American children, practitioners may select tests without empirical research to support their selection or their interpretation of results (Reschly & Grimes, 2004; Saxton, 2001). School psychologists are a primary consumer of the Wechsler scales and other standardized assessments and run the risk of making inaccurate interpretations and recommendations that change the course of students' educational programs and subsequent life opportunities. A salient example of misguided practices

adopted by some school psychologists is to not administer the verbal comprehension subtests of the Wechsler scales when testing Native American children, citing the depressed verbal comprehension subtest scores and discrepancies between the verbal comprehension and perceptual reasoning index scores as inaccurate representations of ability due to language and cultural backgrounds (Tanner-Halverson et al., 1993). Tanner-Halverson, et al. (1993) contended that “the cultural and linguistic experiences of most Native-American children...differ considerably from those of the middle-class, monolingual, English-speaking students upon whom most standardized intelligence tests were normed,” (p. 125) and while this assertion is echoed by others (Jensen, 1980; Naglieri, 1984) including the current researcher, it is still the case that Native American students are educated in and perform in a larger society wherein English is the dominant language, making a measure of verbal ability a rich source of information. Omitting the VCI prevents knowledge of a student’s global intelligence (*g*), which is ultimately the construct of interest. In lieu of best practice recommendations for how to approach assessment of Native American students, even culturally sensitive and well-intentioned school psychologists employ cognitive assessment practices that diverge considerably from the standardized administration procedures defined in the WISC-IV technical manual (Wechsler, 2003b).

### **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 finding 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, such efforts to understand the respective roles that social, cultural, and linguistic factors have on 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 used intelligence measure among Native American students (McShane & Plas, 1982) yet use of this instrument with the Native American population has been a neglected area of study in empirical research. The current study is the sole investigation of the latest Wechsler scale with a Native American sample. While 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 shrouds the cognitive assessment practices of Native American children, leaving implications for best practices still yet to be uncovered. It is the hope of this researcher that the information presented here will be considered by other practitioners who conduct cognitive assessments of Native American children, and that other interested professionals continue the empirical investigation of best practices related to the cognitive assessment of Native American children.

## References

- Allen, J. & Walsh, J. A. (2000). A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 63-86). Mahwah, NJ: Erlbaum.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, *51*(2), 1-38.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Asparouhov & Muthen, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397-438.
- Bartlett, M. S. (1954). A further note on the multiplying factors for various chi-square approximations in factor analysis. *Journal of the Royal Statistical Society, Series B*, *16*, 296-298
- Beiser, M., & Gotowiec, A. (2000). Accounting for Native/Non-Native differences in IQ scores. *Psychology in the Schools*, *37*, 237- 252.
- Ben-Porath, Y. (1990). Cross-cultural assessment of personality: The case for replicatory factor analysis. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 8, pp. 27-48). Hillsdale, NJ: Erlbaum.
- 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.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.
- Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quarterly*, *14*, 208-238.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111-150.

- Butcher, J. N. (1985). Current developments in MMPI use: An international perspective. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in Personality Assessment* (Vol. 4, pp. 83-94). Hillsdale, NJ: Erlbaum.
- Canivez, G.L., & Watkins, M.W. (1998). Long term stability of the WISC-III. *Psychological Assessment*, 10, 285–291.
- Carroll, J. B. (1966). Measurement without numbers. *PsycCRITIQUES*, 11, 118-119.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, 30, 429-452.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122-130). New York, NY: The Guilford Press.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York, NY: World Book.
- Centers for Disease Control and Prevention. (May 29, 2007). *Injuries among Native Americans: Fact Sheet*. Retrieved from <http://www.cdc.gov/ncipc/factsheets/nativeamericans.html>
- 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.
- Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, 30, 15-41.
- Cohen, J. A. (1952a). A factor-analytically based rationale for the Wechsler-Bellevue. *Journal of Consulting Psychology*, 16, 272-277.

- Cohen, J. (1952b) Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *Journal of Abnormal and Social Psychology*, 47, 359-365.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7). Available online <http://pareonline.net/>
- 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.
- Davenport, E. C. (1990). Significance testing of congruence coefficients: A good idea? *Educational and Psychological Measurement*, 50, 289-296.
- Dean, R. S. (1980). Factor structure of the WISC-R with Anglos and Mexican-Americans. *Journal of School Psychology*, 18, 234-239.
- 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.
- Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling*, 16, 295-314.
- Ducheneaux, T. W. (2002). *WISC-III performance patterning differences between Native American and Caucasian children*. (Doctoral dissertation, The University of North Dakota). Available from ProQuest Dissertations and Theses database. (UMI No. 3088043).
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.

- 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.
- Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment of psychology* (Vol. 10, pp. 95-117). Hoboken, NJ: Wiley.
- Georgas, J., Van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III*. San Diego, CA: Academic Press.
- Glutting, J. J., Oh, H. J., Ward, T., & Ward, S. (2000). Possible criterion-related bias of the WISC-III with a referral sample. *Journal of Psychoeducational Assessment*, 18, 17-26.
- Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S. Strack (Ed.) *Differentiating normal and abnormal personality* (2<sup>nd</sup> ed., pp. 209-237). New York, NY: Springer.
- Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka, & W. F. Velicer (Eds.) *Handbook of psychology: Research methods in psychology* (Vol. 2., pp. 143-164). Hoboken, NJ: John Wiley & Sons Inc.
- 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.
- Green, J. P. (2007). Fixing special education. *Peabody Journal of Education*, 82(4), 703-723.
- 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.
- Guadagnoli, E. & Velicer, W. F. (1991). A comparison of pattern matching indices. *Multivariate Behavioral Research*, 26, 323-343.
- Gutkin, T. B. & Reynolds, C. R. (1980). Factorial similarity of the WISC-R for Anglos and Chicanos referred for psychological services. *Journal of School Psychology*, 18, 34-39.

- 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.
- Hocutt, J. M. (1996). Effectiveness of special education: Is placement the critical factor? *The Future of Children*, *6*(1), 77-102.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois, Champaign
- Hunter, J. E., & Schmidt, F. L. (2000). Racial and gender bias in ability and achievement tests: Resolving the apparent paradox. *Psychology, Public Policy & Law*, *6*, 151-158.
- 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.
- Jensen, A. R. (1987). Individual differences in mental ability. In J. A. Glover & R. R. Ronning (Eds.), *Historical foundations of educational psychology* (pp. 61-88). New York, NY: Plenum.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.
- Kamphaus, R. W., Benson, J., Hutchinson, S., & Platt, L. O. (1994). Identification of factor models for the WISC-III. *Educational and Psychological Measurement*, *54*, 174-186.
- Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6-1/2 and 16-1/2 years. *Journal of Consulting and Clinical Psychology*, *43*, 135-147.
- Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn & Bacon, Inc.

- 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., 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.
- Lee, K., & Ashton, M. C. (2007). Factor analysis in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 424-443). New York, NY: Guilford.
- Logerquist-Hansen, S., & Barona, A. (1994). *Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and non-Hispanic white children with learning disabilities*. Paper presented at the meeting of the American Psychological Association, August 12, 1994, Los Angeles, CA.
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*, 57-64.
- Lynn, R., (2006). *Race differences in intelligence: An evolutionary analysis*. Athens, GA: Washington Summit Books.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999) Sample size in factor analysis. *Psychological Methods, 4*, 84-99.
- Mackintosh, N. J. (1998). *IQ and human intelligence*. New York, NY: Oxford University Press.

- 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. Retrieved from [http://www.emich.edu/coe/powerplay/docs/vol\\_2/issue\\_1/pp\\_journal\\_volume\\_2\\_issue\\_1.pdf](http://www.emich.edu/coe/powerplay/docs/vol_2/issue_1/pp_journal_volume_2_issue_1.pdf)
- McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology*, 4, 229-240.
- 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.
- McGrew, K. & Flanagan, D. (1998). *The intelligence test desk reference: Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon, Inc.
- McShane, D. A. (1980). A review of scores of American Indian children on the Wechsler Intelligence Scales. *White Cloud Journal*, 1(4), 3-10.
- McShane, D. A., & Plas, J. M. (1982). WISC-R factor structure of Ojibwa Indian children. *White Cloud Journal*, 2(4), 18-22.
- McShane, D. A. & Plas, J. M. (1984). The cognitive functioning of American Indian children: Moving from the WISC to the WISC-R. *School Psychology Review*, 13, 61-73.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- 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.
- 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.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287-297.
- 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.



- Reschly, D. J., & Grimes, J. P. (2004). Best Practices in Intellectual Assessment. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 1337-1350). Bethesda, MD: The National Association of School Psychologists.
- Reschly, D. J., & Reschly, J. E. (1979). Brief reports on the WISC-R: I. 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. L. (1979). Analysis of test bias in four groups with the regression definition. *Journal of Educational Measurement, 6*, 1-9.
- 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., & Ford, L. (1994). Comparative three-factor structure solutions of the WISC-III and WISC-R at 11 age levels between 6-1/2 and 16-1/2 years. *Archives of Clinical Neuropsychology, 9*, 553-570.
- Reynolds, C. R., & Gutkin, T. B. (1980). A regression analysis of test bias on the WISC-R for Anglos & Chicanos referred for psychological services. *Journal of Abnormal Child Psychology, 8*, 237-243.
- Reynolds, C. R., & Hartlage, L. C. (1979). Comparison of WISC and WISC-R regression lines for academic prediction with Black and with White referred children. *Journal of Consulting and Clinical Psychology, 47*, 589-591.
- 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.
- Reynolds, C. R., & Ramsey, M. C. (2003). Bias in psychological assessment: An empirical review and recommendations. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 67-93). Hoboken, NJ: Wiley.
- 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*, 144-194.
- Saccuzzo, D. P., & Johnson, N. E. (1995). Traditional psychometric tests and proportionate representation: An intervention and program evaluation study. *Psychological Assessment, 7*, 183-194.

- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*, 215-222.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Counseling and Clinical Psychology*, *47*, 919-927.
- Sandoval, J. (1982). The WISC-R factorial validity for minority groups and Spearman's hypothesis. *Journal of School Psychology*, *20*, 198-204.
- Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*, 73-103.
- Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. San Diego, CA: Jerome M. Sattler, Publisher, Inc.
- 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* (5<sup>th</sup> ed., pp. 265-315). San Diego, CA: Jerome M. Sattler, Publisher, Inc.
- Saxton, J. D. (2001). An introduction to cultural issues relevant to assessment with Native American youth. *The California School Psychologist*, *6*, 31-38.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53-61.
- Silverstein, A. B. (1973). Factor structure of the Wechsler Intelligence Scales for Children for three ethnic groups. *Journal of Educational Psychology*, *65*, 408-410.
- Sparks, S. (2000). Classroom and curriculum accommodations for Native American students. *Intervention in School & Clinic*, *35*, 259-264.
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Sue, S., & Okazaki, S. (2009). Asian-American educational achievements: A phenomenon in search of an explanation. *Asian American Journal of Psychology*, *5*(1), 45-55.

- 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* (4<sup>th</sup> ed.). Needham Heights, MA: Allyn and Bacon.
- Tanner-Halversen, 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 Co.
- Tempest, P. (1998). Local Navajo norms for the Wechsler Intelligence Scale for Children – Third Edition. *Journal of American Indian Education*, *37*(3), 18-30.
- 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 October 1, 2009, 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, Inc.
- 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., & Kush, J. C. (2002). Confirmatory factor analysis of the WISC-III for students with learning disabilities. *Journal of Psychoeducational Assessment*, *20*, 4-19.

- 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. (1944). *The measurement of adult intelligence* (3<sup>rd</sup> ed). Baltimore, MD: Williams & Wilkins.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. San Antonio: The Psychological Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1993) *Wechsler Individual Achievement Test*. 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.
- Wechsler, D. (1974). *Wechsler Preschool and Primary Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Weiss, L., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology, 33*, 297-304.
- Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WSC-IV advanced clinical interpretation*. Boston, MA: Elsevier.
- 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.

- Witt, J. C., & Gresham, F. M. (1985). Review of the Wechsler Intelligence Scale for Children-Revised. In J. Mitchell (Ed.), *Ninth mental measurements yearbook* (pp. 1716-1719). Lincoln, NE: Buros Institute.
- Wolff, H. G., & Preising, K. (2005). Exploring item and higher order factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods*, *37*, 45-58.
- 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.

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL OF DATA COLLECTION



---

Office of Research Integrity and Assurance

---

**To:** Marley Watkins  
EDUC - I.

**From:** Mark Roosa, Chair  
Soc Beh IRB

**Date:** 03/25/2009

**Committee Action:** **Exemption Granted**

**IRB Action Date:** 03/25/2009

**IRB Protocol #:** 0903003827

**Study Title:** Psychometric Properties of the WISC-IV Among Arizona Students

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(4) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.