2019

Exploratory Factor Analysis of the Canadian Wechsler Intelligence Scale for Children-Fifth Edition for a Sample of First Nations Students

Jessica Hanson
Eastern Illinois University

The Graduate School
EASTERN ILLINOIS UNIVERSITY

Thesis Maintenance and Reproduction Certificate

FOR: Graduate Candidates Completing Theses in Partial Fulfillment of the Degree
Graduate Faculty Advisors Directing the Theses

RE: Preservation, Reproduction, and Distribution of Thesis Research

Preserving, reproducing, and distributing thesis research is an important part of Booth Library's responsibility to provide access to scholarship. In order to further this goal, Booth Library makes all graduate theses completed as part of a degree program at Eastern Illinois University available for personal study, research, and other not-for-profit educational purposes. Under 17 U.S.C. § 108, the library may reproduce and distribute a copy without infringing on copyright; however, professional courtesy dictates that permission be requested from the author before doing so.

Your signatures affirm the following:

• The graduate candidate is the author of this thesis.

• The graduate candidate retains the copyright and intellectual property rights associated with the original research, creative activity, and intellectual or artistic content of the thesis.

• The graduate candidate certifies her/his compliance with federal copyright law (Title 17 of the U. S. Code) and her/his right to authorize reproduction and distribution of all copyrighted materials included in this thesis.

• The graduate candidate in consultation with the faculty advisor grants Booth Library the nonexclusive, perpetual right to make copies of the thesis freely and publicly available without restriction, by means of any current or successive technology, including but not limited to photocopying, microfilm, digitization, or internet.

• The graduate candidate acknowledges that by depositing her/his thesis with Booth Library, her/his work is available for viewing by the public and may be borrowed through the library's circulation and interlibrary loan departments, or accessed electronically. The graduate candidate acknowledges this policy by indicating in the following manner:

__✓__ Yes, I wish to make accessible this thesis for viewing by the public

______ No, I wish to quarantine the thesis temporarily and have included the ***Thesis Withholding Request Form***

• The graduate candidate waives the confidentiality provisions of the Family Educational Rights and Privacy Act (FERPA) (20 U. S. C. § 1232g; 34 CFR Part 99) with respect to the contents of the thesis and with respect to information concerning authorship of the thesis, including name and status as a student at Eastern Illinois University. I have conferred with my graduate faculty advisor. My signature below indicates that I have read and agree with the above statements, and hereby give my permission to allow Booth Library to reproduce and distribute my thesis. My adviser's signature indicates concurrence to reproduce and distribute the thesis.

Graduate Candidate Signature

Jessica Hanson
Printed Name

School Psychology
Graduate Degree Program

Faculty Adviser Signature

Gary L. Canivez
Printed Name

4/24/2019
Date

Please submit in duplicate.

Exploratory Factor Analysis of the Canadian Wechsler Intelligence Scale

For Children-Fifth Edition for a Sample of First Nations Students

(TITLE)

BY

Jessica Hanson

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Specialist's in School Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

2019

YEAR

I HEREBY RECOMMEND THAT THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

THESIS COMMITTEE CHAIR	4/24/2019 DATE	DEPARTMENT/SCHOOL CHAIR OR CHAIR'S DESIGNEE	4/25/19 DATE
THESIS COMMITTEE MEMBER	4/24/19 DATE	THESIS COMMITTEE MEMBER	DATE
THESIS COMMITTEE MEMBER	4/25/19 DATE	THESIS COMMITTEE MEMBER	DATE

Exploratory Factor Analysis of the Canadian Wechsler Intelligence Scale for Children-Fifth Edition for a Sample of First Nations Students

Jessica Hanson

Eastern Illinois University

Acknowledgements

I would like to acknowledge all individuals that helped this project become a reality. First, I would like to thank my family and friends who encouraged and supported me throughout this project. Second, I would like the thank my thesis chair, Dr. Gary Canivez, for his knowledge, time, effort, and guidance toward this study. I would also like to thank my committee members, Dr. Margaret Floress and Dr. Ronan Bernas, for their feedback, advice, and time they dedicated toward this project. Finally, I would like to thank Registered Psychologist, Merril Dean, for providing data for this study. Without her time and effort, this study would have been unable to move forward.

Table of Contents

List of Tables

Abstract

Native Americans and First Nation students are overrepresented in special education and underrepresented in structural bias research of the intelligence measures that place them there. There are several empirical studies of test bias on the Wechsler scales due to their popularity within the school system, however there is little exploratory factor analysis research on these scales with the Native American Indian population. Further, the Native American Indian and First Nation population is a relatively small minority group compared to other racial and ethnic groups in North America and this group is under-represented in government statistics and overlooked in funding for policies that provide prevention for several risk factors. This study aimed to discover the factor structure of the WISC-V^{CDN} with First Nations students to provide understanding and better interpretation of scores to facilitate ethical data-based decision making and provision of special education services to First Nations students. A total of 102 participant data were collected and a replication of the Canivez, Watkins, and Dombrowski (2016) study was followed to ensure best practice of Exploratory Factor Analysis. Results indicated that a three-factor model was most viable for the First Nations students on the WISC-V^{CDN}, which is dissimilar to previous research. However, results of the dominance of the general intelligence (*g*) factor was similar to previous research of the Wechsler scales using both methods of Exploratory and Confirmatory Factor Analysis. Future research directions and implications for First Nations students, data-based decision making, and special services eligibility is discussed.

The Native American Indian population is a relatively small minority group compared to other racial and ethnic groups in the United States. Unfortunately, due to their small percentage, they are unrepresented in government reported statistics and overlooked in government funding and policies that provide prevention to several risk factors common in all youth and specific factors to the Native American population (Olson & Wahab, 2006). Common risk factors that affect all youth regardless of racial or ethnic identity are mental and addictive disorders, physical or sexual abuse, and recent, severe stressful life events (Olson & Wahab, 2006). Olson and Wahab indicated that there are some risk factors specific to the Native American Indian population due to acculturation, social change, and disruption of tribal unity. These factors can increase the risk for suicide attempts, loss of ethnic identity which can lead to depression, anxiety, and poorer general health. Further, the Indian Health Service (IHS), has reported that Native American Indian and Alaska Natives continually die at higher rates compared to other Americans in several areas such as diabetes mellitus, assault/homicide, and intentional self-harm/suicide. IHS speculates this disparity may be due to inadequate education, poverty, health service discrimination, and cultural differences. Alcohol dependence and substance abuse have been leading health problems among the Native American population. Particularly, marijuana and inhalants have been reported as a more severe problem in the Native American population compared to the general population, whereas alcohol dependence and misuse continues to be the dominant risk factor for this group (Olson & Wahab, 2006).

The Native American population have also been reported to be at a higher risk for mental health disorders than other ethnic groups in the United States. Additionally, Native Americans have been consistently overrepresented for mental health services (Olson & Wahab, 2006). Mental health problems other than depression and anxiety consistently reported for the Native American Indian population include panic disorders, psychosomatic symptoms, and emotional problems. Specifically, Native American youth have been documented to be at higher risk than any other ethnic or racial group in the United States for mental health problems (Olson & Wahab). Contributing factors to these higher rates of mental health problems include poverty, lack of insurance, and steep rates of unemployment. Further, if Native American Indians are given access to adequate mental health care, they have been shown to not utilize it. Research has found that Native Americans harbor negative opinions about non-Native American mental health providers and have higher therapy dropout rates compared to all other ethnic groups in the United States (Olson & Wahab). There is a need for more data about the mental health needs of the Native American Indian population to better demonstrate this increased disparity between this population and other ethnic and racial groups.

There are several studies regarding cognitive assessment instruments used to identify learning needs for students in the schools to accurately understand the needs and capabilities of a student. Validity of tests used to identify and provide services to adequately meet student needs are provided by the assessment publisher in the technical manual. Tests are first tried on a diverse sample of students for the publisher to understand the psychometric structure of the test. Unfortunately, Native American students are vastly underrepresented in the normed "diverse" sample typically collected. Further, independent research of validity in measures are also conducted to replicate publisher results or discover new information for diverse populations of students and specific samples as well. Despite the insufficient sample size for separate study, several risk factors, and overrepresentation of Native American youth in special education, there is still little to no research about the relationship of measures of intelligence (WISC-V and WISC^{CDN}) and the Native American population.

Kush and Watkins (2007) identified three types of validity evidence identified in test bias research: content bias, predictive bias, and construct validity bias. Content bias occurs when test items suggest different statistical properties for groups of individuals with the same underlying skills. Predictive bias exists when there is an error in test score prediction due to membership in a particular group. Lastly, construct validity bias exists when there are not comparable factor structures observed for majority and minority groups (Kush & Watkins, 2007). When a measure fails to adequately assess the underlying constructs across different cultural groups, it can be deduced that the test is not measuring the same constructs for each group and the appropriateness of score interpretation and eligibility must be questioned. Empirical studies of all three types of test bias have occurred most frequently with the Wechsler scales of intelligence due to their popularity within the school system (Suzuki & Valencia, 1997). In the examination of factor structures, two procedures can be used: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA and CFA are complementary procedures that answer different questions about a particular measure. EFA is an analysis technique that explores a larger set of variables to search for a smaller set of latent factors (Henson et al., 2006). CFA, however, is an analysis used to test an *a priori* theory, or used with already set theoretical expectations (Henson et al., 2006). It has been suggested that EFA be performed first due to its nature in generating or suggesting plausible models to test theory. To reiterate the words of Carroll, EFA allows "data speak for themselves" (Carroll, 1993, p. 82). CFA should be performed second to test or confirm a hypothesis that was generated by the initial EFA (Henson et al., 2006). There is little EFA research on the Wechsler scales with the Native American Indian population, with majority of research employing CFA to examine Wechsler scale construct validity (Nakano & Watkins, 2013).

Wechsler scales are among the most commonly used intelligence measures of all time (Nakano & Watkins, 2013). However, a longstanding debate about intelligence tests and perceptions of bias have long affected intelligence measures. Given that intelligence tests, like the Wechsler scales, are used by school psychologists during the special education eligibility process, more than one million students each year are administered these tests (Gresham & Witt, 1997). Due to the increase in diversity within the United States education system, there is concern of the possible over identification of disabilities within minority students. Native American children, in particular, have been found to be more likely to be referred and overrepresented in special education classrooms (Kush & Watkins, 2007). According to the National Center for Education Statistics (NCES), in the school year of 2014-2015, individuals of the American Indian/Alaska Native ethnicity were the highest group with the documented percentage (17%) receiving services under the Individuals with Disabilities Education Act (IDEA). Although Native Americans are significantly overrepresented in special education, they are underrepresented in structural bias research of the intelligence measures that may place them there. This ethno-cultural minority group also has high rates of suicide, school dropout, and environmental deprivation (Nakano & Watkins, 2013). Some scholars question if the overrepresentation of minority students may be due to test bias from the frequent observation that Hispanics, African Americans, and Native Americans have historically scored lower on intelligence tests than the majority White population.

Although concerns about minority overrepresentation and test bias have been observed, little intelligence test research has focused on the Native American population. Schubert and Cropley (1972) were the first individuals to examine test bias for Native American individuals with the Wechsler Intelligence Scale for Children (WISC; Wechsler 1949). Schubert and Cropley explored the abilities of Canadian Indian and White children on two WISC subtests when they were trained by adults in specific strategies to help solve tasks. Results indicated that there was no significant difference between the sociocultural groups, although the White participant group obtained a higher Full Scale IQ score and the Canadian Indian group obtained lower Verbal IQ scores. Shubert and Cropley concluded that the lower IQ scores obtained did not derive from biological defect in intellectual functioning, but instead reflected the differences between cultures and how cultures view the process of intellectual development.

A few years later, the Wechsler Intelligence Scale for Children-Revised (WISC-R Wechsler, 1974) was introduced. The WISC-R retained several aspects of the WISC due to its popularity and acceptance, however five primary changes were made. The first of the changes was the use of a more representative standardization sample, with the inclusion of a proportional number of "nonwhite children" (Murphy, 1978). Second, the WISC-R was provided new administration and scoring criteria, and thirdly, some changes were made in the item content. The subtest administration sequence of the WISC-R was modified and lastly, the age range of the WISC-R increased to six through sixteen years

(Murphy, 1978). Reschly (1978) was the first to examine the construct validity of the WISC-R among four sociocultural groups: Anglos, Blacks, Chicanos, and Native-American Papagos. Results suggested that different factor solutions indicated a better fit for different sociocultural groups. The principal component extraction favored a three factor solution for Anglo and Chicano groups whereas a two factor solution was chosen for Black and Native American groups. However, chi-square tests suggested more than a two-factor solution was needed for the Anglo group, but was satisfactory for all other groups. These results produced the question: how many factors are appropriate in an intelligence measure that was created and used in a diverse nation? Zarske et al. (1981) were motivated to answer this question and challenged the WISC-R three-factor structure with Navajo and Papago Native American children with learning disability. Zarske et al. (1981) results indicated that when data were presented in a two- or three-factor solution, a two-factor solution was a better fit and sufficient to explain participant performance on the WISC-R.

Mishra (1982) was the first to examine the predictive validity of the WISC-R with Navajo Native American students. Item bias of the WISC-R Verbal subtests was explored and compared between Navajo and Anglo participants. Mishra (1982) found that the majority of items were not culturally biased and overall, the WISC-R was fair when used with Navajo Native Americans. Mishra's (1982) conclusions differed from a comparison study conducted by Naglieri and Yazzie (1983). Naglieri and Yazzie compared the WISC-R to the Peabody Picture Vocabulary Test-Revised (PPVT-R) to evaluate the relationship between their standard scores when used with Navajo Native American children. Results suggested PPVT-R standard scores were significantly lower than WISC-R Verbal, Performance, and Full Scale IQ's. Naglieri and Yazzie further cautioned the use and interpretation of the WISC-R Verbal IQ score as a measure of verbal intelligence because it easily subjected to the influence of poor English language skill. Reynolds and Reschly (1983) further challenged WISC-R item content and its potential bias across four sociocultural groups: Anglo-American, Afro-American, Mexican American, and Native Americans. Item bias analysis conducted on six of the 12 WISC-R subtests produced ambiguous results with Native American groups (Reynolds & Reschly, 1983). Reynolds and Reschly (1983) found that the Native American sample item difficulty index (*p* value) differences of adjacent items were particularly lower which suggested that progression of item difficulty was not consistent. Further, outlier analyses suggested that one third of the Verbal Scale subtest's items could be considered biased against Native Americans (Reynolds & Reschly, 1983).

The Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) was created with an updated standardization sample and minor improvements of its predecessor, the WISC-R. Only two empirical studies examined the WISC-III with the Native American population. First, Tempest (1998) examined the WISC-III standardization sample norms compared to norms created with a local Navajo Native American sample. Tempest (1998) hypothesized that the created local Navajo norms would have greater accuracy in identifying Navajo children for eligibility because the norms would mirror the education presented to this population (an education created with an emphasis on nonverbal communication). Tempest's (1998) results indicated that participants had a significant difference between their Verbal and Performance IQ scores when local and standardization norms were compared. Students who were found proficient in the English language had significantly higher Verbal and Performance subtest scores, which suggested that verbal ability influences performance and results on the WISC-III. Kush and Watkins (2007) explored the structural validity of the WISC-III with Native American students. Kush and Watkins (2007) sought to produce an adequately performed study with a large, representative sample of Native American participants. Results mirrored similar research, Native American's obtained lower Verbal subtest scores on the WISC-III. In examination of the factor structure, an oblique four-factor model was found to be the best fit for the Native American sample. This factor structure was similar to the WISC-III's four-factor model obtained with the standardization sample.

The WISC-IV (Wechsler, 2003) added new subtests and deleted several WISC-III subtests, but retained a four first-order factor structure with a higher-order general intelligence factor estimated by FSIQ. The WISC-IV added Word Reasoning, Matrix Reasoning, Picture Concepts, Letter Number Sequencing, and Cancellation subtests. Subtests deleted from the WISC-III included Picture Arrangement, Object Assembly, and Mazes. With the subtests added to the WISC-IV there was an increase in the number of tasks that weren't related to general intelligence. A composite score was created to estimate general intelligence with only verbal and perceptual reasoning subtests. This score was the General Ability Index (GAI), which included the six subtests of the Full Scale IQ derived from the Verbal and Perceptual Reasoning areas and did not include Working Memory and Processing Speed subtests. Only four studies had investigated the structure of the WISC-IV with none solely focused on the Native American population. Due to that lack of research, Nakano and Watkins (2013) explored the factor structure of the WISC-IV with Native American children. Several factor models were presented, along with two hierarchical models to observe general intelligence (*g*). Results indicated the oblique four-factor, higher-order, and bifactor hierarchical models were better fits to data than the other factor models presented. Nakano and Watkins (2013) results supported previous research on factor structure of Wechsler scale measures with the Native American population.

The Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) was created with a five-factor structure. The WISC-V changed the composition of the subtests of each composite from three to two. The number of subtests required for the Full Scale IQ (FSIQ) also decreased in the WISC-V update from 10 subtests to 7. There is an absence of research of the construct validity of the WISC-V with the Native American Indian population. Although previous research on the Wechsler scales and Native American Indians is sparse, results have indicated that scores are reportedly lower among Native Americans on the Verbal subtests. Previous research on the WISC-III and WISC-IV factor structure has supported a four-factor model as best fit for Native American data (Kush & Watkins, 2007; Nakano & Watkins, 2013) but there are no studies yet available to test WISC-V structure with Native American Indians. Due to the overrepresentation of Native Americans currently in special education programs, there is a need for reliable, valid, and diagnostically useful measures of intelligence for accurate identification and placement. Two independent studies Canivez, Watkins, and Dombrowski, (2016, 2017) did not support the five-factor model posited by its publisher and questions exist on best structural model with Native American Indians. Most recently, Watkins, Dombrowski, & Canivez (2018) analyzed the Wechsler (2014) model for the Canadian Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V^{CDN}, Wechsler; 2014) version to identify its reliability and structural validity. Results indicated that the fifth factor (FR) was found to produce negative variance, redundant factors, and low reliabilities, thus rendering it unviable. Similar to other research, Watkins et al. (2018) found that despite the different version, the WISC-V^{CDN} still was better supported structurally with a four factor model, not five as suggested by its publisher.

Due to the absence of research of factor structure on the WISC-V with Native American Indians, this study examined the factor structure of the WISC-V with Native American students. Previous research of both EFA and CFA on the WISC-V has not supported its five factor structure (Canivez et al., 2016, 2017). Further, there is no research to support that the WISC-V factors are adequately represented in the Native American Indian population. This research aimed to explore factor structure of the WISC-V and illuminate the factor structure best suited for the Native American Indian population. With this information, school psychologists can make more informed decisions on what intelligence assessment may be best suited for their particular need while also ensuring the ethical use of the WISC-V on this population.

Initial research goals were to evaluate the U.S. WISC-V with Native American students across states, tribal memberships, and diverse school districts. However, difficulties arose in attempting to obtain WISC-V data. All 183 schools listed under the Bureau of Indian Education (BIE) across 23 states were contacted with data requests and were unresponsive or unwilling to provide data for this necessary research. Schools in Wisconsin, Minnesota, Michigan, and Arizona not associated with BIE, were also contacted and produced no data. Contact was made with Registered Psychologist, Merril Dean, in Canada's Northwest Territories. Ms. Dean, owner of Dean Educational and Psychological Counseling, provides educational and psychological services to students across the Northwest Territories where these services are sparse. Ms. Dean was passionate about this research and provided data on 102 First Nations children whom were administered the WISC-V^{CDN}. Due to the total absence of data on the U.S. WISC-V and the availability of WISC-V^{CDN} data, the factor structure of the WISC-V^{CDN} with First Nations (Canadian equivalent to Native American) children was examined.

First Nations youth have had similar traumatic history of forced cultural assimilation through required education (Latimer et al., 2018). Saddled with previous trauma of assimilation through abusive Indian Residential "schools" and tribal disruption, accumulation of negative health and social outcomes resulted. Over 58% of surveyed First Nations youth reported levels of distress indicative of mental health problems, while higher rates of suicide, depression, addiction, and maladaptive coping strategies have increased (Latimer et al., 2018). Both Native American and First Nation youth have been generationally predisposed to higher risk factors and need for services. However, methods to provide services to these populations have been severely neglected by sample exclusion. In other words, the Native American and First Nations youth have been placed in services by measures that may not adequately represent them and their unique circumstances.

Although the original focus was on the factor structure of the WISC-V with Native American children, First Nations children are similar to Native American and Alaskan Native youth and the WISC-V^{CDN} is quite similar to the WISC-V, and reflects the importance and relevance of this research. There is total lack of research on factor structure of the WISC-V for the Native American Indian and First Nations populations, regardless of version. This complication illustrates the difficulty of determining if the tests being used to identify learning challenges and provide special education services to Native American or First Nations youth are appropriate. Availability of Canadian data on the WISC-V^{CDN} allowed the assessment of the factor structure to identify the most viable solution for First Nations students.

Literature Review

Factorial validity research on the various versions of the Wechsler Intelligence Scale for Children with Native American Indian samples is extremely limited. More research on diverse groups is required to establish invariance and validity of the Wechsler Intelligence Scales for Children editions. There is little research on the Wechsler scales with the Native American Indian population, however each Wechsler scale edition has at least one empirically based study that has provided some insight on its latent factor structure within Native Americans.

Wechsler Intelligence Scale for Children

There appear to be no factor analytic studies of the Wechsler Intelligence Scale for Children (WISC) with Native American Indians. Further, only one study of the WISC was conducted with Native Americans. Schubert and Cropley (1972) explored if WISC subtest scores of two different sociocultural groups of White and Indian children would change when provided training from adults prior to administration (Schubert & Cropley, 1972). Schubert and Cropley (1972) were interested to see if the groups trained by adults would exhibit improvements in their performance on the WISC and verbal regulation behavior scores. Children were trained by adults on more efficient problem solving strategies for two WISC subtests: Similarities and Block Design. The main purpose of this study was not on the initial performance of the children on these subtests, but if there was a significant response to the training.

The WISC Block Design and Similarities subtests were chosen for assessing training ability and performance because they both require a particular technique or strategy for a solution (Schubert & Cropley, 1972). The training procedure taught the participant how to utilize appropriate solutions for similar problems of the two subtests, but did not include coaching during the actual subtest administration. This study consisted of four groups of Canadian Indian and White participants. The first group was composed of 60 Canadian Indian children with ages between 11 and 14 years. This particular group of children spoke their native language at home and had parents with low levels of English speaking. The second group consisted of 66 Canadian Indian children with an age range of 6 to 11 years. These children spoke English for their everyday working language, but were not integrated in common white culture. The third group consisted of 40 White children with ages between 9 to 12 years. Partial data were collected from the last group of 30 participants who were Canadian Indian children between the ages of 11 and 15 years. Full procedure of this study included administration of the WISC to the participants, training on Block Design and Similarities subtests by adults, followed by a retesting on only the trained subtests (Schubert & Cropley, 1972). After retesting, a verbal regulation of behavior test was administered.

A three-way analysis of variance (ANOVA) was performed and results found that the third group of participants, White children, obtained the highest mean FSIQ. The groups with Canadian Indian children obtained higher Performance IQ scores than Verbal IQ scores (Schubert & Cropley, 1972). There was no significant difference between the three groups in test performance after training. Although the White participant group received higher IQ scores, both White and Canadian Indian children had no significant differences in gains as a result of training on the Block Design and Similarities subtest. Schubert and Cropley (1972) concluded that it was unlikely that low IQ scores could result from a biological defect in intellectual functioning. Instead, these scores reflected differences between White and Indian cultures and the processes of intellectual development between the two cultures (Schubert & Cropley, 1972).

Wechsler Intelligence Scale for Children-Revised

In the development of the Wechsler Intelligence Scale for Children-Revised (WISC-R), a three-factor model emerged instead of the two factor model of the WISC. The first of the three factors in the solution is Verbal Comprehension which consisted of Information, Similarities, Vocabulary, and Comprehension subtests. The second factor, Perceptual Organization, included Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes. Lastly, the third factor, Freedom from Distractibility, included Arithmetic, Digit Span, and Coding subtests (Zarske, et al., 1981).

Research conducted on the WISC-R examined its factor structure among Anglo, Black, Chicano, and Native American Papagos groups (Reschly, 1978). Reschly (1978) hypothesized construct validity could provide evidence of the use of measures within different sociocultural groups: the test should measure the same underlying abilities and corresponding scores of these abilities should be similar, if not, the test may be inappropriate or unfair for particular sociocultural group membership. The study examined results of Kaufman's WISC-R factor analysis with three separate non-Anglo groups. The three factors of the WISC-R were labeled as Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. The main purpose of this study was to examine the appropriateness and fairness of the WISC-R for four separate sociocultural groups in terms of comparability of factor structures and the construct validity of the Full Scale IQ (Reschly, 1978).

Participants included 950 children total: 252 identified as Anglo, 235 Black, 223 Mexican, and 240 Native American. Analyses used to explore the number of factors suggested in the WISC-R were the Silverstein (1977) and Kaufman (1975) methods. Principal components analysis was also included with an eigenvalue greater than one as the criterion. Unrestricted maximum likelihood estimation for the two, three, and four factor solutions was allowed and Varimax orthogonal rotation was used for each factor solution (Reschly, 1978). These methods of analysis have been found to be inappropriate when used in this context. Principal components analysis' intent is to summarize several variables into fewer components (data reduction), with the focus not on the latent factors (Henson et al., 2006). Principal components analysis cannot accurately identify how many factors need to be extracted and therefore, researchers run the risk of over or underextraction (Wood et al., 1996). Varimax factor rotation, when paired with principal components analysis, will contain error due to loading distortions, incorrectly identified loadings, and factor splitting (Wood et al. 1996). If overextraction occurs, Varimax rotation has been found to create false factors at the expense of true factors, along with factor splitting and is inappropriate when latent factors are correlated as they are in IQ tests like the WISC (Wood et al., 1996).

Results on the number of factors identified for each sociocultural group using principal component extraction (eigenvalues at or above 1) found that the three-factor solution was better suited for Anglo and Chicano participants, whereas a two-factor solution was suggested for Black and Native American individuals. Chi-square tests suggested that more than two factor solutions were required for the Anglo sample, but was sufficient for the other groups. The highest loadings of the first factor for all groups were Vocabulary, Information, Comprehension, and Similarities (Reschly, 1978). Lastly, the second factor, Perceptual Organization, had similar loadings across all four sociocultural groups. Overall, "the three factor solutions both failed to support the existence of the third (Freedom from Distractibility) factor for Black and Native American Participants" (Reschly, 1978, p. 422). Reschly (1978) identified a significant limitation within his study in that the participant groups varied significantly on socioeconomic status and level of intelligence in addition to their race or ethnicity. The limited sample size made examination of the WISC-R factors impossible.

Another study that examined the factor structure of the WISC-R with Native American individuals (Zarske et al., 1981) explored the factor structure with Native American children with a learning disability. The construct validity of the Full Scale IQ and Verbal and Performance scales for this population was examined.

Each participant was previously diagnosed with a learning disability prior to participation in the study. There were two groups of Native American children in this sample. The Papagos Native American sample consisted of 50 children, whereas the

Navajo Native American sample had 192 children. For data analysis a principal factor analysis with squared multiple correlations, was conducted for each group. After principal factor analysis extraction, Zarske et al. (1981) applied Varimax rotation of the two and three factor solutions. As stated previously, Varmiax rotation has been found to be inappropriate in the evaluation of factors because if forces correlated factors to be uncorrelated and it creates inaccurate factor loadings, loading identification, and can create false factors (Wood et al., 1996). An eigenvalue of greater than or equal to one was the criterion for selection of the appropriate number of factors (Zarske et al., 1981).

Results of two-factor and three-factor solutions were reported. For two-factor solutions the first factor was composed of the Verbal Scale subtests whereas the second factor was composed of the Performance Scale subtests. These results mirrored previous research. Comparisons between Reschly's (1978) study with Papago Native American children and this study found high coefficients of congruence of both factors between Reschly's (1978) sample and the Papagos learning disability group of the Zarske et al. (1981) study. High coefficients of congruence for both factors were also found between Reschly's (1978) sample and the Navajo learning disabled group of Zarske et al. (1981). Results of the three-factor solutions found eigenvalue results preferred a two-factor structure for the Navajo group and a three-factor structure for the Papagos group. The Navajo group when structured with three factors, found the first factor was formed of Vocabulary (V), Similarities (S), Information (I), and Comprehension (C) subtests. The second factor was formed by Object Assembly (OA), Picture Arrangement (PA), Digit Span (DS), and Picture Completion (PC) subtests. Lastly, the third factor was formed by OA and Block Design (BD) subtests. For the Navajo group, the factors appeared to split the second factor into two factors due to the loadings of the OA subtest. It is important to note that the Freedom from Distractibility factor *did not* emerge for either group (Zarske, et al. 1981). Results indicated that "A two factor solution is sufficient to explain the performance of learning disability Navajo and Papago children on the WISC-R" (Zarske, et al. 1981, p. 406).

The only limitation identified by Zarske, et al. (1981), was the small sample size of the Papagos Native American group. This study concluded that the WISC-R was an appropriate measure of intellectual functioning for both Papagos and Navajo children with learning disability and supported the construct validity of the WISC-R for diverse groups (Zarske, et al. 1981).

There is an assumption that cultural and language differences do not have an effect on a measure of intellectual development, such as the WISC-R. Sandoval (1979) argued that minority children's experiences with vocabulary and concepts are different from the majority children and thus WISC-R items may be unfairly difficult for children from minority cultures. In order to accurately examine if items on an intelligence measure were biased, there are three general strategies for information collection. First, content bias must be established through an analysis of item statistics and test means. Second, the internal criteria of the assessment must be examined to determine if the two groups respond to the measure in a similar manner (Sandoval, 1979). Lastly, internal bias parameters such as means, standard deviations, reliability coefficients, correlations, and standard errors of measurement must be evaluated. Sandoval (1979) used these three strategies to examine the performance of majority and minority children on the WISC-R to produce information that could be used to form judgement on the WISC-R and cultural bias.

Participants (*N* = 1,050) were randomly selected from a previous study conducted by Mercer and Lewis (1979). Participant ages ranged from 6 to 11 years and were distributed roughly amongst three ethnic groups: Anglo-American (*n* = 351), Afro-American (*n* = 350), and Mexican American (*n* = 349; Sandoval, 1979). Coefficients alpha were estimated for each WISC-R subtest for each minority group. Item means, rank order correlations for item difficulties, and correlations of differences were compared for all groups. A multivariate factorial analysis of variance was conducted to examine the main effects of ethnic group and socioeconomic status, and the interaction between the two (Sandoval, 1979).

Alpha reliabilities were found to be within .02 across all subtests, with exceptions of Object Assembly, which was more reliable for Afro-Americans. Comprehension and Block Design were less reliable for the Afro-American group and Picture Arrangement and Mazes was less reliable for the Anglo-American group (Sandoval, 1979). Overall, Sandoval (1979) found that the WISC-R had high and comparable reliability estimates for both majority and minority groups. Rank-order correlations for item difficulties were found to be high (.98) for all subtests except Picture Arrangement, Block Design, Mazes for the Anglo-Americans versus Afro-American comparison. Rank-order correlations of differences in item performance were found to be relatively lower (.70), which suggested that few items in each subtest are relatively more difficult for one group or another (Sandoval, 1979). Lastly, Sandoval's (1979) results indicated that the interaction between socioeconomic status and ethnicity was not significant in the determination of performance on the WISC-R. Fifty-nine items in total were found by Sandoval (1979) to be more difficult for Afro-American and Mexican-American groups compared to the Anglo-American group. However, results indicated that WISC-R subtests were essentially equivalent across all three groups and thus, the notion that children from different ethnic groups may have difficulty on particular subtest items was not supported (Sandoval, 1979).

Mishra (1982) examined WISC-R item bias for Native American Navajo's and hypothesized that research of test bias at the item level could lend information about particular items and their relationship to the general intelligence construct for diverse populations while also providing an answer to test fairness when used with minority groups. The purpose of this study was to explore item bias data in the WISC-R Verbal subtests with Anglo and Navajo Native American groups (Mishra, 1982).

Participants included both Anglo and Navajo students that were randomly selected from fourth and fifth grade classrooms. Each group consisted of 40 students with an age range of 9-11 years. Data were analyzed with the log linear model and maximum likelihood estimation. The goodness of fit was also tested with a likelihood chi-square statistic. Three variables were examined: ethnicity, ability, and pass-fail responses to individual items. Ethnicity was described by the two groups of participants: Anglo and Native American Navajo. The ability variable included low and high ability based on the FSIQ. The FSIQ low ability score range was 71 to 101, whereas the FSIQ high ability score range was between 102 to 123. Lastly, individual items were either passed (score of one) or failed (score of zero).

Information, Similarities, and Vocabulary subtest scores were obtained and analyzed between the two groups. Mishra (1982) found that the majority of items were not culturally biased, identifying only 15 items biased against the Native American Navajo group. Results found that six items on the Vocabulary subtests appeared to be more difficult for the Navajo sample compared to the Anglo group, five from the Information subtest, and lastly, four items from Similarities. These results were similar with the Vocabulary subtest for a Mexican-American sample (Mishra, 1982).

Reynolds & Reschly (1983) investigated WISC-R item bias with four sociocultural groups. This investigation was a replication of the previous Sandoval (1979) study, adding a fourth sociocultural group: Native Americans. Data from the Reschly (1978) study were combined with Sandoval (1979). Analyses were conducted on half of the 12 WISC-R subtests: Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Picture Completion. Reynolds and Reschly (1983) used two approaches to analyze the data: internal psychometric characteristics and examination of specific items for bias. The first approach, internal psychometric characteristics, included internal consistency reliability comparisons, rank order of difficulty, difficulty of adjacent items, and the relationship of the item to subtest score. The second approach examined of specific item bias and consisted of an outlier analysis and transformation of item difficulties.

Cronbach alphas were used to estimate internal consistency reliability. Differences between groups were very small, especially on Verbal subtests, which indicated differences in coefficients of .05 or less across all groups (Reynolds & Reschly, 1983). The most notable was that "Verbal scale subtests were found to be more reliable than Performance scale subtests" (Reynolds & Reschly, 1983, p. 145). Analysis using Spearman rho rank order correlations found that item difficulty index (*p* values) for each subtest were similar across all four groups and combinations of groups (≥.97 or higher). Results also found that item difficulty index (*p* values) on the differences of adjacent items were lower across all four groups, particularly lower for the Native American sample (.45 to .65). This indicated a progression of item difficulty was not as consistent as initially expressed in rank order correlations for Native American participants (Reynolds & Reschly, 1983). Outlier analyses found that one third of items on the Verbal Scale subtests could be considered biased against Native Americans, however this may have been due to possible ceiling effects. Native Americans also scored significantly lower than the other three groups on all Verbal scale subtests. Lastly, point biserial correlation results found significant relationships between item responses and total subtest scores. Reynolds and Reschly (1983) found a difference between Native American participants and the three other sociocultural groups. The Native American sample had lower biserial correlations, which could be a result of possible item bias. These results between the two approaches were found by Reynolds and Reschly (1983) to be ambiguous on possible test bias for Native American groups.

Mishra's (1982) research influenced other researchers to explore the validity of the WISC-R with the Native American Indian population. In the year following Mishra's study, the WISC-R was compared to the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Naglieri & Yazzi, 1983). Naglieri (1981) previously found that the PPVT-R and the WISC-R had a significant positive correlation and means of the PPVT-R were not significantly different than the mean WISC-R Verbal (VIQ), Performance (PIQ) and Full

Scale IQ scores with a sample of 38 children with intellectual disabilities. The purpose of the Naglieri and Yazzie (1983) study was to evaluate the relationship between the PPVT-R and WISC-R standard scores with a sample of Navajo children.

In this comparison study, participants included 37 Native American Navajo students who resided in a Navajo Reservation in Arizona. Of these 37 students, 19 were male and 18 were female. The participants were administered both the PPVT-R and the WISC-R by the same examiner in a counterbalanced order to avoid practice effects. The PPVT-R is a measure of receptive vocabulary with a representative normative sample of 4,200 children aged 2-18 years. Results found the mean PPVT-R standard score was significantly lower than the WISC-R Verbal mean, Performance mean, and Full Scale IQ. Further, all correlation coefficients of the PPVT-R and WISC-R were statically significant. Naglieri and Yazzie (1983) concluded that "Verbal IQ should not be used as a measure of verbal intelligence because it too is un-doubtly influenced by poor English language skills" (p. 599).

Wechsler Intelligence Scale for Children-Third Edition

Tempest (1998) conducted a study to develop Navajo norms for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) and to provide insight regarding several questions. Tempest (1998) created local Navajo norms to allow comparisons of Navajo WISC-III scores to those of their peers in the same minority group as well as the general population. Questions this study aimed to answer were first, to find what the average WISC-III Navajo profile consisted of, second, how the Navajo population subtest performance compared to the standardized sample, third, did language proficiency have an impact on the Navajo profile, and lastly, if the residence of the Navajo participant impacted the Navajo WISC-III profile.

Participants were from eighteen elementary schools in New Mexico. A total of 334 students were selected through stratified random sampling, by age and gender, with age ranges from six to eleven years. The participants were administered the WISC-III by trained examiners. WISC-III Navajo norms were developed in a similar fashion as the WISC-III standardized norms. A language assessment was created by Tempest (1998) by selecting subtests of the Preschool Language Assessment Instrument and tasks of the Test of Problem Solving to give an indication of the participant's ability to use expressive and receptive language (Tempest, 1998). Results for the WISC-III profile indicated the participants had a significant difference between VIQ and PIQ scores ($M = 18.3$), with the lowest score on the Vocabulary subtest when standardized norms and local Navajo norms were compared. Students who were found proficient in the English language through the language assessment also had significantly higher scores on Verbal and Performance subtests compared to those who were not found proficient in the English language through the language assessment. These students who were found proficient in the English language also had significantly higher factor scores compared to their Navajo peers (Tempest, 1998). When comparing the residence of the participants, the urban students obtained significantly higher factor scores on Verbal Comprehension (VC), Processing Speed (PS), VIQ, and FSIQ, and performed better on Information, Similarities, Vocabulary, Comprehension, Coding, and Symbol Search subtests compared to those who lived in rural areas.

Overall, results indicated that Navajo students had a higher PIQ's and lower VIQ's. There were no significant differences in the participant's Coding, Block Design, and Mazes scores when compared to those of the standardized sample. Navajo participants had higher Perceptual Organization (PO) and Processing Speed (PS) factor scores. Tempest (1998) hypothesized these results were due to the method of learning being visually orientated over verbal. Reported deficits in VC and FD factor scores with the Navajo sample were possibly due to the verbal/auditory nature of the subtests. Tempest (1998) suggested teaching Navajo students based on their strength of visual informational learning. Tempest (1998) encouraged teachers to be mindful of their student's verbal abilities and that Navajo students may work better with hands-on activities or multi-modality teaching.

Kush and Watkins (2007) examined the structural validity of the WISC-III with a sample of Native American students as only two previous studies had addressed the structural validity of major intellectual tests for the Native American children. Further, these two earlier studies examined only two Native American tribes, leaving the question if structural validity results generalized span across different tribes in the country. Previous factor analytic studies of the WISC-R and WISC-III with Native Americans were methodologically inadequate and included sample sizes too small for adequate estimates. Only two studies that had minimally adequate sample sizes have been published on the WISC-R with Native American participants. Both studies supported factor contiguity, but due to both studies emphases on the WISC-R, results might not generalize to the new version: WISC-III. Further, each study focused on only one Native American tribe, drastically limiting its scope. The purpose of Kush and Watkins (2007) was to use confirmatory factor analysis (CFA) to examine the WISC-III structural validity with the Native American Indian population.

Participants consisted of 344 Native American students who attended the Bureau of Indian Affairs (BIA) schools found in eleven states. Of this sample there were 227 boys and 117 girls ranging from kindergarten through 11th grade. Of these 344 Native American students, twelve BIA Nations were represented as follows: Apache, Arapaho, Cherokee, Chippewa, Navajo, Ojibiwa, Penobscot, Potawatomi, Puyallup, Siboba, Sioux, and Tohono O'odham (Papagos). The WISC-III VIQ, PIQ, and FSIQ were also evaluated in this study. Data on 2,301 Native American students were collected, however some data required exclusion due to Digit Span and Symbol Search subtests not being universally administered. All 12 subtests were required for full factor structure examination, but school psychologists do not routinely administer the optional WISC subtests.

The data analysis consisted of confirmatory factor analyses (CFA) using maximum likelihood estimation on covariance matrices. Kush and Watkins (2007) tested exact fit between the model and observed covariances with comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean residual (SRMR). Test score comparisons mirrored previous research with overall scores lower than the normative WISC-III sample. Verbal scores were particularly lower in the Native American sample. Univariate skewness and kurtosis reflected expected variability of a normal distribution. Results indicated that the normative oblique four-factor model was the best fit for this sample. Factor loadings in this model found that VC and PO reflected the WISC-III normative sample. Kush and Watkins (2007) concluded that the WISC-III normative oblique four-factor structure was found to be the best fit for this sample. This factor structure was similar to the factor structure of the WISC-III used in the normative sample. Further, concurrent and predictive validity evidence for WISC-III FD and PS factors remained generally unsupported due to weak reliability coefficients, and poor long- and short-term stability.

Kush and Watkins (2007) described several limitations to their study. The first limitation was that participants were not put into separate groups based on initial evaluation or periodic evaluation. Further, participants were not separated by special education classification, grade level, or region. Kush and Watkins (2007) stated that an effort was made not to distinguish these qualities of the participants in order to portray a nationally representative sample. Another limitation presented was the administration of the WISC-III was done in English while there was a lack of an English proficiency measure provided to the participants. An English proficiency measure collects information to ensure there is no test or study error in results due to limited English. This information should have been collected because the participants were of a minority culture who may not speak English as a primary language which may have impacted their test, which was administered in English, results. Lastly, all data were collected from archival records so the competency of the examiner could only be assumed.

Wechsler Intelligence Scale for Children-Fourth Edition

One study examined the factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) with referred Native American students (Nakano & Watkins, 2013). To explore the structural validity of the WISC-IV with the Native American population, participants included 176 referred Native

American students between kindergarten and grade 12 and who attended Arizona school districts. Students were selected for the study if the WISC-IV scores with all 10 core subtests were available, if students were Native American, and the primary language of the student was English. Confirmatory factor analysis was the method used for assessing the WISC-IV structural validity. A sample size of over 150 participants was suggested due to prior research stating this size requirement for factor solutions that contain several high loading variables (Tabachnick & Fidell, 2011). Nakano and Watkins (2013) examined four first-order models and two hierarchical models as follows: one factor; two oblique verbal and nonverbal factors; three oblique verbal, perceptual, and working memory/processing speed factors; four oblique verbal, perceptual, working memory, and processing speed factors; an indirect hierarchical model (higher-order) with the first four factors; and a direct hierarchical model (bifactor) with four group factors. The indirect hierarchical model (higher-order) and direct hierarchical (bifactor) model were included to evaluate the effect of general intelligence (*g*) on the first-order factors and specific subtests. Methods of comparative fit index (CFI) and the root mean square error of approximation (RMSEA) were used to indicate fit. Bayesian information criterion (BIC) was used to identify model complexity, with lower values indicating better fit. Higher CFI values and lower RMSEA values suggest a better model fit, Nakano and Watkins (2013) used ΔCFI > +.01, ΔRMSEA > -.015, and ΔBIC > +2 as standards. Results indicated that subtest, factor, and IQ scores of the sample were lower and less variable than the normative WISC-IV sample (Nakano & Watkins, 2013). Results also demonstrated that the general intelligence factor accounted for the majority of variance in the first-order factors. As for indicators of best fit, the first-order models with one through three oblique factors were found inferior to the oblique four-factor model and two hierarchical models (Nakano & Watkins, 2013). Although the oblique four-factor model and two hierarchical models did not have a superior ΔCFI, ΔRMSEA, and ΔBIC values were favored, and results were still interpreted. ΔBIC favored the indirect hierarchical model, ΔCFI was found to be neutral, and ΔRMSEA favored the oblique first-order and indirect hierarchical models. Due to support by two out of three indicators (ΔBIC and ΔRMSEA), the indirect hierarchical model (higher-order) can be suggested as the superior fit to this particular data (Nakano & Watkins, 2013). Nakano and Watkins also found that there was no structural bias evidence within their Native American sample.

Several limitations were presented in this study. First, due to data being collected from an archival source, the competence of the examiner was assumed but could not be known. Second, the sample was from only a small number of Arizona school districts. Another limitation was of the cases that included specific tribal affiliation, nearly all were Navajo. Fourth, some participants lived primarily on the reservation, whereas other participants in the sample lived in rural or urban environments. Previous research has suggested difference in performance on cognitive measures can occur between children who live in rural or urban environments (Nakano & Watkins, 2013). Lastly, no measure of the English-language proficiency within the sample was available. Several research studies have found that English language proficiency can impact Native American performance on cognitive measures (Nakano & Watkins, 2013).

Wechsler Intelligence Scale for Children-Fifth Edition

There is currently no research on the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) or the Canadian Wechsler Intelligence Scale for Children Fifth Edition (WISC-V^{CDN}; Wechsler, 2014) with the Native American Indian or First Nations populations. Other research has been conducted to evaluate the WISC-V factor structure. Based on the previous WISC research, which has resulted in important insights of the interpretation of factor index scores and subtest results when Native Americans are assessed with the Wechsler scales, more exploration and research should be conducted on the newest edition of the Wechsler Intelligence Scale for Children. Due to the overrepresentation of Native American student's in the special education system, research is needed to ethically administer and interpret intelligence scales to minority group children such as Native Americans (Kush & Watkins, 2007).

An exploratory factor analysis (EFA) was conducted on the WISC-V standardization sample by Canivez, Watkins, and Dombrowski (2016). In this study, multiple criteria were used to determine how many factors to retain such as; eigenvalue > 1, scree test, Horn's parallel analysis (HPA), minimum average partials (MAP), Bayesian Information Criterion (BIC), and sample size adjusted BIC (SS-BIC; Canivez et. al., 2016). Principal axis EFA was used for WISC-V standardization sample analysis and retained factors underwent promax oblique rotation. Canivez et al. (2016) set the salient factor pattern coefficients as > .30. Lastly, the Schmid and Leiman (1957) procedure was applied to the second-order EFA solutions.

Results of MAP suggested one factor as best fit, whereas eigenvalue of >1, scree, and HPA suggested two or three factors (Canivez et al., 2016). The BIC and SS-BIC analyses suggested four factors. These findings differ from the WISC-V publisher, which claimed a five-factor model (Canivez et al., 2016). Exploratory factor analysis extraction began with five factors. The five-factor model only produced one salient factor pattern coefficient, Figure Weights (FW) which determined in unviable (Canivez et al., 2016). No other factors were found salient and were psychometrically unsatisfactory. The four-factor model extraction found adequate Verbal Comprehension, Working Memory, Perceptual Reasoning, and Processing Speed factors with their corresponding subtest associations and simple structure was achieved, with none of the subtests found to saliently load on more than one factor (Canivez et al., 2016). Factor correlations were also found to be moderate to high (.387-.747), which suggested the presence of an additional factor, general intelligence (Canivez et al., 2016). In analysis of a three-factor model, Verbal Comprehension and Perceptual Reasoning factors combined, which left Working Memory and Processing Speed factors separate. For the two-factor model, Working Memory merged with the Verbal Comprehension and Perceptual Reasoning factors, which left Processing Speed as its own separate factor. No subtests cross-loaded within the two-factor model, but 13 subtests loaded into only one factor, leaving only three subtests to load onto Processing Speed (Canivez et al., 2016). Results of the EFA selected the four-factor solution and so it was transformed with the Schmid and Leiman (1957) procedure to analyze variance. The hierarchical *g*-factor accounted for 35.5% of total variance and when combined with group factors, a total of 53% common variance was found (Canivez et al., 2016). This indicated that 47% of unique variance remained from WISC scores. Omega-hierarchical (ω_H) and omega-subscale (ω_S) coefficients were analyzed and found that the ω_H for general (*g*) was high and sufficient for interpretation, but ω_S coefficients for the four group factors were low and unsatisfactory. Overall, EFA of the WISC-V did not support the five-factor structure that its publisher claimed. Further, the interpretation of WISC-V scores may be impacted by these results. Canivez et al. (2016) illustrated the conflicting results between their study and the publisher's, and implored users of any intelligence measure to investigate the use of the assessment and its interpretation before utilizing it in practice.

Canivez, Watkins, and Dombrowski (2017) examined the factor structure of the WISC-V using confirmatory factor analysis (CFA). The WISC-V *Technical and Interpretive Manual* (Wechsler, 2014) did not specify the method of estimation, skewness, or kurtosis. Further, maximum likelihood estimation of the WISC-V was not used by the publisher, but weighted least squares was used without justification. In order to evaluate the overall model fit, Canivez et al. (2017) used the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Squared Residual (SRMR), Tucker Lewis Index (TLI), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Higher values indicated a better fit for CFI and TLI, whereas lower values indicated a better fit for SRMR and RMSEA. An adequate model fit was defined as a CFI and TLI less than or equal to .90 along with SRMR less than or equal to .09, with RMSEA less than or equal to .08 (Canivez, et al., 2017). Good model fit required CFI to be greater than or equal to .95 with SRMR and RMSEA less than or equal to .06 (Canivez, et al., 2017). Lastly, Canivez et al. (2017) specified a superior fit as a model that displayed a meaningfully better fit than alternative models (i.e. change in CFI greater than .01, change in RMSEA greater than .015). Canivez, et al. (2017) used the WISC-V standardized sample subtest correlation matrix from the 2,200 participants for this CFA. Results found that every five factor model failed and were rejected due to model misspecification. A bifactor model with four group factors was found to be the best model due to higher CFI and TLI, and lower SRMR, RMSEA, AIC, and BIC (Canivez et al., 2017). Omega-hierarchical (ω_H) and omega-subscale (ω_S) were also analyzed to determine true score variance unique to general (*g*) intelligence and the four group factors. The ω_H coefficient for the general intelligence (*g*) factor was found to be high and sufficient for interpretation. However, ω_S group factor coefficients was significantly lower, indicating that there was not enough true score variance to support interpretation (Canivez et al., 2017).

Most recently, Watkins, Dombrowski, and Canivez (2018) analyzed the reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V^{CDN}, Wechsler, 2014). Previous research suggested that with each revision of an assessment, research must be conducted to evaluate the new measure due to the inability to consider that two instruments are comparable without evidence (Beaujean, 2015). In its revision, like the WISC-V, the WISC-V^{CDN} added three new subtests, deleted two subtests, and created two new factors. Further, changes in instructions and content of remaining subtests occurred. Wechsler (2014) claimed that the revision of the WISC-V^{CDN} version was reliable and valid, however no new evaluative methods of reliability or structure validity were used to measure or support these claims. The WISC-V^{CDN} structure is a higher-order model with an overarching general intelligence (*g*) factor loaded by five general factors. These five factors are then loaded by 16 primary and secondary subtests.

Watkins et al. (2018) stated there were six notable measurement concerns regarding the CFA methods reported by Wechsler (2014) supporting the higher order structure: possibility of confirmation bias, nondisclosed method of latent variables, nonstandard method of parameter estimation, multiple cross-loadings on the Arithmetic subtest, the five-factor model based on chi-square differences, and the use of the global model fit. Further, Wechsler (2014) did not report reliability estimates for the WISC-V^CDN and instead reported estimates from the U.S. standardization sample. Therefore, reliability of the Canadian version are unknown. Split-half reliability coefficients were reported and were stated to range from .83 to .94. This type of reliability is considered a historical approach due to the high possibility of bias but is commonly used in such tests. Instead, Omega coefficients provide a better estimate for multidimensial tests and are the principal coefficients used in current research. High omega values indicate highly reliable scores, leading to the proportion of variance in each subscale score is both general and group factor variance. However, omega coefficients are unable to differentiate specifically between general factor and group factor. Omega coefficients are akin to coefficient alpha and as such should meet the same standard values of .80-.90. In order to distinguish between general and group factor variance, hierarchical omega coefficients can be used. When low, most reliable variance of the group factor is due to the general factor. However, a large hierarchical omega coefficient would suggest opposite: reliable variance of the group factor is due to the general factor. At this time, there is no guideline for acceptable hierarchical omega levels, but researchers state that coefficients should at least exceed .50 at a minimum. Goals of this study were to evaluate the factor structure, variance, and reliability of the WISC-V^CDN in order to identify appropriate structure and discover estimates of model based reliability.

Participants were 880 children aged 6-16 years' old who were considered a good representation of Canadian youth. Correlations, means, and standard deviations of the 16 primary and secondary subtests were analyzed of the WISC-V^CDN. All CFA were from covariance matrices using maximum likelihood estimates via Mplus 7.4. A reference indicator for higher-order models was set to identify latent variable scales and variance was set to identify latency in bifactor models. All CFA were from covariance matrices using maximum likelihood estimates via Mplus 7.4. Models were duplicates of those specified by Wechsler (2014) and included simple structure, bifactor, and global models. They were evaluated with chi-square likelihood ratio, the Tucker-Lewis Index (TLI), standardized root mean square residual (SRMSR), root mean square error of approximation (RMSEA), and Akaike's information criterion (AIC). A "good fit" required TLI to be > .95 and SRMR and RMSEA to be < .06 (Watkins et al., 2018). The lowest AIC value determined the best model.

Results suggested that models created with less than four group factors did not achieve good model fit standards as previously determined. However, models with four- and five-group factors achieved good global fit. Though these models were found to have good fit globally, size parameters and its statistical significance, and interpretability were further analyzed. Three models were found to have negative error variance and were likely to be biased (Watkins, et al., 2018). Further, the higher-order models were found to have improper solutions or have high levels of FR and general intelligence loadings indicating redundant factors. One bifactor model with five group factors resulted in a proper solution (Watkins, et al., 2018). The bifactor model was found to be the best fit based on global fit and simple structure, although several subtests had week loadings and a significant difference between loadings of Block Design and Visual Puzzles. In consideration of variance, the general factor accounted for 33.8% of total variance and 67.7% of common variance (Watkins et al., 2018). No group factors were found to account for large portions of variance. In reliability examination, omega coefficients for the bifactor model indicated that general, Verbal Comprehension, Visual Spatial, and Working Memory factor scores were "reliable" (Watkins, et al., 2018). In other words, some variance was from multiple common factors. However, omega-hierarchical subscale estimates found that only the general factor had good reliability and the group factors were low (Watkins, et al., 2018). This suggests that much of the reliable variance was from the general factor and *not* group factors.

In this most recent study, the Wechsler (2014) model for the WISC-V^{CDN} was analyzed to identify its reliability and structural validity. The higher-order model with the new fifth FR factor was found to produce negative variance, redundant factors, and low reliability estimates (Watkins, et al., 2018). The bifactor model with four group factors and one general factor was found to be the best representation of the structure of this assessment. Results are not surprising as they mirror previous research on the U.S. version of the WISC-V. The bifactor model was found to be favored when there are complexities in the structure, however both higher-order and bifactor models indicated good fit. There is currently no empirical support to distinguish between each model and its estimate of general intelligence, however when specific abilities are required, the bifactor model should be preferred (Watkins, et al., 2018).

Although the research discussed above is not with the Native American Indian population and is instead the standardization sample of the WISC-V or WISC-V^{CDN}, this research is invaluable to help understand if the WISC-V is adequate. The factor model used directly impacts the interpretation of the resulting scores, of which are based in decision making for special education services. Additionally, the WISC-V (Wechsler, 2013) manual did not include an exploratory factor analysis on the standardization sample. Findings from the Canivez, et al. (2016) study indicated that EFA suggested that the WISC-V five-factor model was not supported and instead a four-factor was. Despite the lack of research in this area on the Native American Indian population, previous WISC research has resulted in important insights of the interpretation of factor index scores and subtest results. However, exploration and research should be conducted on the newest edition of the Wechsler Intelligence Scale for Children with Native American Youth. Due to the overrepresentation of Native American student's placement in special education, research is needed to ethically administer and interpret intelligence scales to minority group children such as Native American or First Nations youth (Kush & Watkins, 2007).

There is lack of sufficient support and evidence for a five-factor model on the WISC-V^{CDN} (Watkins et al., 2018). Due to the high risk and overrepresentation of Native American and First Nations youth in special education, methodical and supported best practice exploratory factor analysis (EFA; Watkins, 2018) must be conducted to best understand the latent structure of the WISC-V^{CDN} with First Nations youth. This research aims to discover the factor structure of the WISC-V^{CDN} with this population to provide understanding and better interpretation of scores to facilitate ethical data-based decision making and provision of special education services to First Nations students.

Method

Participants

The sample of this study included 102 diverse First Nations students from several school districts across the Northwest Territories of Canada and were of various First Nations tribal membership. Seven tribes were represented including: Cree (n = 5), Dene (n = 30), Gwichin (n = 13), Inuit (n = 14), Slavey (n = 14), T'licho (n = 1), and Inuvialuktun (n = 25). The sample included 61 males (59%), 40 females (39%), and one non-binary individual (1%) in grades 1-11. Age of participants (M = 11.11, SD = 2.91) were found to be slightly skewed and kurtotic (skew= 1.85, kurtosis = 6.37). Further descriptive statistics on subtest and composite scores of sample are presented in Table 1. Participants data were provided by a single Registered Psychologist who provided psycho-educational assessments for referred client WISC-V^{CDN} scores (including 10 core subtests) from special education evaluations.

Instrument

The WISC-V^{CDN} (Wechsler; 2014) is a measure of general intelligence for individuals between the ages of 6-16 years. A four-level organization of subtest administration is new to this Wechsler version. First, the WISC-V^{CDN} is composed of 10 primary subtests with seven primary subtests that combine to estimate the FSIQ, which across the five factors (VC, VS, FR, WM, and PS). If a subtest of the FSIQ is found invalid, another subtest may be substituted from the secondary level that is within the same factor. The second level, Primary Index Scales includes all 10 primary subtests that estimate the five factor index scores (VCI, VSI, FRI, WMI, PSI) and cannot be substituted by any other level subtest. The Ancillary Index scales are composed of five scales that are not factorially derived, but intelligence oriented: Quantitative Reasoning (QR), Auditory Working Memory (AWM), Nonverbal (NV), General Ability (GA), and Cognitive Proficiency (CP). Each of these five scales have designated subtests used to estimate their intended construct. Lastly, the Complementary Indices consist of three scales: Naming Speed, Symbol Translation, and Storage and Retrieval extracted from new WISC-V subtests: Naming Speed Literacy, Naming Speed Quality, Immediate Symbol Translation, Delayed Symbol Translation, Recognition Symbol Translation, Naming Speed Index and Symbol Translation Index. The Complementary Index scales and associated subtests are not intelligence subtests and are instead created for diagnostic identification. Because of this, Complementary Index scales should not be substituted for Primary or Ancillary subtests.

Procedure and Analyses

This study is a replication of the Canivez et al. (2016) study with a sample of Native American Indian children. Similar procedure and analyses completed by Canivez et al. (2016) were followed for this study's WISC-V exploratory factor analysis (EFA). In addition, all procedures and analysis utilized best practice in exploratory factor analysis as described by Watkins (2018).

Multiple criteria were utilized to examine the number of factors to retain. Criteria included: eigenvalue > 1, the scree test, standard error of scree, Horn's parallel analysis (HPA), and minimum average partials (MAP). Monte Carlo PCA for Parallel Analysis (Watkins, 2000) was used with 100 replications in order to produce stable eigenvalue estimates (Watkins, 2018; Canivez, et al., 2016). Because the scree test is considered a subjective criterion, the SE_{Scree}, programed by Watkins (2007), was used because it is the most accurate objective scree method (Nasser, Benson, & Wisenbaker, 2002; Watkins, 2018). Principal axis exploratory factor analyses were used to analyze the sample of First Nations WISC-V^{CDN} primary subtest scores using SPSS. The extracted factors were subjected to promax oblique rotation and salient factor pattern coefficients defined ≥ .30 (Canivez, et al., 2016). Rotation of factors allows a simpler and more meaningful solution by bringing them "closer" to each variable (Watkins, 2018). Oblique rotation is recommended first to allow factor intercorrelations to emerge, however if there is an absence of relationship between factors, promax will produce orthogonal results (Watkins, 2018). For empirical consistency, a factor cannot be determined unless it is marked by two or more salient subtest loadings (≥.30) and possesses no salient cross-loadings (loading on multiple factors). In best practice, as followed by Canivez, et al. (2016) and explained in Watkins (2018) exploratory factor analysis guide, these processes have the most empirical support for quality examination of all factors.

Due to the possibility that subtest scores may include combinations of both first-order and second-order factors, the second-order factor must be extracted first followed by the Schmid and Leiman (1957) procedure. The Schmid and Leiman (1957) procedure extracts the higher-order factor variance, in order for the lower-order factors to residualize and become orthogonal to the higher-order factor and to each other (Carroll, 1993, 1995, 2003, Canivez, et al., 2016). The factor pattern coefficients from the obliquely rotated first-order EFA solution and its produced second-order EFA solution factor coefficients were subjected to the Schmid and Leiman (1957) procedure as applied by the MacOrtho program (Watkins, 2004). First, common variance was assigned to the higher-order factor and then residual variance was assigned to the group factors. This allowed the examination of unique variance and common variance separately.

To calculate model based reliability estimates of latent factors, omega-hierarchical (ω_H) and omega-subscale (ω_{HS}) were used with the preference of coefficients at .75, but at least exceed .50 (Reise, 2012; Reise, Bonifay & Haviland, 2013). The omega-hierarchical coefficient can be used as a reliability estimate for general intelligence factor separate from the group factor variance. The omega-subscale estimates group factor reliability estimate with all other group and general factors removed (Reise, 2012).

Results

Factor Extraction Criteria Comparisons

Table 2 presents scree plots for Horn's Parallel Analysis (HPA) from the WISC-V^{CDN} First Nations sample. The number of recommended factors from these procedures were as follows: HPA and MAP suggested one factor; eigenvalues > 1, scree, and standard error of scree suggested three factors, and only the publisher theory recommended five factors. Analysis of results from factor extraction procedures indicated fewer factors than suggested by the WISC- V^{CDN} publisher. EFA began with the extraction of five factors as recommended by WISC-V^{CDN} publisher so subtest associations based on the five factor structure could be examined. This process continued to explore factor models with fewer factors (four, three, and two) to determine sufficiency.

Exploratory Factor Analyses

Five factor extraction. Table 3 presents results of the five factor extraction with promax rotation. In the first attempt to extract five factors with the standard 25 iterations, a Heywood case occurred, where a communality estimate was found to be greater than 1.0 and therefore it did not converge. Extraction iteration was increased to the SPSS maximum of 9,999, where another Heywood case resulted. In a final attempt to extract five factors, the Snook and Gorsuch (1989) method of a two-iteration limit for estimating communalities was used. This extraction attempt was successful and resulted in cross loadings of Figure Weights (FW), Picture Span (PS), and Symbol Search (SS) on two or more factors. The fourth factor was found not viable due to the salient loading (>.30) of only one subtest: Symbol Search, which also cross loaded on the third factor (Processing Speed [PS]). General intelligence (*g*) loadings were examined to investigate the association or correlation of a subtest with general intelligence. When five factors were extracted, *g* loadings, based on Kaufman's (1994) criteria were found good (>.70) for the Similarities (SI), Matrix Reasoning (MR), Digit Span (DS), and Figure Weights (FW) subtests; fair (.50-.69) for Block Design (BD), Visual Puzzles (VP), Picture Span (PS), and Symbol Search (SS) subtests; and poor (<.50) for the Coding (CD) subtest. Due to the inadequate fourth factor and numerous cross loading subtests a five-factor solution was not viable.

Four factor extraction. Table 4 presents the results of the extraction of four factors with promax rotation. General intelligence (*g*) loadings ranged from .316 (Coding) to .737 (Digit Span) and were within the good to fair range for all subtests except Coding (CD), Picture Span (PS), and Symbol Search (SS). In the four factor extraction, no more than 25 iterations were needed in order to converge. Perceptual Reasoning (Factor 1), Verbal Comprehension (Factor 2), and Processing Speed (Factor 3) presented consistent salient subtest associations. However, Factor 4 presented inconsistent theoretical associations with salient subtest pattern coefficients for Visual Puzzles (VP), Picture Span (PS), and Symbol Search (SS). Further, several cross-loadings were observed for Visual Puzzles (VP), Picture Span (PS), and Symbol Search (SS) subtests. These cross-loadings suggested that these subtests were not uniquely measuring one specific area when four factors were specified. Due to these cross-loadings and odd configuration of Factor 4, a four-factor solution was not viable.

Three factor extraction. Table 5 presents the results of extracting three factors with promax rotation. The *g* loadings ranged from .303 (Coding) to .741 (Digit Span) and were all in the good to fair range with the exception of the Coding subtest. Table 5 illustrates strong and salient theoretically consistent loadings on the Perceptual Reasoning factor (Block Design, Matrix Reasoning, Figure Weights, and Visual Puzzles), Verbal Comprehension factor (Similarities, Digit Span, Vocabulary), and Processing Speed factor (Coding, Picture Span, Symbol Search). No subtest was found to cross-load onto more than one factor which indicated that three factor extraction results in a simple structure. Interestingly, the Working Memory factor did not emerge, instead Perceptual Reasoning did, indicating the Working Memory factor is not viable. The two subtests that traditionally measure Working Memory: Digit Span and Picture Span split between two different factors (Verbal Comprehension and Processing Speed) respectively instead of forming the Working Memory factor (see Table 5). Despite the disappearance of the

Working Memory factor, eight subtests aligned with the expected three factors structure with reasonable dimension and created a simple structure design.

Oblique rotations allow factors to be correlated and with this rotation of these three factors, all factor correlations produced were positive and moderately sized (see Table 5.) These results indicated that a higher-order or general dimension may explain these relationships between factors. To further investigate, Carrol (2003) argued that a second-order factor analysis must be conducted with the Schmid and Leiman (1957) procedure in order to differentiate between general and group factors through an analysis of correlations of the three extracted factors. Detailed explanation and findings are discussed later in this section.

Two factor extraction. Table 5 presents results from the two factor extraction with promax rotation. General intelligence (*g*) loadings were classified as good (>.70) for Similarities, Matrix Reasoning, Digit Span, Vocabulary, and Figure Weights subtests; fair (.50-.69) for Block Design, Visual Puzzles, Picture Span, and Symbol Search subtests; and poor (<.50) for the Coding subtest. In examination of the two-factor extraction, most subtests loaded saliently (>.30) on the first factor and then two cross loaded on the second factor. Subtests that cross-loaded were: Digit Span and Vocabulary. Due to the subtest cross-loadings the two factor solution was determined not viable.

Hierarchical EFA: SL Bifactor Model

Based on these results, the three-factor EFA solution appeared to be best and was further subjected to second-order EFA of the three-factor correlation matrix (see Table 5) then transformed with the Schmid and Leiman (Schmid & Leiman, 1957) procedure.

Results of the SL procedure are presented in Table 6 and *g* and group factors are uncorrelated with each other and separated. Through the transformation, group factors and *g* are examined for their unique variance contribution. The *g* factor accounted for 32.8% of the total variance and 58.1% of the common variance. Additionally, the general factor (*g*) accounted for between 7.6% (Coding) and 44.9% (Digit Span) of individual subtest variability.

Upon examination of total variance at the group factor level, Perceptual Reasoning (PR) accounted for an additional 9%, Verbal Comprehension (VC) for additional 7%, and Processing Speed (PS) for an additional 7.6%. General (*g*) and group factors combined to measure 56.4% of the common variance in WISC-V^{CDN} scores with the First Nations sample, leaving 43.6% unique variance. Omega-hierarchical (ω_H) and omega-subscale (ω_{HS}) were then estimated based on Schmid Leiman (1957) results and are illustrated in Table 6. Omega-hierarchical (ω_H) estimations identify the unique true score variance of an equally unit-weighted score from the indicators, whereas the omega-hierarchical subscale (ω_{HS}) estimates identify the unique true score variance of a unit-weighted score for the group factor, with variance of *g* factor *and* other group factors removed. Coefficients larger than .5 are recommended. The ω_H coefficient for general intelligence (*g*) was .729 which was high and appreciable for interpretation. However, ω_{HS} coefficients for the three group factors (Perceptual Reasoning, Verbal Comprehension, and Processing Speed) were much lower (.257-.385) and did not meet criterion of sufficient unique true score variance (>.5). This indicated that the three group factors did not account for appreciable and unique variance; and variance of the WISC-V^{CDN} was mainly *g* factor variance with this First Nations sample.

Discussion

Research on the Wechsler scales with Native American Indian and First Nations samples is sparse. Although each new version of the WISC had at least one empirically based study of its factor structure or its relation to Native American Indians, there is still not nearly enough support to understand its true implications for ethical data-based decision making. Beaujean (2015) noted the need to conduct new and empirically sound research for each revision of tests and creation of new versions. Unfortunately, in the revision of the WISC-V[CDN], Wechsler (2014) did not re-examine validity or other measurement characteristics to ensure empirically sound assessment (Watkins, et al., 2018). Previous research of the U.S. WISC-V found inconsistency in the publisher claimed latent structure (five factors) due to cross-loadings, variance distribution, and poor saliency (Watkins, et al., 2018; Canivez, Watkins, & Dombrowski, 2017; and Canivez, Watkins, & Dombrowski, 2016).

The original version of the WISC (Wechsler, 1949) was examined by Shubert and Cropley (1972) for test bias with the Native American population. First Nation and White children were compared on their two WISC subtests when trained by adults to utilize specific strategies to solve tasks. Results indicated no significant difference between the two groups, however the First Nation group obtained lower Verbal IQ scores (Shubert and Cropley, 1972). Shubert and Cropley (1972) concluded that the lower IQ scores did not derive from intellectual functioning, but instead reflected differences between cultures. Years later, the revised WISC (WISC-R; Wechsler, 1972) was created which allowed more exploration of its structure with sociocultural groups. In one study regarding the construct validity of the WISC-R, results indicated that a two-factor model was most appropriate for Native American and Black groups, whereas a three-factor structure was best for Anglo and Chicano groups (Reschly, 1978). This suggested that there may be differences in test results between different sociocultural groups. In another study conducted on the WISC-R with Navajo and Papagos Native American Indian students, results indicated that a two-factor model was the best fit for Navajo youth, but a three-factor structure was best for the Papagos (Zarske et al., 1981). This indicated that different structures may fit better for some cultural memberships than others. Research to examine this concept and possible item bias was conducted by Mishra (1982) with Navajo students. Results suggested that the majority of items on the WISC-R were not culturally biased and overall and was fair when used with the Native American Indian population (Mishra, 1982). These results differed from Naglieri and Yazzie (1983) whose study results strongly cautioned the use and interpretation of the WISC-R Verbal IQ score as a measure of verbal intelligence due to its easy subjection to influence of poor English language skill. Lastly, Reynolds and Reschly (1983) further examined possible bias of the WISC-R with sociocultural groups and found that the Native American sample item difficulty was not consistent and some items on the Verbal Scale subtest could be biased against this population.

Similar to results of the WISC-R, Tempest's (1998) research on the WISC-III (Wechsler, 1991) found Native American students had significant differences between their Verbal and Performance IQ scores. Specifically, Native American Indians with higher English language proficiency performed significantly higher on both Verbal and Performance scores which implied that verbal ability influences performance and results on the WISC-III (Tempest, 1998). Structural validity with the Native American population on the WISC-III was examined by Kush and Watkins (2007). Results mirrored previous research of lower Verbal IQ estimates and a four-factor model was found to be the best fit for the Native American sample (Kush and Watkins, 2007). The WISC-IV factor structure with four factors and a higher-order factor, was supported for the publisher claim of its factor structure (Nakano & Watkins, 2013). Further, examination of the WISC-III structure indicated a preference for four factors, not three, for Native American students (Kush & Watkins, 2007). Although this finding is discrepant with the current study, it demands further examination and exploration of the latent structures of assessment and effective measurement of ability for First Nation students.

There is a lack of research regarding the WISC-V factor structure with Native American children. Previous research did not support the five-factor structure claimed by its publisher, although the WISC-V is most commonly used to determine special education services and facilitate data-based decision making, but the factor structure of the WISC-V is unknown among Native Americans (Watkins et al., 2018; Canivez et al., 2017; Canivez et al., 2016). Further, Native American youth are overrepresented in special education, despite the lack of publisher and independent research. First Nations youth, similar to Native American, have high risk factors and need for services, but tests used with them (like the WISC-V^{CDN}) have been poorly studied.

Exploratory factor analysis provides understanding and improved interpretation of scores through the specification of true score variance of each factor and global scale. Clinician interpretation and decision making can be negatively impacted when the factors do not adequately measure what they claim. The WISC-V^{CDN} did not provide a large sample, which is required to produce proportional sampling for representative norms. It is necessary to understand the underlying measurement constructs of the assessment utilized to truly interpret and assign services appropriately.

The WISC-V^{CDN} publisher claimed CFA support for a five-factor model structure with questionable research and statistical methods (Watkins et al., 2018). Present research on the WISC-V^{CDN} performed by Watkins et al. (2018) found that the fifth factor (Fluid Reasoning) produced negative variance, redundant factors, and low reliability. These findings were similar to the U.S. WISC-V and other versions (French, Spanish, UK) factor structure (Watkins et al., 2018; Kush & Canivez, 2018; Canivez, Watkins, & McGill, 2018; Lecerf & Canivez, 2018). Independent studies on the CFA and EFA of the WISC-V structure indicated poor five-factor structure and suggested a four-factor structure as best fit (Canivez, et al., 2016; Canivez et al., 2017).

In this study, in attempts to extract five factors, Heywood cases resulted indicating the structure was not appropriate. Using the Snook and Gorsuch (1989) method (limiting iterations to 2 in extraction) to extract five factors the results indicated further problems. Cross-loadings of Figure Weights (FW), Picture Span (PS), and Symbol Search (SS) occurred across two or more factors and the fourth factor produced only one subtest with salient loadings (>.30). These findings indicated the five-factor structure was psychometrically unsatisfactory. The five-factor model was also found not viable in previous WISC-V and WISC-V^{CDN} research and has historically provided poor subtest loadings (Watkins et al., 2018; Canivez et al., 2017; and Canivez et al., 2016).

In the extraction of four factors, cross-loadings between factors and inconsistent salient and theoretical subtest associations emerged, rendering Working Memory not viable. However, the three factors that remained: Perceptual Reasoning, Verbal Comprehension, and Processing Speed, produced satisfactory results with salient and theoretically consistent subtest associations. Subtests Visual Puzzles (VP), Picture Span (PS), and Symbol Search (SS) cross-loaded between factors, which suggested that they were not measuring a unique intelligence concept. These results are not consistent with WISC-V and WISC-V^{CDN} EFA and CFA research, which had instead, supported the four-four factor structure and claimed it viable (Kush & Watkins, 2007; Nakano & Watkins, 2013; Watkins et al., 2018; Canivez et al., 2017; and Canivez et al., 2016). These findings are also dissimilar to other WISC-V versions including the UK, French, and Canadian (Watkins et al., 2018; Kush & Canivez, 2018; Canivez et al., 2018; Lecerf & Canivez, 2018).

The most viable model in the present study was three first-order factors. Although some similarity of results in this study were found, the superiority of the three factor structure was in contrast to prior research. However, in the preferred three factor structure, Working Memory subtests migrated to Verbal Comprehension and Processing Speed factors. Specifically, Digit Span (DS) loaded on Verbal Comprehension and Picture Span migrated to Processing Speed. Subtest migration of the Working Memory factor has not been observed in other versions of the WISC-V. The reason for three-factor preference is unknown, but may be due to limited sample size and homogeneity of the sample. When the three-factor model was subjected to second-order EFA with transformation by the Schmid and Leiman (1957) procedure, WISC-V^{CDN} general (*g*) factor accounted for more than half of the total variance (56.4%) compared to combined group factor variance. Omega-hierarchical (ω_H) coefficient for general intelligence (*g*) was found to be high and valid for interpretation, which indicated there is enough true score variance independent from other factors, to imply general intelligence (*g)* ability. However, omega-hierarchical subscale (ω_{HS}) coefficients did not yield similar results, and instead indicated that the Perceptual Reasoning, Verbal Comprehension, and Processing Speed factors did not account for any appreciable or unique variance independent from the general intelligence (*g*) factor.

Two-factor extraction resulted in inadequate solutions. Most subtests were found to saliently load onto the first factor and subsequently cross load onto the second. Specific subtests that were found to cross load were Digit Span and Verbal Comprehension. Result of the insufficient two-factor model is similar to other research of the WISC-V and previous WISC's (Kush & Watkins, 2007; Nakano & Watkins, 2013; Watkins, et al., 2018; Canivez, et al., 2017; and Canivez, et al., 2016).

Results of the dominance of general intelligence (*g*) measurement in this study is similar to previous research and study of Wechsler Scales using both EFA and CFA methods of factor analysis across U.S., Canadian, French, U.K., and Spanish versions (Canivez et al., 2018; Watkins, et al., 2018; Canivez, et al., 2017; Canivez, et al., 2016; Kush & Canivez, 2018; Canivez et al., 2018; Lecerf & Canivez, 2018). This indicated that dominance of general intelligence factor (*g*) measurement of WISC-V are consistent with broader literature in this field. Further, ω_{HS} coefficients demonstrated that unique variance captured by the three group factors were low and did not meet the criteria to indicate sufficient unique variance due to each separate factor (>.5). Instead, general intelligence (*g*) was found to be high and viable for interpretation of a composite score based on the 10 subtest indicators that would capture sufficient unique true score variance. These findings were consistent with current research on the WISC-V, WISC-V^{CDN}, French WISC-V, WISC-V UK, and WISC-V Spain, and that it mainly measures the general intelligence (*g*) factor (Canivez, et. al., 2018; Watkins, et al., 2018; Canivez, et al., 2017; Canivez, et al., 2016).

Limitations

The present study examined EFA of the WISC-V^{CDN} with a relatively small sample of First Nations children in the Northwest Territories of Canada. Data were provided by a single psychologist, who provided educational and psychological services to students in this region due to low accessibility. The sample size ($N = 102$) was deemed minimally appropriate for the EFA based on the examination of the correlation matrix, Bartlett's (1954) test of Sphericity, and Kaiser-Meyer-Olkin (KMO; Kaiser, 1974). However, the sample of this study was likely homogeneous in its makeup of only special education serviced students in a limited geographical location (Northwest Territories of Canada). Due to geographical limitations, generalization of these findings should not be extended. Further, because data were only derived from students with special education services, it is not possible to generalize these results to all First Nation students. To strengthen the study, additional data should be obtained and added for both general education and special education students. In future research of the WISC-V^{CDN}, an attempt to collect a wider range of First Nations students across Canada, and a larger sample size, may produce different results. Further, results from a larger and more multivaried group of First Nations youth may then better represent their population on the WISC-V^{CDN}. Care should also be taken in the interpretation of results of each sociocultural group, like the First Nations, and then the makeup of smaller groups within it (for example, tribal membership). Previous research has supported different findings between cultures, sociocultural groups, and also within specific memberships who make up each group (Shubert & Cropley, 1972; Zarske et al., 1981; Reschly, 1978). Additional research should be conducted to investigate the impact of sociocultural groups and different tribal memberships on the factor structure of the WISC-V^{CDN}. Collection of data from more than one source should be considered to increase sample variety and increase generalization. Examination of the fourth factor, Working Memory, and its associated subtests should also be investigated due to their inability to form the Working Memory factor structure. Other analyses should also be conducted on the construct validity of the WISC-V^{CDN}, as this study only examined the latent structure. Lastly, examination of the diagnostic utility of the WISC-V^{CDN} for the use of clinical decision making is recommended (Canivez, 2013). Due to the popularity of the WISC-V and the high percentages of Native Americans and First Nations youth classified with special needs and who receive special education services, the tests used must be able to accurately identify and assess needs and services. This study and previous research suggests the small portions of unique true score variance at the group factor level indicates the group factor scores are unable to provide meaningful value (Canivez et al., 2018; Watkins, et al., 2018; and Canivez, et al., 2016).

Conclusions

Based on the findings of this study, the WISC-V^{CDN} is over-factored when five factors are extracted, which mirrors previous research (Canivez et al., 2018; Watkins, et al., 2018; and Canivez, et al., 2016). These results reinforce the demand for more research in assessment factor structure for all assessment versions and the need for a diverse sample. Specifically, research on cognitive assessments and their relationship with First Nation individuals must be conducted to better understand measurement utility for this population. Further, this study suggests extreme caution in interpretation of the WISC-V^{CDN} for First Nation students beyond the Full Scale IQ (FSIQ) due to high true score variance in the general intelligence (*g*) factor and low portions of unique variance in group factors. Interpretation of these scores beyond FSIQ risks unethical interpretation of scores (Canivez et al., 2018). The inability to produce and maintain a salient fourth factor (Working Memory), and absence of the Fluid Reasoning factor must also be taken into consideration during interpretation. Fluid Reasoning (FR) and Working Memory (WM) of the WISC-V^{CDN} are potentially misleading for this specific population. There is a need for creation or inclusion of more or better indicators for these factors in order to distinguish them from other factors and general intelligence (*g*). Findings from this study provide much needed information about the WISC-V^{CDN} with First Nation children. Results replicated previous research and refute publisher claims of uniqueness or importance of scores beyond general intelligence (*g*) through use of the WISC and its other versions. Data-based decision-making is directly affected by these results and present caution for clinicians to provide ethically based eligibility and special services to their students. Extreme caution must be taken for First Nation students as they are underrepresented or ignored in several samples of assessment, but overrepresented in special education (Olson & Wahab, 2006). With more research on cognitive assessments and hopefully greater inclusion of Native Americans and First Nations youth in the collection of the standardized sample, better understanding can be achieved to best provide services for students in schools. Until then, professionals must adequately read and understand the technical manual as well as independent research on its measurement in order to avoid misinterpretation of scores and maintain an ethical practice.

References

Beaujean, A. A. (2015). Adopting a new test edition: Psychometric and practical considerations. *Research and Practice in the Schools, 3*, 51-57.

Disparities | Fact Sheets. (2017, April). Retrieved November 09, 2017, from https://www.ihs.gov/newsroom/factsheets/disparities/

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessments* (pp. 84- 112). New York, NY: Oxford University Press.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 28*, 975-986. http://dx.doi.org/10.1037/pas0000238

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29*, 458-472. http://dx.doi.org/10.1037/pas0000358

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2018, June). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1111/bjep.12230

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, England: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780 511571312

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30*, 429 – 452. http://dx.doi.org/10 .1207/s15327906mbr3003_6

Carroll, J. B. (2003). The higher–stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). New York, NY: Pergamon Press. http://dx.doi.org/ 10.1016/B978-008043793-4/50036-2

Children and Youth with Disabilities. (2017, May). Retrieved November 05, 2017, from https://nces.ed.gov/programs/coe/indicator_cgg.asp

Gresham, EM. & 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.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. doi:10.1007/BF02291575

Kaufman, A. Factor analysis of the WISC-R at 11 age levels between 6 and 16 years. *Journal of Consulting and Clinical Psychology, 1975, 43*, 135- 147.

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., & Canivez, G. L. (2018, September). The higher order structure of the WISC–IV Italian adaptation using hierarchical exploratory factor analytic procedures. *International Journal of School and Educational Psychology*. http://dx.doi.org/10.1080/21683603.2018.1485601

Latimer, M., Sylliboy, J. R., MacLeod, E., Rudderham, S., Francis, J., Hutt-MacLeod, D., … Finley, G. A. (2018). Creating a safe space for First Nations youth to share their pain. *Pain reports, 3*(Suppl 1), e682. doi:10.1097/PR9.0000000000000682

Lecerf, T., & Canivez, G. L. (2018). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment, 30*, 793-808. http://dx.doi.org/10.1037/pas0000526

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

Naglieri, J.A. (1982). Use of the WISC-R and the PPVT-R with mentally retarded children. *Journal of Clinical Psychology, 38*(3), 635-637.

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

Nakano, S., & Watkins, M. (2013). Factor Structure of the Wechsler Intelligence Scales for Children-Fourth Edition Among Referred Native American Students. *Psychology in the Schools, 50*(10), 957-968.

Nasser, F., Benson, J., & Wisenbaker, J. (2002). The performance of regression–based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurement, 62*, 397–419. http://dx.doi.org/10.1177/0016440206200300 1

Olson, L. M., & Wahab, S. (2006). American Indians and Suicide: A Neglected Area of Research. *Trauma, Violence & Abuse, 7*(1), 19-33. doi:10.1177/1524838005283005

R Development Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Reschly, D. (1978). WISC-R Factor Structures Among Anglos, Blacks, Chicanos, and Native-American Papagos. *Journal of Consulting and Clinical Psychology, 46*, 417-22.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. Multivariate *Behavioral Research, 47*, 667–696. http://dx.doi.org/10 .1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140. http://dx.doi.org/ 10.1080/00223891.2012.725437

Ross-Reynolds, J. & Reschly, D. (1983). An Investigation of Item Bias on the WISC-R with Four Sociocultural Groups. *Journal of Consulting and Clinical Psychology, 51*, 144-146.

Sandoval, J. The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology, 1979, 47*, 919-927.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53– 61. http://dx.doi.org/10.1007/ BF02289209

Schubert, J., & Cropley, A. J. (1972). Verbal Regulation of Behavior and IQ in Canadian Indian and White Children. *Developmental Psychology, 7*, 295-301.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106*(1), 148-154. http://dx.doi.org/10.1037/0033-2909.106.1.148

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

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

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

Watkins, M. W. (2000). Monte Carlo PCA for Parallel Analysis [Computer Software]. State College, PA: Ed.& Psych Associates.

Watkins, M. W. (2004). MacOrtho [Computer Software]. State College, PA: Ed.& Psych Associates.

Watkins, M. W. (2007). SEscree [Computer Software]. State College, PA: Ed.& Psych Associates.

Watkins, M. W. (2013). Omega [Computer Software]. Phoenix, AZ: Ed.& Psych Associates.

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*, 219-246.

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–Fifth Edition. *International Journal of School and Educational Psychology, 6*, 252–265. doi:10.1080/21683603.2017.1342580

Wechsler, D. (2014). Wechsler Intelligence Scale for Children-Fifth Edition: Canadian manual. Toronto, Canada: Pearson Canada Assessment.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354-365. doi: 10.1037/1082-989X.1.4.354

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.

Table 1

Wechsler Intelligence Scale for Children-Fifth Edition: Canadian (WISC–V^{CDN}) Descriptive Statistics for the First Nations Sample (N = 102)

Subtest/Composite	*M*	*SD*	Skewness	Kurtosis
Block Design	9.04	2.56	-0.03	-0.18
Similarities	6.10	3.64	0.38	-0.68
Matrix Reasoning	7.50	2.45	-0.06	0.26
Digit Span	6.60	2.46	0.16	0.70
Coding	7.06	2.39	0.34	-0.31
Vocabulary	6.31	2.86	0.89	1.43
Figure Weights	7.38	2.80	0.49	0.62
Visual Puzzles	8.72	2.47	0.00	-0.12
Picture Span	7.46	3.12	0.67	1.07
Symbol Search	7.44	2.41	0.08	0.30
Verbal Comprehension Index	77.47	18.07	-0.06	1.80
Visual Spatial Index	93.21	20.05	2.87	13.85
Fluid Reasoning Index	86.01	20.40	2.85	13.58
Working Memory Index	80.97	13.56	0.68	0.72
Processing Speed Index	84.34	11.87	0.49	0.26
Full Scale IQ	78.56	12.96	0.32	-0.03

Table 2 (Report results in text rather than a separate table?)
Number of Factors Suggested for Extraction Across Six Different Criteria

	Number of WISC-V Canadian Factors Suggested
Extraction Criterion	10 Primary Subtests First Nations Sample (N = 102)
Eigenvalue > 1	3
Scree Test (Visually Examined)	1 or 3
Standard Error of Scree (SE_{Scree})	3
Horn's Parallel Analysis (HPA)	1
Minimum Average Partials (MAP)	1
Publisher (Theory) Proposed	5

Table 3

Wechsler Intelligence Scale for Children-Fifth Edition: Canadian (WISC–V^{CDN}) Exploratory Factor Analysis: Five Oblique Factor Solution for the First Nations Sample (N = 102)

WISC–V^{CDN}	General	F1: PR		F2: VC		F3: PS		F4: Inadequate		F5: WM		
Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
BD	.668	**.769**	.798	.043	.449	.047	.322	-.105	-.014	.001	.370	.649
SI	.716	.079	.505	**.913**	.859	-.058	.212	-.061	.126	-.087	.551	.756
MR	.724	**.674**	.766	.027	.487	.137	.401	.061	.167	.052	.493	.620
DS	.742	.182	.546	.279	.667	.126	.441	-.030	.250	**.357**	.689	.579
CD	.306	.071	.220	-.096	.122	**.644**	.615	.065	.032	-.053	.246	.389
VC	.706	-.060	.408	**.811**	.831	.001	.252	.106	.320	.035	.649	.706
FW	.717	***.459***	.653	-.005	.559	-.120	.265	-.023	.285	***.496***	.647	.579
VP	.656	**.723**	.721	-.015	.431	-.084	.169	.291	.366	.011	.466	.612
PS	.534	-.146	.250	.124	.461	***.387***	.561	.032	.267	***.398***	.614	.490
SS	.501	.041	.285	.059	.337	***.515***	.547	***.439***	.457	-.019	.516	.504
Eigenvalue		4.52		1.31		1.11		0.70		0.57		
% Variance		41.05		7.74		7.19		1.95		0.95		
Factor Correlations		F1: PR		F2: VC		F3: PS		F4: I		F5: WM		
F1: Perceptual Reasoning (PR)		–										
F2: Verbal Comprehension (VC)		.542		–								
F3: Processing Speed (PS)		.342		.312		–						
F4: Inadequate (I)		.103		.248		.019		–				
F5: Working Memory (WM)		.484		.723		.464		.531		–		

Note. WISC–V Canadian Subtests: BD = Block Design, SI = Similarities, MR = Matrix Reasoning, DS = Digit Span, CD = Coding, VC = Vocabulary, FW = Figure Weights, VP = Visual Puzzles, PS = Picture Span, SS = Symbol Search. *S* = Structure Coefficient, *P* = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (*g* loadings). Salient pattern coefficients presented in bold (pattern coefficient ≥ .30). Pattern coefficients presented in italics cross-loaded.

Table 4

Wechsler Intelligence Scale for Children-Fifth Edition: Canadian (WISC–V^{CDN}) Exploratory Factor Analysis: Four Oblique Factor Solution for the First Nations Sample (N = 102)

	General	F1: Perceptual Reasoning		F2: Verbal Comprehension		F3: Processing Speed		F4: ?		
WISC–V^{CDN} Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
Block Design	.691	**.862**	.837	.033	.456	.127	.353	-.244	.085	.761
Similarities	.726	.059	.521	**.962**	.871	-.110	.168	-.190	.270	.799
Matrix Reasoning	.719	**.661**	.753	.048	.508	.119	.348	.084	.350	.592
Digit Span	.737	.212	.570	**.450**	.693	.141	.372	.160	.472	.561
Coding	.316	.059	.222	-.154	.153	**.699**	.691	.123	.195	.497
Vocabulary	.701	-.055	.441	**.840**	.825	-.050	.203	.069	.444	.689
Figure Weights	.706	**.481**	.665	.238	.584	-.014	.231	.149	.422	.519
Visual Puzzles	.677	***.798***	.772	-.122	.431	-.212	.050	***.325***	.496	.714
Picture Span	.531	-.117	.282	.313	.510	***.369***	.487	***.317***	.494	.463
Symbol Search	.509	.021	.311	-.024	.377	***.355***	.460	***.584***	.644	.537
Eigenvalue		4.52		1.31		1.11		0.70		
% Variance		41.52		8.66		7.94		3.20		
Promax Based Factor Correlations		F1: PR		F2: VC		F3: PS		F4: ?		
F1: Perceptual Reasoning (PR)		–								
F2: Verbal Comprehension (VC)		.581		–						
F3: Processing Speed (PS)		.302		.306		–				
F4: ? (?)		.336		.479		.181		–		

Note. *S* = Structure Coefficient, *P* = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (*g* loadings). Salient pattern coefficients presented in bold (pattern coefficient ≥ .30). Pattern coefficients presented in italics cross-loaded.

Table 5

Wechsler Intelligence Scale for Children-Fifth Edition: Canadian (WISC–V^{CDN}) Exploratory Factor Analysis: Two and Three Oblique Factor Solutions for the First Nations Sample (N = 102)

WISC–V^{CDN}	Two Oblique Factors				Three Oblique Factors				
Subtest	g^1	F1: *g*	F2: ?	h^2	g^1	F1: PR	F2: VC	F3: PS	h^2
BD	.670	**.868** (.745)	-.192 (.362)	.577	.669	**.847** (.785)	-.044 (.442)	-.070 (.336)	.622
SI	.675	**.489** (.647)	.248 (.560)	.456	.721	.083 (.542)	**.900** (.849)	-.191 (.329)	.746
MR	.726	**.743** (.755)	.017 (.492)	.570	.727	**.744** (.788)	-.025 (.497)	.116 (.480)	.630
DS	.751	***.398*** (.6810	***.443*** (.697)	.580	.741	.225 (.607)	**.420** (.687)	.242 (.579)	.565
CD	.296	-.060 (.210)	**.424** (.385)	.150	.304	.074 (.241)	-.251 (.132)	**.636** (.540)	.328
VC	.671	***.350*** (.607)	***.403*** (.626)	.465	.711	-.076 (.476)	**.890** (.845)	.005 (.439)	.718
FW	.719	**.660** (.727)	.106 (.527)	.536	.709	**.518** (.692)	.241 (.588)	.050 (.440)	.522
VP	.657	**.805** (.719)	-.135 (.379)	.527	.652	**.753** (.738)	.009 (.452)	-.041 (.345)	.546
PS	.554	-.156 (.381)	**.841** (.741)	.564	.540	-.150 (.325)	.267 (.500)	**.613** (.679)	.501
SS	.495	-.011 (.376)	**.605** (.598)	.358	.502	.020 (.360)	.007 (.372)	**.664** (.678)	.460
Eigenvalue		4.52	1.31			4.52	1.31	1.11	
% Variance		40.34	7.49			41.15	7.85	7.38	
Factor Correlations		F1	F2			F1	F2	F3	
	F1	–			F1	–			
	F2	.639	–		F2	.617	–		
					F3	.506	.531	–	

Note. WISC–V Canadian Subtests: BD = Block Design, SI = Similarities, MR = Matrix Reasoning, DS = Digit Span, CD = Coding, VC = Vocabulary, FW = Figure Weights, VP = Visual Puzzles, PS = Picture Span, SS = Symbol Search. *S* = Structure Coefficient, *P* = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (*g* loadings). Salient pattern coefficients presented in bold (pattern coefficient ≥ .30). Pattern coefficients presented in italics cross-loaded.

Table 6

Sources of Variance in the Wechsler Intelligence Scale for Children-Fifth Edition: Canadian (WISC–V^CDN) for the First Nations Sample (N = 102) According to a Schmid-Leiman Orthogonalized Higher-Order Factor Model with Three First–Order Factors

	General		F1: Perceptual Reasoning		F2: Verbal Comprehension		F3: Processing Speed				
WISC–V Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2	ECV
Block Design (BD)	.568	.323	**.543**	**.295**	-.026	.001	-.053	.003	.621	.379	.522
Similarities (SI)	.660	.436	.053	.003	**.536**	**.287**	-.143	.020	.746	.254	.603
Matrix Reasoning (MR)	.627	.393	**.477**	**.228**	-.015	.000	.087	.008	.628	.372	.633
Digit Span (DS)	.670	.449	.144	.021	***.250***	***.063***	.182	.033	.565	.435	.878
Coding (CD)	.275	.076	.047	.002	-.150	.023	**.478**	**.228**	.329	.671	.249
Vocabulary (VC)	.660	.436	-.049	.002	**.530**	**.281**	.004	.000	.719	.281	.608
Figure Weights (FW)	.624	.389	**.332**	**.110**	.144	.021	.038	.001	.522	.478	.779
Visual Puzzles (VP)	.558	.311	**.483**	**.233**	.005	.000	-.031	.001	.546	.454	.572
Picture Span (PS)	.504	.254	-.096	.009	.159	.025	***.461***	***.213***	.501	.499	.544
Symbol Search (SS)	.459	.211	.013	.000	.004	.000	**.499**	**.249**	.460	.540	.458
Total Variance		.328		.090		.070		.076			
Explained Common Variance		.581		.160		.124		.134			
ω		.895		.840		.844		.670			
ω_H / ω_{HS}		.729		.314		.257		.385			
Relative ω		.814		.373		.304		.574			
H		.840		.533		.462		.473			
PUC		.733									

Note. b = loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness, ω_h = Omega hierarchical, ω_s = Omega subscale. Bold type indicates coefficients and variance estimates assigned to group factors and generally consistent with the theoretically proposed factor. Italic type indicates coefficients and variance estimates associated with a factor not theoretically related. The highest subtest loading with the specific group factor was used in omega subscale estimates.