A Multiple Group Confirmatory Factor Analysis of the Structural Invariance of the Cattell-Horn-Carroll Theory of Cognitive Abilities Across Matched Canadian and U.S. Samples

Sean Locke
University of Windsor

Kevin S. McGrew
Woodcock-Muñoz Foundation

Laurie Ford
University of British Columbia

WMF PRESS
THE WOODCOCK-MUÑOZ FOUNDATION

Publisher: Fredrick A. Schrank, PhD

Editor: Melanie Pammer Maerz

The WMF Press® is an imprint of the Woodcock-Muñoz Foundation.

Reference Citations

To cite this document, use:

Locke, S., McGrew, K.S., Ford, L. (2011). A Multiple Group Confirmatory Factor Analysis of the Structural Invariance of the Cattell-Horn-Carroll Theory of Cognitive Abilities Across Matched Canadian and U.S. Samples. *WMF Press Bulletin*, *No. 1*. Retrieved from the Woodcock-Muñoz Foundation website: http://woodcock-munoz-foundation.org/press/pressbulletins.html

This research was funded by the Woodcock–Muñoz Foundation. The opinions expressed in this manuscript do not necessarily reflect those of the Woodcock–Muñoz Foundation.

A Multiple Group Confirmatory Factor Analysis of the Structural Invariance of the Cattell-Horn-Carroll Theory of Cognitive Abilities Across Matched Canadian and U.S. Samples

Sean Locke
University of Windsor

Kevin S. McGrew
Woodcock-Muñoz Foundation

Laurie Ford
University of British Columbia

*In order to compare an individual's test scores across time or to examine the pattern of correlations between variables, equivalence of measurement must be established. Determining a psychological instrument's factorial invariance provides this necessary empirical foundation. In the recently published Woodcock-Johnson® III Normative Update (WJ III® NU) Tests of Cognitive Abilities (WJ III COG) and Tests of Achievement (WJ III ACH), the authors provide evidence for the Cattell-Horn-Carroll (CHC) factor structure of the entire battery based on the U.S. population. In Canada, although the WJ III is not used as frequently as the Wechsler® and Stanford-Binet scales, it is still one of the three most commonly used batteries of individually-administered tests of cognitive ability and achievement. The purpose of this study was to investigate the factorial invariance of the nine CHC constructs (*Gf, Gc, Gv, Ga, Gsm, Glr, Gs, Grw*, and* Gq*) as measured by 30 WJ III tests. The invariance of the WJ III CHC factor structure was examined with multiple-group confirmatory factor analysis (MGCFA) structural equation methods (SEM) in matched samples of 310 Canadian and U.S. school-age subjects. The results provide empirical support for the same CHC-based interpretation of the WJ III battery in Canada as has been validated and used in the United States.*

The Woodcock-Johnson® III Normative Update (WJ III® NU) Tests of Cognitive Abilities (WJ III COG) and Tests of Achievement (WJ III ACH) represent a normative update of the third edition of this widely-used battery of cognitive and achievement tests (Woodcock & Johnson, 1977, 1989; Woodcock, McGrew, & Mather, 2001, 2007). Over the past decade, use of this battery has increased in North America. The increased use of the WJ III battery is most likely due to the use of a test design blueprint based on the Cattell-Horn-Carroll (CHC) theory of cognitive abilities, the consensus psychometric model of the structure of human cognitive abilities (Ackermann & Heggestad, 1997; Carroll, 1993; Flanagan, McGrew & Ortiz, 2000; McGrew, 2005, 2009; McGrew & Flanagan, 1998; Messick, 1992; Stankov, 2000).

Consistent with the Standards on Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 1999), a test specification "blueprint" is a test design framework that maps the development of new or revised measures (the measurement domain) to the constructs in the theoretical domain. The WJ-R® test specification blueprint (Horn, 1991) was grounded in the Cattell-Horn multiple intelligences theory of fluid (*Gf*) and crystallized (*Gc*) abilities (Horn, 1965, 1968, 1985, 1986, 1988, 1994). This was in contrast to other intelligence test batteries at the time, such as the Wechsler scales, that used a more eclectic atheoretical measurement model to account for the instruments' latent factor structure. In its latest revision, the WJ III NU's test design blueprint was based on the CHC theory of cognitive abilities, an overarching integration of the Carroll three-stratum (Carroll, 1993, 1997) and Cattell-Horn *Gf-Gc* (Horn, 1991) models under a common theoretical umbrella (Kaufman, 2000; McGrew, 2005, 2009).

Overview of the WJ III NU and the Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities

The Woodcock-Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001; 2007) is the most recent edition of the Woodcock-Johnson battery originally published in 1977 (Woodcock & Johnson, 1977). The WJ III is based on the CHC theory of cognitive abilities (Schrank, Miller, Wendling, & Woodcock, 2010). The WJ III was published in 2001, and the norms were recalculated with a normative update in 2007. In brief, the original 2001 WJ III norms were based on year 2000 U.S. Census projections available at the time the standardization of the WJ III commenced (1996). Census projections are estimates of the population for future dates and are subsequently replaced by census statistics. The 2000 census final statistics produced a somewhat different description of the U.S. population than was available from the last projections issued in 1996. The WJ III NU updated the WJ III norms to reflect the final U.S. 2000 census statistics. In addition, innovative bootstrap re-sampling methods were used in the development of the WJ III NU norms; methods not fully developed at the time of the 2001 publication of the WJ III (see McGrew, Dailey, & Schrank, 2007 for details).

McGrew (1997) was the first to synthesize Cattell-Horn's *Gf-Gc* and Carroll's three-stratum models in an attempt to provide a comprehensive integrative framework for interpreting human cognitive abilities. The result is the CHC (Cattell-Horn-Carroll) theory, which served as the theoretical blueprint for the WJ III (McGrew & Woodcock, 2001). The latest updates of contemporary CHC theory can be found in McGrew (2005, 2009). The theoretical underpinnings of the WJ III are different from the venerable Wechsler batteries (e.g., WISC-IV®; WAIS-IV®), although more recent revisions of the Stanford-Binet, Kaufman Assessment Battery for Children, and the Differential Ability Scales (i.e., SB5; KABC-II, and DAS®-II) have either been explicitly or implicitly grounded in the CHC framework (see Keith & Reynolds, 2010; McGrew, 2009). In order to accurately interpret the WJ III, an understanding of the CHC model is needed.

The CHC model incorporates Carroll's (1993) tri-stratum theory of intelligence, organizing cognitive abilities and Horn-Cattell *Gf-Gc* theory into an integrated three-level hierarchy.

Carroll (1993) identified over 69 specific, or narrow cognitive abilities, at Stratum I. The narrow abilities are subsumed under the broad (Stratum II) cognitive ability domains of Fluid Intelligence or Reasoning (*Gf*), Crystallized Intelligence or Comprehension-Knowledge (*Gc*), Broad Visual-Spatial Processing (*Gv*), Broad Auditory Processing (*Ga*), and Processing Speed (*Gs*). At the apex of his model (Stratum III), Carroll identified a higher-order factor above the broad factors, which he interpreted as general intelligence, or *g* (for a more extensive discussion of the CHC model and Carroll's tri-stratum theory, see Carroll, 1993; McGrew, 2005, 2009). On the WJ III COG, clusters represent the broad abilities (e.g., *Gf*, *Gc*, *Gv*) and the individual tests (e.g., Verbal Comprehension, Retrieval Fluency) are intended to represent the narrow abilities.

The WJ III NU (Woodcock, McGrew, & Mather, 2001, 2007) is a comprehensive measure of cognitive abilities and achievement organized into three distinct, co-normed test batteries: The Woodcock-Johnson III NU Tests of Cognitive Abilities (WJ III COG); the Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities (WJ III DS; Woodcock, McGrew, Mather, & Schrank, 2003; Woodcock, McGrew, Schrank, & Mather, 2007) and the Woodcock-Johnson III NU Tests of Achievement (WJ III ACH). The WJ III is designed to measure a wide array of cognitive, oral language, and academic achievement abilities for individuals from preschool (2 years) through the geriatric (90+ years) age levels.

Both the WJ III COG and WJ III ACH are organized into Standard and Extended batteries which can be used independently, together, or in conjunction with other tests, including tests from the WJ III DS. In addition to the CHC clusters, the complete set of 31 cognitive tests (20 in WJ III COG plus 11 in the WJ III DS) are also organized by three broader categories representing cognitive performance (Cognitive Performance Model, Woodcock, 1997). The three broad categories are Verbal Ability, Thinking Ability, and Cognitive Efficiency; and the five clinical clusters are Broad Attention, Executive Functioning, Working Memory, Cognitive Efficiency, and Phonemic Awareness. Twenty-two achievement tests are organized by curricular area (reading, mathematics, written language, academic knowledge, oral language) and by clusters within these areas (e.g., Basic Reading Skills, Math Reasoning), with additional groupings for special purpose clusters (e.g., Academic Skills, Phoneme/Grapheme Knowledge). These batteries have particular diagnostic utility in that examiners can be selective in their testing and choose different evaluation tools based on the unique needs of each referral. The Appendix provides descriptions of the WJ III COG and WJ III ACH subtests and clusters used in the analyses.

Like the earlier versions of the WJ, the WJ III has been viewed as state of the art in the individual measurement of cognitive abilities and achievement (Cizek, 2003; Cummings, 1995; Hicks & Bolan, 1996; Lee & Stefany, 1995; Sandoval, 2003). The WJ Tests of Achievement have long been one of the most widely-used, individually-administered academic achievement batteries. Furthermore, the WJ III COG is taught as a primary measure of intelligence in over 1/3 of all school psychology training programs across the United States and Canada (Braden & Alfonso, 2003; Ford, Percy, & Negreiros, 2010). Its strong psychometric properties, the co-normed tests of cognitive abilities and achievement, its utility for individuals at almost any age, and empirical procedures which help identify processing strengths and weaknesses, contribute to its increasing use in Canada. The widespread use of the WJ III NU in the absence of norm transportability research heightens the importance of the current investigation.

The Present Study

The question arises as to whether a U.S. normed test is appropriate for use with Canadian populations. While a number of U.S. normed batteries of cognitive and achievement abilities are used extensively throughout Canada, surprisingly few comprehensive validation and/or standardization studies are reported in the literature. Of the limited published studies on U.S. normed cognitive and achievement tests being used in Canada, the majority published to date have examined differences in the various versions and editions of the Wechsler scales. All studies have pointed to significant score differences across the Canadian and U.S. populations with Canadian samples scoring on average 2 to 5 standard score points higher than the U.S. sample, depending on the factor or subtest (Hildebrand & Saklofske, 1996; Wechsler, 1996; 2001; 2003; 2004; 2008). These findings have suggested the need for a Canadian standardization of the Wechsler scales.

Given the widespread use of many U.S. normed cognitive and achievement batteries with Canadian populations for diagnosis, treatment, and program planning, it is evident more research is needed. There is a need to determine if the U.S. norms are "transportable" and applicable to Canadian populations and, if not, whether additional norming with a Canadian sample is needed, or if special adjustments are necessary to the norms for tests standardized in the United States. A review of the literature reveals that to date only one WJ III study has evaluated the transportability of the WJ III NU norms across the border for use with school-aged Canadian students (Ford, Swart, Negreiros, Lacroix, & McGrew, 2010).

Ford and Swart, et al. (2010) found that although the U.S. sample typically scored slightly higher than the Canadian sample on the WJ III NU COG clusters, the differences were not statistically significant, with one exception—the Long-Term Retrieval cluster. The mean difference of the General Intellectual Ability-Extended (GIA-Ext) score for the U.S. sample (M = 100.74, SD = 15.77) was not significant ($t(309) = 1.84$, $p > .05$) compared to the Canadian sample (M = 98.88, SD = 13.73). While the Canadian sample scored slightly higher (M = 101.30, SD = 14.16) than the U.S. sample (M = 100.90, SD = 15.37) on the Total Achievement cluster, the difference was not statistically significant, $t(309) = .37$, $p = .715$. However, at the test level, five statistically significant differences were noted: the Canadian sample scored significantly higher on the Reading Fluency, Quantitative Concepts, and Oral Comprehension tests while the U.S. sample scored statistically higher on the Reading Vocabulary and Editing tests. Thus, these findings support, with some caution, the use or transportability of the WJ III NU U.S. based norms with Canadian populations.

The Ford and Swart, et al. (2010) findings lay the foundation for further research on the transportability of the U.S. normed WJ III to Canadian populations. Mean score statistical comparisons (t-test) can only tell us whether or not there is a statistically significant mean difference between two samples. Such analyses do not suggest hypotheses regarding what the scores mean and whether the scores mean the same thing; for that we turn to factorial construct validity methods, such as a confirmatory factor analysis (CFA).

In support of the construct validity of the WJ III CHC measurement model, McGrew and Woodcock (2001) presented an extensive set of CFAs across five broad age groups (spanning ages 6 through 100), as well as for a combined sample across all ages. Taub and McGrew (2004) also reported support for the invariance of the CHC factor structure of the WJ III throughout the entire age range (ages 6 to 100 years). However, for use in Canada the dilemma arises: this research is based on a U.S. population; the factor structure might vary in a Canadian population. Although Ford and Swart, et al. (2010) found general support for the U.S. based norms in a Canadian population, a handful of significantly different scores were noted, which might be the result of differences in the factor structure between the populations.

The purpose of the present study was to examine the comparability of the WJ III NU factor structure in matched school-age Canadian and U.S. samples. The goal was to investigate the measurement or metric invariance of the WJ III NU between Canadian and U.S. samples. This question focuses on the extent to which nine CHC constructs (*Gf*, *Gc*, *Gv*, *Ga*, *Glr*, *Ga*, *Gsm*, *Gs*, *Grw*, and *Gq*), as operationally measured by the WJ III COG and WJ III ACH tests, are equally valid indicators of the same ability constructs between Canadian and U.S. samples. The results of these analyses will assist in the identification of any construct score differences that might suggest different interpretations of WJ III NU cluster scores in Canada.

Method

Participants

The current study is comprised of two matched samples: one strategically sampled from Canada and a matched sample of WJ III NU U.S. standardization subjects. This section describes the sample selection and comparison procedures.

Canadian Sample. The Canadian sample consisted of 341 English-speaking school-aged children from three geographical areas (Western Canada, Central Canada, and Atlantic Canada). The sampling procedures mirrored those used in the standardization of the WJ III (McGrew & Woodcock, 2001). A three-stage procedure of sampling communities, then schools, and finally subjects, was used to identify and select a sample that would be broadly representative of the English-speaking Canadian school-age population. Communities were sampled by census region and type of community, as defined by Statistics Canada (1996). Participants were obtained from six Provinces (British Columbia, Saskatchewan, Manitoba, Ontario, Prince Edward Island, and Newfoundland). Communities were targeted for selection within each of the three geographical areas based on the characteristics of geographic distribution, size of community, and socioeconomic status (SES) characteristics (high, average, and low SES communities). School board participation was then solicited from the targeted communities. When school board participation was not obtained for a targeted community, a similarly-matched community from the same geographic area was identified and school board participation was subsequently sought. In summary, final inclusion of a community in the sample reflects: a) a targeted community based on geographical area, community size, and community SES, and b) school board agreement to participate in the study.

In small school communities, testing was conducted in all schools. In larger communities, testing was conducted in a subset of schools. The general guideline for selecting the subset of schools was to obtain an equal distribution of schools in high and low SES areas. This guideline was specified to avoid any potential selection bias. To best represent a cross-section of students in the community, Catholic schools were included in communities where these schools were available and agreement to participate was obtained. In all, 34 schools were included in the study.

Sampling of participants was based on a quota-by-grade level criterion. The solicitation of subjects was entirely at random. The permission forms included subject identifying information (e.g., date of birth, grade, sex, and ethnic origin), parent education level, and mother and father's current occupation. Any subject who had less than one year of experience in an English-only classroom was excluded from the sample.

From among the returned permission forms, subjects were selected based on the identified subject-level variables needed to fill the sampling plan (male vs. female, highest grade completed by parents, ethnic origin) and were subsequently tested at school. Although the total sample was 341 in grades kindergarten through 12, only 310 students in grades 1 through 12 were included in the present study due to missing data from some tests or clusters. The 310 Canadian children ranged in age from 6 years, 8 months to 19 years, 5 months (M = 149.41 months, SD = 40.37) and are closely distributed by sex (148 males and 162 females).

U.S. Matched Sample. The 310 Canadian subjects served as the foundation for the U.S. matched sample that was selected from the 8,782 participants in the WJ III standardization sample. A U.S. subject that best matched each Canadian subject was selected from the complete WJ III standardization sample.

Subject matching was based on a hierarchical sequence of matching variables—age (in months), parent education (highest level of either mother or father's education), race/ethnicity (white or non-white), and sex. If more than one U.S. subject met the match criteria, a U.S. subject was randomly selected from the available pool. Although a concerted effort was made to collect common demographic indicators across the two samples, an exact match was not possible given differences in the way census information and demographic variables are defined in Canada and in the United States. For example, Statistics Canada (1996) defines ethnic groups according to ancestry (e.g., British, French or European, Multiple Origins, or Other); the U.S. Census categorizes individuals according to race (e.g., White, Black/African-American, American Indian, Asian/Pacific Islander) and Hispanic origin (Hispanic or non-Hispanic). The 310 U.S. subjects selected ranged in age from 6 years, 8 months to 19 years 5 months (M = 149.74 months, SD = 40.11) and were closely distributed by sex (150 males and 160 females).

Comparison. A comparison of the two samples on the matching and other variables revealed a high degree of comparability. Chi-square (χ^2) revealed no significant differences in frequencies of subjects in the Canadian and U.S. samples as a function of parent education level ($\chi^2 = 4.56$; $df = 4.0$, $p = 0.34$), race ($\chi^2 = 0.37$; $df = 1.0$, $p = 0.54$), or gender ($\chi^2 = 0.03$; $df = 1.0$, $p = 0.87$). Comparison of the ages (in months) of the Canadian and U.S. samples (t-test) revealed no significant difference ($M_{diff} = 0.40$, $t(618) = 0.12$, $p = 0.90$). A similar t-test comparison of grade placement in tenth of years ($M_{diff} = 0.29$, $t(618) = 1.09$, $p = 0.27$) was also not significant. Comparisons of the distributional characteristics of the two samples also suggested strong comparability of the two samples (see Ford and Swart, et al., 2010 for additional details). Summary statistics for the matching variables are presented in Table 1.

Analyses

Figure 1 presents both the WJ III NU theoretical factor structure and measurement models (when tests 1–7 and 11–17 of the WJ III COG and 1–2, 5–11, and 13–19 of the WJ III ACH are administered) that were specified for investigation.

The variance-covariance matrices from both samples served as input data for the analyses and were compared using multiple-group CFA methods via the AMOS™ 7.0 program (Arbuckle, 2006). Given that the evaluation of measurement invariance can take several forms (e.g.,

Table 1

Sample Characteristics of Canadian and U.S. Participants

	Canadian		U.S.	
	N	Percentage	*N*	Percentage
Sex				
Male	148	47.7	150	48.4
Female	162	52.3	160	51.6
Grade				
One through Four	91	29.4	99	31.6
Five through Eight	105	33.8	112	35.8
Nine through Twelve	114	36.8	99	31.6
Father Education Level				
< H.S Diploma	76	24.5	5	1.6
High School Diploma	62	20.0	50	16.1
Post Secondary/Diploma	75	24.2	80	25.8
University Degree	80	25.8	90	29.1
Not Reported	17	5.5	85	27.4
Mother Education Level				
< H.S Diploma	65	21.0	47	15.2
High School Diploma	61	19.7	84	27.1
Post Secondary/Diploma	105	33.8	107	34.5
University Degree	75	24.2	72	23.2
Not Reported	4	1.3	0	0
Ethnic Origin				
White/Anglo/European	247	79.7	252	81.3
Asian-Pacific Islander	46	14.8	13	4.2
First Nations/Aboriginal/ Native American	10	3.2	7	2.3
Black/African /African American	7	2.3	38	12.2
Hispanic[a]	4	1.3[b]	33	10.6[b]
Community Size				
Central Place	89	28.7	81	26.1
Urban Fringe	76	24.5	75	24.2
10,000 to 50,000	38	12.3	73	23.6
<10,000	107	34.5	81	26.1

Note: N = 310.

[a] Calculated independently of the other ethnic categories; not included in total percentage.

[b] Significant differences.

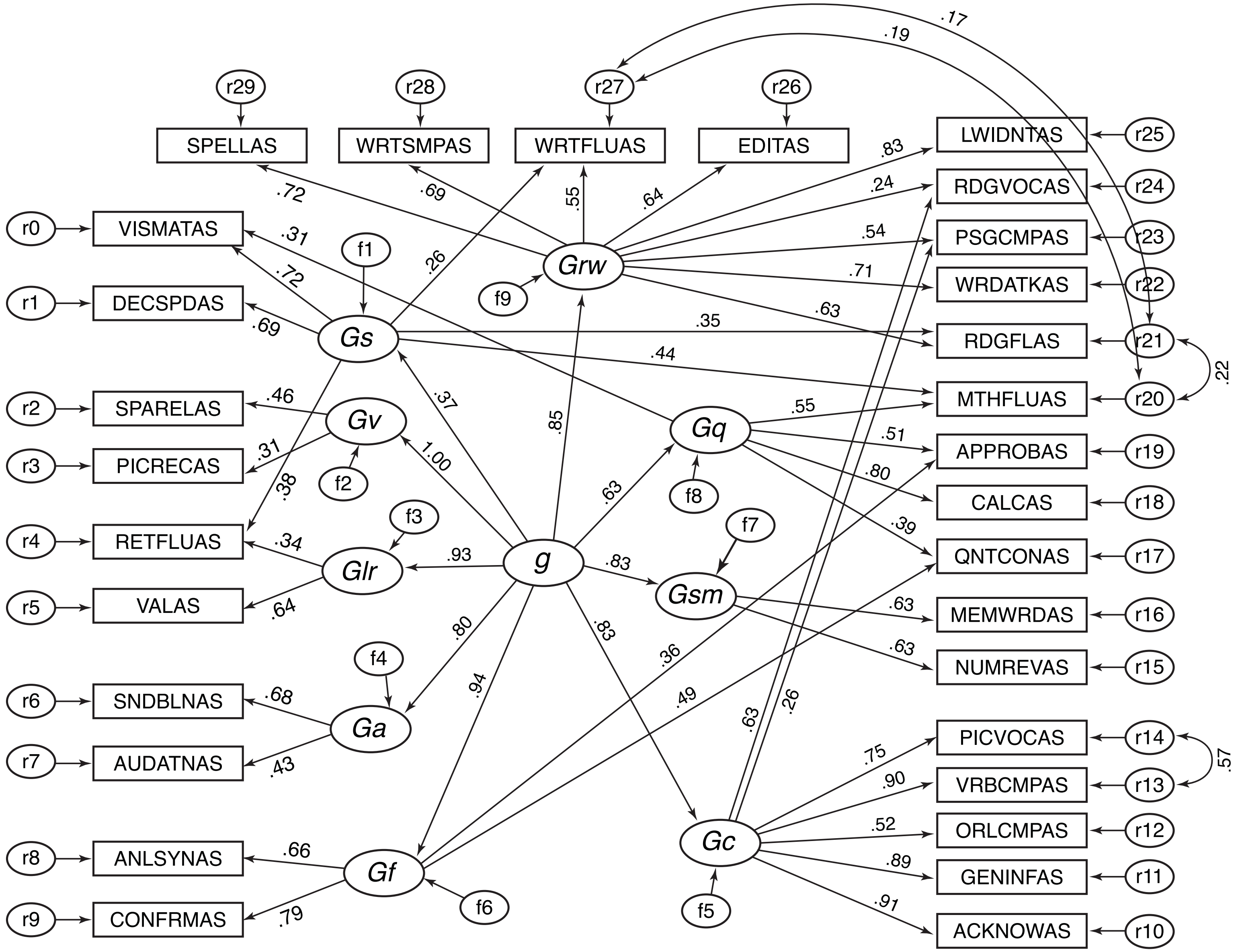

Figure 1. *Represents the Average First- and Second-Order Standardized Path Coefficients (across the Canadian and U.S. samples) of the WJ III CHC Theoretical Factor Structure. Short forms of the test names are used in the model due to space limitations; refer to the Appendix for the corresponding short forms and full test names.*

configural, metric, and/or scalar invariance; Horn, McArdle, & Mason, 1983; Meredith, 1993), three different sets of multiple group CFA analyses were conducted.[1]

In the first set of analyses (Configural), the number of first- and second-order factors and the assignment of the 30 tests to the first-order factors were investigated. This is known as configural invariance and involves fitting a structural model that specifies the same factor structure across groups (Horn et al., 1983). In the Configural model, the pattern of all path coefficients leading from the second-order general factor (*g*) to the 9 first-order broad CHC factors and from the first-order CHC factors to the 30 manifest WJ III COG and WJ III ACH tests were specified to be the same across the two countries. The purpose of this set of analyses was to determine if the 30 selected tests of the WJ III COG and WJ III ACH measure the same latent CHC constructs in Canada as they do in the United States.

Metric, or factorial invariance, is a more restrictive test and is present when configural invariance is extended to include the condition that all factor loadings are equal across all groups (Bollen, 1989). The fit of the metric invariance model is then compared to the fit of the configural invariance model. The finding of a non-significant change in fit (as determined by the difference in the respective model chi-squares and degrees of freedom) supports the null hypothesis that there is not a difference between models and supports the interpretation of metric invariance. In this study, metric invariance was investigated via a two-stage process. In the first test of metric invariance (Invariance 1), the paths from the second-order broad CHC factors (*Gf*, *Gc*, *Glr*, *Gsm*, *Ga*, *Gv*, *Gs*, *Grw*, and *Gq*) to the first-order WJ III COG tests were fixed to be invariant (equal), but the path loadings from the third-order (*g*) factor to the second-order (broad CHC) factors were allowed to be free or vary. The second stage of analysis was the Invariance 2 model, in this model the Invariance 1 model was further constrained to require the factor loadings from the third-order (*g*) general factor to the second-order factors to be invariant across all age groups.[2]

Results

The statistical significance of each model was tested via the obtained chi-square. It is well known that inflated chi-square statistics are often produced in studies that use large sample sizes. This phenomenon is the main reason why a number of additional fit statistics have been developed (Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988). Inflated chi-square statistics in large samples often result in the rejection of an otherwise excellent fitting model.

The results from the three sets of analyses were evaluated using "goodness of fit indices" that provide empirical evidence of the degree of correspondence between the proposed theoretical model and the standardization data from all five age groups (Keith, 1997). The Goodness of

---

[1] Prior to running the invariance models, the modification indices were inspected for possible modifications to the measurement model. As a result, three correlated residual terms were determined to make logical or theoretical sense and were specified and retained across all age groups and models. The three correlated residuals were: Writing Fluency/Reading Fluency; Reading Fluency/Math Fluency; Math Fluency/Writing Fluency. These correlated residual parameters were statistically significant in both the Canadian and U.S. samples and most likely represent shared academic fluency or speed variance.

[2] Although an even more restrictive set of analyses that would test the invariance of error variances and covariances across groups is possible, this degree of invariance is widely accepted as being of little importance and represents an overly restrictive test of the data (Bentler, 1995; Byrne, 2001).

Fit Index (GFI), the Tucker-Lewis Index (TLI, also called the non-normed fit index), and the Comparison Fit Index (CFI) (Keith, 1997; Keith & Witta, 1997; Robles, 1995) were used to evaluate the fit of the models. Values for these indices can range from 0.00 to 1.00, with values of ≥ .95 indicating an excellent fit and fit indices ≥ .90 indicating an adequate fit (Hu & Bentler, 1999).

A final fit index, the Root Mean Square Error of Approximation (RMSEA) statistic takes into account the error of approximation in the population and answers the question "How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?" (Browne & Cudek, 1989; pp. 137–138). Additional advantages of the RMSEA are (a) its sensitivity to the number of estimated model parameters (model complexity) and (b) the provision of 90% confidence intervals that assess the precision of the RMSEA estimates (Byrne, 2001). RMSEA values range from 0.00 to 1.00 with zero indicating no error (a perfect fit). Typically, RMSEA values equal to or less than .05 indicate good fit and values up to .10 suggest adequate or mediocre fit (Byrne, 2001). A wide 90% RMSEA confidence interval suggests that the estimated RMSEA is imprecise, while a very narrow confidence interval suggests a precise RMSEA value (Byrne, 2001).

During the analysis of Model 1, a path coefficient greater than 1.0 was found between general Visual Processing (*Gv*) and *g* in both the Canadian and U.S. groups. This finding, which suggests that *Gv* is isomorphic with *g*, represents a "Heywood" case. Heywood cases are not uncommon in structural equation modeling due to a variety of reasons (Loehlin, 1992; Long, 1983). The most likely cause of the Heywood cases in the current investigation was the inherent and necessary design focus of the investigation—the evaluation of the invariance of a number of *two-test* WJ III COG factor clusters. Standard factor analytic rules-of-thumb recommend three or more indicators per factor to properly "identify" a factor model (Floyd & Widaman, 1995; Marsh, Hau, Balla, & Grayson, 1998; Raykov & Widaman, 1995). Three or more indicators per factor was not possible given the set of 30 tests administered. As a result, these 1.0+ values most likely reflect "empirical under-identification," a situation where a model is nearly identified (Long, 1983). To provide for proper model identification in the current investigation, the error variance associated with the latent *Gv* factor was fixed to zero in both the Canadian and U.S. samples. All the above model specifications were maintained in all subsequent analyses.

The results from the test of configural invariance (Configural) are reported in Table 2. The GFI fit index was .84. Both the CFI and TLI fit indices reported for the Configural model are at approximately .90 (.91 and .90). Collectively the GFI, CFI, and TLI fit indices indicate that the theoretical model provided a very good fit to the data across all age groups. Notable is the RMSEA of .045 (lower and upper 90% confidence interval values of .042 and .048). The hypotheses that the theoretical WJ III CHC model in Figure 1 fits the data from both groups could not be rejected ($p > .05$). The small RMSEA 90% confidence interval (.042–.048) indicates that the RMSEA value of .045 is a precise estimate of good model fit. The results for the two additional metric invariance analyses (Invariance 1 and Invariance 2) are also summarized in Table 2.

As described previously, the fit of the nested and successfully more constrained metric invariance models was evaluated via the difference in the chi-square statistic ($\Delta\chi^2$) and degrees of freedom ($p < .01$). The first test of metric invariance (Invariance 1) evaluated the invariance of the second-order broad factors. As reported in Table 2, the difference between the Configural model and the Invariance 1 model's chi-square ($\Delta\chi^2 = 12.15$) was not significant ($p > .01$), and therefore, the hypothesis that the second-order factor structure was invariant across both groups was not rejected. Furthermore, all fit indices for the Invariance

Table 2

Results from the Tests of Stability and Invariance of the WJ III's CHC Factor Structure across the U.S. and Canadian matched samples

Model	χ^2	(*df*)	GFI	CFI	TLI	RMSEA	(Low-High)	$\Delta\chi^2$	(*df*)	*p*
Configural	1719.06	(768)	.84	.91	.90	.045	(.042-.048)	—	—	—
Invariance 1	1731.21	(776)	.84	.91	.90	.045	(.042-.047)	12.15	8	p > .01
Invariance 2	1786.68	(810)	.84	.91	.90	.044	(.041-.047)	55.47	34	p > .01

Note: Critical values (p =.01) for 8 and 34 *df* are 20.09 and 61.16, respectively.

1 model confirm a strong fit (CFI and TLI values of .91 and .90; RMSEA = .045). The final test of metric or factorial invariance was the most restrictive test of the factor structure of the WJ III and specified that the proposed factor structure (Figure 1) fit the data and that the second-order and third-order path coefficients (i.e., loadings of broad CHC factors on higher-order *g*-factor) are identical across both groups. The hypothesis that the factor structure of the WJ III was invariant across both groups could not be rejected since the change in chi-square ($\Delta\chi^2 = 55.47$) of Invariance 2 model was not significant ($p > .01$). This finding is further supported by the goodness of fit indices presented in Table 2. Finally, although multiple-group CFA requires the constraining of unstandardized parameters in order to allow for formal statistical tests, unstandardized parameter estimates are often difficult to interpret. Thus, the average standardized invariant values across both samples are presented in Figure 1.[3]

Discussion

Since the publication of the WJ III, the transportability of the WJ III NU U.S.-based norms was theoretical and lacking empirical evidence; it was only recently that this was studied and supporting evidence was found. Similarly, until this study, Canadians had been using the WJ III NU without empirical support for the factor structure in a Canadian population. The factor structure of the WJ III of the nine first-order broad CHC factors and from the first-order CHC factors to the 30 manifest WJ III COG and WJ III ACH tests was specified to be the same across the two countries. The purpose of this set of analyses was to determine if the selected 30 tests of the WJ III COG and WJ III ACH measure the same latent CHC constructs in Canada as they do in the United States.

The results of this study, which are based on a matched Canadian/U.S. sample, support the same pattern of loadings across the nine first-order broad CHC factors and from the first-order CHC factors to the 30 manifest WJ III COG and WJ III ACH tests across the two countries. Two significant implications arise from the results of this study.

First, the WJ III's factor structure in a Canadian population was supported. Canadian practitioners can be confident that the nine broad CHC factor scores, which are based on

[3] Given the presence of metric invariance across both samples, it was reasoned that the estimation of the model with the sample variance-covariance matrix derived from *both samples* would provide the most accurate picture of the average standardized loadings across both samples.

U.S. norms, are measuring the same constructs as the test was designed to measure in the United States.

Second, the finding of metric invariance suggests that this particular WJ III COG and WJ III ACH test administration scenario meets Standards 7.1 and 7.8 of the Standards on Educational and Psychological Testing (AERA et al., 1999) which recommend that test scores only be interpreted as having similar meaning across different subgroups if evidence supports the invariant meaning of the scores across the groups. Such evidence was found for the 30-test, nine-CHC factor measurement model in the current investigation. The significance of the metric invariance in the WJ III test-to-factor measurement model across the two countries is important. Factorial invariance has been a fundamental topic of research and debate in psychometrics for decades (Horn et al., 1983; Labouvie & Ruetsch, 1995; Reise, Widaman, & Pugh, 1993). Why? To accurately compare an individual's (or a group's) test scores on the same measures across time, to compare performance on trait measures in different age cohorts, to examine the pattern of correlations between variables in age differentiated groups, or to compare scores across different populations (U.S. vs. Canadian), the tests must measure the same traits across groups. If scores are found not to be comparable across groups (i.e., lack of measurement or metric invariance), then score comparisons may potentially be an artefact and substantively misleading (Reise et al., 1993).

The nine WJ III CHC clusters evaluated in the current investigation meet the basic measurement invariance prerequisite for studying individual and group differences in Canadian populations—that is, the interpretation of these nine WJ III cognitive and achievement clusters have the same meaning in the Canadian context as outlined in the various U.S. WJ III technical and examiner manuals.

References

Ackermann, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Arbuckle, J.L. (2006). *AMOS Users Guide Version 7.0*. United States of America: SPSS Inc.

Bentler, P. M. (1995). *EQS: Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Braden, J., & Alfonso, V. (2003). The *WJ III Tests of Cognitive Abilities* in cognitive assessment courses. In F. Schrank, & D. Flanagan, *WJ III Clinical Use and Interpretations: Scientist-Practitioner Perspectives*. San Diego: Academic Press.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445–455.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahway, NJ: Lawrence Erlbaum Associates.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. New York: Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 122–130). New York: Guilford Press.

Cizek, G. J. (2003). Review of the Woodcock-Johnson III. In B. S. Plake, & J. C. Impara, *The Fifteenth Mental Measurements Yearbook* (pp. 1020–1024). Lincoln, NE: Buros Institute of Mental Measurement.

Cummings, J. (1995). Review of the Woodcock-Johnson Psycho-Educational Battery-Revised. In J. Conoley, & J. Impara, *The Twelfth Mental Measurements Yearbook* (pp. 1113–1116). Lincoln, NE: Buros Institute of Mental Measurement.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler Intelligence Scales and* Gf-Gc *Theory: A Contemporary Approach to Interpretation*. Needham Heights, MA: Allyn & Bacon.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286–299.

Ford, L., Percy, A., & Negreiros, J. (2010, June). Canadian cognitive assessment training. Paper presented at the annual meeting of the Canadian Psychological Association, Winnipeg, Manitoba.

Ford, L., Swart, S., Negreiros, J., Lacroix, S., & McGrew, K. (2010). Woodcock-Johnson III NU tests of cognitive abilities and tests of achievement with Canadian populations. *Woodcock-Johnson III Assessment Service Bulletin Number 12*, Rolling Meadows, IL: Riverside Publishing.

Hicks, P., & Bolan, L. (1996). Review of the Woodcock-Johnson Psycho-Educational Battery-Revised. *Journal of School Psychology*, *6*, 93–102.

Hildebrand, D., & Saklofske, D. H. (1996). The Wechsler Adult Intelligence Scale-Third Edition. *Canadian Journal of School Psychology*, 74–76.

Horn, J. L. (1965). Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities. Unpublished doctoral dissertation, University of Illinois, Urbana, IL.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*, 242–259.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of Intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L. (1986). Intellectual ability concepts. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*, Vol. 3 (pp. 35–77). Hillsdale, NJ: Lawrence Erlbaum.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade (Ed.), *Handbook of Multivariate Psychology* (pp. 645–685). New York: Academic Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *WJ-R Technical Manual* (pp. 197–232). Rolling Meadows, IL: Riverside.

Horn, J. L. (1994). The theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *The Encyclopedia of Intelligence*. New York: Macmillan.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist*, *1*, 179–188.

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

Kaufman, A. S. (2000). Intelligence tests and school psychology: Predicting the future by studying the past. *Psychology in the Schools*, *37*, 7–16

Keith, T.Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 373–402). New York: Guilford Press.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*, 635–650.

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

Labouvie, E., & Ruetsch, C. (1995). Testing for equivalence of measurement scales: Simple structure and metric invariance reconsidered. *Multivariate Behavioral Research*, *30*, 63–76.

Lee, S., & Stefany, E. (1995). Review of the Woodcock-Johnson Psycho-Educational Battery-Revised. In J. Conoley, & J. Impara (Eds.), *The Twelfth Mental Measurement Yearbook* (pp. 1116–1117). Lincoln, NE: The University of Nebraska Press.

Loehlin, J. C. (1992). *Latent Variable Models: An Introduction to Factor, Path and Structural Analyses*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Long, J. S. (1983). *Confirmatory Factor Analysis*. Beverly Hills, CA: Sage.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410.

Marsh, H. W., Hau, K. T., Balla, J.R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181–220.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (2nd ed., pp. 151–179). New York: Guilford.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, (2nd ed., pp. 136–182). New York: Guilford Press.

McGrew, K. S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.

McGrew, K. S., Dailey, D., & Schrank, F. A. (2007). *Woodcock-Johnson III/Woodcock-Johnson III normative update score differences: What the user can expect and why (Woodcock-Johnson III Assessment Service Bulletin No. 9)*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Flanagan, D. P. (1998). *The Intelligence Test Desk Reference (ITDR):* Gf-Gc *Cross Battery Assessment*. Boston: Allyn & Bacon.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual. Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's Frames of Mind and Sternberg's Beyond IQ in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, *3*, 365–384.

Raykov, T., & Widaman, K. F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling*, *2*, 289–328

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.

Robles, J. (1995). Confirmation bias in structural equation modeling. *Structural Equation Modeling*, *3*, 73-83.

Sandoval, J. (2003). Review of the Woodcock-Johnson III. In B. Plake, & J. Impara, *The Fifteenth Mental Measurements Yearbook* (pp. 1024-1027). Lincoln, NE: Buros Institute of Mental Measurements.

Schrank, F. A., Miller, D. C., Wendling, B. J., & Woodcock, R. W. (2010). *Essentials of the WJ III Tests of Cognitive Abilities Assessment* (2nd ed.). New York: John Wiley.

Stankov, L. (2000). The theory of fluid and crystallized intelligence—New findings and recent developments. *Learning and Individual Differences*, *12*, 1–3.

Statistics Canada (1996). *1996 Census of Population*, Ottawa, ON: Statistics Canada.

Taub, G. E., & McGrew, K. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly*, *19*, 72–87.

Wechsler, D. (1996). *Wechsler Intelligence Scale for Children–Third Edition, Canadian*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Adult Intelligence Scale–Third Edition, Canadian*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *The Wechsler Preschool and Primary Scale of Intelligence–Third Edition, Canadian*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children–Fourth Edition, Canadian*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition, Canadian*. San Antonio, TX: The Psychological Corporation.

Woodcock, R. W. (1997). The Woodcock-Johnson Tests of Cognitive Ability—Revised. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 230–246). New York: Guilford.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised*. Rolling Meadows, IL: Riverside

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003, 2007). *The Diagnostic Supplement to the Woodcock-Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside.

Appendix

Descriptions of WJ III NU COG Clusters and Tests Used in the Analyses

CHC Factor and Description	Test and Description
Comprehension-Knowledge (*Gc*): The depth and breadth of a person's acquired knowledge. This factor is analogous to the traditional notion of crystallized intelligence.	**Verbal Comprehension** (Test 1): Comprised of four subtests which together provide a measure of general language development, lexical knowledge, and ability to apply this knowledge on verbal reasoning tasks (VBCMPAS).[i] **General Information** (Test 11): A measure of general acquired (verbal) knowledge (GENINFAS).
Long-Term Retrieval (*Glr*): The ability to store and retrieve—often through association—information, concepts, or facts fluently from memory.	**Visual-Auditory Learning** (Test 2): A paired-associative memory task that measures the ability to encode and retrieve visual-auditory symbolic information. A controlled learning task with corrective feedback (VALAS). **Retrieval Fluency** (Test 12): A set of three open-ended probes that measure the ability to fluently retrieve words within a specified limited period of time (RETFLUAS).
Visual-Spatial Thinking (*Gv*): The ability to store and recall visual stimuli and to synthesize, analyze, manipulate, and perceive visual patterns.	**Spatial Relations** (Test 3): A task requiring the ability to identify which two or three parts that, when combined, form a target visual figure (SPARELAS). **Picture Recognition** (Test 13): A measure of visual recognition memory of figures of common objects (PICRECAS).
Auditory Processing (*Ga*): The ability to discriminate, analyze, and synthesize auditory stimuli.	**Sound Blending** (Test 4): A measure of the ability to synthesize auditory stimuli (phonemes) (SNDBLNAS). **Auditory Attention** (Test 14): A measure of the ability to discriminate sounds in the presence of increasingly distracting auditory stimuli (AUDATNAS).
Fluid Reasoning (*Gf*): Problem-solving in relatively novel situations, particularly those requiring deductive and inductive thinking.	**Concept Formation** (Test 5): An inductive concept rule formation task that also requires mental flexibility. A controlled learning task with corrective feedback and reinforcement (CONFRMAS). **Analysis-Synthesis** (Test 15): A mathematically-based deductive reasoning task that requires the application of rules from a key to the solving of logic problems. A controlled learning task with corrective feedback and reinforcement (ANLSYNAS).
Processing Speed (*Gs*): Speed of mental processing when performing relatively simple cognitive tasks under conditions requiring sustained attention and concentration.	**Visual Matching** (Test 6): A task measuring the ability to rapidly discriminate and identify two identical numbers within a line of numbers (VISMATAS). **Decision Speed** (Test 16): A measure of the ability to rapidly identify pictures of two objects that are conceptually related, from within a line of object pictures (DECSPDAS).
Short-Term Memory (*Gsm*): The ability to consciously store, maintain, and use information presented within a few seconds.	**Numbers Reversed** (Test 7): A working memory task requiring the retention and mental manipulation of a sequence of numbers (NUMREVAS). **Memory for Words** (Test 17): A memory span test requiring the ability to retain and repeat a sequence of unrelated words (MEMWRDAS).

[i] The parentheses at the end of each test description contain the short form of the test's name, which is used in Figure 1 (WJ III CHC Theoretical Factor Structure).

Descriptions of WJ III NU ACH Clusters and Tests Used in the Analyses

CHC Factor and Description	Test and Description
Reading (*Grw*): The ability to identify and comprehend written language.	**Letter-Word Identification** (Test 1): A test requiring the identification and pronunciation of printed letters and words; sight word recognition is used (LWDINTAS).
	Reading Fluency (Test 2): A measure of the ability to read printed statements rapidly and determine whether true or false (RDGFLAS).
	Passage Comprehension (Test 9): A task requiring the identification of a missing key word that makes sense in the context of a written passage (PSGCMPAS).
	Word Attack (Test 13): A measure of the ability to pronounce nonwords that conform to English spelling rules (WRDATKAS).
	Reading Vocabulary (Test 17): A measure of the ability to provide synonyms, antonyms, and completing analogies for words that are read (RDGVOCAS).
Math (*Gq*): The ability to manipulate symbols and to reason procedurally with quantitative information and relations.	**Calculation** (Test 5): A test measuring mathematical ability by performing various mathematical calculations (CALCAS).
	Math Fluency (Test 6): A measure of calculation speed, by adding, subtracting, and multiplying facts rapidly (MTHFLUAS).
	Applied Problems (Test 10): Analyzing and solving practical mathematical problems presented orally (APPROBAS).
	Quantitative Concepts (Test 18): Identifying math terms and forms; identifying number patterns (QNTCONAS).
Written Language (*Grw*): The ability to write meaningfully while applying lexical, grammatical, and syntactical rules.	**Spelling** (Test 7): Spelling letter combinations that are regular patterns in written English (SPELLAS).
	Writing Fluency (Test 8): Formulating and writing simple sentences rapidly (WRTFLUAS).
	Writing Samples (Test 11): Writing meaningful sentences for a given purpose (WRTSMPAS).
	Editing (Test 16): Identifying and correcting errors in written passages (EDITAS).
Oral Language: The ability to understand and comprehend the English Language orally.	**Picture Vocabulary** (Test 14): Requires naming pictures that are pointed to as directed (PICVOCAS).
	Oral Comprehension (Test 15): Listening to a short passage and providing the missing final word (ORLCMPAS).
Special Purpose Cluster	**Academic Knowledge** (Test 19): A measure of information in curricular areas of science, social science, and humanities (ACKNOWAS).