School Psychology Quarterly, Vol. 19, No. 1, 2004, pp. 72-87

A Confirmatory Factor Analysis of Cattell–Horn–Carroll Theory and Cross-Age Invariance of the Woodcock–JohnsonTests of Cognitive Abilities III

Gordon E. Taub
University of Central Florida

Kevin S. McGrew
University of Minnesota

Establishing an instrument's factorial invariance provides the empirical foundation to compare an individual's score across time or to examine the pattern of correlations between variables in differentiated age groups. In the recently published Woodcock–Johnson Tests of Cognitive Ability (WJ COG) and Achievement (WJ ACH) Third Edition (111)the authors provide evidence for the factor structure of the entire battery, but they did not report the formal testing of the factorial invariance of the battery across age groups. In practice, all WJ III tests are generally not administered to a single examinee. The purpose of this study was to investigate the factorial invariance of the WJ COG under one of the most frequent testing scenarios: the calculation of an examinee's General Intellectual Ability Score-Extended (GIA-EXT; a single, global or full-scale score of intelligence) and performance on the seven latent cognitive processing or Cattell–Horn–Carroll (CHC) clusters. The overall results from this study provide support for the factorial invariance of the WJ COG when the **14** tests contributing to the calculation of an examinee's GIA and CHC factors scores are administered. Support is provided for the WJ COG theoretical factor structure across five age groups (ages 6 to **90+** years).

The Woodcock–Johnson Tests of Cognitive Abilities III (WJ III COG) represents the third edition of this widely used battery of cognitive and achievement tests (Woodcock & Johnson, 1977; 1989; Woodcock, McGrew & Mather, 2001). Although the WJ III Tests of Achievement is one of the most frequently used achievement batteries, the WJ III COG, and all other individually administered intelligence batteries, historically are not used as frequently as the Wechsler trilogy (Alfonso, Oakland, LaRocca, & Spanakos, 2000; Kamphaus, Petoskey & Rowe, 2000; Kaufman, 2000; Wilson & Reschly, 1996). However, recently there has been increased interest in the WJ III COG (Alfonso et al., 2000).

The authors thank Measurement/Learning/Consultants and Richard Woodcock for making **data** from the WJ III standardization sample available for this research (http://www iapsych.com).

Address correspondence to Gordon E Taub, PhD, University **of** Central Florida, College of Education, School of Psychology, Department of Child, Family, and Community Sciences, Orlando, FL **32765.** E-mail: gtaub@mail.ucf.edu

A primary reason for the increasing popularity of the WJ COG is the fact that the last two revisions used test design blueprints based on what many believe to be the most empirically supported and theoretically sound model of the structure of human intelligence (Ackermann & Heggestad, 1997; Carroll, 1993; Flanagan, McGrew & Ortiz, 2000; 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], & 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 (McGrew, Werder, & Woodcock, 1991) was grounded in the Cattell–Horn multiple intelligences theory of fluid *(GJ)* and crystallized *(Gc)* abilities (Horn, 1965, 1968, 1985, 1986, 1988, 1994). This is in contrast to other intelligence tests, such as the Wechsler scales, that use an atheoretical measurement model to account for the instruments' latent factor structure. In the latest revision, the WJ III COG test design blueprint is based on the Cattell–Horn–Carroll (CHC) theory of cognitive abilities, an overarching integration of the Carroll Three-Stratum (Carroll, 1993, 1997) and Cattell–Horn *Gf-Gc* models under a common theoretical umbrella.

The CHC model is a hierarchical model of intelligence that consists of three levels or strata. The first level includes over 70 narrow cognitive abilities, which in turn are subsumed by nine to 10 broad abilities. At the apex of the model is a third-order general factor (i.e., Spearman's *g*). The primarily interpretive structure of the WJ III COG operationalizes seven broad CHC abilities via cognitive cluster scores comprised of two tests each designed to measure a different narrow cognitive ability within the respective broad ability domain (McGrew & Woodcock, 2001).

Although the theoretical organization of the WJ III COG remains the same, the WJ III COG represents a significant departure and improvement from the WJ-R. Of the 20 WJ III COG tests, eight are new tests and two were significantly revised (McGrew & Woodcock, 2001). Similar to the WJ-R, the WJ III COG incorporates seven broad factor scores intended to serve as indicators of seven of the broad stratum II CHC abilities. Although the same seven theoretical constructs (*Gf, Gc, Glr, Gv, Ga, Gsm,* and *Gs*) are present in both the WJ-R and WJ III, the composition of all but one (Fluid Reasoning or *GJ*) of the seven WJ III COG cognitive clusters has changed.

Table 1 provides a brief description of each of the seven broad CHC abilities

TABLE 1. WJ III CHC Factor and Test Descriptions

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: Comprised of four subtests, which together provide a measure of general language development, lexical knowledge and the ability to apply this knowledge on verbal reasoning tasks.
	General Information: A measure of general acquired (verbal) knowledge.
Long-Term Retrieval (Glr): The ability to store and retrieve, often through association, information, concepts, or facts fluently from memory.	Visual–Auditory Learning: A paired-associative memory task that measures the ability to encode and retrieve visual–auditory symbolic information. A controlled learning task with corrective feedback.
	Retrieval Fluency: A set of three open-ended probes that measure the ability to fluently retrieve words within a limited period of time.
Visual-Spatial Thinking ***(Gv)*:**The ability to store and recall visual stimuli and to synthesize, analyze, manipulate, and perceive visual patterns.	Spatial Relations: A task requiring the ability to identify which two or three parts that, when combined, form a target visual figure.
	Picture Recognition: **A** measure of visual recognition and memory of common stimuli.
Auditory Processing *(Ga)*:The ability to discriminate, analyze, and synthesize auditory stimuli.	Sound Blending: A measure of the ability to synthesize auditory stimuli (phonemes).
	Auditory Attention: A measure of the ability to discriminate sounds in the presence of increasingly distracting auditory stimuli.
Fluid Reasoning (*Gf*): Problem-solving in relatively novel situations, particularly those requiring deductive and inductive thinking.	Concept Formation: **An** inductive concept rule formation task that also requires mental flexibility. A controlled learning task with corrective feedback and reinforcement,
	Analysis Synthesis: **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.
Processing Speed (Gs): Speed of mental processing when performing relatively simple cognitive tasks under conditions requiring sustained attention and concentration.	Visual Matching: A task measuring the ability to rapidly discriminate and identify two identical numbers within a line of numbers.
	Decision Speed: A measure of the ability to rapidly identify the two objects, from within **a** row of object pictures, that are the most related conceptually.
Short-Term Memory: The ability to consciously store, maintain, and use information presented within a few seconds.	Numbers Reversed: A working memory task requiring the retention and mental manipulation of a sequence of numbers.
	Memory for Words: A memory span test requiring the ability to retain and repeat a sequence of unrelated words.

(continues)

TABLE 1. WJ III CHC Factor and Test Descriptions

measured by the WJ III COG. Table 1 also identifies and describes the 14 tests that contribute to the seven CHC COG clusters and the General Intellectual Ability–Extended (GIA-Ext) cluster score. Although it is possible to obtain a General Intellectual Ability–Standard (GIA-Std) cluster score with the administration of only seven tests from the WJ III COG, to ensure adequate construct representation of the complete CHC model, all 14 tests listed in Table 1 must be administered to obtain both the seven broad CHC cluster scores and a GIA-Ext score. The diagram in Figure 1 depicts the hierarchical factor structure of the WJ III COG based on the administration of the 14 tests contributing to the calculation of a GIA-Ext score. The change in the composition of the seven WJ III COG clusters was driven by the goal to increase the construction representation (and therefore, the construct validity) of the WJ III CHC cognitive cluster scores (McGrew & Woodcock, 2001).

In addition to expanding the breadth of the narrow abilities measured by each WJ III cognitive cluster score, the equally weighted WJ-R Broad Cognitive Ability full-scale IQ cluster was changed to differentially weighted GIA-Std and GIA-Ext cluster scores derived from principal components analysis (McGrew & Woodcock, 2001). The two GIA scores are more operationally consistent with the theoretical characteristics of a stratum III general intelligence (*g*) factor (Carroll, 1993). With these revisions, the WJ III COG now represents an operational measurement model that is closely aligned with contemporary CHC theory (Carroll 1993, 1997). The WJ III COG presumes a hierarchical factor structure comprised of single tests of narrow stratum I abilities, two-test clusters of broad stratum II abilities, and a differentially weighted g-factor composite cluster at stratum III. The WJ III COG is the only individually administered battery of cognitive tests specifically designed to represent an operational measurement model for the CHC theoretical domain.

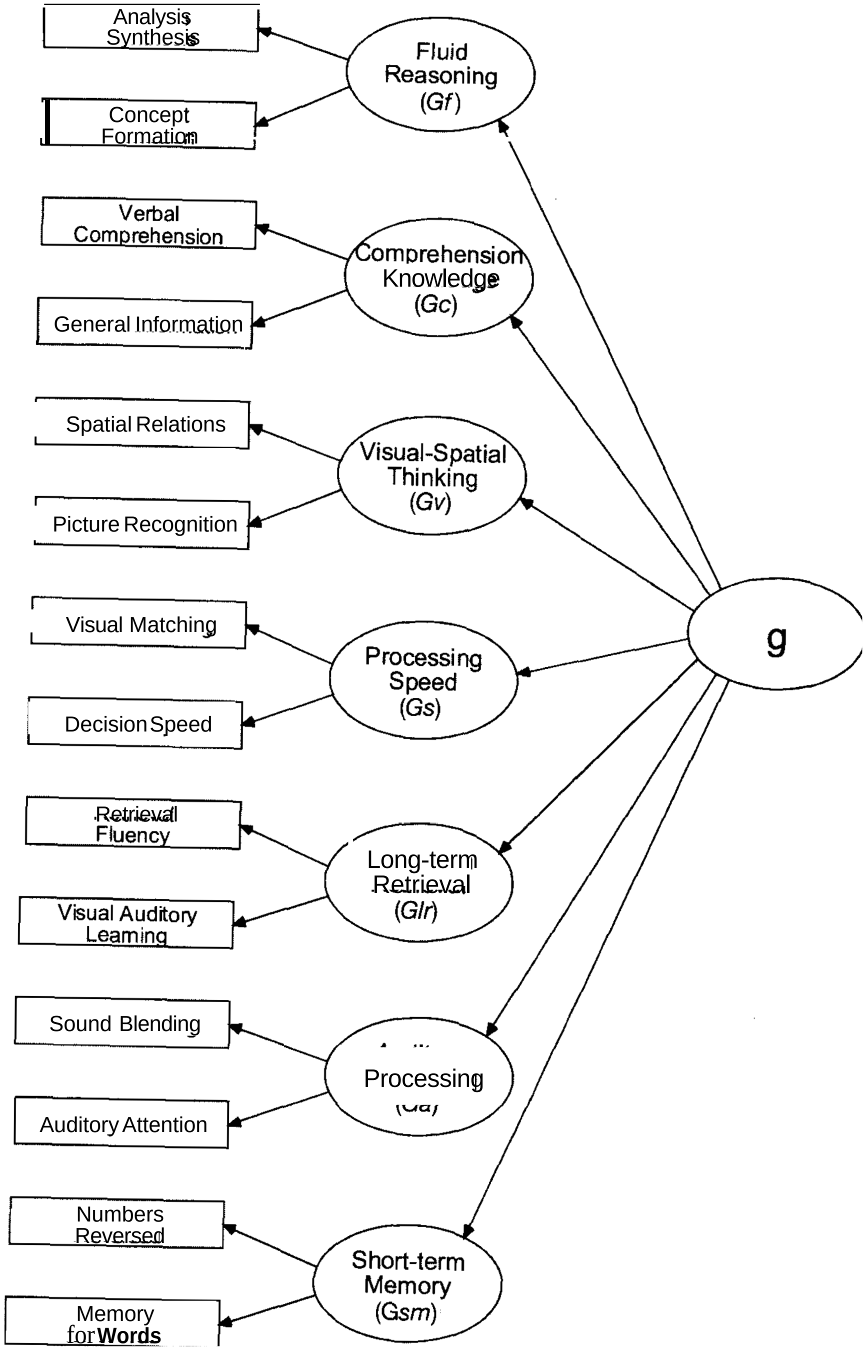

FIGURE 1. Hypothesized CHC Theoretical Factor Structure of the WJ III COG.

In support of the construct validity of the WJ III CHC measurement model, McGrew and Woodcock (2001) presented an extensive set of confirmatory factor analyses (CFA) across five broad age groups (spanning ages *6* through 90+), in addition to a combined sample across all ages. Despite strong structural or internal validity evidence, the WJ III COG examiner's (Mather & Woodcock, 2001) and technical manuals (McGrew & Woodcock, 2001) leave a practical question unanswered. Specifically, "Is there adequate construct representation of the **7** broad CHC factors, across the entire age range, if a clinician administers the 14 tests contributing to the calculation of an examinee's GIA-Ext score?' In the case of the WJ-R, although not provided in the technical manual (McGrew et al., 1991), Bickley, Keith, and Wolfe (1995) found support for the developmental invariance of a hierarchical three-stratum organization of cognitive abilities from ages 2 to 90 and older in the WJ-R norm data.

In practice, the entire 20-test WJ-III COG battery is not generally administered. To obtain an individual's GIA-Ext cluster score and the seven broad CHC cognitive cluster scores, a clinician administers 14 of 20 WJ III COG tests. Data supporting the construct validity of the WJ III COG's CHC factor structure based on this "real world" test administration scenario are not provided in the WJ III technical manual. Furthermore, although the test authors provide multiple group CFA evidence that supports the configural invariance of the complete cognitive and achievement battery, similar evidence is not provided for the configural invariance of the CHC measurement model when only the 14 tests contributing to the calculation of an examinee's GIA-Ext score are administered. More importantly, no empirical evidence is provided to support the interpretation of the 14 primary WJ III COG tests as measures of the same intellectual constructs across the entire age range of the battery (i.e., factorial invariance).

The purpose of the current study was to address several simple, yet significant questions about the structural validity of the 14-test option of the WJ III COG. The primary goal was to investigate the measurement or metric invariance of the WJ III CHC cognitive clusters from age **6** to age 90 or more. This question focuses on the extent to which the 14 primary WJ III COG tests (the measurement model) are equally valid indicators of the seven respective CHC cognitive ability domains across all age groups. The results of these analyses will assist in the identification of any age-related differences present for those tests that are not developmentally invariant. A secondary goal was to investigate the stability or invariance of the seven-factor CHC theoretical model throughout the entire age range of the instrument.

METHOD

Participants

The sample for this investigation was the section of the WJ III standardization sample (ages *6* through 90+) used by McGrew and Woodcock (2001) in their reported CFA studies. As described in the WJ III technical manual, the entire sample was stratified according to race, gender, geographic region, education, and age to ensure that the norm sample mirrored the population characteristics of children, adolescents, and adults in the United States, as described by the U.S. Census projections for the year 2000. The entire WJ III COG was standardized on 8,818 individuals. Given that all of the 14 primary WJ III COG tests investigated in this study do not provide scores during the preschool years, the analyses were restricted to the norm group participants above the chronological age of **5** years. Data from 7,485 of these individuals, the portion of the standardization sample between ages **6** through 90 and older, were included in the present analyses. The five age-differentiated subsamples included ages **6** through 8, ages 9 through 13, ages 14 through 19, ages 20 through 40, and ages 40 to 90 and older.’

Analyses

The present study investigated the stability of the CHC theoretical factor structure for one of the most likely WJ III COG test administration scenarios (see Figure 1). Figure 1 presents both the theoretical factor structure and measurement models of the WJ COG when tests 1 to 7 and 11 to 17 are administered. The variance-covariance matrices for all five age groups served as input data for the analyses and were compared using multiple group CFA methods via the AMOS program (Arbuckle & Wothke, 1999). Given that the evaluation of measurement invariance can take several forms (e.g., configural, metric, or scalar invariance; Horn, McArdle, & Mason, 1983; Meredith, 1993), three different sets of multiple group CFA analyses were conducted.[2]

In the first set of analyses the number of first- and second-order factors and the assignment of the 14 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 seven first-order broad CHC factors and from the first-order CHC factors to the 14 manifest WJ III COG tests was specified to be the same across the five age groups. The purpose of this set of analyses was to determine if the 14-test WJ COG measures the same latent CHC constructs from ages **6** through 90 and older.

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 nonsignificant change in fit (as determined by the difference in the respective model χ^2 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 first-order broad CHC factors (*Gf,* Gc, *Glr,* **Gsm,** *Ga,* **Gv,** Gs) to the manifest WJ III COG tests were fixed to be invariant (equal), but the path loadings from the second-order (*g*) factor to the first-order (broad CHC) factors were allowed to be free or to 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 second-order (g) general factor to the first-order CHC factors to be invariant across all age groups.[3]

[1] The reader should consult the WJ III technical manual (McGrew & Woodcock, 2001) for additional details regarding the tive sample groups.

[2] Before performing the invariance models, the modification indices were inspected for possible modifications to the measurement model. As a result, four correlated residual terms were determined to make logical or theoretical sense and were specified and retained across all age groups and models. The four correlated residuals were: (a) Visual Matching/Numbers Reversed (similar stimuli – numerals); (b) Sound Blending/Memory for Words (common method – use of audio cassette to administer items); (c) Retrieval Fluency/Decision Speed (both require speed of lexical access); (d) Visual Matching/Retrieval Fluency (processing speed).

RESULTS

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 Fit Index (GFI), the Tucker-Lewis Index (TLI, also called the non-normed fit index), the Comparative Fit Index (CFI), and the Adjusted Goodness of Fit Index (AGFI; 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 >.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, whereas a very narrow confidence interval suggests a precise RMSEA value (Byme, 2001).

[3] 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).

In addition to using the fit indices reported above to evaluate the overall fit of the configural invariance model, the statistical significance of each model was tested via the obtained χ^2. It is well known that inflated χ^2 statistics are often produced in studies that use large sample sizes. This phenomenon is the main reason that a number of additional fit statistics have been developed (Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988). Inflated χ^2 statistics in large samples often result in the rejection of an otherwise excellent fitting model. To avoid the rejection of potentially good models within the large samples used in this study (total $n = 7485$), the Differential Fit Value (DFI), a conversion of the χ^2 statistic based on a sample size of 1000, was used to evaluate all models (Keith & Witta, 1997).[4] During the analysis of Model 1, a path coefficient greater than 1.0 was found between general Long-Term Retrieval *(Glr)* and g in all age groups. This finding, which suggests that *Glr* 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 the two-test WJ III COG 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 were not possible given the goal of the investigation, which was to fit models to data containing only two-tests per cluster. 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 *Glr* factor was fixed to zero in all five samples. A similar sample-specific finding occurred in the oldest age group for the *Gv* loading on *g*. As a result, the *Gv* error variance was modified in a similar manner in the oldest age group. These model specifications were maintained in all subsequent analyses.

The results from the test of configural invariance (configural) are reported in Table 2. The GFI, CFI and TLI fit indices reported for the configural model are all above the .95 criteria and indicate that the theoretical model provided an excellent fit to the data across all age groups. Notable is the RMSEA of .025 (lower and upper 90% confidence interval values of .024 to .026. The hypotheses that the theoretical model in Figure 1 fits the data from all age groups of the WJ COG could not be rejected ($p > .05$). The very small RMSEA 90% confidence interval (.024 to 0.25) indicates that the RMSEA value of .025 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 χ^2 statistic and degrees of freedom. The finding of a nonsignificant χ^2 difference supports the interpretation of metric invariance with the one exception of the previously noted loading of +1.0 for *Gv* on g in the oldest adult sample.

TABLE 2. Results from the Tests of Stability and Invariance *of* the WJ III COG's CHC Factor Structure across Age Groups.

Model	χ^2	*(df)*	GFI	CFI	TLI	**RMSEA**	(Low–High)	$\Delta\chi^2$	*(df)*	*p*
Configural	249.22	(337)	.965	.964	.951	.025	(.024–026)	—	—	—
Invariance 1	265.94	(361).	961	.959	.948	.025	(.024–026)	16.72	26	<.05
Invariance 2	304.70	(388)	.958	.956	.948	.025	(.024–026)	55.48	51	<.05

Note. GFI = Goodness **of** Fit Index; CFI =Comparison fit index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. Configural = Unrestricted model; Invariance 1= Invariant **on** first-order factor; Invariance *2* = Invariant *on* first- and second-order factors.

As described previously, the first test of metric invariance (Invariance 1) evaluated the invariance of the first-order broad factors. As reported in Table 2, the difference between the Configural model and the Invariance 1 model's χ^2 (16.72) was not significant ($p > .05$), and therefore, the hypothesis that the first-order factor structure was invariant across all age groups was not rejected. The hypothesis that the proposed model fit the data and the first-order path loadings are identical (with exception of *Gv* loading +1.0 on g in the oldest sample) across all age groups could not be rejected ($p > .05$). Furthermore, all fit indices for the Invariance 1 model confirm an excellent fit (GFI, CFI, and TLI values all greater than .90; RMSEA = .025). The final test of metric or factorial invariance was the most restrictive test of the factor structure of the WJ COG and specified that the proposed factor structure (Figure 1) fit the data and that the first- and second-order path coefficients are identical across all age groups.

It was expected that first- and second-order cognitive ability metric invariance across such a wide age range (6 to 90+) in this investigation would be rejected. However, contrary to expectations, the hypothesis that the factor structure of the WJ COG was invariant across age groups could not be rejected, since the change in χ^2 of the Invariance 2 model (55.49) was not significant ($p > .05$). 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 to allow for formal statistical tests, unstandardized parameter estimates are often difficult to interpret. The average standardized values across all five samples are presented in Figure 2.[5]

[4] The DFV was obtained by applying the formula $((\chi^2) / (n - 1)) \times ((1000 - 1))$. For example, the actual χ^2 for Model 1 was 1867.04. The DFV was calculated by applying the formula $((1867.04)/ (7485 - 1)) \times (1000 - 1)) = 249.22$.

[5] Given the presence of metric invariance across all samples (with the one exception **of** the *Gv/g* loading in the adult 40- to 100-year-old sample), it was reasoned that the estimation of the model with the sample variance-covariance matrix derived from all, would provide the most accurate picture of the average standardized loadings across all five samples. The standardized and unstandardized **path** coefficients for each of the five age-differentiated samples can be obtained by contacting the authors **or** by visiting the website at **http://www.iapsych.com/resrpts.htm.**

DISCUSSION

It is expected that the recently released WJ III, like its predecessors, will generate significant applied and research interest. Current interest in the WJ III is likely due to the battery being an operational measurement model of the CHC Theory of Cognitive Abilities. The technical manual of the WJ 111 COG provides empirical evidence for the structural validity of the battery across five broad age groups. Furthermore, multiple-group confirmatory factor analysis supports the configural invariance of the battery as a function of gender and race in the technical manual (McGrew & Woodcock, 2001). Yet, evidence of the developmental invariance of one of the most likely battery configurations used to interpret an examinee's performance (administering the 14 cognitive tests that provide the 7 broad CHC cognitive ability clusters and the GIA-Ext global intelligence score) is not provided in the test's manuals. The current study was designed to investigate the invariance of this common WJ 111 COG assessment scenario.

The results of this study, which are based on the same samples used by McGrew and Woodcock (2001), support the same pattern of loadings of the 14 primary COG tests on the seven CHC latent factors (configural invariance) in five age-differentiated samples (spanning the age ranges of 6 to 90+). More importantly, the hypothesis that the 14 WJ III COG tests have identical factor loadings on their respective latent CHC factors across the five age groups (metric invariance) was not rejected. The 14-test, seven CHC COG cluster administration of the WJ III appears to be remarkably consistent in what it measures from age 6 through late adulthood.

The results of this investigation have two significant implications. First, the results support the WJ 111 authors' assertion that the WJ 111 can be used across a wide age range (Mather & Woodcock, 2001; McGrew & Woodcock, 2001). Practitioners can be confident that the seven CHC COG cluster scores, which are based on two tests each, are measuring the same constructs from age six through late adulthood. Second, the finding of metric invariance suggests that this particular WJ III COG test administration scenario meets Standards 7.1 and 7.8 of the Standards *on* Educational and Psychological Testing (AERA et al., 1999). These two test standards 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 14-test, 7-CHC factor measurement model in the current investigation.

The significance of the metric invariance in the WJ III test-to-factor measurement model across a wide developmental range should not be lost on the reader. 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, or to examine the pattern of correlations between variables in age differentiated groups, 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 be potentially artifactual and substantively misleading (Reise et al., 1993). The seven WJ III CHC COG clusters meet the basic measurement invariance prerequisite for studying individual and group differences across time.

A comparison of the same respective test factor loadings across all samples (Figure 2) with those reported by McGrew and Woodcock (2001) in their 50-test indicator, nine-factor CFA model 6 for the same age range, found the relative magnitude and pattern of factor loadings to be very similar.[6] For all seven comparable CHC factors across both analyses, the same test was always the highest loading test on its respective CHC factor (Spatial Relations/*Gv*; Visual Matching/Gs; Visual-Auditory Learning/*Glr*; Sound Blending/*Ga*; Concept Formation/*Gf*; Verbal Comprehension/*Gc*; Number Reversed/*Gsm*). These seven tests appear to be the best single indicator measures of each of the seven WJ III COG CHC factors. Not surprisingly, given the design goals of the WJ Ill COG (McGrew & Woodcock, 2001), these seven tests comprise the WJ 111 COG standard battery. The WJ 111 COG GIA-Std cluster appears to be comprised of the best WJ III indicators of each theoretical CHC factor construct.

Although the primary focus of this investigation was not on the invariance of the CHC theoretical model, the structural portion of the models tested (i.e., the loadings of the first-order CHC factors on the second-order g-factor) provides partial, yet tentative, support for Carroll's (1993) conclusion that the CHC model is largely invariant across most of the lifespan.[7] With the exception of a different Gv loading (1.0) on g in the 40 and older adult sample, the relative contribution of each of the first-order CHC factors on the second-order g-factor were identical from ages 6 through late adulthood. As reported in Figure 2, the broad CHC factors most associated with g were Glr (1.0), *Gf* (.92), and Gv (.91). These three factors were followed next by Gsm (.85), Gc (.84), and Ga (.82). Finally, *Gs* (.64) had a noticeably lower g-loading than the other six CHC factors. Although the relative magnitude and pattern of some of these second-order g-loadings are consistent with the extant literature (e.g., high g-loading for *Gf*; lower g-loading for Gs), others are not (e.g., Gc loading of .84 was less than Gv of .91; Carroll, 1993). Given the hypothesized empirical under-identification (only two indicators per first-order factor) that resulted in the 1.0 g-loading for Glr, plus the fact that the relative magnitude and pattern of the current g-loadings differs from those reported in McGrew and Woodcock's (2001), more comprehensive analyses (analyses that had five to 12 tests loading on the different latent factors), we contend that the current structural invariance results are of tentative theoretical

[6] The **CFA** models reported by McGrew and Woodcock (2001) also included broad reading and writing *(Grw)* and quantitative knowledge *(Gq)* factors and indicators.

[7] **If** the primary purpose of the investigation had been on testing the invariance of the CHC theoretical model, the complete **set** of 50 test indicators used in the analyses reported by McGrew and Woodcock (2001) should have been used in this investigation.

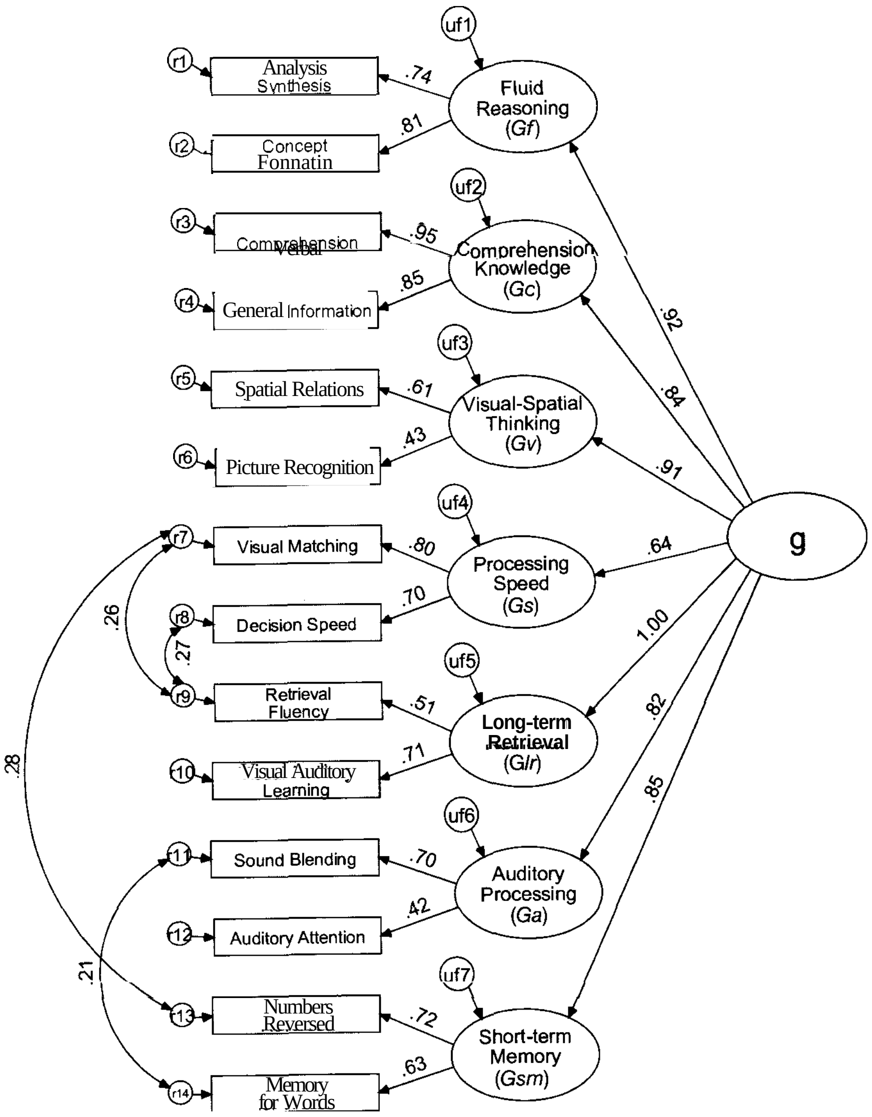

FIGURE 2. Average First- and Second-Order Standardized Path Coefficients (Across All Five Samples; Ages 6 to 100 years) of the WJ III COG CHC Theoretical Factor Structure.

value. Formal testing of the developmental invariance of the CHC theoretical model with all 50-test indicators used by McGrew and Woodcock (2001) is necessary before reaching firm conclusions regarding the developmental invariance of the latent CHC theoretical constructs and overall model.

There are a number of study limitations that suggest room for additional research. The first limitation was the use of only two indicators per factor in the current investigation. This was a forced limitation given the practical focus of the investigation—evaluating the invariance of the two-test WJ III COG clusters. Similar multiple-group CFA invariance evaluations of the same fourteen tests together with additional indicators is a recommended next step. Second, a specific explanation for the 1.0 *Gv* loading on g in the sample of oldest adults (40 to 90+ years of age) is currently undetermined. Research is needed to determine if this finding is a function of the under-identification of the *Gv* factor or a reflection of a fundamental difference in the nature of *Gv* abilities in this age group. Finally, discrepancies between the relative magnitude and pattern of CHC factor loadings on the second-order *g* factor in the current investigation from those reported by McGrew and Woodcock (2001) and those values reported in the extant literature (Carroll, 1993) beg for additional investigation with the complete complement of WJ III indicators. Not only would such an investigation shed additional light on the metric invariance of all the WJ III tests, but also such a properly designed study would make a valuable contribution to the theoretical literature concerning the invariance of the CHC taxonomic framework across the lifespan.

REFERENCES

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

Alfonso, V. C., Oakland, T. D., LaRocca, R., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review, 29,* 52–64.

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

Arbuckle, J. L & Wothke, W. (1999). *AMOS users guide version 4.0.* Chicago: SmallWaters.

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.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M . (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the lifespan. *Intelligence,* 20, 309–328.

Bollen, K. A. (1989). Structural equations with latent variables. New **York:** Wiley.

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.* Mahwah, NJ: Erlbaum .

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1997). The three-stratrum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellecual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford.

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*(3), 286–299.

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 thepsychology of human intelligence* (Vol. 3, pp. 35–77). Mahwah, NJ: Erlbaum.

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

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 indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology Research and Practice,* 31, 155–164.

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. 373402). New York: Guilford.

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 ofmeasurement scales: Simple structure and metric invariance reconsidered. *Multivariate Behavioral Research,* 30(1), 63–76.

Loehlin, J. C. (1992). Latent variable models: An introduction to factor, path and structural analyses. Mahwah, NJ: Erlbaum.

Long, J. S. (1983). *Confirmatoryfactor 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*(3), 391410.

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.

Mather, N., & Woodcock, R. W. (2001). *Examiner's manual: Woodcock-Johnson III Tests of Cognitive Abilities.* Itasca, 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., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock-Johnson psycho-educational battery revised technical manual.* Chicago: Riverside.

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

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's Frames of Mind and Stemberg'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.

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

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review,* ***25,*** 9–23.

Woodcock, R. W., & Johnson, M. B. (1977). Woodcock-Johnson psycho-educational battery. Chicago: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). Woodcock-Johnson psycho-educational battery—Revised. Chicago: Riverside.

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

Action Editor: Cecil Reynolds

Gordon E. Taub, Ph.D., is a faculty member in the School Psychology Program at the University of Central Florida. His major areas of scholarly interest include the structure and development of intelligence, individual differences in real-world success, measurement of educational and psychological constructs, practical intelligence, and issues in school psychology.

Kevin S. McGrew, Ph.D., is a Visiting Professor in the Department of Educational Psychology at the University of Minnesota, the Director of the Institute on Applied Psychometrics (IAP), and coauthor of the Woodcock-Johnson Tests of Achievement and Tests of Cognitive Abilities-Third Edition (WJ-III). Dr. McGrew's specialization is in the areas of applied psychometrics, educational and psychological measurement, intelligence, psychoeducational assessment, personal competence, and the use of educational indicators for policy research in special education.