School Psychology Quarterly, Vol. 20, No. 1, 2005, pp. 66-88

The General (*g*), Broad, and Narrow CHC Stratum Characteristics of the WJ III and WISC-III Tests: A Confirmatory Cross-Battery Investigation

LeAdelle Phelps
State University of New York at Buffalo

Kevin S. McGrew
University of Minnesota

Susan N. Knopik
St. Cloud Community Schools

Laurie Ford
University of British Columbia

One hundred, forty-eight randomly selected children (grades three–five) were administered the WISC-III, WJ III Tests of Cognitive Abilities, WJ III Tests of Achievement, and seven research tests selected from the WJ III Diagnostic Supplement. The validity of the existing WISC-III and WJ III *broad* Cattell-Horn-Carroll (CHC) test classifications was investigated via the application of CHC-organized, broad-factor, cross-battery confirmatory factor analyses (CFA). Likewise, the validity of the WISC-III and WJ III *narrow* CHC ability classifications was investigated via the evaluation of a three-stratum hierarchical (*narrow+broad+g*) CHC CFA cross-battery model. The Tucker-Lewis Index, the Comparison Fit Index, and the Root Mean Square Error of Approximation evaluated the fit for the resulting models. All statistical values indicated good to excellent fit.

Theories of intelligence have been proposed and investigated since the 19th century (Cattell, 1998). For example, Francis Galton (1822–1911) was one of the first to suggest a theory of human ability and the measurement thereof (Thorndike, 1997). In 1904, Spearman proposed a two-factor theory of intelligence consisting of *g* (general) and *s* (specific) abilities. Cattell (1941, 1957) refined Spearman's concept of *g* by identifying two types of general abilities: fluid (Gf) and crystallized (Gc). As further research results accumulated, the Cattell-Horn theory developed when Horn recognized additional broad cognitive abilities: visual perception or processing (Gv), short-term acquisition and retrieval (SAR or Gsm), tertiary storage and retrieval (TSR or Glr), speed of processing (Gs), auditory processing (Ga), quantitative ability or knowledge (Gq), facility with reading and writing (Grw), and correct decision speed (CDS; Horn, 1988, 1989; Horn & Noll, 1997; Horn & Stankov, 1982).

Address correspondence to LeAdelle Phelps, Ph.D., 409 Baldy Hall, Graduate School of Education, SUNY-Buffalo, Buffalo, NY 14260; E-mail: phelps@buffalo.edu

Although other theorists had suggested hierarchical models of intelligence (e.g., Eysenck, 1939, Vernon, 1950), Carroll (1993, 1997, 1998) was the first to complete systematic and exhaustive exploratory factory analyses of over 460 data sets that supported a three-stratum model. At the highest level (Stratum III) is a general factor commonly referred to as general intelligence or *g*. The middle level (Stratum II) consists of *broad* factors including: fluid (Gf) and crystallized (Gc) intelligence, general memory and learning (Gy), visual perception (Gv), auditory perception (Ga), retrieval ability (Gr), cognitive speediness (Gs), and processing speed (Gt). The bottom level (Stratum I) consists of over 60 first-order *narrow* abilities grouped under their respective Stratum II abilities. (To assist the reader, Table 1 lists all factors and related abbreviations and Appendix A identifies common terms utilized in CHC research).

Although the Cattell-Horn and Carroll models are remarkably similar in structure and organization, notable differences are present. In order to create a single Gf-Gc taxonomy to evaluate and interpret intelligence tests, McGrew (1997) proposed an integration of the two models that integrated *g* (general), ten broad abilities (i.e., short-term memory [Gsm], crystallized intelligence [Gc], quantitative knowledge [Gq], reading/writing [Grw], visual processing [Gv], auditory processing [Ga], long-term storage and retrieval [Glr], fluid intelligence [Gf], processing speed [Gs], and decision/reaction time or speed [Gt; McGrew & Flanagan, 1998; Woodcock, 1998]), and 73 narrow abilities. The integrated Cattell-Horn-Carroll (CHC) framework provides a common theoretical nomenclature by which to identify and understand the ability constructs measured by major intelligence batteries.

Currently there are two primary methods for conducting CHC-based intellectual assessments. The first is to use the WJ III (Woodcock, McGrew, & Mather, 2001), a battery of cognitive and achievement tests that were normed simultaneously and used the CHC theory as the theoretical blueprint. The second option is to utilize the CHC-based Cross Battery (CB) assessment procedures (McGrew & Flanagan, 1998). Both the WJ III and CB approaches focus on sampling selectively from all ten of the broad CHC factors.

The validity of CHC inferences drawn from the tests used in the WJ III or CB methods hinges on a series of joint confirmatory factor analytic (CFA) studies. Collectively these empirically based broad CHC CFA studies are referred to as the CHC Broad Confirmatory Factor Analyses (CHC BCFA; Flanagan, McGrew & Ortiz, 2000; McGrew & Flanagan, 1998). All CHC BCFA studies have focused on the classification of individual intelligence tests

TABLE 1. Factors and Abbreviations

Abbreviation	Factor Represented	CHC Broad Abilities	CHC Narrow Abilities
CDS	Correct Decision Speed	Horn	
Ga	Auditory Processing	Horn, Carroll	X
Gc	Crystallized Intelligence	Cattell	X
Gf	Fluid Intelligence	Cattell	X
Glr	Long-term Storage and Retrieval	Horn	X
Gq	Quantitative Knowledge	Horn	X
Grw	Facility with Reading and Writing	Horn	X
Gr	Retrieval Ability	Carroll	
Gs	Processing Speed	Horn	X
Gs	Cognitive Speediness	Carroll	
Gsm	Short-term Memory	Horn	X
Gt	Decision/Reaction time or Speed	Carroll	X
Gv	Visual Processing	Horn, Carroll	X
Gy	General Memory and Learning	Carroll	
KO	General Information		X
LD	Language Development		X
LS	Listening Ability		X
MA	Associative Memory		X
MS	Memory Span		X
MW	Working Memory		X
NA	Naming Facility		X
PC	Phonetic Coding		X
RQ	Quantitative Knowledge		X
SR	Spatial Relations		X
SAR	Short-term Acquisition/Retrieval	Horn	
TSR	Tertiary Storage and Retrieval	Horn	
US	Speech Discrimination		X
VL	Lexical Knowledge		X
VZ	Visualization		X
R3	Rate of Speech Discrimination		X

Note. CHC = Integrated Cattell-Horn-Carroll model (McGrew, 1997). Authors refer to Carroll (1993, 1997, 1998), Cattell (1941, 1957), Horn (1988, 1989), and Spearman (1904).

at the *broad* CHC ability stratum. Conversely, the *narrow* ability classifications underlying both assessment approaches rest on content validity (Flanagan et al., 2000; McGrew & Flanagan, 1998; McGrew & Woodcock, 2001). Collectively, these studies are referred to as the CHC Narrow Ability Classification (CHC NAC). The lack of CHC-designed exploratory or confirmatory factor analytic studies that focus on *both* the broad and narrow classifications of tests is currently a weak link in the validity argument proposed by the authors of the WJ III and CB methods.

One exception to this weak link is the presentation of a tentative and *illustrative* hierarchical three-stratum (narrow, broad, general) factor model in the WJ III norm sample (McGrew & Woodcock, 2001). Although McGrew and Woodcock's three-stratum CFA model provided evidence in support of some narrow WJ III ability test classifications, the hierarchical model could not resolve all narrow ability test classifications. For example, in the broad Gc domain, the three narrow factors all displayed Gc factor loadings that approached 1.0 (Listening Ability [LS] = .96, General Information [K0/K2] = .97, Language Development/Lexical Knowledge [LD/VL] = .98). McGrew and Woodcock concluded that CFA methods, although extremely useful in establishing the construct validity of individual tests, are limited and may be unable to mathematically differentiate highly correlated, yet different, narrow abilities within certain domains. McGrew and Woodcock suggested that other forms of nonfactor analytic evidence must be examined to support the plausibility of complex hierarchical CHC models. For example, McGrew and Woodcock noted that the tests contributing to the narrow Listening Ability (LS) factor demonstrated a distinctly different pattern of growth than the tests that defined the General Information (K0/K2) and Language Development/Lexical Knowledge (LD/VL) factors. Thus, McGrew and Woodcock concluded that despite high Gc factor loadings (.96–.98), the developmental evidence could be viewed as supporting the conclusion that these three highly correlated narrow Gc abilities are not necessarily measuring the same construct. This use of different types of validity evidence (e.g., content, developmental, internal structure) is consistent with recent recommendations that a coherent integration of multiple sources of validity be used to evaluate the validity of interpretations from a test or test battery (AERA, 1999). Nevertheless, McGrew and Woodcock suggested that additional research is needed to support the WJ III *narrow+broad+g* hierarchical model presented in the WJ III technical manual.

In fairness to the authors of the WJ III and CB assessment methods, the sheer number of test indicators required to conduct a proper multivariate study of both narrow and broad characteristics of all tests would be a daunting task. A single definitive study is impractical. Instead, a series of joint test-battery studies, much like those that served as the foundation for the current broad CHC classifications of tests, is needed with an eye toward greater factor specification. The current study represents one step toward this end as it embodies the cross-battery analyses of the WJ III and WISC-III.

The purposes of the current study are threefold. First, the validity of the existing WISC-III and WJ III broad CHC test classifications is investigated via the application of CHC-organized, broad-factor, cross-battery CFA. Second, the validity of the current logically based WISC-III and WJ III narrow CHC ability classifications is investigated via the specification and evaluation of a three-stratum hierarchical (*narrow+broad+g*) CHC CFA cross-battery model. This represents the first-ever three-stratum, CFA CHC-based analyses of a Wechsler/Woodcock data set and should augment the McGrew and Woodcock (2001) three-stratum WJ III model vis-à-vis the inclusion of additional external ability indicators or markers. Finally, a secondary focus of this study, which is possible due to the inclusion of seven of the ten WJ III Diagnostic Supplement tests

(Woodcock, Mather, & Schrank, 2003) in the dataset, is to provide independent evidence bearing on the validity of these measures.

METHOD

Participants

All participants were children who comprised a portion of the third- through fifth-grade samples of the nationally representative standardization sample of the WJ III, which included 8,818 individuals ranging in age from 24 months to more than 95 years of age (McGrew & Woodcock, 2001). A description of the study, supporting letter by the school district superintendent, and a permission form were sent home with all children in grades three through five in three elementary schools in a suburban school district in western New York State. Approximately 90% of the parents consented in writing to allow their children to participate. From these permission forms, 148 children were randomly selected for testing.

The children ranged in age from 8 years, 3 months to 12 years, 4 months (M = 117.5 months, SD = 10.7 months) and consisted of 65 males (43.9%) and 83 females (56.1 %). The final sample was predominately Caucasian (White = 98.6%, American Indian = 0.7%, Asian/Pacific Islander = 0.7%, Hispanic = 2.0%), primarily of middle socioeconomic status (23% high school graduates, 44% one to three years college, 30% bachelors degree of higher), and residing in a suburban area (54% of populations lived in non-urbanized areas)[1].

Chi-square comparisons between the 148 children and all other grade three through five WJ III norm participants revealed no difference by gender (χ^2 = 0.72, df = 1, p. = 0.40).[2] The predominance of Caucasian (98.6%) and non-Hispanic (98.0%) participants differed significantly (Race χ^2 = 45.5., df = 3, p. = 0.001; Hispanic origin χ^2 = 8.20., df = 1, p. = 0.004) from the remaining WJ III grade three through five norm group that were, collectively, more ethnically diverse (Race = 73.1 % Caucasian, 17.5% black, 3.1% Indian, 6.3% Asian/Pacific Islander; Hispanic origin = 9.0 % Hispanic). The education level of the participants' parents was significantly higher (Father Education χ^2 = 25.3., df = 4, p. = 0.001; Mother Education, χ^2 = 21.48., df = 4, p. = 0.001) than the remaining norm group (e.g., 11.5/11.6% of fathers and mothers in the remaining sample were classified as having less than a fifth-grade education or high school diploma, while the research sample had 2.6/3.4%, respectively). Finally, the residence of the research sample participants (as described above) differed significantly (Census location size χ^2 = 202.70., df = 5, p. = 0.001) from the remaining third- through fifth-grade norm group (e.g., only 12.4% and 17.5 % of the remaining norm sample subjects lived in nonurbanized areas populated by less than 49,000 people). In summary, the current sample was biased primarily toward Caucasian students living in suburban areas with parents having above average levels of education.

Instruments and Procedures

The children in this study were administered: (a) 12 subtests (excluding Mazes) of the Wechsler Intelligence Scale of Children-Third Edition (WISC-III; Wechsler, 1991); (b) 18 tests of the WJ III Tests of Cognitive Abilities (WJ III COG), 14 of which produce the General Intellectual Ability Extended (GIA-Ext) score and four of which are clinical tests; (c) seven WJ III Diagnostic Supplement tests; and, (d) the sixth Oral Language and Math tests from the WJ III Tests of Achievement (WJ III ACH). Given that (a) at least two variables (preferably three or more) are required to properly identify a factor (Tinsley & Tinsley, 1987), and (b) the WISC-III includes the Arithmetic test as its only apparent math-related measure, the WJ III Calculation, Math Fluency, and Applied Problems achievement tests were included in the analyses to allow proper identification of the abilities measured by the WISC–III Arithmetic subtest. Furthermore, given that the WJ III ACH Oral Language tests are Gc markers (McGrew & Woodcock, 2001), these tests were included to gain a better understanding of the nature of the WISC-III Verbal and WJ III Comprehension and Knowledge tests. Finally, because neither the WISC-III nor the WJ III COG includes tests that require reading or writing, the WJ III Reading and Writing tests were not included in these analyses.

The WISC-III and WJ III were administered in counter-balanced order over a one-month period. Advanced school psychology graduate students who were supervised by a certified school psychologist employed by the school district completed all testing.

Data Analyses

Prior to the cross-battery analyses, multiple regression analyses were completed to estimate each test's shared variance with the complete set of remaining tests (commonly referred to as a test's communality). These screening procedures were completed because of the acknowledged small subject-to-variable ratio (3.1 to 1) in the current study. As per MacCallum, Wideman, Zhang, and Hong's (1999) conclusion that sample size is less of an issue if the variables in the analyses have relatively high communalities, the communalities for the 44 tests were inspected. Two WJ III tests (Picture Recognition and Visual Closure) had communalities noticeably lower (<.40) than the remaining variables, and were thus deleted from the analyses. In addition, the WJ III Understanding Directions test had a communality approaching unity (.99), a condition that can result in nonconvergence and improper solutions; hence, the Understanding Directions test was also eliminated from the analyses. The final set of 41 tests is listed in Table 2.

---

[1] All sample demographic variables reported for the current sample are based on the U.S. Census category variables employed in the standardization of the WJ III (McGrew & Woodcock, 2001)

[2] The final WJ III grade three through five norm sample was weighted to provide a close approximation to the time period relevant U.S. Census gender figures (see McGrew & Woodcock, 2001).

TABLE 2. Test Means and Standard Deviations

Test	Mean	*SD*
WISC-III		
Picture Completion	10.36	2.85
Information	11.95	2.51
Coding	11.29	2.68
Similarities	11.59	2.75
Picture Arrangement	10.67	3.56
Arithmetic	10.99	2.88
Block Design	11.02	3.39
Vocabulary	10.56	3.20
Object Assembly	10.18	2.75
Comprehension	11.20	3.20
Symbol Search	11.95	3.30
Digit Span	11.09	3.02
WJ III Tests of Cognitive Abilities		
Verbal Comprehension	107.18	13.03
Visual-Auditory Learning	102.64	12.82
Spatial Relations	100.20	14.29
Sound Blending	98.82	13.81
Concept Formation	101.95	10.84
Visual Matching	103.12	14.71
Numbers Reversed	103.46	12.42
Incomplete Words	96.94	16.17
Auditory Working Memory	102.51	11.85
General Information	104.66	11.22
Retrieval Fluency	101.15	12.57
Auditory Attention	100.26	12.70
Analysis- Synthesis	105.67	11.50
Decision Speed	102.59	13.51
Memory for Words	101.21	13.67
Rapid Picture Naming	100.41	10.23
Planning	97.86	12.64
WJ III Diagnostic Supplement Tests		
Memory for Names	98.65	15.61
Sound Patterns-Voice	105.65	13.56
Number Series	103.19	13.86
Number Matrices	104.97	14.40
Memory for Sentences	104.47	13.41
Block Rotation	103.70	16.71
WJ III Tests of Achievement		
Story Recall	107.95	17.92
Calculation	106.99	12.26
Applied Problems	112.08	14.85
Oral Comprehension	106.48	12.22
Math Fluency	104.06	11.48
Academic Knowledge	104.56	11.04

The relatively small subject-to-variable ratio in the current investigation (3.1 to 1) warrants additional comment. Although the common lore surrounding recommended subject-to-variable ratios reveals values ranging from 5:1 to 10:1, a considerable amount of unrecognized diversity of contradictory opinions and evidence exists (Floyd & Widaman, 1995; Guadagnoli & Velicer, 1988; Kenny & McCoach, 2003; MacCallum, Browne & Sugawara,1996; MacCallum, et al.,1999). The wide range of recommendations regarding sample size has typically been stated in terms of minimum necessary sample size (*N*) or the minimum ratio N to the number of variables (*p*) in the analyses (MacCallum et al., 1999).[3] It is important to recognize, however, that explicit subject-to-variable sample size guidelines have "... always been in flux, passed down from generations of factor analysts in an oral tradition" (Floyd & Widaman, 1995, p. 289). A recommended ratio of as low as 5:1 (Streiner, 1994) has been suggested as adequate for sample sizes of 100 or more. According to Kenny and McCoach (2003) "... the effect of number of variables in the model depends on a host of factors such as the type of model being estimated, the type and degree of misspecification, distributions of the variables, the sample size, the estimation method, and the specific measure of fit" (p. 349).

Guadagnoli and Velicer (1988) and Raykov and Widaman (1995) have suggested that the issue is more complex than a simple fixed subject-to-variable ratio and that there is no clear theoretical and/or empirical foundation for most subject-to-variable rules of thumb. Raykov and Widaman (1995) have indicated that other, often overlooked study characteristics, may be even more important than the *n:p* ratio. For example, their research suggests that "variable saturation" with the factors (as indicated by the size of the factor loadings), *together* with sample size *and* the number of variables is more important. Raykov and Widaman (1995) have reported that when most factor loadings are relatively high (e.g., .80 or above), highly stable factor analysis solutions can be found in samples as small as 50, regardless of the number of manifest variables. In addition, MacCallum et al. (1996) presented a useful framework for conducting power analyses for close, not close, and exact model fits in covariance structure models. Briefly, MacCallum et al.'s (1996) power analysis framework indicated that statistical power in structural equation modeling is consistently low when degrees of freedom (*df*) are small, even when sample size (*N*) is relatively large. In contrast, reasonable power is attained in studies with moderate to large *df* and moderate sample sizes, while very high power is achieved in large samples. For instance, MacCallum et al. (1996) reported that power is well above .90 with a sample $n = 200$ and a model with $df = 100$. Extrapolating from the power estimates for select *df* and *N* columns presented in

---

[3] Space does not allow for a thorough review of the recommended sample size literature in factor analysis. The reader is encouraged to review Kenny and McCoach (2003), MacCallum, Browne and Sugawara (1996), and MacCallum, Widaman, Zangh and Hong (1999).

MacCallum et al.'s (1996) Table 2, the current studies *df* (>750) and *n*(148) produce power estimates >.99.[4]

We concur with MacCallum et al. (1996) that standard subject-to-variable rules of thumb need to be replaced with contemporary and sophisticated methods given that more studies with larger numbers of variables are likely to become the norm with the widespread adoption of structural equation modeling and the increasing speed and memory of computers. Subject response burden will likely result in the deliberate design of studies that will require various constraint tradeoffs (time, money, etc.) and will, thus, push the limits of subject-to-variable rules of thumb. "Satisficing" designs[5], rather than perfect designs, will require the presentation of the rationale (as presented above for the current investigation) for the adequacy of a study's sample size and number of variables employed.

The Amos 4.0 computer program was used to specify and evaluate all CFA models (see Keith, 1997 for an overview of the use of CFA and AMOS in the evaluation of the internal structural validity of psychoeducational test batteries). Maximum-likelihood estimation of age-based standard scores was used for all analyses. The two primary a priori models (*broad+g* and *narrow+broad+g*) were specified and evaluated.

The a priori *broad+g* factor-to-test specifications for the WISC-III and WJ III tests were based primarily on the most recent CHC BCFA and NCA literature summaries (Flanagan et al., 2000; Flanagan & Ortiz, 2001) and WJ III CHC insights provided more recently by McGrew (2002) as well as McGrew and Evans (2002). Due to space limitations, copies of both the fully specified a priori models and the subsequent final models are not presented. Figure 1 (*broad+g* model) and Figure 2 (*narrow+broad+g* model) present the final result of the complete modeling process described in the body of this article. The exact a priori model specified for Figure 1 also included the following additional factor-to-test paths: Gf→Similarities; Gc→Picture Arrangement; Gs→Arithmetic; Gf→Block Design; Glr→Retrieval Fluency (instead of Gc→Retrieval Fluency); Glr→Rapid Picture Naming (instead of Gs→Rapid Picture Naming); Gc→Memory for Sentences. The additional Wechsler paths, in particular, were added in light of discrepancies between the historical clinical lore regarding the interpretation of the Similarities, Picture Arrangement, Arithmetic, and Block Design (see Kaufman, 1994) and contemporary CHC-based confirmatory factor studies (refer to Flanagan et al., 2000). The reasons for the changes from these initial specifications to the final models presented in Figures 1 and 2 are discussed below in the Results section.

RESULTS

Broad+g Model

The a priori *broad+g* model was first reviewed for non-significant parameters (*p.* > .05). Factor loadings with critical values (estimate/standard error) less than 1.96 (p. > .05) were eliminated (e.g., Glr→WJ III Retrieval Fluency). The second model "tweaking" step focused on the possibility of adding new test-to-factor paths, but only if a new parameter path made logical or theoretical sense. For example, adding a Gc→Retrieval Fluency was judged consistent with the fact that Retrieval Fluency is a task that requires rapid access and retrieval from a person's lexicon (e.g., vocabulary). The model was then re-estimated and again submitted to another round of review for possible model re-specification. In addition, of interest in the final *broad +g* model was the *deletion* of non-significant paths from Gf→WISC-III Block Design and Similarities, Gs→WISC-III Arithmetic, Gc→WISC-III Picture Arrangement and WJ III Memory for Sentences, Glr→ WJ III Retrieval Fluency and Rapid Picture Naming, and Gv→WJ III Planning. Significant *additional* paths included Gc→WJ III Retrieval Fluency and Gs→WJ III Rapid Picture Naming. The final *broad+g* post hoc test-to-factor specifications are summarized in Figure 1.

A review of all test-to-factor specifications reveals the measurement model for each broad CHC factor. For example, the latent Gc factor was defined by the WISC-III Information, Vocabulary, Similarities, Comprehension, and Picture Completion tests, as well as the WJ III Story Recall, Oral Comprehension, General Information, Academic Knowledge, and Verbal Comprehension tests. The latent Gq factor was defined by four tests (WISC-III Arithmetic, WJ III Calculation, Applied Problems, and Math Fluency). All first-order broad CHC factors were in turn specified to be indicators of a second-order general intelligence *g* factor.

Narrow+broad+g Model

Given the complexity of the mathematical estimation involved in a hierarchical three-stratum model, a decision was made not to start with the specification of a complex model with numerous factorially complex tests (i.e., tests that load on more than one broad factor). Instead, the final results of the *broad +g* model presented in Figure 1 were used to inform the specification of the initial *narrow+broad +g* model. With the exception of the WJ III Retrieval Fluency and Rapid Picture Naming tests, all WJ III and WISC-III tests were specified as

---

[4] Power estimates were obtained from the MacCallum et al.'s (1996) tabled values for *close, not close,* and *exact* fits for sample sizes (*n*) of 100 and 200. All tabled values were entered into a polynomial curve fitting program, and the best fitting polynomial equation (to the tabled values) was then used to generate extrapolated power estimates for the *df* and *n* in the current study. All estimates were well beyond .999.

[5] Economist Herbert Simon first introduced the term *satisficing* (Simon, 1957, 2003), a word that is a combination of the words *satisfying* and *sufficing*. To satisfice is to seek solutions and designs that are "good or satisfactory solutions instead of optimal ones" (Petroski, 2003, p. 8). According to Simon, decision makers (in the current context, researchers) must make choices between optimal decisions for an imaginary simplified world or decisions that are "good enough" (that satisfice) in that they allow a reasonable approximation of the complexity of reality within given constraints (Petroski, 2003).

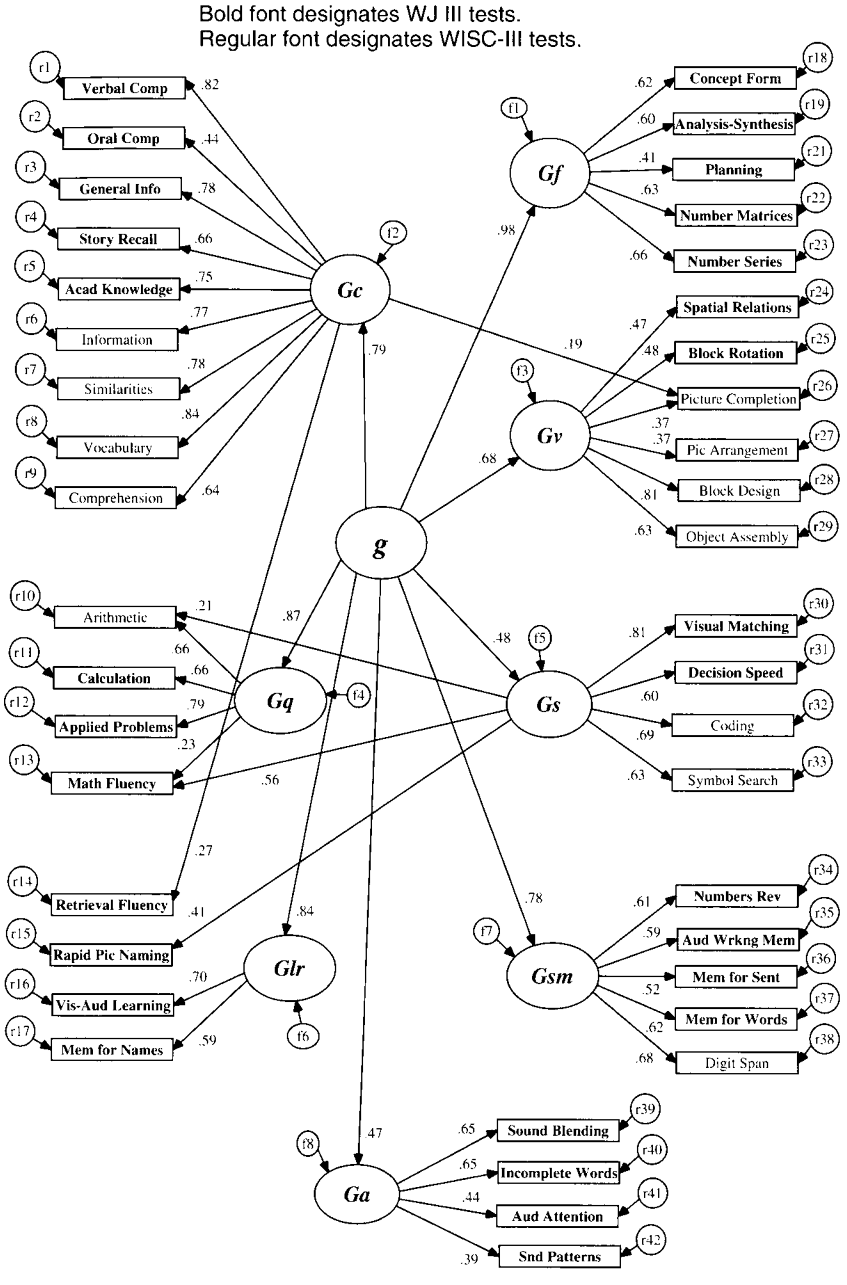
Bold font designates WJ III tests.
Regular font designates WISC-III tests.

FIGURE 1. *Broad+g* Model.

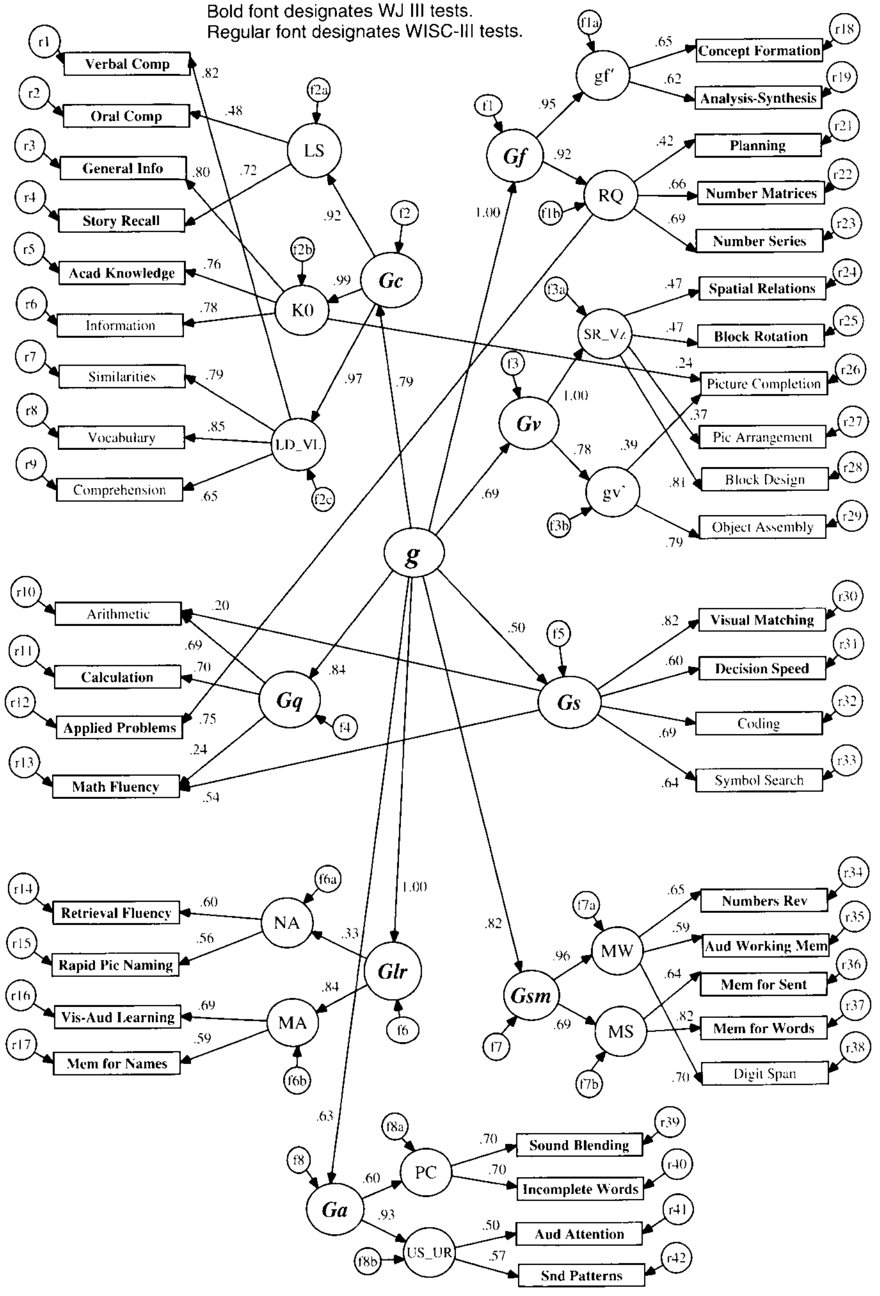
Bold font designates WJ III tests.
Regular font designates WISC-III tests.

FIGURE 2. *Narrow+Broad+ g* Model

indicators of narrow abilities only in the broad domains where they were significant in the *broad+g* final model. Based on the *narrow+broad+g* model presented by McGrew and Woodcock (2001), the first *narrow+broad+g* model specified here included WJ III Retrieval Fluency and Rapid Picture Naming as indicators of Naming Facility (NA) under Glr.

Because of an insufficient number of indicators for some narrow abilities believed to be measured by certain WISC-III and WJ III tests (based on logical content analyses and expert opinion), special "placeholder" narrow factors were constructed. These special factors are designated by lower case italic broad ability codes with a prime notation (e.g., *gf'*) in Figure 2. For example, under the broad Gf factor a *gf'* subfactor was specified that is comprised of the WJ III Concept Formation and Analysis Synthesis tests. These placeholder narrow factors are factorially complex (and uninterpretable) mixtures of narrow abilities within a broad domain.

Similar to the *broad+g* model, post hoc model refinement occurred via the elimination of non-significant paths and the evaluation (and possible inclusion) of alternative factor-to-test specifications. The resultant post hoc model was re-estimated and again submitted to another round of review for possible model re-specification. Four re-specified model iterations were necessary. Figure 2 presents the final *narrow+broad+g* model.

Another issue with the *narrow+broad+g* model was the need to constrain a number of parameters to unity (1.0). First, in the initial model the narrow P factor loading on Gs exceeded 1.0. When constrained to 1.0 (by fixing the residual variance term to 0.0) in the next iteration, the Gs factor loading for *gs'* exceeded 1.0. These findings argued against the specification of separate narrow Gs factors, and thus, resulted in the specification of a single broad Gs factor in the subsequent models. Three other model parameters exceeded unity (Gf and Glr factor loadings on *g* and SR/Vz factor loading on Gv). Thus they were constrained to 1.0 (via the fixing of the respective factor residual error terms [f1, f6, f3a]) to 0.0 in the final model.

A comparison of the significant factor loadings in the final *narrow+broad+g* post hoc model with the original a priori test-to-factor model specification indicates that, similar to the final *broad+g* post hoc model, significant secondary factor loadings were found for Picture Completion (Gc-K0 = .24), Arithmetic (Gs = .20) and Math Fluency (Gs = .54). The WJ III Retrieval Fluency and Rapid Picture Naming tests remained significant indicators of the narrow ability of Naming Facility (NA; .60 and .56, respectively) under the broad Glr factor (not Gc or Gs).

Model Evaluation

Multiple "goodness of fit indices" were used to judge the overall fit of the models. CFA fit statistics provide empirical evidence of the degree of correspondence between the proposed theoretical model and the underlying structure of

the sample data (Keith, 1997). The Tucker-Lewis Index (TLI, also called the nonnormed fit index) and the Comparison Fit Index (CFI; Keith, 1997; Keith & Witta, 1997) 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), was also used to judge the fit of each model. The RMSEA has a number of distinct advantages over the other fit statistics. The RMSEA takes into account the error of approximation in the population and indicates how well the model, with unknown but optimally chosen parameter values, fits the population covariance matrix (Browne & Cudeck, 1989). In addition, the RMSEA is sensitive to the number of estimated model parameters (model complexity) and provides a 90% confidence interval that allows for the evaluation of the precision of the RMSEA estimates (Byrne, 2001). A wide 90% RMSEA confidence interval suggests that the estimated RMSEA is imprecise, whereas a very narrow confidence interval suggests a more precise RMSEA value. RMSEA values range from 0.00 to 1.00 with zero indicating no error (a perfect fit). Typically, RMSEA values ≥ .05 indicate good fit and values up to .10 suggest adequate or mediocre fit (Byrne, 2001).

The fit statistics for the final *broad+g* post hoc and *narrow+broad+g* post hoc models are presented in Table 3. For both models, the TLI and CFI indices were .98, indicating a very good fit of the hypothesized models to the underlying structure in the sample data. Similarly, for both models, the RMSEA values were approximately .04, with the 90% confidence interval lower-bound values slightly lower and upper-bound values of .054 and .055, again suggesting a good fit to the data. The difference in respective AIC values does suggest that the addition of the narrow factors (i.e., *narrow+broad+g* post hoc model) does improve the model fit.

DISCUSSION

The current study offers a research-based model for the hierarchical three-stratum CHC-based, cross-battery interpretation of individual tests from the Wechsler (WISC-III) and Woodcock-Johnson (WJ III) assessment batteries. The implications of the current results are primarily three-fold. First, these data provide consistent validity evidence for the previously identified WISC-III and WJ III *broad* CHC test classifications (see Flanagan et al., 2000). Second, the specification and evaluation of the *g+broad+narrow* CHC model presents much needed empirical information for evaluating the logically based narrow ability WISC-III/WJ III test classifications. Finally, the results of the current study provide new information regarding the *narrow* and *broad* CHC abilities measured by the new WJ III Diagnostic Supplemental tests. The implications of the results are discussed below by shared and unshared WJ III/WISC-III CHC domain cover-

TABLE 3. Model Fit Statistics

Model	χ^2	p	TLI	CFI	AIC	RMSEA	RMSEA (Lo-Up)
Broad+g post hoc model	1031.605 (768)	<.001	0.986	0.987	1299.605	0.048	.040–.055
Narrow+broad+g post hoc model.	999.474 (758)	<.001	0.987	0.988	1287.474	0.046	.038–.054

Note: TLI = Tucker-Lewis Index. Values of ≥ .95 indicate excellent fit.
CFI = Comparison Fit Index. Values of ≥ .95 indicate excellent fit.
RMSEA = Root Mean Square Error of Approximation. Values of ≤ indicate good fit.
RMSEA (Lo-Up) = 90% confidence interval. Narrow intervals indicate estimated RMSEA value is precise.

age. It is critical to note that all narrow ability interpretations that follow hinge on accepting the plausibility of the model presented in Figure 2.

Shared WJ III/WISC-III CHC Domain Measurement

This study documented five areas in which tests from *both* the WISC-III and the WJ III indicated factor loadings (Gc, Gq, Gs, Gsm, Gv). By comparison, Ga (Auditory Processing), Gf (Fluid), and Glr (Long-term Storage and Retrieval) had loadings from only WJ tests (see Figure 1). The CHC domains with *shared* loadings will be discussed first.

Gc. Nine WISC-III and WJ III tests (WJ III Verbal Comprehension, General Information, Oral Comprehension, Story Recall, Academic Knowledge; WISC-III Information, Similarities, Vocabulary, Comprehension) assessed Gc. Of these, eight (all except Oral Comprehension) displayed strong loadings (above .60) on the broad Gc factor. In this study, the only WJ III Gc test classification at variance with prior CHC factor studies was Story Recall with a primary loading on Gc (.61) and not Glr as previously reported (McGrew & Woodcock, 2001).

Regarding the interpretation of more narrow Gc abilities, three of the four WISC-III Verbal tests (i.e., Similarities, Vocabulary, Comprehension) were measures of the narrow abilities of language development (LD) and lexical knowledge (VL; see Figure 2). The remaining WISC-III Gc test (Information), together with WJ III General Information and Academic Knowledge, provided strong indicators of general information (K0; see Figure 2). Hence, both of the WISC-III and WJ III COG primary Gc clusters (i.e., WISC-III Verbal Comprehension Index, WJ III Verbal Ability) appear to tap the same narrow Gc abilities, namely, language development (LD), lexical knowledge (VL), and general information (K0).

Nonetheless, the fact that the three narrow Gc factors (LS, K0, LD/VL) intercorrelated at extremely high levels (.92–.97) suggests that the narrow ability hypothesis offered here should be viewed with considerable caution. Additional nonfactor analytic investigations (e.g., developmental growth curve analyses; neurocognitive studies) are recommended to evaluate the accuracy of the narrow CHC ability interpretations suggested by the model in Figure 2.

Gv. The second CHC broad domain with the largest number of joint indicators was Gv (see Figure 1). At the broad factor level, the WISC-III included a greater proportion of Gv tests (Picture Completion, Picture Arrangement, Block Design, Object Assembly) than the WJ III (Spatial Relations and Block Rotation). In both final models (Figures 1 and 2), the WISC-III Block Design test appeared to be the strongest single indicator of both Gv (factor loading = .81) and the combined narrow abilities of spatial relations and visualization (SR/VZ loading = .81). WISC-III Object Assembly also appeared to be a strong indicator of Gv (factor loading = .63), although its interpretation at the narrow ability level was indeterminate in the current study.

Interestingly, the two WJ III tests hypothesized to measure SR/VZ (Spatial Relations, Block Rotation) displayed moderate factor loadings (.47 or .48) in both final models, suggesting that each test contributes unique ability variance within the Gv domain, and they are not interchangeable. Also, these two WJ III spatial tests appear to measure other sources of ability variance than those measured by the WISC-III Block Design test. Additional research, such as that described in the case of Gc, is needed to further uncover the Gv narrow ability nuances of the WISC-III (Block Rotation) and WJ III (Spatial Relations; Block Rotation) spatial tests.

Finally, the relatively low Gv loadings for WISC-III Picture Completion and Picture Arrangement, plus the additional Gc loading for Picture Completion, confirm the conclusion that these two tests are not strong or relatively pure indicators of Gv abilities. In the context of CHC-defined assessments, the use of the WISC-III Picture Arrangement and Picture Completion tests is discouraged as their scores, when combined with other and better Gv indicators, may confound ability profile interpretation of Gv (see McGrew & Flanagan, 1998 and Flanagan et al., 2000).

Gsm. Five strong Gsm indicators (factor loadings from .59–.82) in both final models (Figures 1 and 2) were present across the WISC-III and WJ III, although the WISC-III included only one measure (Digit Span). The results of the *narrow+broad+g* model (see Figure 2) supported the interpretations of Flanagan et al. (2000) and McGrew and Woodcock (2001) that WJ III Numbers Reversed and Auditory Working Memory Tests are measures of working memory (MW), whereas Memory for Words and Memory for Sentences are measures of memory span (MS). Likewise, the validity of the new WJ III Diagnostic Supplement clusters of Auditory Memory Span (Memory for Words and Memory for Sentences) and Working Memory (Numbers Reversed and Auditory Working Memory) was supported by the final model in Figure 2. As was the case in the prior discussion of the WISC-III and WJ III spatial tests, the fact that the respective factor loadings of each WJ III test of MW (.65 and .59) and MS (.64 and .82; see Figure 2) were high yet sufficiently different indicates that each test contributes additional unique ability variance.

Despite different forward and backward digit recall components, which argue for the logical classification of the WISC-III Digit Span test as a mixed measure of MS and MW (see Flanagan et al., 2000), in the current study Digit Span was found to be a strong indicator (.70) of MW. It is possible that this finding is due to differential stimulus characteristics of the remaining MW and MS factor indicators. The MS narrow factor tests require the repetition of language (words or sentences), whereas the MW indicators primarily require the processing of numerals. Further research on the possible influence of content "facets" on Gsm test performance is recommended. The Berlin Intelligence Structure Model (BIS; Beauducel, Brocke, Liepmann, 2001; Sü, Oberauer, Wittman, Wilhelm, & Schulze, 2002) appears particularly relevant to this type of analysis.

Gs. The presence of relatively strong Gs factor loadings for WJ III Visual Matching and Decision Speed and WISC-III Coding and Symbol Search indicated that both the WJ III and WISC-III processing speed composite measures can be interpreted as valid measures of broad Gs abilities. However, the failure to differentiate Gs at the narrow ability level (Figure 2) leaves unanswered the question of the specific narrow abilities measured by each test. Because empirical evidence suggests that cognitive speed may be more complex than specified by Carroll or Horn (Ackerman, Beier, & Boyle, 2002; O'Connor & Burns, 2003; Stankov, 2000), additional research within the WJ III and WISC-III Gs tests, possibly together with indicators of the broad domain of Gt (Decision/Reaction Time or Speed), is recommended to determine if the respective narrow abilities can be further differentiated.

Gq. Quantitative ability was the last shared CHC broad domain of the WJ III and WISC-III. As with Gs, only a broad Gq model was viable in the current investigation. In both Figures 1 and 2, the WJ III Math Fluency test was found to be a factorially complex measure of Gs (.54 and .56) and Gq (.23 and .24), a finding consistent with prior empirical and logical analyses (Flanagan et al., 2000). The WJ III Calculation test was found to be a strong indicator of Gq (.66 and .70).

The findings for the WJ III Applied Problems and WISC-III Arithmetic tests were of particular interest. Depending on which model is embraced, the Applied Problems test could be interpreted as a measure of quantitative ability (Gq loading = .79) in the broad factor model (Figure 1), or, alternatively, as a strong indicator (.75) of quantitative reasoning (RQ) within the Gf domain in the broad and narrow model (Figure 2). We believe that the most accurate interpretation is most likely similar to McGrew's (1997) expert-based consensus interpretation of Applied Problems as a blended measure of both Gq and RQ abilities. Additional research is clearly warranted.

Finally, at variance with clinical interpretative lore was the finding that the WISC-III Arithmetic test was not an indicator of quantitative reasoning (under Gf). This finding is also at variance with Flanagan et al.'s (2000) classification of Arithmetic as a mixed measure of Gq and Gf. Instead, the Arithmetic test appeared to be a mixed measure of Gq (.66 and .69) and Gs (.20 and .21). The Gs variance may reflect the fact that up to two bonus points are awarded for quick and perfect performance on items 19 through 24.

Unique WJ III CHC Domain Measurement: *Gf, Glr, Ga*

Consistent with the extant Wechsler CHC CFA research (Flanagan et al., 2000), the WJ III provides valid measures of three additional broad CHC domains (Gf, Glr, Ga) not covered by the WISC-III. These are now discussed.

Gf. The significant and strong Gf factor loadings (.60–.65 across the models in Figures 1 and 2) for the WJ III Concept Formation and Analysis-Synthesis tests continued to support the interpretation of these tests (and the WJ III Fluid Reasoning cluster) as strong measures of Gf. A unique contribution of the current study is the support provided for the plausibility of a narrow quantitative reasoning (RQ) ability domain measured by the WJ III Diagnostic Supplement Number Matrices (.66) and Number Series (.69) tests (see Figure 2). These four WJ III Gf tests (Concept Formation, Analysis-Synthesis, Number Matrices, Number Series) provide a diverse array of different Gf indicators by which to supplement the WISC-III vis-à-vis CHC CB procedures.

A potentially important new finding is the consistent significant loading (.41) of the WJ III Planning test on the broad Gf factor (Figure 1), and on the narrow RQ (.42; Figure 2) under Gf. In the WJ III norm-based CFA studies, the Planning test primarily loaded on Gv, but also displayed a tendency to load occasionally on the Gf factor at some age levels (McGrew & Woodcock, 2001). It is hypothesized that the presence of a broader array of Gf and Gv indicators in the current joint data set has produced operational Gf and Gv latent factors with greater breadth and construct validity. Thus, it appears possible that the WJ III Planning test may require fewer Gv abilities (spatial scanning in particular), and more Gf. One hypothesis, based on the consistent finding that working memory may be closely related to Gf (Kyllonen, 1996; Kyllonen & Christal, 1990), is that the WJ III Planning test requires working memory abilities because performance is enhanced if an examinee mentally tries out and evaluates different "forward thinking" solutions prior to implementation. In addition, Planning's loading on the RQ factor suggests that some form of mental counting of the line segments may occur during performance on the test. However, the relatively moderate (.41 and .42) loadings suggest that a significant portion of unexplained Planning test variance remains to be discovered for proper interpretation of this test.

Ga. The moderate to high (.39–.65) Ga factor loadings in Figure 1 provided additional support for interpreting the WJ III Sound Blending, Incomplete Words, Auditory Attention, and Sound Patterns tests as valid indicators of related, yet different, aspects of Ga. The hierarchical Ga structure in Figure 2 suggests the plausibility that the four WJ III Ga tests measure two respective narrow abilities: Phonetic Coding (PC) as indicated by Sound Blending and Incomplete

Words, and Sound Discrimination/Resistance to Auditory Stimulus Distortion (US/UR) with loadings from Auditory Attention and Sound Patterns. This suggests that CB supplementation of the WISC-III should consider selecting one test from each narrow Ga ability domain for adequate Ga construct representation. Although both the Auditory Attention and Sound Patterns tests display moderate to strong US/UR factor loadings (.50 and .57), the Auditory Attention and Sound Patterns tests still measure unique aspects of human functioning that require further study.

Glr. The divergence of factor loadings for two of the WJ III Glr tests (Retrieval Fluency and Rapid Picture Naming) across the models in Figures 1 and 2 continues to suggest the need for further exploration of abilities measured by these two tests. With the *narrow+broad+g* hierarchical model, support is provided for the interpretation of a common NA (naming facility) ability between Retrieval Fluency (NA factor loading = .60) and Rapid Picture Naming (NA factor loading = .56), as both tests share the requirement to rapidly retrieve names from memory (Figure 2). Conversely, the broad factor model (Figure 1) suggests that Retrieval Fluency may be influenced by knowledge (Gc) whereas Rapid Picture Naming is more influenced by processing speed (Gs). Consistent with the WJ III norm-based CFA studies, two WJ III tests (Visual-Auditory Learning and Memory for Names) are strong indicators of Glr (factor loadings of .70 and .59 in Figure 1, respectively), and associative memory (MA) in particular (factor loadings of .69 and .59 in Figure 2, respectively).

Glr. The divergence of factor loadings for two of the WJ III Glr tests (Retrieval Fluency and Rapid Picture Naming) across the models in Figures 1 and 2 suggested the need for further exploration of abilities measured by these two tests. When a theoretically driven hierarchical Glr structure is hypothesized (Figure 2; also see McGrew and Woodcock, 2001), support is provided for the interpretation of a common NA (naming facility) ability between Retrieval Fluency (NA factor loading = .60) and Rapid Picture Naming (NA factor loading = .56), as both tests share the requirement to rapidly retrieve names from memory. Conversely, the broad factor model (Figure 1) suggests that Retrieval Fluency may be influenced more by a person's store of Gc knowledge, while Rapid Picture Naming is more influenced by general cognitive processing speed (Gs). Consistent with the WJ III norm-based CFA studies, the WJ III Visual-Auditory Learning and Memory for Names tests are significant and strong indicators of Glr (Glr factor loadings of .70 and .59 in Figure 1, respectively), and associative memory (MA) in particular (MA factor loadings of .69 and .59 in Figure 2, respectively).

With regard to WISC-III CB assessments, the Glr findings suggest that practitioners should consider selecting either Visual-Auditory Learning or Memory for Names, together with either Retrieval Fluency or Rapid Picture Naming, in order to draw inferences about the broad ability of Glr, a domain that appears to have both rate and level indicators (see Carroll, 1993). The specific combination of tests selected should be guided by referral-specific questions (Flanagan et al., 2001).

Concluding Comments and Caveats

It is important to remember that, as is the case with all CFA research, the strong fit statistics upon which this study's conclusions are based suggest that the current models are plausible—they do not prove that these are *the* correct models. Other caveats regarding these findings are in order. Given the homogeneity (i.e., age, ethnicity, geographic) and number of participants ($N = 148$) in this sample, cross validation is essential (DeVellis, 1991). Until such occurs, the broad generalizability of these results should be made with caution. Likewise, there are notable subtest alterations with the recently published WISC-IV (Wechsler, 2003). Confirmatory factor analyses with the WISC-IV subtests Letter-Number Sequencing, Matrix Reasoning, Picture Concepts, Word Reasoning, and Cancellation will provide much needed refinement to the Wechsler/Woodcock cross-battery approach. It is hypothesized that the WISC-IV Word Reasoning subtest will load on Gc, Matrix Reasoning and Picture Concepts on Gf, Cancellation on Gs, and Number-Letter Sequencing on Gsm. Completing a new set of confirmatory factor analyses with a large diverse population could test these hypotheses. Finally, the next step in CHC research and the cross-battery approach is the linkage between assessment findings and intervention planning (McGrew, Flanagan, Keith, & Vanderwood, 1997). As is true with the entire field of assessment, the linkage of results to treatment utility is sorely lacking. We now need to determine if, and how, the use of the CHC model and CB assessment results in enhanced remedial interventions and possible aptitude-to-treatment interactions.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology, 131*, 567–589.

American Educational Research Association. (AERA). (1999). *Standards on educational and psychological testing.* Washington, DC: Author.

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences, 30*, 977–994.

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: 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.

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–24). Mahwah, NJ: Lawrence Erlbaum Associates.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38*, 592.

Cattell, R. B. (1957). *Personality, motivation structure and measurement.* New York: World Book.

Cattell, R. B. (1998). Where is intelligence? Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice.* (pp. 29–38). Mahwah, NJ: Lawrence Erlbaum Associates.

DeVellis, R. F. (1991). *Scale development: Theory and application.* Newbury Park, CA: Sage.

Eysenck, H. J. (1939). Primary mental abilities. *British Journal of Educational Psychology, 9,* 270–275.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment.* NY: Wiley & Sons.

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

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265–275.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (Rev. ed., pp. 645–685). New York: Academic Press.

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–180). New York: Guilford Press.

Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6,* 65–185.

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. (1994). *Intelligence testing with the WISC-III.* New York: Wiley.

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

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10,* 333–351.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis, & P. Tapsfield (Eds.), *Human Abilities: Their nature and measurement* (pp. 49–76). Mahwah, NJ: Erlbaum.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14,* 389–433.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis anddetermination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

MacCallum, R. C., Wideman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.

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* (pp. 151–180). New York: Guilford Press.

McGrew, K. S. (2002, February). *Advanced interpretation of the Woodcock-Johnson III.* Workshop presented at the meeting of the National Association of School Psychologists,Chicago, IL.

McGrew, K. S., & Evans, J. J. (2002). *Exploring the structure of the WJ III.* (IAP Research Report No. 8). Cloud, MN: Institute for Applied Psychometrics.

McGrew, K. S. & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR) Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 189–210.

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

O'Connor, T. A., & Burns, N. R. (2003). Inspection time and general speed of processing. *Personality and Individual Differences, 35,* 713–724.

Petroski, H. (2003). *Small things considered: Why there is no perfect design.* New York: Knopf.

Raykov, T., & Widaman, K. F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling, 2,* 289–318.

Simon, H. A. (1957). *Models of man.* New York: Wiley.

Simon, H. A. (2003). Nobel prize in economic sciences, 1978. *American Psychologist, 58,* 753–755.

Sü, H. M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261–288.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12,* 35–51.

Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry, 39,* 135–140.

Thorndike, R. M. (1997). The early history of intelligence testing. In D. P. Flanagan, J. L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3–16). New York: Guilford Press.

Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34,* 414–424.

Vernon, P. E. (1950). *The structure of human abilities.* New York: Wiley.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-third edition.* San Antonio, TX:Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-fourth edition.* San Antonio, TX: Psychological Corporation.

Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137–156). Mahwah, NJ: Lawrence Erlbaum Associates.

Woodcock, R. W., Mather, N., & Schrank, F. A. (2003). *WJ III Diagnostic Supplement to the Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishers.

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

Action Editor: John Kranzler

LeAdelle Phelps, Ph.D., is Associate Dean for Academic Affairs with the Graduate School of Education, and Co-Director of the Ph.D. program in Combined Counseling/School Psychology at the University at Buffalo, SUNY, and Editor of *Psychology in the Schools.* She has published numerous articles and book chapters on such diverse health-related topics as eating disorders, prenatal alcohol and cocaine exposure, and lead poisoning. She authored the *Phelps Kindergarten Readiness Scale II,* edited *Health-related Disorders in Children and Adolescents: A Guidebook for Understanding and Educating* (APA, 1998), and co-authored *Pediatric Psychopharmacology: Combining Medical and Psychosocial Interventions* (APA, 2002). The Combined Counseling/School Psychology program at the University of Buffalo, SUNY is accredited by the APA.

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

Susan N. Knopik is a school psychologist for District 742 Community Schools in St. Cloud, Minnesota. Her degrees are in psychology, music therapy (B.A./B.A., College of St. Teresa, Winona, MN) and school psychology (M.S.E., University of Wisconsin-River Falls). She has also been an Adjunct Professor in the Department of Special Education at St. Cloud State University. Her research interests include cognitive abilities, personality, and academic achievement.

Laurie Ford, Ph.D., is an Associate Professor and Coordinator of the School Psychology Program at the University of British Columbia. Her major areas of research interest include cognitive and social-emotional assessment of young children; family, school and community factors that promote early school readiness; interdisciplinary practice; and family and diversity issues.

APPENDIX A: COMMON TERMS AND ABBREVIATIONS UTILIZED IN CHC STUDIES

Abbreviation	Term
CB	Cross-Battery assessment procedure
CHC BCFA	Broad Confirmatory Factor Analyses of the Cattell-Horn-Carroll model
CHC NAC	Narrow Ability Classifications based on the Cattell-Horn-Carroll model
CFA	Confirmatory Factor Analysis
narrow+broad+g	McGrew and Woodcock (2001) three-stratum hierarchical model