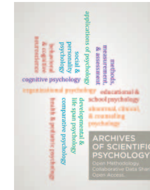
Archives of Scientific Psychology 2018, 6, 1–13
DOI: http://dx.doi.org/10.1037/arc0000039

2169-3269

American Psychological Association

Archives of Scientific Psychology

www.apa.org/pubs/journals/arc



An Alternative Conceptualization of the Theoretical Structure of the Woodcock-Johnson IV Tests of Cognitive Abilities at School Age: A Confirmatory Factor Analytic Investigation

Stefan C. Dombrowski
Rider University

Ryan J. McGill
The College of William & Mary

Gary L. Canivez
Eastern Illinois University

ABSTRACT

Confirmatory factor analysis produced a different theoretical conceptualization for the Woodcock-Johnson IV Tests of Cognitive Abilities (hereinafter, WJ IV Cognitive) in the age range between 9 to 19 years than the actual 7-factor theoretical structure presented in the instrument's technical manual. Using maximum likelihood confirmatory factor analysis, the results of this study indicate that the WJ IV Cognitive measures 4 factors (Verbal, Working Memory, Perceptual Reasoning, and Processing Speed) and aligns with the prior Wechsler intelligence scales conceptualization of cognitive ability. The results also suggest that some caution should be exercised when moving to index-level interpretation as proposed in the *WJ IV Technical Manual*. Subtest alignment is different than what is proposed in the manual and the various indices do not account for sufficient variance for independent interpretive emphasis. The results of this study, therefore, have implications not only for direct CHC-index-level interpretation, but also for clinical interpretive approaches such as cross-battery assessment and processing strengths and weaknesses.

SCIENTIFIC ABSTRACT

The actual 7-factor theoretical structure of the Woodcock-Johnson Fourth Edition (WJ IV) Cognitive was not investigated by the test publisher. Instead, the structure for the WJ IV Cognitive battery was extrapolated from analyses of the full WJ IV test battery. The present study investigated the theoretical structure of the WJ IV Cognitive in isolation, using maximum likelihood confirmatory factor analysis applied to 2 standardization sample age groups (Age 9 to 13 and Age 14 to 19). The results of this study propose an alternate theoretical structure for the WJ IV Cognitive—one that aligns with the prior 4-factor conceptualizations of the Wechsler Scales (i.e., Verbal Ability, Working Memory, Processing Speed, and Perceptual Reasoning) rather than the 7-factor Cattell-Horn-Carroll (CHC) structure posited in the *WJ IV Technical Manual*. Implications for direct application of CHC-index-level interpretation and approaches to clinical interpretation, such as cross-battery assessment and processing strengths and weaknesses, are discussed.

Keywords: Woodcock-Johnson IV, confirmatory factor analysis, bifactor model, CHC theory, intelligence

This article was published February 26, 2018.

Stefan C. Dombrowski, Department of Graduate Education, Leadership & Counseling, Rider University; Ryan J. McGill, School of Education, The College of William & Mary; Gary L. Canivez, Department of Psychology, Eastern Illinois University.

Correspondence concerning this article should be addressed to Stefan C. Dombrowski, Department of Graduate Education, Leadership & Counseling, Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648. E-mail: sdombrowski@rider.edu

The Woodcock-Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014a) is the latest rendition of the Woodcock-Johnson family of tests. Use of the WJ IV in clinical practice is ubiquitous, and it is likely to become one of the more frequently cited cognitive ability instruments in the professional literature. It comprises separate tests of cognitive ability, achievement, and oral language (Schrank, Mather, & McGrew, 2014a, 2014b; Schrank, McGrew, & Mather, 2014a, 2014b). No other cognitive ability instrument contains as many co-normed interrelated instruments. The WJ IV is unique in its use of imputation in creating normative data and other novel analytical techniques to establish the instrument's reliability and validity. The WJ IV battery of tests endeavors to evaluate what is essentially the periodic table of human cognitive abilities, and is utilized for such practices as cross-battery assessment (XBA) and processing strengths and weakness (PSW) analyses (McGrew, LaForte, & Shrank, 2014) in school and clinical psychology. Although not without controversy (e.g., Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016; McGill & Busse, 2017), these practices portend, as suggested by some (e.g., Flanagan, Ortiz, & Alfonso, 2013), to augment the field's understanding of the processes undergirding learning and cognition, allowing for a bridging of the nexus between assessment and intervention. Because the WJ IV also plays a central role in Cattell-Horn-Carroll (CHC) theory building McGrew et al. (2014), and clinical assessment and interpretive practice, it is critically important to understand fully the theoretical/factor structure of the WJ IV Cognitive.

The WJ IV structure was aligned with CHC theory, which McGrew and Woodcock (2001) indicated is an amalgam of Carroll's (1993) three-stratum (3S) theory of cognitive abilities, the work of Horn and Cattell (1966), along with contemporary neuroscience research on memory.[1] Importantly, Cucina and Howardson (2016) noted that Carroll's 3S theory and Horn and Cattell's theory emanate from different theoretical traditions. Carroll's 3S theory posits the existence of a general factor. Cattell and Horn's theory, on the other hand, overlooks the presence of the general factor in favor of lower order factors (initially, Gc and Gf, but later, Gsm, Gs, and others). Thus, Cucina and Howardson suggest that the merging of these two disparate theoretical traditions is an awkward and unnecessary one. Instead, Cucina and Howardson view the third edition of WJ (WJ III; McGrew & Woodcock, 2001) as being more aligned with Carroll's 3S theory, and even support Carroll's own questioning of the need for the creation of CHC theory. In fact, in his last publication, Carroll (2003) explained that the WJ III introduced

> a so-called CHC (Cattell-Horn-Carroll) theory of cognitive abilities that supplemented Horn's Gf-Gc theory with essentially a three-stratum theory similar to that proposed by the present writer (Carroll, 1993). Even though I was to some extent involved in this change (as an occasional consultant to the authors and publisher), I am still not quite sure what caused or motivated it. (p. 16)

The *WJ IV Technical Manual* (McGrew et al., 2014) indicates that the WJ IV Cognitive was designed to measure a hierarchically ordered general intellectual ability factor (i.e., *g*) along with the lower order CHC factors of Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Short-Term Working Memory (Gwm), Cognitive Processing Speed (Gs), Auditory Processing (Ga), Long Term Retrieval (Gltr), and Visual-Processing (Gv). However, McGrew et al. (2014) never separately subjected the WJ IV Cognitive to exploratory or confirmatory factor analytic procedures, or if they were conducted, the results were not reported in the manual. Users of the WJ IV Cognitive must therefore discern the structure of the WJ IV Cognitive from analyses reported for the 47 subtest total battery structure. It may be argued that a latent variable exists regardless of whether it has been tested among a smaller subset of variables. However, the relationship among the variables may change in the presence of a different subset of variables, and so it is important to place greater emphasis on empirical evaluation rather than theoretical conceptualization (Youngstrom & Van Meter, 2016).

An understanding of the internal structure (i.e., structural or factorial validity) of an instrument is critically important. The factor structure of an instrument is more than an esoteric musing of the psychometric researcher. Rather, it provides a statistical rationale for how an instrument should be interpreted. Without full explication of, and factor analytical justification for, structural validity, a clinician or researcher will be less able to properly understand and interpret the scores provided by that instrument.

Because neither exploratory nor confirmatory factor analyses of the WJ IV Cognitive was reported in the *WJ IV Technical Manual* (McGrew et al., 2014), an implied structure for the WJ IV Cognitive was extrapolated from the factor analytic evidence presented in the manual for the full WJ IV test battery. Although extrapolation of factors may be acceptable when attempting to build theory—Carroll (1993) never tested his 3S theory on a single battery, but rather created it by investigating separate sets of tests—extrapolating the structure of a narrower instrument from a broader instrument that contains overlapping but additional subtests that will not be administered by users may require empirical evaluation. It is quite possible, even likely, that patterns of subtest loadings, reproduction of covariance structure, and manifestation of latent factors will be different when analyzed from a subset of a larger correlation/covariance matrix (18 indicators in the WJ IV Cognitive vs. 47 indicators for the WJ IV full battery). Even if it were unequivocally acceptable to extrapolate the structure of an applied cognitive ability instrument from a broader psychoeducational assessment battery, concerns have been raised about the factor structure of the two most recent Woodcock-Johnson batteries of tests (i.e., WJ III and WJ IV) and their potential linkage with the theory said to underlie these measures (CHC).

In a series of articles on the WJ III, Dombrowski and colleagues (e.g., Dombrowski, 2013, 2014a, 2014b, 2015a; Dombrowski & Watkins, 2013) found that the WJ structure *did not* fully align with CHC theory. Specifically, the full test battery was determined to be a six-factor instrument at Age 9 to 13 (Gc, Grw, Gs, Combined Gf/Gq, Ga and Glr) and a five-factor instrument at Age 14 to 19 (Gc, Ga, Gs, Gq and Glr), in contrast to the nine-factor structure promulgated by the publisher. Further, a more focal analysis of the WJ III Cognitive by Dombrowski (2013), using factor extraction decision-making rules, suggested that the measure contained four factors (for example, perceptual reasoning [combined Gf/Gv], verbal ability [Gc], processing speed [Gs], and memory [Glr]) at Age 9 to 13 and not the theoretically posited seven factors. Disregarding factor extraction decision rules and forcing the publisher's proposed theoretical structure, an exploratory factor analysis (EFA) practice that may not be generally accepted as statistically robust resulted in the location of six CHC factors (Glr, Gv, Ga, Gs, Gc, and Gsm) across both age ranges. Regardless of whether four or six factors were uncovered, variance apportionment and omega statistics raised concern about the independent interpretation of the WJ III Cognitive group factors. Thus, the theoretically proposed structure was not adequately supported by independent exploratory factor analyses of the WJ III whether the full

[1] It is noted in McGrew (2009) that the CHC taxonomy was not necessarily created with the intent to become a theory of cognitive abilities, but rather a framework from which subtests might be selected as part of a battery or cross-battery of tests.

test battery, the cognitive test battery, or the achievement test battery were investigated.

Dombrowski and colleagues (Dombrowski, 2013, 2014a, 2014b, 2015a; Dombrowski & Watkins, 2013; McGill & Busse, 2017) also pointed out omissions in the confirmatory factor analysis (CFA) methodology undertaken to establish the evidentiary basis for CHC theory and the WJ III (e.g., lack of analysis of competing CFA models, absence of important CFA fit statistics, and absence of standardized parameter estimates for the measurement model adopted; Dombrowski & Watkins, 2013).

Some of these psychometric omissions are similarly repeated in the *WJ IV Technical Manual* (McGrew et al., 2014). Specifically, there have been a number of concerns raised about the choice of both EFA and CFA procedures used by the WJ IV test authors to examine the structure of the WJ IV full test battery (for thorough discussions, see Canivez, 2016, and Dombrowski, McGill, & Canivez, 2017a, 2017b). Concerns include choice of EFA and CFA, including omission of important statistical information such as variance apportionment, omega coefficients, and lack of analysis of competing models. Also, the validity evidence (i.e., Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013) from the extant literature that supported a different perspective compared with the publisher-provided models was generally overlooked. When using CFA, it is important to use both theory *and* empirical evidence to guide the testing of competing models. On this basis, it appears there is need for additional investigation into the structure of the WJ IV Cognitive.

The need for additional analyses from those presented in the *WJ IV Technical Manual* (McGrew et al., 2014) led Dombrowski et al. (2017a) to examine the factor structure of the WJ IV Cognitive across two standardization sample age ranges (9–13 and 14–19) using EFA (i.e., principal axis factoring with an oblique [promax] rotation) followed by the Schmid-Leiman (SL) orthogonalization procedure. The WJ IV Cognitive was found to have a higher order (*g*) factor with four first-order group factors (presumably, Working Memory, Verbal Ability, Processing Speed, and Perceptual Reasoning). Dombrowski et al.'s EFA-SL analysis did not provide evidence for the seven theoretically proposed factors by the test publisher. The evidence from variance apportionment and omega estimates suggested that the WJ IV Cognitive offers strong measurement of general intelligence but limited specificity at the group factor level across both age ranges. As a result, Dombrowski et al. concluded that a degree of caution should be exercised before interpreting beyond *g*, as the general factor was quite strong and only nominal residual variance remained in the group factors after the general factor variance was accounted for, with the possible exception of the verbal ability (Gc) factor at Age 9 to 13. The CHC factors of fluid reasoning (Gf), visual spatial (Gv), auditory processing (Ga), and long-term memory and retrieval (Glr) were unable to be located. Instead, Dombrowski et al. identified four factors reminiscent of the four-factor structure found in the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV; Wechsler, 2003) that is, Verbal Ability, Working Memory, Processing Speed, and Perceptual Reasoning. Incidentally, Canivez, Watkins, and Dombrowski (2016, 2017) and Dombrowski, Canivez, and Watkins (2017) identified these same four group factors when analyzing the WISC-V (Wecshler Intelligence Scale for Children–Fifth Edition; Wechsler, 2014), rather than the five theoretically proposed CHC factors. The finding of distinct Verbal, Memory, Processing Speed, and Perceptual Reasoning factors is consistent with independent research on the WJ III Cognitive, in which Dombrowski's (2013) EFA-SL analysis identified only four factors (Verbal, Memory, Perceptual Reasoning, Processing Speed) at Age 9 to 13 and three factors (Verbal, Processing Speed, Memory) at Age 14 to 19.

It has long been posited that EFA and CFA are complementary procedures that provide answers to different empirical questions. In fact, it is typical to use EFA procedures to suggest a possible structural disposition for an instrument with a subsequent CFA analyses to further investigate the plausibility of the structure uncovered by EFA. This is the practice suggested by many resources in the factor analysis literature (Brown, 2016; Gorsuch, 1983; Thompson, 2004). However, since its publication, the WJ IV Cognitive has only experienced one independent EFA investigation and has yet to be subjected to an independent CFA investigation, suggesting that understanding of the underlying structure for the measurement instrument remains incomplete.

Accordingly, the present study used CFA with maximum likelihood estimation to examine the theoretically proposed WJ IV Cognitive higher order seven-factor structure, while also examining rival bifactor, higher order, and oblique models across three-, four-, and seven-factor models. The three-factor model is consistent with Woodcock's cognitive processing model (CPM; Taub & McGrew, 2014), and the four-factor model was suggested by Dombrowski et al.'s (2017a) EFA-SL study. The seven-factor model is consistent with that theoretically proposed within the *WJ IV Technical Manual* (McGrew et al., 2014). The factor alignment of subtests for Woodcock's CPM and the publisher's proposed seven-factor model is presented in Table 1. The subtest-factor alignments for the four-factor tested models are presented in the subsequent tables.

Also presented are model-based reliability estimates (omega-hierarchical and omega-hierarchical subscale) and Hancock's and Mueller (2001) H, a measure of construct replicability. These statistics enable investigation of the plausibility of interpreting group and general factors. Given the role that the WJ has played in the development of CHC and other closely related interpretive approaches (e.g., XBA) over the last decade, it is believed that the present results are instructive for establishing evidence-based interpretive procedures for the WJ IV as well as its potential utility as a CHC reference instrument in the school and clinical psychology research literature.

Method

Participants

The *WJ IV Technical Manual* (McGrew et al., 2014) reported information relative to the age groups of 9 to 13 and 14 to 19 years. The WJ IV was normed on a nationally representative sample of 7,416 participants from Age 2 to 90 plus controlling for census region, gender, country of birth, race, community type, parent education, and occupational level. (For further detailed demographic information please refer to the *WJ IV Technical Manual*). In the present study, two school-aged (9 to 13 years and 14 to 19 years) subtest correlation matrices (18 × 18) and descriptive statistics were obtained from the technical manual.[2] The 9 to 13 and 14 to 19 age groups contained an average of 1,572 and 1,685 participants, respectively.

Analyses

Mplus 7.4 (Muthén & Muthén, 1998–2015) was used to conduct CFA using maximum likelihood estimation. Covariance matrices were produced for CFA using the correlation matrices from the WJ IV standardization sample for Ages 9 to 13 and 14 to 19. Some first-order factors, particularly for the publisher's suggested seven-factor WJ IV CHC models, were underidentified because they were measured by

[2] The Age 9 to 13 and Age 14 to 19 correlation matrices are publically available in McGrew et al. (2014, pp. 311–312, Tables E-3 and E-4).

Table 1
Subtest Alignment for the Three- and Seven-Factor Tested Models (Oblique, Higher Order, and Bifactor)

Seven-factor publisher proposed (oblique, higher order, and bifactor)			
Gc	**Gf**	**Gsm**	**Gv**
Oral Vocabulary	Analysis-Synthesis	Memory for Words	Visualization
General Information	Concept Formation	Object-Number Sequence	Picture Recognition
	Number Series	Numbers Reverse	
	Verbal Attention		
Ga	**Gs**	**Glr**	
Nonword Repetition	Letter-Pattern Matching	Visual-Auditory Learning	
Phonological Processing	Number-Pattern Matching	Story Recall	
	Pair Cancellation		

Woodcock's cognitive processing model (three-factor oblique, higher order and bifactor)		
Gc	**Gf**	**Gs**
Verbal Attention	Visualization	Number Series
Memory for Words	Visual-Auditory Learning	Letter-Pattern Matching
Object-Number Sequence	Picture Recognition	Number-Pattern Matching
Oral Vocabulary	Analysis-Synthesis	Pair Cancellation
General Information	Concept Formation	
Nonword Repetition	Story Recall	

Note. Gwm = working memory; Ga = auditory processing; Gv = visual-spatial thinking; Glr = long-term retrieval; Gf = fluid reasoning; Gs = processing speed; Gc = comprehension-knowledge/crystallized ability.

only two subtests. In those cases, the two subtests were constrained to equality prior to estimating bifactor models (Brown, 2016; Little, Lindenberger, & Nesselroade, 1999). For the oblique and higher order models, the unstandardized loadings for the initial group factor indicator were scaled to 1.0. For the bifactor models, the unstandardized loading for the general factor and initial group factor indicators was freed, the general and group factor variance was set to 1.0, and the covariance among the group factors was set to zero.

Several models were tested, including the seven-CHC-factor higher order model that was theoretically proposed, but not examined, in the *WJ IV Technical Manual* (McGrew et al., 2014). Additionally, Woodcock's three-factor CPM (Taub & McGrew, 2014) and the four-factor EFA-SL models identified by Dombrowski et al. (2017a) were evaluated. Oblique, higher order, and bifactor structures were evaluated across the three-, four-, and seven-factor models.

Overall model fit was evaluated using the comparative fit index (CFI), standardized root mean squared residual (SRMR), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). Although there are no universally accepted metrics of model fit (McDonald, 2010), higher values indicate better fit for the CFI and TLI, whereas lower values indicate better fit for the SRMR and RMSEA.Hu and Bentler's (1999) criteria for adequate model fit were inspected (i.e., CFI and TLI $\geq$ .90 along with SRMR $\leq$ .09 and RMSEA $\leq$ .08). Good model fit required CFI and TLI $\geq$ 0.95, with SRMR and RMSEA $\leq$ 0.06 (Hu & Bentler, 1999). Models that were considered superior had to exhibit adequate to good overall fit along with meaningfully better fit (ΔCFI $>$ .01 and ΔRMSEA $>$ .015) than alternative models (Chen, 2007; Cheung & Rensvold, 2002). The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were also considered in determining model superiority. The AIC and BIC do not have meaningful scales, but the model with the smallest AIC values is most likely to replicate (Kline, 2016) and would be considered superior. To determine meaningful model differences ΔAIC $>$ 10 (Burnham & Anderson, 2004) was also referenced.

Model-based reliabilities were estimated with coefficients omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}; Reise, 2012). Omega-hierarchical is the model based reliability estimate for the General Intelligence factor with the variability of group factors removed. Omega-hierarchical subscale is the model based reliability estimate of a group factor with all other group factors *and* general factor removed (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012). Omega-hierarchical and omega-hierarchical subscale coefficients were estimated using Watkins's (2013) Omega program. The omega coefficients should exceed .50, but .75 is preferred (Reise, 2012; Reise, Bonifay, & Haviland, 2013) to indicate sufficient construct based reliability for independent interpretation of a group or hierarchical factor.

Additionally, an index of construct reliability or replicability (called H; Hancock & Mueller, 2001) that furnishes an estimate of the reliability of the underlying factor by reflecting the proportion of variability in the construct explained by its own indicators was utilized. High H values ($>$.80) suggest a well-defined latent variable that portends stability across studies. Rodriguez et al. (2016) indicated that it is difficult to specify group factors within a single instrument and it should only be done when H values are higher than .70. Further, when H values are large, it might be useful to utilize a weighted composite score instead of a unit-weighted composite score. H was estimated using the formula offered by Hancock and Mueller (2001).

Results

Age 9 to 13

CFA results for the 18 WJ IV Cognitive subtests are presented in Table 2. These results included a unitary g factor model, and three-, four-, and seven-factor models according to oblique, higher order, and bifactor structures. The three-factor model was examined to consider Woodcock's CPM (Taub & McGrew, 2014). The publisher's espoused higher order model that included seven first-order CHC factors resulted in an inadmissible solution (i.e., Heywood case and a negative variance estimate for the Ga factor), suggesting possible misspecification of the model. Imposing a constraint of zero on the Ga variance estimate allowed the model to converge, but Hair, Anderson, Tatham, and Black (1998) suggested that employing such procedures "only masks the underlying problem" (p. 610), indicating that this model "should not be trusted" (Kline, 2016, p. 237). As a result, fit

Table 2
CFA Fit Statistics for the Age 9 to 13 Group

Model	χ^2	*df*	CFI	TLI	SRMR	RMSEA	RMSEA 90% CI	BIC	AIC
g	4,084.860	135	.679	.636	.078	.132	[.128, .135]	144,444	144,152
Three-factor (Gc, Gf, Gs) bifactor	Inadmissible model—not positive definite covariance matrix, negative residual variance OV –.14								
Three-factor oblique (Gc, Gf, Gs)	2,700.708	132	.791	.758	.060	.108	[.104, .111]	143,083	142,774
Three-factor (Gc, Gf, Gs) higher order Woodcock cognitive processing model	2,794.331	132	.784	.749	.068	.11	[.106, .113]	142,867	143,177
Seven-factor oblique	Inadmissible model—not positive definite covariance matrix, negative residual variance (Gwm loads Ga 1.041; OV loads Gc .989)								
Seven-factor higher order (theoretically proposed but not tested by publisher)	Inadmissible model—not positive definite covariance matrix, Ga loads g 1.02 & neg residual variance –.039								
Seven-factor bifactor	2,289.285	124	.824	.783	.059	.102	[.098, .106]	142,730	142,378
Four-factor oblique	2,073.499	129	.842	.813	.062	.095	[.091, .098]	142,478	142,152
Four-factor higher order	2,077.775	131	.842	.815	.062	.094	[.090, .098]	142,467	142,153
Four-factor bifactor (Dombrowski, McGill, & Canivez, 2017a; Age 9 to 13)	1,665.445	118	.874	.837	.046	.088	[.085, .092]	142,152	141,766

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval; BIC = Bayesian information criterion; AIC = Akaike information criterion; *g* = general intelligence. Gwm = working memory; Ga = auditory processing; Gv = visual-spatial thinking; Glr = long-term retrieval; Gf = fluid reasoning; Gs = processing speed; Gc = comprehension-knowledge/crystallized ability; OV = Oral Vocabulary.

index information and standardized estimates for this post hoc adjusted model are not included.

The bifactor model with four group factors suggested by the Dombrowski et al. (2017a) proposed structure based upon principal axis factoring with promax rotation and Schmid-Leiman orthogonalization produced an admissible solution and was also superior to all other models, including the three-factor and seven-factor oblique, higher order, and bifactor models. Although some of the fit statistics (e.g., CFI, TLI) produced by the bifactor model were not well fitting based on Hu and Bentler's (1999) criteria, they were superior to the same fit statistics from all of the models presented in the *WJ IV Technical Manual* (McGrew et al., 2014). Also, the four-bifactor model produced SRMR and RMSEA fit statistics that could be characterized as good and acceptable, respectively. Considering that the SRMR is fairly low, this may suggest that the difference between the model implied and observed correlation matrices are not too large; thus, the misfit of the CFI/TLI likely does not come from that source. Instead, it is feasible that the baseline to which the fitted model is compared in the CFI/TLI does not fit too poorly, so relative to the baseline model, the fitted model has less room to improve.

Figure 1 presents standardized loadings for the bifactor measurement model with four group factors. Table 3 presents sources of variance for the 18 WJ IV Cognitive subtests according to the bifactor model with four group factors. As indicated, most subtest variance is associated with the General Intelligence factor and substantially smaller portions of variance are uniquely associated with the four WJ IV group factors (presumably, Working Memory, Perceptual Reasoning, Processing Speed, and Verbal Ability). Omega-hierarchical and omega-hierarchical subscale coefficients were estimated based on the bifactor results from Table 3. The ω_H coefficient for General Intelligence (.803) was sufficiently high for confident general factor interpretation of a unit-weighted composite based on the specified indicators. The ω_{HS} coefficients for the four WJ IV cognitive group factors (Working Memory, Perceptual Reasoning, Processing Speed, Verbal Ability), however, were considerably lower, ranging from .221 (Working Memory) to .444 (Verbal Ability). Thus, unit-weighted composites based on the specified indicators of the four WJ IV Cognitive group factors would contain too little true score variance to support clinical interpretation, with the possible exception of the Verbal Ability factor (Reise, 2012; Reise et al., 2013). Results from Hancock and Mueller's (2001) H index were similarly low for confident clinical interpretation of any score except for the general factor.

Age 14 to 19

CFA results for the 18 WJ IV Cognitive subtests in the Age 14–19 group are presented in Table 4. These results included a unitary *g* factor model, and three- (i.e., Woodcock's CPM), four-, and seven-factor models according to an oblique, higher order, and bifactor models. The publisher's espoused higher order model that included seven CHC group factors resulted in an inadmissible solution (i.e., Heywood case and a negative variance estimate for the Ga factor) potentially caused by misspecification of the model. As with the Age 9 to 13 analysis, constraining the Ga variance estimate to zero allowed the model to converge, but this solution is problematic. Hair et al. (1998) note that this "only masks the underlying problem" (p. 610), indicating that this model "should not be trusted" (Kline, 2016, p. 237). Accordingly, fit index information and standardized estimates for this post hoc adjusted model are not included.

The bifactor model suggested by the Dombrowski et al. (2017a) Age 14 to 19 proposed structure based upon principal axis factoring with promax rotation and Schmid-Leiman orthogonalization produced an inadmissible solution due to a negative residual variance attributable to Oral Vocabulary. However, the bifactor structure with four group factors from Dombrowski et al.'s Age 9 to 13 EFA-SL analysis, in which Phonological Processing loads on the Gwm factor and Number Series loads on the Gs factor, converged and was superior to all other models tested, including the three-factor and seven-factor oblique, higher order, and bifactor models. Like the Age 9 to 13 Group CFA, the resulting four bifactor-model displayed CFI/TLI fit statistics that were poor, but SRMR fit statistics that were good and RMSEA that was acceptable. Thus, the structural model from Dombrowski et al.'s (2017a) EFA-SL Age 9 to 13 results applied to the Age 14 to 19 group yielded superior model fit compared with the other models and suggests that the bifactor model with four group factors is consistent across both the 9 to 13 and the 14 to 19 age ranges. Figure 2 presents standardized loadings for the four-group factor Age 14 to 19 bifactor model. A majority of subtest variance is associated with the General Intelligence factor, with substantially smaller portions of variance uniquely associated with the four WJ IV group factors (presumably, Working Memory, Perceptual Reasoning, Processing Speed, and Verbal Ability). The bifactor results from Table 5 were used to calculate omega-hierarchical and omega-hierarchical subscale coefficients. The ω_H coefficient for the General Intelligence factor (.829) was high and exceeded the threshold for confident scale interpretation of a unit-weighted composite composed of the specified indicators. The ω_{HS}

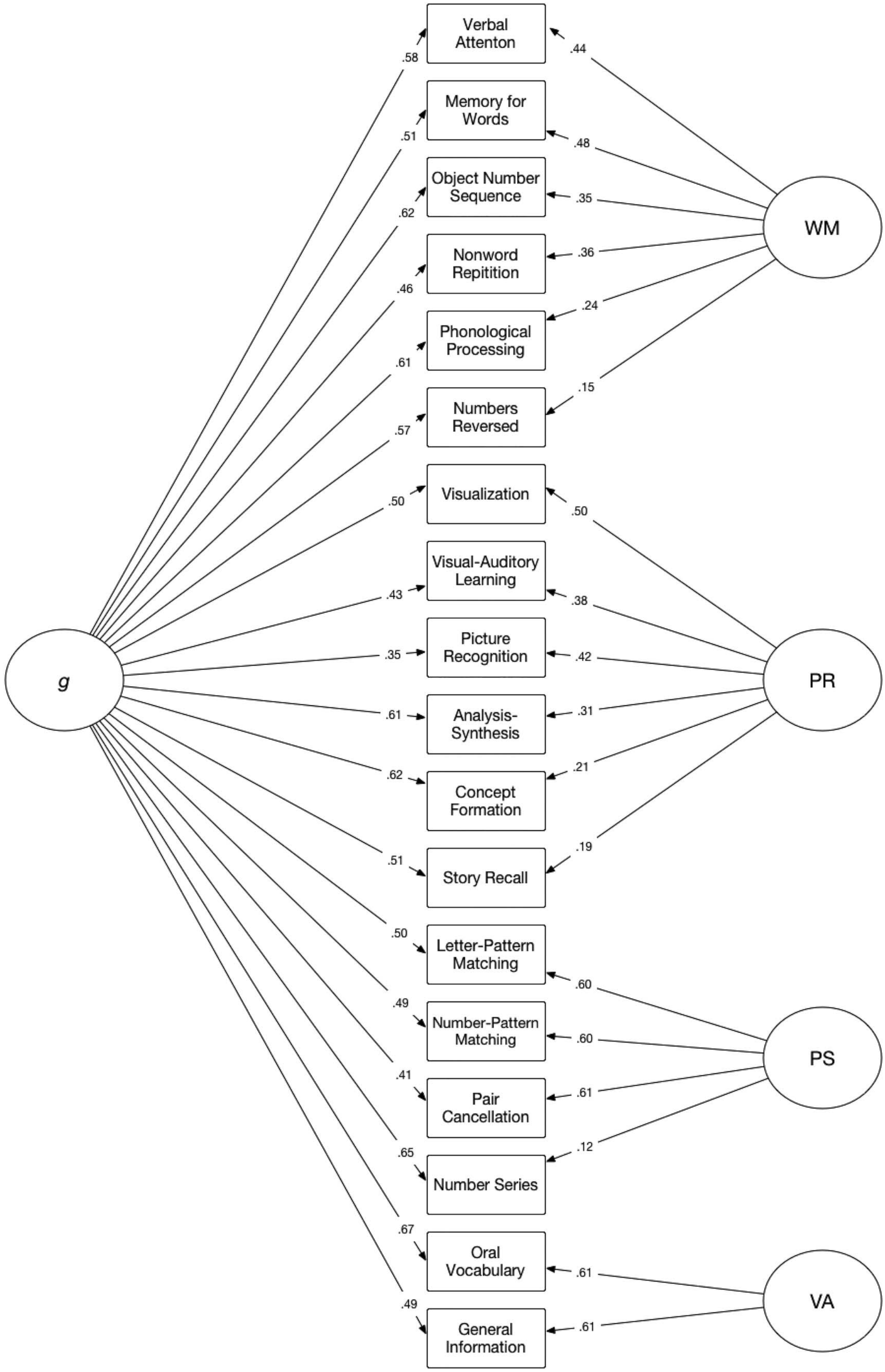


Figure 1. Four-factor bifactor (direct hierarchical) measurement model with standardized loading coefficients for the Woodcock-Johnson IV Cognitive Age 9–13. For the sake of parsimony, disturbance terms are omitted. *g* = general intelligence; WM = working memory; PR = perceptual reasoning; PS = processing speed; VA = verbal ability.

Table 3
Sources of WJ IV Cognitive Subtest Variance According to a Bifactor CFA (Ages 9–13)

			First-order factors									
	General *g*		Working Memory (Gwm)		Perceptual Reasoning		Processing Speed (Gs)		Verbal Ability (Gc)			
Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
Verbal Attention (Gwm)	.589	.347	.446	.199							.546	.454
Memory for Words (Aud Mem)	.518	.268	.480	.230							.499	.501
Object Number Sequence (Gwm)	.628	.394	.358	.128							.522	.478
Nonword Repetition (Ga)	.460	.212	.366	.134							.345	.655
Phonological Processing (Ga)	.619	.383	.240	.058							.441	.559
Numbers Reversed (Gwm)	.574	.329	.159	.025							.355	.645
Visualization (Gv)	.506	.256			.506	.256					.512	.488
Visual-Auditory Learning (Glr)	.433	.187			.382	.146					.334	.666
Picture Recognition (Gv)	.350	.123			.426	.181					.304	.696
Analysis-Synthesis (Gf)	.616	.379			.314	.099					.478	.522
Concept Formation (Gf)	.626	.392			.214	.046					.438	.562
Story Recall (Glr)	.516	.266			.190	.036					.302	.698
Letter-Pattern Matching (Gs)	.502	.252					.604	.365			.617	.383
Number-Pattern Matching (PerSpd)	.493	.243					.601	.361			.604	.396
Pair Cancellation (Gs)	.413	.171					.612	.375			.545	.455
Number Series (Gf)	.657	.432					.129	.017			.448	.552
Oral Vocabulary (Gc)	.672	.452							.613	.376	.828	.172
General Information (Gc)	.497	.247							.613	.376	.623	.377
Common variance		.610		.089		.087		.128		.086	.486	.514
Total variance		.296		.043		.042		.062		.042		
ω_H/ω_{HS}		.803		.221		.242		.385		.440		
H		.889		.48		.48		.64		.55		

Note. Alignment of subtests with respective Cattell-Horn-Carroll (CHC) Stratum I or II factors posited in the Woodcock-Johnson IV Technical Manual is indicted following each subtest name. b = factor loading; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ω_H = ω hierarchical (g); ω_{HS} = ω hierarchical subscale (group factors); H = construct replicability.

coefficients for the four WJ IV cognitive factors (Working Memory, Perceptual Reasoning, Processing Speed, Crystalized/Verbal Ability), however, were considerably lower, ranging from .211 (Working Memory) to .373 (Crystalized/Verbal Ability). Thus, unit-weighted composite scores based on specified indicators for the four WJ IV Cognitive CHC group factors likely contain insufficient true score variance to support accurate clinical interpretation (Reise, 2012; Reise et al., 2013). Results from Hancock and Mueller's (2001) H

Table 4
CFA Fit Statistics for the Age 14 to 19 Group

Model	χ^2	*df*	CFI	TLI	SRMR	RMSEA	RMSEA 90% CI	BIC	AIC
g	3,896.338	135	.726	.689	.072	.129	[.125, .132]	142,863	142,570
Three-factor (Gc, Gf, Gs) bifactor	Inadmissible model—not positive definite covariance matrix, negative residual variance								
Three-factor oblique (Gc, Gf, Gs)	2,719.723	132	.811	.781	.056	.108	[.104, .112]	141,709	141,400
Three-factor (Gc, Gf, Gs) higher order Woodcock cognitive processing model	2,929.152	132	.796	.763	.066	.112	[.109, .116]	141,918	141,609
Seven-factor oblique	Inadmissible model—not positive definite covariance matrix, Ga loads 1.00 on Gwm; Glr loads .922 with Gwm; OV loads .991 Gc								
Seven-factor higher order (theoretically proposed, but not tested by publisher)	Inadmissible model—not positive definite covariance matrix, negative residual variance Ga –.04 & 1.02 loading on *g*)								
Seven-factor bifactor	2,312.099	124	.84	.803	.057	.102	[.098, .106]	141,361	141,008
Four-factor oblique	2,240.087	129	.846	.817	.058	.099	[.091, .098]	141,252	140,926
Four-factor oblique reassigned PP & NS to Gwm & Gs, respectively	2,233.607	129	.846	.818	.061	.098	[.095, .102]	141,245	140,919
Four-factor higher order	2,245.243	131	.846	.82	.058	.098	[.090, .098]	141,242	140,927
Four-factor higher order reassigned PP & NS to Gwm & Gs, respectively	2,236.473	131	.846	.821	.061	.098	[.094, .101]	141,233	140,918
Four-factor bifactor (Dombrowski, McGill, & Canivez, 2017a; Age 14 to 19) Four subtest Gc	Inadmissible model—not positive definite covariance matrix, negative residual variance OV								
Four-factor bifactor (Dombrowski, et al., 2017a; Same model as Age 9 to 13) PP to Gwm, NS to Gs	1,680.885	118	.886	.852	.045	.089	[.085, .092]	140,774.132	140,388.882

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; SRMR = standardized root mean square; RMSEA = root mean square error of approximation; CI = confidence interval; BIC = Bayesian information criterion; AIC = Akaike information criterion; *g* = general intelligence; PP = phonological processing; NS = number series; Gwm = working memory; Ga = auditory processing; Gv = visual-spatial thinking; Glr = long-term retrieval; Gf = fluid reasoning; Gs = processing speed; Gc = comprehension-knowledge/crystallized ability; OV = oral vocabulary.

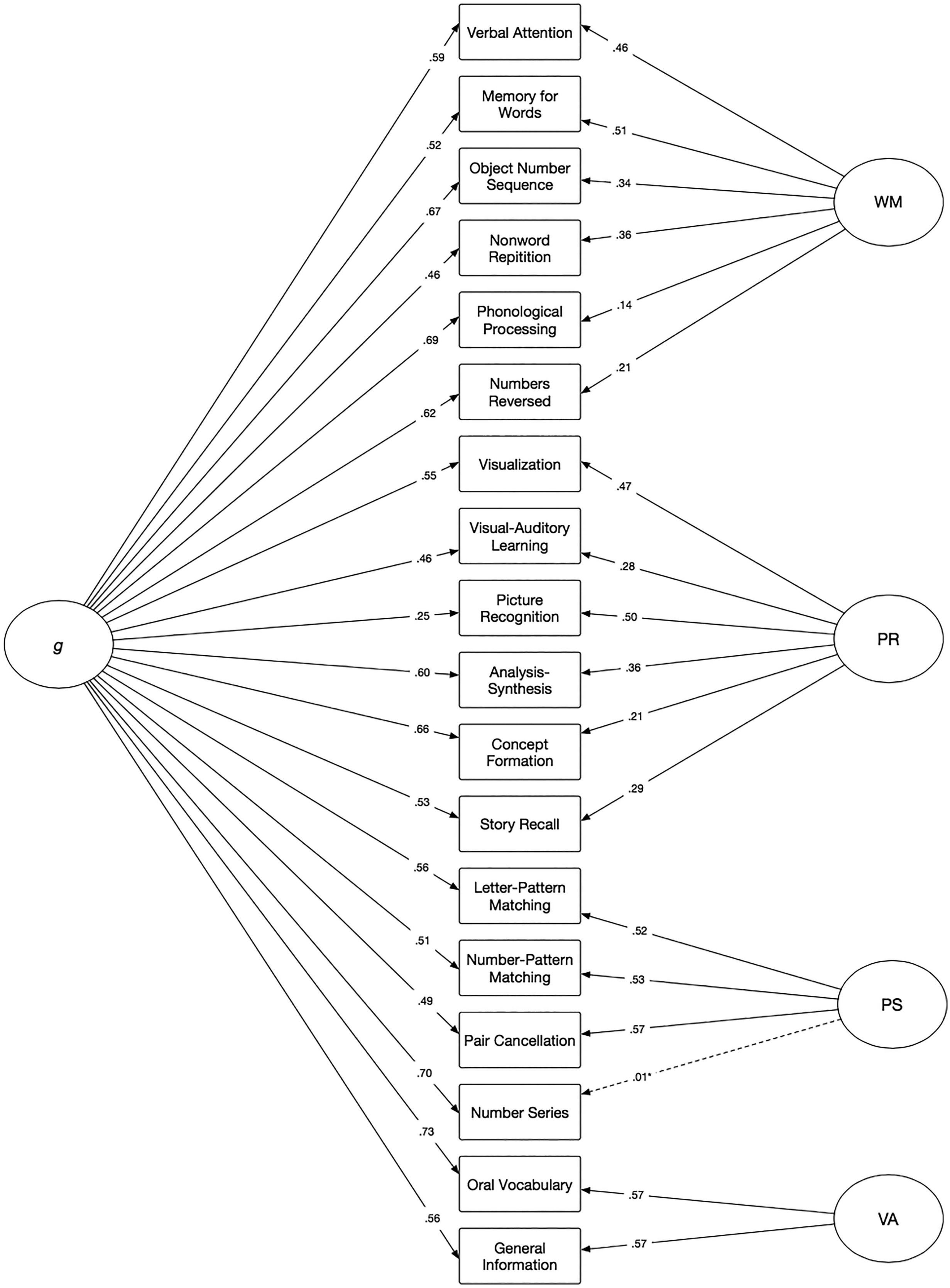


Figure 2. Four-factor bifactor (direct hierarchical) measurement model with standardized loading coefficients for the Woodcock-Johnson IV Cognitive Age 14–19. For the sake of parsimony, disturbance terms are omitted. *g* = general intelligence; WM = working memory; PR = perceptual reasoning; PS = processing speed; VA = verbal ability. * Nonsignificant standardized loading.

Table 5

Sources of WJ IV Cognitive Subtest Variance According to a Bifactor CFA (Ages 14–19)

			First-order factors										
	General *g*		Working Memory (Gwm)		Perceptual Reasoning		Processing Speed (Gs)		Verbal Ability (Gc)				
Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
Verbal Attention (Gwm)	.590	.348	.463	.214							.563	.437
Memory for Words (Aud Mem)	.525	.276	.515	.265							.542	.458
Object Number Sequence (Gwm)	.671	.450	.344	.118							.568	.432
Nonword Repetition (Ga)	.461	.213	.365	.133							.346	.654
Phonological Processing (Ga)	.692	.479	.145	.021							.500	.500
Numbers Reversed (Gwm)	.628	.394	.219	.048							.443	.557
Visualization (Gv)	.553	.306			.473	.224					.530	.470
Visual-Auditory Learning (Glr)	.469	.220			.289	.084					.303	.697
Picture Recognition (Gv)	.256	.066			.503	.253					.319	.681
Analysis-Synthesis (Gf)	.605	.366			.364	.132					.499	.501
Concept Formation (Gf)	.665	.442			.218	.048					.490	.510
Story Recall (Glr)	.535	.286			.291	.085					.371	.629
Letter-Pattern Matching (Gs)	.564	.318					.525	.276			.594	.406
Number-Pattern Matching (PerSpd)	.512	.262					.539	.291			.553	.447
Pair Cancellation (Gs)	.492	.242					.572	.327			.569	.431
Number Series (Gf)	.705	.497					.017[a]	.000			.497	.503
Oral Vocabulary (Gc)	.731	.534							.570	.325	.860	.140
General Information (Gc)	.568	.323							.570	.325	.647	.353
Common variance		.655		.087		.090		.097		.071	.511	.489
Total variance		.335		.044		.046		.050		.036		
ω_H/ω_{HS}		.829		.211		.260		.282		.373		
H		.908		.50		.50		.56		.49		

Note. Alignment of subtests with respective CHC stratum I or II factors posited in the WJ IV Technical Manual is indicted following each subtest name. b = factor loading; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ω_H = ω hierarchical (*g*); ω_{HS} = ω hierarchical subscale (group factors); H = construct replicability.

[a] Number series loading nonsignificant on Gs.

index were also too low for confident clinical interpretation of any score except for the general factor.

Discussion

The present results yield an alternate factor structure for the WJ IV Cognitive to that published in the WJ IV *Technical Manual.* The present results support the outcomes produced from a recent independent EFA investigation of the WJ IV Cognitive (Dombrowski et al., 2017a) that provided evidence of a four-factor solution (e.g., Verbal, Working Memory, Processing Speed, and Perceptual Reasoning) across the 9 to 19 age range. Results also suggest that when modeling seven first-order factors and the higher order *g* factor with all 18 WJ IV Cognitive subtests, as presented in the *WJ IV Technical Manual* (McGrew et al., 2014), inadmissible results were produced including a Heywood case (1.02 loading of Ga on g) and a negative variance estimate (−0.04) for the Auditory Processing (Ga) factor within both age groups. Modeling an oblique seven-factor structure also yielded an inadmissible solution. The Ga factor produced a Heywood case and was linearly dependent upon the Gwm factor (1.041 loading at Age 9 to 13; 1.006 loading at Age 9 to 14). Oral Vocabulary was essentially isomorphic with the Gc factor (loading of .989 at Age 9 to 13; .991 at Age 14 to 19). All of these results are suggestive of possible over factoring and model misspecification (Kline, 2016). A bifactor representation of the WJ IV Cognitive with general intelligence (*g*) and seven group factors produced admissible results but the fit indices were inferior to most of the four-factor structures (oblique, higher order, bifactor) that were examined. Certainly, one could have attempted extensive post hoc model modification with the various seven factor models post hoc to improve model fit and possibly achieve a desired outcome. However, as Horn (1989) noted, "the statistical demands of structure equation theory are stringent. If there is tinkering with results to get a model to fit, the statistical theory, and thus the basis for strong inference, goes out the window" (p. 39). Horn also warned that when there is excessive model tinkering "one should not give any greater credence to results from modeling analyses than one can give to results from comparably executed factor analytic studies of the older variety" (e.g., EFA; p. 40).

Most models that included four group factors (oblique, higher order, and bifactor) converged and produced admissible results (see Tables 2 and 4). At Age 14 to 19, the reassignment of Phonological Processing and Number Series to be identical with the Age 9 to 13 structure, in which Gc contains only Oral Vocabulary and General Information, and in which Phonological Processing loads on Gwm and Number Series loads Gs, produced the best model fit for Age 14 to 19. It is apparent that the Number Series subtest is predominantly *g* loaded and perhaps a poor indicator of any group factor. The Oral Vocabulary subtest also essentially reflects the Gc factor with a loading higher than .97 on the Gc factor.

The bifactor model with four group factors (depicted in Figures 1 and 2) replicated the four-factor structure reflected in the EFA-SL study (Dombrowski et al., 2017a) and was consistent across the 9 to 13 and 14 to 19 age ranges. Incidentally, the resulting bifactor structure found with the WJ IV Cognitive in the present study is similar to structural models found in independent WISC–V research (Canivez et al., 2016, 2017; Dombrowski, Canivez, Watkins, & Beaujean, 2015; Dombrowski, Canivez, & Watkins, 2017; Watkins, Dombrowski, & Canivez, 2017), in which results suggested a four-factor structure including Verbal Ability, Working Memory, Processing Speed, and

Perceptual Reasoning. This same four-factor theoretical alignment is similar to prior findings on the WJ III Cognitive by Dombrowski (2013), who investigated its exploratory and hierarchical structure.

The latent factors supported in the present study indicate that the WJ IV Cognitive does not fully align with CHC theory with respect to the identification of separate Gf, Gv, Glr, and Ga factors. There was modest evidence for Gc, Gwm, and Gs, and a fused Gf/Gv factor (which may be better conceptualized using the Wechsler nomenclature of "perceptual reasoning"). The composition of subtests under a four-factor model does not comport with the theoretically proposed composition of CHC factors, as many subtests migrated away from their theoretically posited factors. The Glr subtests Story Recall and Visual Auditory Learning loaded together with the Gf and Gv subtests to form a perceptual reasoning factor. The Ga subtests of Phonological Processing and Nonword Repetition paired with the Gwm subtests. The Gs subtests loaded together along with Number Series. Number Series, a newly added subtest to the WJ IV, appears to be heavily *g* loaded, leaving little residual variance for group factor alignment. Finally, the Gc factor contained the two theoretically proposed factors, but Oral Vocabulary loaded .971 on the Gc factor in the Age 9 to 13 analysis and .991 in the Age 14 to 19 analysis, suggesting that it captured a large percentage of the Gc factor variance.

Model-based reliability coefficients (ω_H and ω_{HS}) estimated unique portions of true score variance that would be captured by unit-weighted composite scores from the WJ IV Cognitive general and group factors. Omega-hierarchical estimates indicated that although the broad *g* factor permitted individual interpretation (18 subtest ω_H = .803 for Age 9 to 13; .829 for Age 14 to 19), the ω_{HS} estimates for the four WJ IV Cognitive group factors were generally low (<.50; see Tables 3 and 5), suggesting that the group factors should not be independently interpreted (Brunner et al., 2012; Reise, 2012). Similarly, the H index (Hancock & Mueller, 2001) furnished evidence for individual interpretation of the general factor (H = .889 for Age 9–13; H = .908 for Age 14–19) but not the respective group factors (i.e., H < .65 across all group factors with most <.50).

The conclusion that interpretation should reside primarily at the general factor stratum is not unique to the WJ IV Cognitive and has also been observed in both EFA and CFA studies of the WISC–V (Canivez et al., 2016, 2017; Dombrowski, Canivez, & Watkins, 2017; Dombrowski, Canivez, Watkins, & Beaujean, 2015; Watkins et al., 2017), WISC–IV (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Dombrowski & Noonan, 2004; Keith, 2005; Watkins, 2006, 2010; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), and other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Canivez, Watkins, Good, James, & James, 2017; Gignac, 2005, 2006; Golay & Lecerf, 2011; Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Lecerf & Canivez, in press; McGill & Canivez, 2016; Watkins & Beaujean, 2014; Watkins, Canivez, James, Good, & James, 2013). Further, these results are also not unique among Wechsler scales, as similar results were also observed with the Differntial Abilities Scales, Second Edition (DAS-II; Canivez & McGill, 2016), Stanford-Binet, Fifth Edition (SB-5; Canivez, 2008; DiStefano & Dombrowski, 2006; Dombrowski, Distefano, & Noonan, 2004), Wechsler Abbreviated Scale of Intelligence (WASI) and Wide Range Intelligence Test (WRIT; Canivez, Konold, Collins, & Wilson, 2009), Reynolds Intellectual Assessment Scales (RIAS; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007), Cognitive Assessment System (CAS; Canivez, 2011), and WJ III (Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Strickland, Watkins, & Caterino, 2015). In sum, the results from this study and the extant structural validity literature coalesce to suggest that a host of commercial ability measures such as the WJ IV Cognitive may be overfactored with insufficient target construct variance at the group-factor level for confident clinical interpretation of those indices (Frazier & Youngstrom, 2007).

Because the standardization data was unavailable, the correlation matrices presented in the *WJ IV Technical Manual* (McGrew et al., 2014) were relied upon for analyses. Additionally, the same correlation matrices used by Dombrowski et al. (2017a) were used in this study. Ideally, a different sample would have been used by this study. Nevertheless, the present results suggest alternative rival structures for the WJ IV Cognitive that can be cross-validated on different clinical samples in future research. Replication of these results would provide greater support for these structures.

Conclusion and Potential Implications for the Application of CHC Theory in Applied Practice

The results of this study suggest that the WJ IV Cognitive, as presented in the *WJ IV Technical Manual* (McGrew et al., 2014), appears to be overfactored. Although the publisher proposes a theoretical linkage of the WJ IV Cognitive with seven CHC factors, the evidence provided in the present study does not support this proposed linkage. Instead, results suggest a more parsimonious four-factor solution and offer a different theoretical conceptualization for the WJ IV Cognitive—one that is more consistent with the prior four-factor based Wechsler models (i.e., General Ability along with Verbal Ability, Working Memory, Processing Speed, and Perceptual Reasoning).

As the WJ IV Cognitive has replaced its predecessor as an important measure for making future refinements to the CHC theory/model, these results have implications beyond the present measurement instrument. The evidence base is accumulating to suggest that existing measures of cognitive ability that have attempted to align with CHC theory have experienced some difficulties locating distinct CHC factors other than factors that have been well established in the psychometric literature (e.g., verbal ability, processing speed, and memory). Beyond these group factors, the alignment of various instruments with CHC theory has been questioned in some independent analyses but not all (Keith & Reynolds, 2010). Even if there were perfect alignment of commercial ability measures with the theoretically proposed CHC structure presented in respective technical manuals, the evidence from construct reliability (i.e., omega statistics; H) and variance apportionment suggest that there is generally insufficient group factor variance for independent interpretation of CHC-based group factors regardless of whether a higher order model or a bifactor model is adopted.

Equally important, if the structure proposed within an instrument's technical manual is not replicated by independent studies, then engaging in accurate CHC-based profile analytic interpretive procedures (for example, XBA and PSW analyses) will be challenging. The foundation for such practice rests upon a theoretical/factor structure that has consistent and replicated empirical support. This study and the body of literature cited above suggests that the support for the linkage of CHC theory with numerous instruments (e.g., DAS-II; Kaufman Assessment Battery for Children, Second Edition (KABC-2); SB-5; WISC-V; WJ III full test battery and WJ III Cognitive; WJ IV full test battery and WJ IV Cognitive) may not be as strong as is commonly advocated in the professional literature or test technical manuals. Therefore, the field is advised to exercise a degree of caution when attempting to interpret the various CHC broad factor indices or when engaging in interpretive approaches such as XBA and PSW using the WJ IV Cognitive and other instruments linked to CHC theory until the empirical literature provides more consistent support for these approaches (Cucina & Howardson, 2016). The present study adds to this literature base and suggests a different theoretical structure for the WJ IV Cognitive at school age. The WJ IV Cognitive for Age 9 to 19 appears to measure four abilities—verbal ability (Gc), memory (Gwm), processing speed (Gs), and perceptual reasoning—but primarily measures general intelligence (*g*). As a result, users are encouraged to incorporate a more circumspect appraisal of the scores provided by the WJ IV Cognitive to ensure an interpretive approach that is guided by the presently available empirical evidence (Dombrowski, Ambrose, & Clinton, 2007; Dombrowski & Gischlar, 2014; Dombrowski, Kamphaus, et al., 2006) rather than just theoretical expectations.

References

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417–424. http://dx.doi.org/10.1080/09297040802603661

Brown, T. A. (2016). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80,* 796–846. http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33,* 261–304. http://dx.doi.org/10.1177/0049124104268644

Canivez, G. L. (2008). Orthogonal higher-order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. *School Psychology Quarterly, 23,* 533–541. http://dx.doi.org/10.1037/a0012884

Canivez, G. L. (2011). Hierarchical factor structure of the Cognitive Assessment System: Variance partitions from the Schmid-Leiman (1957) procedure. *School Psychology Quarterly, 26,* 305–317. http://dx.doi.org/10.1037/a0025973

Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly, 29,* 38–51. http://dx.doi.org/10.1037/spq0000032

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247–271). Gottingen, Germany: Hogrefe.

Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler Abbreviated Scale of Intelligence and Wide Range Intelligence Test: Convergent and structural validity. *School Psychology Quarterly, 24,* 252–265. http://dx.doi.org/10.1037/a0018030

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales-Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment, 28,* 1475–1488. http://dx.doi.org/10.1037/pas0000279

Canivez, G. L., & Watkins, M. W. (2010a). Investigation of the factor structure of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment, 22,* 827–836. http://dx.doi.org/10.1037/a0020429

Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly, 25,* 223–235. http://dx.doi.org/10.1037/a0022046

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

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

Canivez, G. L., Watkins, M. W., Good, R., James, K., & James, T. (2017). Construct validity of the Wechsler Intelligence Scale for Children – Fourth UK Edition with a referred Irish sample: Wechsler and Cattell-Horn-Carroll model comparisons with 15 subtests. *British Journal of Educational Psychology, 87,* 383–407. http://dx.doi.org/10.1111/bjep.12155

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

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

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Cucina, J. M., & Howardson, G. N. (2016). Woodcock-Johnson-III, Kaufman Adolescent and Adult Intelligence Test (KAIT), Kaufman Assessment Battery for Children (KABC), and the Differential Ability Scales (DAS) support Carroll but not Cattell-Horn. *Psychological Assessment, 29,* 1001–1015. http://dx.doi.org/10.1037/pas0000389

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet-Fifth Edition. *Journal of Psychoeducational Assessment, 24,* 123–136. http://dx.doi.org/10.1177/0734282905285244

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly, 28,* 154–169. http://dx.doi.org/10.1037/spq0000010

Dombrowski, S. C. (2014a). Exploratory bifactor analysis of the WJ-III Cognitive in adulthood via the Schmid-Leiman procedure. *Journal of Psychoeducational Assessment, 32,* 330–341. http://dx.doi.org/10.1177/0734282913508243

Dombrowski, S. C. (2014b). Investigating the structure of the WJ-III Cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment, 32,* 483–494. http://dx.doi.org/10.1177/0734282914530838

Dombrowski, S. C. (2015a). Exploratory bifactor analysis of the WJ-III Achievement at School Age via the Schmid-Leiman orthogonalization procedure. *Canadian Journal of School Psychology, 30,* 34–50. http://dx.doi.org/10.1177/0829573514560529

Dombrowski, S. C. (2015b). *Psychoeducational assessment and report writing*. New York, NY: Springer Science. http://dx.doi.org/10.1007/978-1-4939-1911-6

Dombrowski, S. C., Ambrose, D. A., & Clinton, A. (2007). Dogmatic insularity in learning disabilities diagnosis and the critical need for a philosophical analysis. *International Journal of Special Education, 22,* 3–10.

Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2017). Factor Structure of the 10 WISC–V Primary Subtests in Four Standardization Age Groups. *Contemporary School Psychology*. Advance online publication http://dx.doi.org/10.1007/s40688-017-0125-2

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53,* 194–201. http://dx.doi.org/10.1016/j.intell.2015.10.009

Dombrowski, S. C., DiStefano, C., & Noonan, K. (2004). Review of the Stanford-Binet, Fifth Edition. *Communiqué, 33,* 12–15.

Dombrowski, S. C., & Gischlar, K. L. (2014). Ethical and empirical considerations in the identification of learning disabilities. *Journal of Applied School Psychology, 30,* 68–82. http://dx.doi.org/10.1080/15377903.2013.869786

Dombrowski, S. C., Kamphaus, R. W., Barry, M., Brueggeman, A., Cavanagh, S., Devine, K., . . . Vess, S. (2006). The Solomon effect in learning disabilities diagnosis: Can we learn from history? *School Psychology Quarterly, 21,* 359–374. http://dx.doi.org/10.1037/h0084128

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017a). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment, 29,* 394–407. http://dx.doi.org/10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017b). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology*

Quarterly. Advance online publication. http://dx.doi.org/10.1037/spq0000221

Dombrowski, S. C., & Noonan, K. (2004). Review of the WISC-IV. *Communiqué, 33,* 35–38.

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school-aged analysis. *Psychological Assessment, 25,* 442–455. http://dx.doi.org/10.1037/a0031335

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment, 27,* 494–507. http://dx.doi.org/10.1177/0734282909333179

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182. http://dx.doi.org/10.1016/j.intell.2006.07.002

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment, 12,* 320–329. http://dx.doi.org/10.1177/1073191105278118

Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences, 27,* 73–86. http://dx.doi.org/10.1027/1614-0001.27.2.73

Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment, 23,* 143–152. http://dx.doi.org/10.1037/a0021230

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25,* 496–508. http://dx.doi.org/10.1037/a0030676

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and Future* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

Horn, J. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–75). Urbana, IL: University of Illinois Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253–270. http://dx.doi.org/10.1037/h0023816

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581–614). New York, NY: Guilford Press.

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

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Cross-battery assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School & Educational Psychology, 4,* 146–157. http://dx.doi.org/10.1080/21683603.2016.1192855

Lecerf, T., & Canivez, G. L. (in press).Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment*.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192–211. http://dx.doi.org/10.1037/1082-989X.4.2.192

McDonald, R. P. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science, 5,* 675–686. http://dx.doi.org/10.1177/1745691610388766

McGill, R. J., & Canivez, G. L. (2016). Orthogonal higher order structure of the WISC-IV Spanish using hierarchical exploratory factor analytic procedures. *Journal of Psychoeducational Assessment, 34,* 600–606. http://dx.doi.org/10.1177/0734282915624293

McGill, R. J., & Spurgin, A. R. (2017). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children-Second Edition (KABC-II) school-age battery. *Assessment, 24,* 540–552. http://dx.doi.org/10.1177/1073191115614081

McGrew, K. S. (2009). CHC theory and the Human Cognitive Abilities Project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10. http://dx.doi.org/10.1016/j.intell.2008.08.004

McGrew, K. S., LaForte, E. M., & Shrank, F. A. (2014). *Technical manual: Woodcock–Johnson IV*. Rolling Meadows, IL: Riverside.

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

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.) Los Angeles, CA: Author.

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment, 24,* 129–140. http://dx.doi.org/10.1037/a0024878

Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology, 45,* 439–456. http://dx.doi.org/10.1016/j.jsp.2007.03.003

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

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

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98,* 223–237.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock–Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock–Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock–Johnson IV*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock–Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment, 27,* 689–697. http://dx.doi.org/10.1037/pas0000052 (Correction published 2015, *Psychological Assessment, 27,* p. 697)

Taub, G. E., & McGrew, K. S. (2014). The Woodcock-Johnson Tests of Cognitive Abilities III's cognitive performance model: Empirical support for intermediate factors within CHC theory. *Journal of Psychoeducational Assessment, 32,* 187–201. http://dx.doi.org/10.1177/0734282913504808

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understandingconcepts and applications*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10694-000

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children—Fourth edition. *Psychological Assessment, 18,* 123–125. http://dx.doi.org/10.1037/1040-3590.18.1.123

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment, 22,* 782–787. http://dx.doi.org/10.1037/a0020043

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

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition. *School Psychology Quarterly, 29,* 52–63. http://dx.doi.org/10.1037/spq0000038

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC-IVUK with a large referred Irish sample. *International Journal of School & Educational Psychology, 1,* 102–111. http://dx.doi.org/10.1080/21683603.2013.794439

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2017). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–5th edition. *International Journal of School and Educational Psychology*. Advance online publication. http://dx.doi.org/10.1080/21683603.2017.1342580

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred students. *Educational and Psychological Measurement, 66,* 975–983. http://dx.doi.org/10.1177/0013164406288168

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children–Fifth Edition*. San Antonio, TX: NCS Pearson.

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice, 23,* 327–347. http://dx.doi.org/10.1111/cpsp.12172

Received July 19, 2017
Revision received October 24, 2017
Accepted October 27, 2017 ■