Please use the following citation when referencing this work:

Dombrowski, S. C., McGill, R. J., & Casey, C. J. (in press). The Woodcock-Johnson IV Tests of Cognitive Abilities: A paradox for evidence-based assessment. In G. L. Canivez (Ed.), Assessing psychometric fitness of intelligence tests: Toward evidence-based interpretation practices (pp. xx-xx). Rowman & Littlefield.

Chapter 4

The Woodcock-Johnson IV Tests of Cognitive Abilities: A Paradox for Evidence-Based Assessment

Stefan C. Dombrowski, Ryan J. McGill, and Corinne J. Casey

The Woodcock-Johnson IV (WJ IV; Schrank et al., 2014a) comprises separate co-normed batteries of tests of cognitive ability (18 subtests; Schrank et al., 2014b), achievement (20 subtests; Schrank et al., 2014a), and oral language (12 subtests; Schrank et al., 2014b). The WJ IV authors report in the technical manual (McGrew et al., 2014) that the instrument's structure was guided by Carroll's (1993) Three Stratum Theory of Cognitive Ability (3S), the work of Horn and Cattell (1966), and contemporary neuroscience research on memory (e.g., McGrew et al., 2014). It has some very positive features including the linkage of its development with theory and the creation of three distinct yet symbiotic test batteries. The WJ IV Cognitive, Achievement, and Oral Language can be used independently or in any combination for selective testing. The Woodcock-Johnson Online Scoring and Reporting System (Schrank & Dailey, 2014) makes scoring easy and provides an array of composite, index scores, and other narrow ability and clinical cluster scores associated with the Cattel-Horn-Carroll Theory of Human Cognitive Abilities (CHC; Schneider & McGrew, 2018). Additionally, the WJ IV Interpretation and Instructional Interventions Program (WIIIP; Schrank & Wendling, 2015a) is available and provides an online option for generating narrative interpretive text and related interventions, which the authors of this program claim are useful for linking assessment to intervention. The WJ IV, including the WJ IV Cognitive, is heavily featured in the cross-battery assessment (XBA; Flanagan et al., 2013) approach and other assessment approaches (e.g., patterns of strengths and weaknesses [PSW]) marketed to be useful for the identification of specific learning disabilities

(SLDs) given its purported linkage to CHC theory. The various WJ battery of tests across the decades also have been used to make additional clinical and neuropsychological decisions (Miller et al., 2016).

Although this chapter focuses on the WJ IV Cognitive, it will be necessary to discuss the WJ IV full test battery due to the mutualism among the three independent batteries and the lack of information furnished by the test publisher regarding the independent psychometric structure of the WJ IV Cognitive. In fact, the WJ IV Cognitive was never separately evaluated for either standalone reliability or validity evidence. Instead, the evidence from the WJ IV full test battery was extrapolated from, and extended to, the WJ IV Cognitive. One other point deserves mention. The WJ IV is unique in its co-norming of three separate tests via imputation of missing data. In other words, not all participants were administered all 50 subtests. Instead, scores were imputed for participants who did not take a particular subtest. As a result of what the test publisher refers to as “planned missingness,” test-retest results could not be calculated for many of the WJ IV tests. This information is commonly furnished in all tests of cognitive ability and permits an evaluation of the stability of the individual subtest and composite scores across repeated administrations.

The WJ IV Tests of Cognitive Abilities

The technical manual 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 broad abilities of Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Short-Term Working Memory (Gwm), Cognitive Processing Speed (Gs), Auditory Processing (Ga), Long-Term Retrieval (Glr), and Visual-Processing (Gv). It is presently the only commercial ability measure to furnish composite scores for all of the seven consensus CHC broad abilities. The first ten subtests of the WJ IV Cognitive comprise the Standard Battery, and subtests 11–18 comprise the Extended Battery. There are numerous, additional scores that may be produced by the WJ IV scoring system using either selected subtests from the standard battery or a combination of standard and extended battery subtests. For instance, the first three subtests are used to produce the Brief Intellectual Ability composite score (BIA) while subtests 1–10 provide the General Intellectual Ability composite score (GIA; i.e., *g;* subtests 1–7) and the following three CHC index scores: Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), and Short-Term Working Memory (Gwm). Additionally, a Cognitive Efficiency index and Gf-Gc composite scores may be produced from the standard 10 subtest battery. Subtests from the Extended Battery may be used to produce the additional four CHC index scores (e.g., Gs, Ga, Glr, and Gv) in addition to a Perceptual Speed (P), Quantitative Reasoning (RQ), and Number Facility (N) score. The WJ IV Cognitive subtest alignment with respective CHC theoretical factor scores and additional scores produced by the scoring system is presented in Table 4.1.

<T04_001>

An Extrapolated Factor Structure

To comport with CHC theory, the publisher proposed a hierarchical seven factor theoretical structure for the WJ IV Cognitive. However, that model was not directly investigated for the instrument by the test publisher or, if it was, those results were not disclosed. Instead, the test publisher extrapolated the structure from the WJ IV full test battery structure from analyses reported for its 47[1] subtests and extended the results to the WJ IV Cognitive as well as the other two remaining batteries (McGrew et al., 2014). It may be contended that this is acceptable practice; however, Cattell (1978) argued that it cannot be assumed that a previously identified latent variable will automatically be located by a smaller subset of variables. Nevertheless, this validation approach often occurs during the initial stages of theory building. However, the WJ IV Cognitive is more than a simple tool used to build theory.[2] It is used the world over to evaluate individuals for various clinical and educational conditions some of which portend to change the trajectory of that individual's life by providing access to remedial services. Consequently, the instrument's factor structure would benefit from direct empirical validation.

For this reason, it is potentially inappropriate for the test publisher to claim that an instrument measures an attribute and then go on to produce various scores purporting to measure that attribute without furnishing empirical evidence in the form of factor analysis featuring the actual measured variables used to produce those scores. In the case of the WJ IV Cognitive, the test publisher claims that the WJ IV Cognitive measures attributes resembling a general factor, seven CHC factors, and additional cluster scores that are reported by the scoring system (see Table 4.1). The publisher's claims are based on the supposition that the structural alignment of an 18-subtest instrument nested within a 50-subtest battery will not be impacted when indicators are removed or added. However, the relationship among the variables (i.e., subtests) may change in the presence of a different subset of variables and so it is important to place greater emphasis on empirical evaluation (using factor analysis) than on theoretical conceptualization (Youngstrom & Van Meter, 2016; and see Dombrowski, McGill, Benson et al., 2019). Consequently, a case can be made that the some of the CHC index scores and the myriad additional scores produced by the WJ IV scoring system for the WJ IV Cognitive lack appropriate evidentiary support.

Why is an understanding of the internal structure (i.e., structural or factorial validity) of an instrument critically important? The factor structure of an instrument provides the foundational validity evidence upon which all other validity evidence rests (e.g., incremental and diagnostic utility). Importantly, it also provides a statistical rationale for how an instrument should be scored and interpreted. Without full explication of, and factor analytical justification for, structural validity, a clinician or researcher will be less able to understand and properly interpret the scores provided by that instrument. Thus it is more than just the theoretical musing of an academic researcher; it provides the statistical rationale for an instrument's scores.

Put simply, even if it were 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 Woodcock-Johnson IV full test battery and its potential alignment with CHC theory. It is noted that some of these same concerns were raised previously during independent exploratory factor analyses (EFA) of the WJ III full test battery (Dombrowski & Watkins, 2013).

Lack of Regard for Prior WJ III Post-Publication Research?

One of the concerns surrounding the WJ IV relates to a lack of regard of prior independent factor analytic literature. When developing the WJ IV, the test authors did not appear to consider prior, relevant exploratory structural validity research on the Woodcock-Johnson III (WJ III; e.g., Dombrowski & Watkins, 2013). Several studies from the literature published around the same time as the WJ IV specifically analyzed the WJ III full battery (e.g., Dombrowski & Watkins, 2013) and WJ III Cognitive in isolation (e.g., Dombrowski & Watkins, 2013; Dombrowski, 2013, 2014a, 2014b, 2015).[3] For instance, Dombrowski and Watkins (2013) evaluated the structure of the WJ III full test battery and found that it did not fully align with hypothesized CHC theory. Specifically, the full test battery was determined to measure six factors at ages 9 to 13 (Gc, Grw, Gs, Combined Gf/Gq, Ga and Glr) and five factors at ages 14 to 19 (Gc, Ga, Gs, Gq and Glr), in contrast to the nine factors promoted by the publisher. An evaluation of the WJ III Cognitive by Dombrowski (2013) suggested that the instrument measured four factors (e.g., perceptual reasoning [combined Gf/Gv], verbal ability [Gc], processing speed [Gs], and memory [Glr]) at ages 9–13 and not the theoretically posited seven factors. Regardless of whether four or seven factors were uncovered, variance apportionment and subtest misalignment with theory raised concerns about whether and how to interpret the WJ III Cognitive group-specific factors.

In totality, the theoretically proposed structure was not adequately supported by independent exploratory factor analyses of the WJ III whether the full test battery or the cognitive was investigated. In addition to concerns about exploratory structure, the series of articles published by Dombrowski and colleagues raised concerns about the CFA analyses conducted by the test publisher and its linkage with CHC theory (Dombrowski & Watkins, 2013; Dombrowski, 2013; Dombrowski, 2014a, 2014b; Dombrowski, 2015). This included a lack of analysis of competing CFA models, an absence of important CFA fit statistics and standardized parameter estimates for the measurement model that was adopted (Dombrowski & Watkins, 2013).

Even so, post-publication CFA research conducted by Taub and McGrew (2004, 2013) have provided stronger support for the proposed theoretical structure of the WJ III Cognitive. For example, in the 2004 study, the proposed CHC model was found to be invariant across age using data obtained from the normative sample. Additionally, Taub and McGrew (2014) evaluated whether an additional hierarchy in the three-stratum model was plausible featuring elements of the Cognitive Performance Model proposed by Richard Woodcock. The retained model featured CPM dimensions atop the broad abilities mediating the influence of *g* to the previously second-order CHC broad abilities. However, Dombrowski, McGill, and Canivez (2018a) investigated this model using the normative sample from the WJ IV Cognitive and found that it was inferior to a seven factor bifactor conceptualization and the four factor model retained by these authors. Finally, Strickland et al. (2015) conducted an EFA and CFA of the WJ III Cognitive featuring a sample of referred elementary school participants. Whereas EFA results were consistent with the previous EFAs conducted by Dombrowski and colleagues, CFA results supported the seven factor model posited by the test publisher. In sum, these conflicting WJ III Cognitive results and competing models in subsequent research suggest that any singular one-shot factor analysis, including those within tests' technical manuals, should be viewed cautiously until there is consistent replication across samples and contexts (Dombrowski, McGill, Farmer et al., 2022). Unfortunately, replication is something that is rarely undertaken in assessment psychology as there are few researchers with either the inclination or expertise to conduct such analyses (Dombrowski & McGill, 2024).

Recapitulation of CHC Structural Validity Concerns

Some of the same omissions evident in the WJ III Technical Manual were also evident in the WJ IV Technical Manual. First, there is an absence of information on the factor structure of the WJ IV Cognitive as the test publisher did not conduct a separate factor analysis of the WJ IV Cognitive. Additionally, there are concerns about the choice of both exploratory and confirmatory factor analytic (CFA) procedures used by the WJ IV test authors to examine the instrument's structure and its linkage with theory (Canivez, 2016; Dombrowski et al., 2017, 2018a, 2018b). Some of the more salient concerns are next addressed.

EFA Problems and Omissions

When analyzing the structure of the full WJ IV battery using principal axis factoring (PAF) with an oblique rotation, the test publisher stated that their analyses produced Heywood cases, impermissible factors, and a lack of convergence of the full test battery structure (McGrew et al., 2014). However, the technical manual did not report the details of these analyses, which may be attributable to overextraction (McGrew et al., 2014). The test publisher then proceeded to conduct a different yet what can be argued as a technically less appropriate method of analysis (i.e., principal components analysis [PCA]). Since the technical manual (McGrew et al., 2014) did not furnish the results of the PAF analysis, a determination of whether there was any practical difference in the patterns of loadings between the two analyses (i.e., PCA or PAF) was not possible.

Why would this comparison have been important? Some have argued that PCA is not a factor analytic procedure (Fabrigar & Wegener, 2012; Gorsuch, 1983) and thus it would have been useful to discern whether the choice of data analysis had any practical ramifications. Osborne (2015) explained, PCA is "not considered a true method of factor analysis and there is disagreement among statisticians about when it should be used, if at all" (p. 1). PCA is regarded as a mathematically simplified version of the general class of dimension reduction analyses. It computes the analysis without consideration of the underlying latent structure of the variables, using all the variance in the manifest variables, and therefore it does not discriminate between different dimensions of variance (e.g., shared and unique variance). Consequently, the components derived from PCA should not be interpreted as a reflection of latent dimensions such as a general factor or subordinate lower order factors (Bentler & Kano, 1990; Jensen, 1998; Preacher & MacCallum, 2003; Widaman, 1993). Instead, the measured variables in PCA are of interest rather than a hypothetical latent construct such as general intelligence or group-specific factors (i.e., Gc, Gf). Widaman (1993) noted that salient loadings tend to be higher in PCA than in factor analysis and that such inflation is magnified when the salient loadings are more moderate in value (e.g., 0.40 in the population) rather than high (e.g., 0.80). Fabrigar et al. (1999) explained that PAF removes random error from the factors so the relation among factors in a PAF analysis are more likely to approach the population values. Additionally, PAF includes only common variance whereas PCA includes common and specific variance, which can inflate factor loadings and give the misleading appearance of a stronger group factor structure (Snook & Gorsuch, 1989).

Representing a further possible problem with the EFA analyses undertaken, the test publisher used a type of rotation (e.g., varimax) following factor extraction that some consider less appropriate when evaluating tests of cognitive ability (Dombrowski, McGill, Canivez et al., 2021). Gorsuch (1983) commented that "varimax is inappropriate if the theoretical expectation suggests a general factor may occur" (p. 185). Indeed, the WJ IV was created with presumption of a general factor and specifically referenced 3S as the basis for its implementation of CHC theory (Beaujean, 2015) so the choice of varimax rotation is questionable (and possibly inappropriate) given the fact that rotation mathematically precludes identification of a general factor. Varimax rotation is also considered inappropriate when factors are highly correlated as in the case of tests of cognitive ability such as the WJ IV Cognitive. In these circumstances, an oblique rotation (e.g., promax) is considered necessary (Thompson, 2004). Although an oblique rotation is necessary, an additional step is often suggested. Gorsuch (1983) commented that higher-order factors are implicit in all oblique rotations so it is recommended that these factors be extracted and examined. The test publisher did not include this additional step (McGrew et al., 2014) because varimax rotation, as with all orthogonal rotations, does not produce correlations between rotated factors and therefore eliminates the option of undertaking a second-order factor analysis.

One elegant approach recommended for extracting and examining higher-order factors is through the Schmid-Leiman transformation (SL; Schmid & Leiman, 1957) procedure.[4] This procedure was used by Carroll (1993) when he created 3S. Because 3S was cited as being highly influential in the development of CHC theory and the WJ IV, it was surprising that the SL procedure was overlooked. The SL procedure involves making first-order factors orthogonal to second-order factors by first extracting the variance explained by the second-order factors. The next step in the procedure is to residualize the first-order factors of all the variance present in the second-order factors. Schmid and Leiman (1957) argued that this process "preserves the desired characteristics of the oblique solution" and "discloses the hierarchical structure of the variables" (p. 53). Carroll (1995) emphasized that "orthogonal factors should be those produced by the SL (1957) orthogonalization procedure" (Carroll, 1995, p. 437) noting in fact that this transformation produces "an orthogonal factor pattern very similar to the Spearman-Holzinger bi-factor pattern" (Carroll, 1993, p. 90). The SL procedure permits variance partitioning (i.e., determining the variance accounted for by higher and lower order factors), which assists when attempting to make interpretive decisions in clinical practice (Dombrowski, McGill, Canivez et al., 2021). The SL procedure was used to examine the structure of the WJ III Cognitive, Achievement, and full test battery (Dombrowski, 2014a, 2014b, 2015b; Dombrowski & Watkins, 2013) and the WJ IV Full test battery and Cognitive (Dombrowski, McGill, & Canivez, 2017, 2018b). It is a widely used and well-established procedure that has also been used in many other analyses of cognitive ability including the Cognitive Assessment System (e.g., Canivez, 2011), the Differential Abilities Scales–Second Edition (e.g., Dombrowski, McGill, Canivez et al., 2019), Reynolds Intellectual Assessment Scales (e.g., Dombrowski et al., 2009; Nelson & Canivez, 2012), the Stanford-Binet–Fifth Edition (e.g., Canivez, 2008; DiStefano &

Dombrowski, 2006), and Wechsler scales (e.g., Canivez & Watkins, 2010; Canivez et al., 2016; Dombrowski, Canivez, & Watkins, 2018; Dombrowski, McGill, Watkins et al., 2021; Kush & Canivez, 2019). Finally, McGrew (2012) indicated that he conducted the SL procedure on the 50 subtest WJ III battery but the results of those analyses have yet to be disclosed.

There are additional problems with the exploratory factor analyses (EFA) presented in the WJ IV Technical Manual (McGrew et al., 2014). The technical manual did not present rudimentary factor analytic statistics including percentage of variance accounted for by higher- and lower-order factors, communality estimates, and estimates of clinical relevance including omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}; Canivez, 2016; Reise, 2012; Rodriguez et al., 2015). The body of literature on EFA methodology (e.g., Carroll, 1993, 1995, 2003; Dombrowski, McGill, Canivez et al., 2021; Gorsuch, 1983; Thompson, 2004) in concert with indices of clinical relevance (e.g., Reise et al., 2013; Rodriguez et al., 2016) recommends the inclusion of this information when evaluating instruments with higher- and lower-order dimensions as it can aid test users in determining how an instrument's scores should be scored and interpreted to make high-stakes decisions about individuals.

CFA Problems and Omissions

Along with problems with the EFA analyses undertaken, there were also problems associated with the confirmatory factor analytic (CFA) procedures reported (see Canivez, 2017 for a detailed review). First, the technical manual reported only testing a few competing CFA models (Model 1: single *g* factor; Model 2: 9 broad CHC higher-order model; and Model 3: broad plus narrow CHC higher-order factor model). It would seem to have behooved the test publisher to also examine rival models such as Woodcock's Cognitive Performance Model (e.g., Taub & McGrew, 2014) model and Dombrowski and Watkins's (2013) WJ III full test battery

SL models. Second, best-fitting initial models and cross-validation models across each age group had comparative fit indexes (CFI; .603–.700), Tucker–Lewis Indexes (TLI; .607–.684), or root mean square errors of approximation (RMSEA; .115–.123) that did not approach levels considered to be adequate in the psychometric literature (CFI, TLI > .90; Hu & Bentler, 1999; RMSEA ≤ .08; Hu & Bentler, 1999). Thus the WJ IV structural models tested were not well-fitting. Third, the WJ IV data were multivariate non-normal (Mardia's 1970 multivariate kurtosis estimate was 27.6) and produced Heywood cases. Consequently, maximum likelihood (ML) estimates should not have been relied upon and robust ML estimates should instead have been used (Byrne, 2006, 2012). Although the test publisher claimed that they "left no stone unturned" (McGrew et al., 2014; p. 179), the scholarly evidence presented in the technical manual does not seem to match this claim. The omission of a separate examination of the structure of the WJ IV Cognitive is a major problem. Likewise, the less-than-optimal choice of EFA methodology, the lack of inclusion of rudimentary EFA statistics, the omission of clinical relevance estimates, the omission of variance partitioning, the poor CFA results, and the incomplete review of the WJ III structural validity literature suggests that simple extrapolation of the factor structure from the WJ IV full test battery to the WJ IV Cognitive is problematic. For users of the instrument who wish to consider the information in the technical manual alongside independent factor analytic results, there is research available that will help guide an evidence-based approach to interpretation of the WJ IV Cognitive.

Independent Factor Analytic Research Findings

Dombrowski, McGill, and Canivez (2017) examined the factor structure of the WJ IV Cognitive across two standardization sample age ranges (9–13; 14–19) using EFA (i.e., principal axis factoring with an oblique [promax] rotation) followed by the SL procedure. The results generally yielded a recommendation of four factors. Additionally, the seven-factor solution proposed by the test publisher was investigated using these same procedures and found untenable as three of the factors had only a single salient subtest loading indicating they were mathematically impermissible. Across both age ranges there was significant subtest migration to theoretically different factors from that proposed by the test publisher. This rendered the seven factor solution across both age ranges nonviable. When extracting four factors across both age ranges, the WJ IV Cognitive was found to have a hierarchical (*g*) factor with four first-order group factors (presumably working memory, verbal ability, processing speed, and perceptual reasoning) at both age ranges (Dombrowski et al., 2017). However, the composition of the four factors deviated from that proposed by the test publisher with the exception of the verbal ability factor (Gc), processing speed (Gs), and working memory (Gwm) factors at ages 9–13 and working memory and processing speed factors at ages 14–19. Interestingly, the four factor solution located by Dombrowski and colleagues (2017) was reminiscent of the prior WISC–IV factor structure that contained verbal ability, working memory, processing speed, and perceptual reasoning factors.

In a second study, Dombrowski, McGill, and Canivez (2018a) used CFA to compare a variety of structural models including the publisher proposed seven factor structure, Woodcock's Cognitive Processing three factor model, and the four factor EFA/SL structure suggested by Dombrowski and colleagues (2017). The results of the CFA study found most evidentiary support for the four-factor EFA structural solution suggested by Dombrowski et al. (e.g., Gc, Gwm, Gs and Perceptual Reasoning [Gf/Gv]) across the 9 to 19 age range). The CFA results also suggested that when modeling seven first-order factors and the higher-order *g* factor with all 18 WJ IV Cognitive subtests, as presented in the technical manual, inadmissible results were produced including a Heywood case (1.02 loading of Ga on *g*) and a negative variance estimate (-0.04) for the 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 ages 9–13; 1.006 loading at ages 9–14). Oral Vocabulary was essentially isomorphic with the Gc factor (loading of .989 at ages 9–13; .991 at ages 14–19). 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), largely replicating the four-factor structure presented in the EFA-SL study by Dombrowski et al. (2017). The resulting four-factor structure found with the WJ IV Cognitive is similar to structural models found in independent WISC–V research (Canivez et al., 2016, 2017; Canivez, McGill, Dombrowski et al., 2020; Dombrowski, Canivez, & Watkins, 2018; Dombrowski, McGill, Watkins et al., 2022). These findings were also replicated by Dombrowski, McGill, & Morgan (2021) who used Monte Carlo simulation to study the structure of the WJ IV Cognitive for the 9–13 age range.

The conclusion from several independent factor analytic studies suggests that the WJ IV Cognitive may not fully align with CHC theory with respect to the identification of separate Gf, Gv, Gltr, and Ga factors. There was modest evidence for Gc, Gwm, Gs, and a fused Gf/Gv factor. However, it should be noted that the composition of subtests under a four-factor model does not comport with the publisher-proposed subtest composition as many subtests migrated away from their theoretically posited factors. The Glr subtests of 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 .97 on the Gc factor in the age 9 to 13 analysis and .99 in the age 14 to 19 analysis suggesting that it captured a large percentage of the Gc factor variance entirely. It is noted that most of the research regarding the WJ IV Cognitive focuses on the extended battery. However, when CFA was used to evaluate the 10-subtest standard battery from the WJ IV Cognitive across the 9-13 and 14-19 age ranges—the only study to date to investigate the standard battery's factor structure—the results provided support for the scoring structure posited by the test publisher (e.g., a general ability factor [GAI] and three group factors [Gc, Gf & Gwm]; McGill, 2023).

Evidence-Based Interpretation in the Face of Multidimensional Complexity

There is an abundance of interpretive books, manuals, and book chapters available for the WJ IV battery of tests, and there are frameworks (e.g., school neuropsychological assessment Miller & Maricle, 2019) that use the WJ IV battery of tests to make additional interpretative and diagnostic decisions. For the sake of parsimony, we will not provide further explication of how to use these interpretive procedures as they are adequately explained elsewhere (e.g., Flanagan et al., 2013); rather, this section will provide a discussion of whether to engage in these practices given the available empirical evidence outlined to this point in the chapter.

As previously discussed, the structural validity of the WJ IV Cognitive has been questioned (Sattler, 2018). As a result, a case can be made that users should interpret the panoply of composite and cluster scores produced by the WJ scoring system with caution. Despite these evidentiary shortcomings, this chapter will forge ahead and consider additional aspects of validity and issues that might influence use of the instrument in clinical practice.

Predictive Validity

Intelligence test scores have been used to predict relationships with other outcomes such as academic achievement. This aspect of construct validity is associated more broadly with predictive validity. It is important to be aware that the establishment of structural validity is a necessary prerequisite to these evaluations. If there is evidence that the factor structure of an instrument does not measure what it claims to measure then the promoted scoring structure is likely nonsensical and using that scoring structure to study other aspects of validity will arguably produce specious outcomes.

The results of three recent WJ IV Cognitive multiple regression studies contend that selected CHC cluster scores account for significant portions of academic achievement across early lifespan and thus should be the focal point of interpretation on the test (e.g., Cormier, Bulut, McGrew, & Frison, 2016; Cormier, Bulut, McGrew, & Singh, 2017; Cormier, McGrew, Bulut, & Funamoto, 2017). However, some of these studies de-emphasize the influence of *g* (Cormier, Bulut, McGrew, & Singh, 2017; Cormier, McGrew, Bulut, & Funamoto, 2017) or omit its influence altogether (Cormier, Bulut, McGrew, & Frison, 2016).

The Cormier et al. (2016) article used a series of multiple regression analyses to examine the linear relationship between the seven WJ IV broad CHC cluster scores and the two WJ IV ACH writing clusters at each of the 14 age groups (ages 6–19, inclusively). The regression models incorporated all seven broad CHC cluster scores (i.e., Gc, Gf, Gwm, Gs, Ga, Glr, and Gv) as predictor variables with the WJ IV ACH writing clusters of Basic Writing Skills (two subtests: Spelling and Editing) and Written Expression (two subtests: Writing Samples and

Sentence Writing Fluency) as criterion variables. Although the article claims the results are useful for predicting the relationship of selected CHC cognitive clusters with basic writing skills and written expression, the methodological approach used within this study may be questioned. The analyses used simultaneous multiple regression to investigate the influence of all seven CHC cognitive factors but did not consider the effects of the general factor (as measured by the GIA). Additionally, the CHC factors were not investigated separately to determine the *R*-squared change for each factor. Without these additional analytical steps (i.e., investigation of *g* only; *g* plus seven CHC factors; *g* plus each CHC factor separately), the article's conclusion that Gc, Gs, and Gf are "especially important predictors of basic writing skills and written expression" (p. 787) may be questioned. While it is acknowledged that the authors have found an association, it is contended that the importance of that association is better investigated using an incremental rather than simultaneous approach to regression.

In the other two studies (e.g., Cormier, Bulut et al., 2017; Cormier, McGrew et al., 2017), the influence of *g* was considered against the backdrop of the seven CHC factors simultaneously entered into the regression models in the respective prediction of reading and writing. In both studies the incremental inclusion of each separate CHC factor in the regression analyses was not investigated but would have permitted a more granular examination of the predictive capacity of the factors and whether they contained sufficient variance to warrant further interpretive consideration (Keith, 2015; Pedhazur, 1997). Instead, it is noted that the change in *R*-squared reflects that of the seven CHC factors in combination. Moreover, the presentation of additional statistical information (e.g., *F* ratios, significance level, degrees of freedom, unstandardized beta weights, and standard errors) commonly found in multiple regression studies would have been helpful. Since both studies omitted participants without complete cognitive and achievement data, a comparison to the imputed standardization sample would have also been worthwhile. In fairness, it is recognized that the authors undertook a significant number of analyses and perhaps had to be selective in their reporting of the results.

The Cormier, McGrew et al. (2017) reading prediction study reported that the variance accounted for by the general factor ranged from 29% to 60% depending upon the age range. The results further indicated a medium effect and that the broad CHC clusters account for a moderate amount of additional variance in Reading Rate and Reading Fluency with the broad CHC cluster *R*-squared change values of .13 and .08, respectively. Following these findings, the authors moved to an investigation of the standardized regression coefficients to make the case that the Gf, Gc, and Gs clusters have interpretive relevance because their standardized regression coefficients were sufficiently high (> .10 = moderate; > .30 = strong). As mentioned, the author's case for this position would have been better supported had they actually investigated and presented the *R*-squared change values following the incremental addition of each factor (i.e., Gc, Gf, Gs, etc.).

In the Cormier, Bulut et al. (2017) math achievement predictive study, the authors noted that the broad CHC abilities of Gf, Gc, and Gs demonstrate significant relations above and beyond those accounted for by general intelligence. Specifically, the authors concluded that Gf, Gc, and Gs appear to have significant relations with math calculation skills throughout the school years.[5] In addition, Gf and Gc also demonstrate consistent relations with math problem solving throughout the school years. However, it is important to consider the results of the analysis furnished in the study. First, unlike Cormier, McGrew et al. (2017), the variance accounted for by the GIA was not presented. The degree of variance accounted for by the GIA would have been worthwhile to understand. Second, and as with the Cormier, McGrew et al. (2017) study, the authors did not present the individual CHC factor *R*-squared change and instead used the *R*-squared change for the seven CHC factors in combination. The reported change in *R*-squared value was .05 to .10 for math calculation skills and .04 to .13 for math problem solving across 14 age groups. However, a review of the *R*-squared change values revealed that, with the exception of ages 6 and 12 for Math Problem Solving and age 9 for Math Computation, the variance accounted for by the CHC clusters in the prediction of both math clusters was generally below .08 suggesting a small but non-trivial incremental addition of variance. Following this finding, the author's inspected the standardized regression coefficients of each of the CHC clusters, which led them to the assertion that Gf, Gc, and Gs appear to provide additional contribution of variance above that accounted for by the general factor. Again, an appropriate approach to determine the predictive relevance of the specific CHC factors would have been to investigate the *R*-squared change with each specific CHC factor singularly, something that was not done in any of the Cormier studies.

When looking to predictive validity studies on instruments aligned with CHC theory, including the WJ III, the empirical evidence suggests that the general factor's predictive relevance exceeds that of any group factor for an outcome such as academic achievement (Zaboski et al., 2018). And incremental predictive studies that account for the influence of general intelligence mostly find that the residual predictive influence of CHC factors is generally trivial with the exception of Gc predicting Oral Expression (e.g., change in *R*-squared Gc = .23, McGill & Busse, 2015).

What does this all mean in terms of applied interpretive relevance? The conclusions produced by the Cormier et al. (2016) analyses should be given less emphasis when making generalizations regarding the predictive relevance regarding the writing clusters given that the analysis contained the analytical omissions described previously. The other two Cormier et al. (2017) studies demonstrated that while it is technically correct that there is an additional contribution of variance (when all the CHC clusters are included) for reading rate, reading fluency, math calculation, and math problem solving, it appears that this incremental amount of variance is generally small in comparison to that provided by the GIA. However, it remains less clear whether and which factors actually make an incremental contribution as each CHC cognitive factor was not investigated separately for its incremental variance contribution. Even if it is assumed that there is non-trivial incremental variance attributable to Gc, Gf, and Gs—which should only tentatively be concluded because this analysis was not undertaken—it may not be worthwhile to move to a consideration of these factors as targets for the interpretive prediction of achievement. What does it actually mean to say that all seven CHC factors in combination offer a nominal amount of additional predicative variance? Ultimately this is a value judgment. However, considering the small amount of variance attributable to all seven CHC factors in combination, it appears that these factors are less important in comparison to the general factor and may not be worth the time, effort, and expense to administer and interpret.

Of course, the Cormier and colleagues series of studies are predicated upon evidence for the structural validity of the WJ IV Cognitive; yet, the results of independent examinations of both the full battery (Dombrowski et al., 2018a) and the WJ IV Cognitive (Dombrowski et al., 2017; 2018b) demonstrate that evidence for the publisher posited structural model is questionable. Even if independent factor analytic results for the myriad scores on the WJ IV suggested the attainment of simple structure and a hierarchical regression analysis is employed, it is possible that the predictive validity of the lower-order factors would follow the same pattern of results from the extended cognitive ability literature where the general factor variance accounts for the majority of variance in predicting a specific outcome such as achievement. Unfortunately, the results from Dombrowski et al. (2017; 2018a; 2018b) suggest a lack of simple structure and that the factors proposed by the technical manual should be considered empirically fragile. Simply put, the veracity of the publisher-proposed CHC factors and the other composites for the WJ IV Cognitive (and achievement [Dombrowski, McGill, Benson et al., 2019]) has yet to be definitively established so using them as a basis for predictive validity is problematic and prone to producing specious clinical outcomes.

Diagnostic Utility

Diagnostic utility (DU) is another consideration. It is likewise predicated upon the attainment of structural validity. DU employs a type of analysis that determines whether, and how much greater than chance, a score or combination of scores will be when making diagnostic decisions. In the case of tests of cognitive ability, combinations of index level scores have been used to make diagnostic decisions such as SLD. For instance, Flanagan et al. (2013) extensively used the WJ IV battery of tests (and other batteries aligned with CHC theory) as part of their cross-battery assessment SLD diagnostic approach. There are other approaches as well that use a type of PSW approach for diagnosing SLD (e.g., Hale & Fiorello, 2004; Schultz & Stevens, 2015), but the empirical foundation for these interpretative algorithms have been called into question (Benson et al., 2019a, 2019b; Dombrowski, Benson & Maki, 2024; McGill et al., 2018). Most of the evidence for the viability of PSW comes from studies that show the association of a specific CHC factor with an academic achievement area (see Fenwick et al., 2016; Fiorello et al., 2006 for examples). Although these types of associative studies may be an important first step toward the understanding of a linkage between cognitive and academic achievement within a PSW model, a more authoritative approach would be to employ diagnostic utility methodology.

As mentioned, this procedure will permit an understanding of whether these group differences are diagnostic (i.e., how much more accurate the procedure is in offering a diagnosis than a simple coin flip). Although presently accumulating, the available diagnostic utility statistics for permutations of these approaches featuring the WJ III/WJ IV suggest that diagnostic decisions rendered in these models, featuring these data, operate at chance and in some cases less than chance levels (Kranzler et al., 2016, 2019; Maki et al., 2022). For instance, a recently published systematic review details the lack of evidentiary support for the diagnostic accuracy of the various PSW models and suggests the field should abandon the procedures (Dombrowski, Benson & Maki, 2024).

Additionally, it is important to consider the lack of research regarding the longitudinal stability of PSW profiles. If the stability of these profiles is unknown or inadequate, then this creates problems for their use for diagnostic utility purposes. As previously noted, the utilization of the scores undergirding the PSW profiles are predicated upon structural validity evidence so there could be a cascading problem with validity and then diagnostic utility. Future research regarding the longitudinal stability of the various PSW profiles is clearly necessary but only after issues with the structural validity of the WJ IV Cognitive are addressed.

Direct Subtest, Index, and Cluster Level Interpretation

Direct subtest, index, and cluster level interpretation is likely the most commonly used approach to interpretation of the WJ IV Cognitive. Although there is consensus within the scientific community that subtest level interpretation in its various forms for diagnostic purposes should be eschewed (McGill et al., 2018; Watkins, 2000), there are still calls to engage in clinically astute detective work (Kaufman & Lichtenberger, 2006) or to use clinical judgment to guide this practice (Sattler, 2018). It is surprising that we still need to invoke this admonition after more than 30 years of empirical rebuke of this practice. At best this suggests that clinical assessment is slow to self-correct; at worst, but perhaps too Draconian, it raises concern about whether clinical assessment should be regarded as a pathological science (Michell, 1999) and may be impervious to scientific reform (Dombrowski & McGill, 2024). Nonetheless, training and subsequent practice are generally guided by the available assessment texts, which may well be responsible for promoting practices that can be regarded as low-value (Dombrowski, McGill, Farmer et al., 2022; Farmer et al., 2021).

As mentioned throughout, even in some of the factors that have been located for the WJ IV Cognitive, they often contained a different subtest composition than what was proposed by the publisher. These findings would preclude interpretation of the index for any of the factors that did not attain concordance between the structure proposed by the test publisher and that found within factor analysis.

Even if the WJ IV Cognitive attained simple structure (i.e., subtests clearly load theoretically proposed factors without any cross loading), which it does not, there are several metrics of interpretability that ought to be considered before engaging in index level interpretation (see Dombrowski, 2020 for a discussion and example). These metrics have not been incorporated into the technical manuals of commercial ability measures or in interpretive guides that foster contemporary test interpretation (e.g., Flanagan & McDonough, 2018). They include explained common variance (ECV), explained total variance (ETV), omega coefficients, and *H* (an index of factor replicability). Omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}) estimate the unit-weighted portion of reliable variance in latent factors. The ω_H coefficient is the estimate for the general intelligence factor with variability of group factors removed while the ω_{HS} coefficient is the estimate of a group factor with all other group and the general factor removed (Rodriguez et al., 2016). Although subjective, it has been suggested that omega coefficients should at a minimum exceed .50, but .75 is preferred (Reise et al., 2013). Additionally, it is important to consider ECV, ETV, and construct replicability (*H*). If the WJ IV factor or subtest scores fail to capture meaningful portions of true score variance they will likely be of limited clinical utility. The technical manual for the WJ IV battery of tests would have benefited from this information although in fairness the WJ IV was published before the widespread dissemination of these metrics.

Omega-hierarchical estimates for Dombrowski et al.'s (2018a) CFA study (see Table 4.2) indicate that while the broad *g* factor contained sufficient target construct variance to permit individual interpretation (18 subtest ω_H = .80 for age 9–13), the ω_{HS} estimates for the four WJ IV Cognitive group factors were generally low (< .50), suggesting that the group factors should be cautiously interpreted if at all (Brunner et al, 2012; Reise, 2012). Similarly, the *H* index (Hancock & Mueller, 2001) furnished evidence for individual interpretation of the general factor (*H* = .89 for ages 9–13) but not the respective group factors (i.e., *H* < .65 across all group factors with most < .50).[6] Similarly, ECV and ETV scores indicate that the general factor contains approximately five to seven times more variance than the respective group factors. Put simply, it appears that the WJ IV Cognitive is likely overfactored and contains insufficient target construct variance at the group-factor level for confidant clinical interpretation of those indices (Frazier & Youngstrom, 2007).

<T04_002>

Conclusion

Although the publisher proposed a theoretical linkage of the WJ IV Cognitive with seven CHC factors, the research evidence provided at the 9 to 19 age range does not fully support this proposed linkage at the present time. Instead, results suggest a more parsimonious four-factor solution and offers a different theoretical conceptualization for the WJ IV Cognitive, one 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). Additionally, the panoply of additional scores, which were created by the test publisher by simply combining subtests without structural validity evidence, are to be questioned until they are afforded adequate empirical support.

One final point deserves consideration. If the structure proposed within an instrument's technical manual is not established and, further, cannot be consistently replicated by independent studies, then engaging in accurate CHC–based profile analytic interpretive procedures (e.g., XBA, PSW analyses) will be challenging if not potentially contraindicated (Beaujean et al., 2018). The foundation for such practice rests upon a theoretical/factor structure that has consistent and replicated empirical support. The evidence presented in this chapter suggests that this support is not as strong as is frequently conveyed in the non-empirical literature and challenges the interpretive practices posed within the commercially marketed literature, test technical manuals, or workshops claiming to be evidence based.

Therefore, the field is advised to exercise 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 until more consistent, empirical support for these approaches appears (Cucina & Howardson, 2017). Consequently, users are encouraged to interpret the scores provided by the WJ Cognitive circumspectly to ensure that interpretive practice is guided by the presently available empirical evidence (Dombrowski et al., 2007; Dombrowski et al., 2006) rather than theoretical or intuitive considerations. Although calls to use to clinical judgment in regard to cognitive ability test interpretation within technical manuals and other interpretive guidebooks are intuitively appealing,[7] they often offer little additional insight leading to diagnosis at better than chance levels. To persist with interpretation based upon intuitive considerations thinking that one has the capacity to make clinically astute decisions using myriad IQ test profiles (despite the empirical evidence contraindicating this practice) should be considered a form of interpretive hubris. By contrast, the extant empirical literature suggests that there is no pot of clinical gold at the end of these interpretive practices. Continuing to interpret the WJ IV Cognitive (or, for that matter, any test of cognitive ability) without consideration of the totality of the empirical evidence risks characterizing assessment psychology as a pathological discipline (Langmuir & Hall, 1989) and little more than a professional granfalloon.[8]

References

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*(4), 121–136. https://doi.org/10.3390/jintelligence3040121

Beaujean, A. A., Benson, N., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ Scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence*, *6*(3), 1–25. https://doi.org/10.3390/jintelligence6030036

Benson, N. F., Beaujean, A., McGill, R. J., & Dombrowski, S. C. (2018). Critique of the core-selective evaluation process. *The DiaLog*, *47*(2), 14–18.

Benson, N. F., Beaujean, A. A., McGill, R. J., & Dombrowski, S. C. (2019). Rising to the challenge of SLD identification: A rejoinder. *The DiaLog*, *48*(1), 17–18.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25*(1), 67–74. http://doi.org/10.1207/s15327906mbr2501_8

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.14676494.2011.00749.x

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge.

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*(4), 533–541. https://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*(4), 305–317. http://doi.org/10.1037/a0025973

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). Hogrefe.

Canivez, G. L. (2017). Review of the Woodcock-Johnson IV. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurements yearbook* (pp. 875–882). Buros Center for Testing.

Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC–V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, *27*(2), 274–296. https://doi.org/10.1177/1073191118811609

Canivez, G. L., & Watkins, M. W. (2010). Investigation of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV): Exploratory and higher-order factor analyses. *Psychological Assessment*, *22*(4), 827–836. http://doi.org/10.1037/a0020429

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*(8), 975–986. https://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*(4), 458–472. http://doi.org/10.1037/pas0000358

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, *30*(3), 429–452. http://doi.org/10.1207/s15327906mbr3003_6

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

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Plenum.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment*, *35*(8), 731–754. https://doi.org/10.1177/0734282916659208

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet, Fifth Edition. *Journal of Psychoeducational Assessment, 24*(2), 123–136. https://doi.org/10.1177%2F0734282905285244

Dombrowski, S. C. (2013). Investigating the structure of the WJ–III Cognitive at school age. *School Psychology Quarterly*, *28*(2), 154–169. https://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*(4), 330–341. https://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*(6), 483–494. https://doi.org/10.1177%2F0734282914530838

Dombrowski, S. C. (2015). Exploratory bifactor analysis of the WJ–III Achievement at school age via the Schmid-Leiman orthogonalization procedure. *Canadian Journal of School Psychology*, *30*(1), 34–50. https://doi.org/10.1177%2F0829573514560529

Dombrowski, S. C. (2020). A newly proposed framework and a clarion call to improve practice. In S. C. Dombrowski, *Psychoeducational assessment and report writing* (2nd. ed., pp. 9–59). Springer Nature.

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*(1), 3–10.

Dombrowski, S. C., Beaujean, A. A., Schneider, J. W., McGill, R. J., & Benson, N. (2019). Using exploratory bifactor analysis to understand the latent structure of multidimensional psychological measures: An applied example featuring the WISC-V. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(6), 847–860. https://doi.org/10.1080/10705511.2019.1622421

Dombrowski, S. C., Benson, N. F., & Maki, K. E. (2024). A Systematic Review of the PSW Diagnostic Accuracy Evidence for SLD Identification: Is It Time to Abandon PSW? *School Psychology Review*, 1–19. https://doi.org/10.1080/2372966X.2024.2369494

Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2018). Factor structure of the 10 WISC–V primary subtests across four standardization age groups. *Contemporary School Psychology*, *22*, 90–104. https://doi.org/10.1007/s40688-017-0125-2

Dombrowski, S. C., Golay, P., McGill, R. J., & Canivez, G. L. (2018). Investigating the theoretical structure of the DAS–II core battery at school age using Bayesian structural equation modeling. *Psychology in the Schools*, *55*(2), 190–207. https://doi.org/10.1002/pits.22096

Dombrowski, S. C., Kamphaus, R. W., et al. (2006). The Solomon Effect in learning disabilities diagnosis: Have we not yet learned from history? *School Psychology Quarterly*, *21*(3). 359–373. https://doi.org/10.1037/h0084128

Dombrowski, S. C., & McGill, R. J. (2024). Clinical assessment in school psychology: Impervious to scientific reform? *Canadian Journal of School Psychology*, *39*(4), 297-306. https://doi.org/10.1177/08295735231224052

Dombrowski, S. C., McGill, R. J., Benson, N., & Beaujean, A. A. (2019). The Woodcock-Johnson IV Tests of Achievement provides too many scores for clinical interpretation. *Journal of Psychoeducational Assessment, 37*(7), 819–836. https://doi.org/10.1177/0734282918800745

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Cognitive at school age. *Psychological Assessment*, *29*(4), 294–407. https://doi.org/10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018a). An alternative conceptualization of the theoretical structure of the Woodcock-Johnson IV Tests of Cognitive Abilities at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology*, *6*(1), 1–13. https://doi.org/10.1037/arc0000039

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018b). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology Quarterly*, *33*(2), 235–250. https://doi.org/10.1037/spq0000221

Dombrowski, S. C., McGill, R. J., Canivez, G. L., & Peterson, C. H. (2019). Investigating the theoretical structure of the Differential Ability Scales-Second Edition through hierarchical exploratory factor analysis. *Journal of Psychoeducational Assessment*, *37*(1), 94–104. http://dx.doi.org/10.1177/0734282918760724

Dombrowski, S. C., McGill, R. J., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2021). Factor analysis and variance partitioning in intelligence research: Clarifying misconceptions. *Journal of Psychoeducational Assessment*, *39*(1), 28–38. https://doi.org/10.1177/0734282920961952

Dombrowski, S. C., McGill, R. J., Farmer, R. L., Kranzler, J. H., & Canivez, G. L. (2022). Beyond the rhetoric of evidence-based assessment: A framework for critical thinking in clinical practice. *School Psychology Review*, *51*(6), 771–784. https://doi.org/10.1080/2372966X.2021.1960126

Dombrowski, S. C., McGill, R. J., & Morgan, G. B. (2021). Monte Carlo modeling of contemporary Intelligence Test (IQ) factor structure: Implications for IQ assessment, interpretation, and theory. *Assessment*, *28*(3), 977–993. https://doi.org/10.1177/1073191119869828

Dombrowski, S. C., McGill, R. J., Watkins, M. W., Canivez, G. L., Jacobson, L., & Pritchard, A. (2022). Will the *real* theoretical structure of the WISC–V please stand up? An exploratory factor analytic replication study. *Contemporary School Psychology*, *26*, 492–503. https://doi.org/10.1007/s40688-021-00365-6

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*(2), 442–455. https://doi.org/10.1037/a0031335

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, *27*(6), 279–286. https://doi.org/10.1177/0734282909333179

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. http://doi.org/10.1037/1082-989X.4.3.272

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2021). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology*, *36*(2), 98–114. https://doi.org/10.1177/0829573520978111

Fenwick, M. E., Kubas, H. A., Witzke, J. W., Fitzer, K. R., Miller, D. C., Maricle, D. E., ... & Hale, J. B. (2016). Neuropsychological profiles of written expression learning disabilities determined by concordance-discordance model criteria. *Applied Neuropsychology: Child*, *5*(2), 83–96. https://doi.org/10.1080/21622965.2014.993396

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools*, *43*(8), 835–853. https://doi.org/10.1002/pits.20192

Flanagan, D. P., & McDonough, E. M. (Eds.) (2018). *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed.). Guilford Press.

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

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155–171. https://doi.org/10.1002/pits.10083

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

Gustafsson, J. E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107–135). Greenwood Press.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* Guilford Press.

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—A festschrift in honor of Karl Joreskog* (pp. 195–216). Scientific Software International.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. http://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), 1–55. https://doi.org/10.1080/ 10705519909540118

Jensen, A. R. (1998). *The* g *factor: The science of mental ability.* Praeger.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). John Wiley.

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York: Routledge.

Kline, R. B. (2016). Methodology in the social sciences. *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kranzler, J. H., Floyd, R. G., Benson, B., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth. *International Journal of School & Educational Psychology, 4*(3), 124–136. https://www.doi.org./10.1080/21683603.2016.1155515

Kranzler, J. H., Gilbert, K., Robert, C. R., Floyd, R. G., & Benson, N. F. (2019). Further examination of a critical assumption underlying the Dual-Discrepancy/Consistency approach to specific learning disability identification. *School Psychology Review*, *48*(3), 207–221. htpps://www.doi.org./10.17105/SPR-2018-0008.V48-3

Kush, J. C., & Canivez, G. L. (2019). The higher order structure of the WISC-IV Italian adaptation using hierarchical exploratory factor analytic procedures. *International Journal of School & Educational Psychology*, *7*(Sup1), 15–28. https://doi.org/10.1080/21683603.2018.1485601

Langmuir, I., & Hall, R. N. (1989). Pathological science. *Physics Today*, *42*(10), 36–48. https://doi.org/10.1063/1.881205

Maki, K. E., Kranzler, J. H., & Moody, M. E. (2022). Dual discrepancy/consistency pattern of strengths and weaknesses method of specific learning disability identification: Classification accuracy when combining clinical judgment with assessment data. *Journal of School Psychology*, *92*, 33–48. https://doi.org/10.1016/j.jsp.2022.02.003

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519–530. https://doi.org/10.1093/biomet/57.3.519

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology, 22*(3), 384–393. https://doi.org/10.1007/s40688-017-0168-4

McGill, R. J. (2023). Confirmatory factor analysis of the WJ IV Cognitive: What does the standard battery measure at school age? *Journal of Psychoeducational Assessment, 41*(4), 461-468. **https://doi.org/10.1177/07342829231159440**

McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly*, *30*(3), 353–365. https://doi.org/10.1037/spq0000094

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, *71*, 108–121. https://doi.org/10.1016/j.jsp.2018.10.007

McGrew, K. S. (2012). *Implications of 20 Years of CHC cognitive-achievement research: Back to the future and beyond CHC*. Woodcock-Muñoz Foundation Press.

McGrew, K. S., LaForte, E. M., & Shrank, F. A. (2014). *Woodcock–Johnson IV technical manual.* Riverside Publishing.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge University Press.

Miller, D. C., & Maricle, D. E. (2019). *Essentials of school neuropsychological assessment* (3rd ed.). John Wiley.

Miller, D. C., McGill, R. J., & Johnson, W. L. B. (2016). Neurocognitive applications of the Woodcock-Johnson IV. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 355–388). Academic Press. https://doi.org/10.1016/B978-0-12-802076-0.00013-X

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*(1), 129–140. https://doi.org/10.1037/a0024878

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation*, *20*(2). https://doi.org/10.7275/hb2g-m060

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Prediction and explanation* (3rd ed.). New York, NY: Holt, Rinehart & Winston.

Pratkanis, A. R. (1995). How to sell a pseudoscience. *The Skeptical Inquirer*, *19*, 19–25.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

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*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). La Mesa, CA: Sattler Publishing.

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

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Schrank, F. A., & Dailey, D. (2014). Woodcock–Johnson online scoring and reporting [Online format]. Riverside.

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

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

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock-Johnson IV.* Riverside.

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

Schrank, F. A., & Wendling, B. J. (2015a). WJ IV Interpretation and Instructional Interventions Program (WIIIP). Riverside.

Schultz, E. K., & Stephens, T. L. (2015). Core-Selective Evaluation Process: An efficient and comprehensive approach to identify students with SLD using the WJ IV. *The DiaLog*, *44*(2), 5–12.

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

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*(2), 689–697. https://doi.org/10.1037/pas0000052

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll Theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly*, *19*(1), 72–87. https://doi.org/10.1521/scpq.19.1.72.29409

Taub, G. E., & McGrew, K. S. (2014). The Woodcock-Johnson Tests of Cognitive Abilities Cognitive Performance Model: Empirical support for intermediate factors within CHC theory. *Journal of Psychoeducational Assessment*, *32*(3), 187–201. https://doi.org/10.1177/0734282913504808

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association. https://doi.org/10.1037/10694-000

Vonnegut, K. (1976). *Wampeters, Foma & Granfalloons (opinions).* Dial Press.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, *15*(4), 465–479. https://doi.org/10.1037/h0088802

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 251–268). Guilford Press.

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–Fifth Edition. *International Journal of School and Educational Psychology*, *6*(4), 252–265. https://doi.org/10.1080/21683603.2017.1342580

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*(3), 263–311. https://doi.org/10.1207/ s15327906mbr2803_1

Youngstrom, E. A., & Van Meter, A. R. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice,* 23(4), 327–347. https://doi.org/10.1111/cpsp.12172

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of School Psychology*, *71*, 42–56. https://doi.org/10.1016/j.jsp.2018.10.001

Table 1

Subtest Alignments by CHC Factor and Narrow Ability

Gc	**Gf**	**Gwm**
Oral Vocabulary (1)	Number Series (2)	Verbal Attention (3)
General Information (8)	Concept Formation (9)	Numbers Reversed (10)
	Analysis-Synthesis (15)*	Object-Number Sequencing (16)*
Gs	**Ga**	**Glr**
Letter-Pattern Matching (4)	Phonological Processing (5)	Story Recall (6)
Pair Cancellation (17)	Nonword Repetition (12)	Visual-Auditory Learning (13)
Gv	**Quantitative Reasoning**	**Number Facility**
Visualization (7)	Number Series (2)	Numbers Reversed (10)
Picture Recognition (14)	Analysis-Synthesis (15)	Number-Pattern Matching (11)
Perceptual Speed	**Cognitive Efficiency**	
Letter-Pattern Matching (4)	Verbal Attention (3)*	
Number-Pattern Matching (11)	Letter-Pattern Matching (4)	
	Numbers Reversed (10)	
	Number-Pattern Matching (11)	

Note. *Additional tests are required to create an extended version of the cluster listed.
Adapted from the Technical Manual by McGrew et al., (p. 9, 2014).

Table 2

Sources of WJ IV Subtest Variance According to a Bifactor CFA (Ages 9-13)

	General	First Order Factors			
	g	Working Memory (Gwm)	Perceptual Reasoning (Gf/Gv)	Processing Speed (Gs)	Verbal Ability (Gc)
Subtest	*b*	*b*	*b*	*b*	*b*
Verbal Attention (Gwm)	.59	**.45**			
Memory for Words (Aud Mem)	.52	**.48**			
Object Number Sequence (Gwm)	.63	**.36**			
Nonword Repetition (Ga)	.46	**.37**			
Phonological Processing (Ga)	.62	**.24**			
Numbers Reversed (Gwm)	.57	**.16**			
Visualization (Gv)	.51		**.51**		
Visual-Auditory Learning (Glr)	.43		**.38**		
Picture Recognition (Gv)	.35		**.43**		
Analysis-Synthesis (Gf)	.62		**.31**		
Concept Formation (Gf)	.63		**.21**		
Story Recall (Glr)	.52		**.19**		
Letter-Pattern Matching (Gs)	.50			**.60**	
Number-Pattern Matching (PerSpd)	.49			**.60**	
Pair Cancellation (Gs)	.41			**.61**	
Number Series (Gf)	.66			**.13**	
Oral Vocabulary (Gc)	.67				**.61**
General Information (Gc)	.50				**.61**
Explained Common Variance	.61	.09	.09	.13	.09
Explained Total Variance	.30	.04	.04	.06	.04
ω_H/ω_{HS}	.80	.22	.24	.39	.44
H	.89	.48	.48	.64	.55

Note. *b* = factor loading, ω_H = Omega hierarchical (*g*), ω_{HS}= Omega hierarchical subscale (group factors). *H* =construct replicability. Alignment of subtests with respective CHC stratum I or II factors posited in the WJ IV Technical Manual is indicted following each subtest name. Adapted from Dombrowski et al. (2018).

[1] The WJ IV Technical Manual reports 50 subtests yet only the 47-subtest correlation matrix was furnished.

[2] Dombrowski and Watkins (2013) noted that there is an element of circularity—a psychometric tautology, if you will—as it was used to initially provide evidence for CHC Theory and subsequently serve as the main reference instrument for making future refinements to the theory.

[3] In fairness, the test authors may have been in the final stages of developing the WJ IV and could have understandably overlooked this body of research that, incidentally, represented the first independent examination of the factor structure of the WJ III battery of tests 10 years after the instrument was published.

[4] It is noted that the variance apportioned to higher- and lower-order dimensions may be accomplished in other ways as well and regardless of whether a higher-order or bifactor conceptualization of intelligence is presumed (see Dombrowski, Golay et al., 2018, Table A4 for an example).

[5] It should be noted that in some PSW workshops, it is suggested that Gc cannot be the primary weakness identified in a confirmatory pattern associated with a learning disability.

[6] Metrics of interpretability are similarly available in Dombrowski et al. (2017) and consistent with those reported herein.

[7] The value of clinical judgment is not to be discounted as part of the broader psychological/psychoeducational assessment process. There are times when this is necessary and important. Extreme caution is suggested when using clinical judgment to decipher the myriad IQ test profile analytic scores. The history of this practice has not yielded meaningful clinical insight (see Dombrowski, McGill, Farmer et al., 2022; McGill et al., 2018; Watkins, 2000; Watkins et al., 2005) and is prone to a phenomenon known as the “Alchemist’s Fallacy.”

[8] Adapted from Vonnegut (1976). Dombrowski, McGill, Farmer et al. (2022) expand on his conceptualization by describing a professional granfalloon as a group of individuals who share an identity organized around a common cause that is potentially meaningless and scientifically unsupported. Pratkanis (1995) discusses one of the ways in which a pseudoscience may be marketed is via the creation of a granfalloon.