Contemporary School Psychology (2022) 26:492–503
https://doi.org/10.1007/s40688-021-00365-6

Will the *Real* Theoretical Structure of the WISC-V Please Stand Up? Implications for Clinical Interpretation

Stefan C. Dombrowski[1] · Ryan J. McGill[2] · Marley W. Watkins[3] · Gary L. Canivez[4] · Alison E. Pritchard[5] · Lisa A. Jacobson[5]

Accepted: 11 February 2021 / Published online: 8 April 2021

Abstract
The Wechsler Intelligence Scale for Children's (WISC) factorial\theoretical structure has undergone numerous substantive changes since it was first developed, and each of these changes has subsequently been questioned by assessment experts. Given remaining questions about the structure of the latest revision, the WISC-V, the present study used three different exploratory factor analytic techniques to investigate the factor structure of the 10 primary subtests in a large clinical sample ($N = 5359$). Results revealed that the WISC-V likely contains four factors (e.g., Gc, Gwm, Gs, and either Gv (via exploratory bifactor analysis) or a complexly determined perceptual reasoning factor (via the Schmid-Leiman procedure and oblique/higher-order factor analysis)), *not* the five factors proposed by the test publisher. Variance apportionment and omega estimates indicate that only secondary interpretive emphasis should be placed upon group factors, but only when there is structural validity support. This study suggests that the WISC-V measures four, not five, factors, although the four-factor configuration may be different than previously reported in the technical literature.

Keywords WISC-V · Evidence-based test interpretation · Assessment · Intelligence · IQ test interpretation

The factor structure of an intelligence test, or, for that matter, any psychological assessment instrument, has important implications for psychologists, assessment trainers, and researchers as they seek to better understand how a test should be interpreted in clinical practice. Evaluating an instrument's structural validity is also important because the results from such analyses provide the statistical rationale for the scores that are developed for that measure (Brunner et al., 2012). On a theoretical level, factor analytic results are necessary for determining whether an instrument is linked to the theory it purports to measure and have been critical in the evolution of intelligence theory (McGill & Dombrowski, 2019; Schneider & McGrew, 2018). It is important to note that structural validity is necessary, but singularly insufficient for establishing broader construct validity for test scores that are furnished by the test publisher. Stated another way, it is a necessary first step for construct validation, though contradictory evidence discovered at this stage makes it difficult to conclusively establish other important forms of construct validity (e.g., predictive validity, diagnostic validity, treatment utility; Keith & Kranzler, 1999).

The theoretical structure of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) has demonstrated considerable lability since its initial derivation from the Wechsler-Bellevue Intelligence Scale (Wechsler, 1946) over 75 years ago. For instance, the WISC-III (Wechsler, 1991) purported to measure three strata of abilities including *g*, verbal and performance IQs, and four additional factorially derived composite scores: Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index (FDI), and Processing Speed Index (PSI). The WISC-IV (Wechsler, 2003) eliminated the verbal and performance IQs as well as the FDI and claimed to measure a general factor in addition to verbal abilities, perceptual reasoning, working memory (essentially a renamed version of the FDI), and processing speed.

✉ Stefan C. Dombrowski
sdombrowski@rider.edu

[1] Department of Graduate Education, Leadership, and Counseling, Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648, USA

[2] William & Mary, Williamsburg, VA, USA

[3] Baylor University, Waco, TX, USA

[4] Eastern Illinois University, Charleston, IL, USA

[5] Kennedy Krieger Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

However, in a 2013 special issue of the *Journal of Psychoeducational Assessment* dedicated to understanding the latent structure of the WISC-IV, Weiss et al. (2013) reexamined the WISC-IV normative data and concluded that a five-factor model more consistent with Cattell-Horn-Carroll (CHC) abilities provided a better fit to the data than the publisher-preferred model. In that same issue, Canivez and Kush (2013) questioned the value of that model due to miniscule improvements in statistical fit when compared to the four-factor model, and evidence of potential model misspecification in the form of a perfectly redundant second-order loading between fluid reasoning and general intelligence. The results also suggested that identification of fluid reasoning was only possible through the specification of an intermediary factor (Induction), further complicating clinical interpretation as this indictor appears to have been created to improve resulting fit statistics and is not available to users of the WISC-IV. Despite these concerns, the results furnished by Weiss et al. (2013) appear to have been influential for the subsequent revision of the instrument to match the CHC theory.

The most recent rendition, the Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V; Wechsler, 2014), was developed to measure five Cattell-Horn-Carroll (CHC; Schneider & McGrew, 2018) abilities: crystalized ability (*Gc*), fluid reasoning (*Gf*), visual-spatial ability (*Gv*), working memory (*Gwm*), and processing speed (*Gs*). The most consequential and controversial change to the theoretical structure of the test was the decision to split the former PRI into standalone fluid reasoning (*Gf*) and visual-spatial (*Gv*) factors. Of note, the test publisher relied primarily on the 16-subtest full-test battery to determine the instrument's structure, even though it can be argued that the 10-subtest battery is more frequently administered in clinical practice settings (Benson et al., 2019; Oakland et al., 2016), and should have been investigated independent of the 16-subtest battery. Fit statistics for the publisher-preferred model for 10-subtest configuration are reported in the *Technical and Interpretive Manual* (Wechsler, 2014); however, no rival models were explored. Additionally, the standardized coefficient between fluid reasoning (FR) and *g* reported in Figure 5.2 (p. 84) approaches unity (.99), suggesting that the discriminant validity between those constructs is likely questionable, and that the FR factor may be empirically redundant (Brown, 2015).

Although linkage with theory is laudable, the theoretical structure of an instrument requires empirical validation rather than simple extrapolation from a battery with more numerous subtests, especially when the broader 16-subtest battery's theoretical structure has been questioned (Dombrowski, McGill & Morgan, 2019). Exploratory factor analytic techniques are worthwhile for this purpose as they let the data speak for themselves, furnish more complete information about the interrelationship among variables and latent constructs that may better reflect reality than confirmatory factor analytic models constrained by identification, and may better avoid errors in scientific decision-making such as confirmation bias (Horn, 1989). Of course, the ultimate arbiter of a final, validated theoretical structure is replication across samples (i.e., clinical and standardization samples) using multiple methods and models of factor analysis (Gorsuch, 1983). When multiple methods of factor analysis converge, then we can be more confident in a derived theoretical model and resulting scoring structure for an instrument.

Few, if any, studies investigating the WISC-V factor structure have furnished empirical support for the publisher-proposed five-factor theoretical structure (see Dombrowski et al., 2020 for a psychometric meta-analysis). Instead, independent studies evaluating (and focusing primarily on) the 16-subtest full battery, whether on the US-based standardization sample (e.g., Beaujean, 2016; Canivez & Watkins, 2016; Canivez et al., 2016, 2017; Dombrowski et al., 2015, Dombrowski, Beaujean at al., 2019) international standardization samples (e.g., Canivez et al., 2019; Watkins et al., 2018; Lecerf & Canivez, 2017), or simulated samples (e.g., Dombrowski et al., 2019), have suggested that the WISC-V measures four factors reminiscent of the WISC-IV ((e.g., *Gc*, *Gwm*, *Gs*, and perceptual reasoning (PR; fusion of *Gf*/*Gv* subtests)). The only structural validity study to date that has supported a five-factor CHC structure was that offered by Reynolds and Keith (2017) who investigated the 16-subtest full battery, and engaged in a series of model adjustments (~17) to arrive at their final, adopted model. This model was subsequently criticized in the scientific literature (e.g., Canivez, Mcgill, Dombrowski et al., 2020; Dombrowski, McGill et al., 2021) as lacking a practical basis for proper use and interpretation by psychologists.

Although the 10 primary WISC-V subtest battery is likely more frequently administered than the full battery (Benson et al., 2019; Oakland et al., 2016), only a few studies to date have investigated the theoretical structure of the 10 primary subtest battery, with two using the standardization sample (e.g., Canivez & Watkins, 2016; Dombrowsk, Canivez et al., 2017) and one using a clinical sample (Canivez, McGill, Dombrowski et al., 2020). Additionally, Graves et al. (2020) and Dombrowski, Watkins et al. (2021) used a clinical sample, but the emphasis of that study was on the invariance of the scoring structure of the instrument that did not consider the general factor. The study by Canivez, McGill, Dombrowski et al. (2020) reported a four-factor structure reminiscent of the prior WISC-IV (e.g., PR, *Gc*, *Gwm*, and *Gs*). Using the standardization sample, Dombrowski, Canivez et al., (2017) partitioned the 10 primary subtest battery into four age ranges (6 to 8; 9 to 11; 12 to 14; and 15 to 16), and found support for the five-factor CHC structure at ages 15 to 16, but not ages 6 to 14 (four factors found). Thus, the majority of studies on the 10 subtest primary battery provide support for a four-factor WISC-V theoretical structure, though questions still remain about the primary battery's linkage with theory.

To date, most of the studies have focused on the 16-subtest primary and secondary battery, offering different four-factor and five-factor configurations, with only the four-factor WISC-IV theoretical model having any evidence of replication. Few studies have focused exclusively on the 10 primary subtest battery, the battery most commonly used by researchers and practicing psychologists. Evidence from the test publisher for a five-factor theoretical structure for the 10 primary subtest battery is underdeveloped. As noted by Cattell (1978), it should not be assumed that the constructs located by a more complex measurement model (i.e., the 16-subtest battery) will automatically be located by a more parsimonious model (i.e., the 10-subtest battery). While the test publisher used confirmatory factor analysis and investigated their preferred five-factor model (model 5e in the technical manual; Wechsler, 2014, p,84), no comparisons were made to any plausible competing models; instead, the test publisher provided evidence for the final, adopted model solely by looking at the magnitude of fit indexes, a practice not entirely supported in the structural equation modeling literature (Kline, 2016). In fact, the *Technical and Interpretive Manual* devoted only 115 words, less than this article's abstract word count, to substantiate the theoretical structure of the WISC-V 10-subtest primary battery (see below).

> The model selected to represent the WISC-V test structure (Model 5e) was fitted to the ten primary subtests for the entire age range. At this stage, the issue being investigated no longer concerned the number of factors influencing subtests. Rather, this analysis addressed how well the identified factors account for the intercorrelations among the reduced set of primary subtests. Goodness-of-fit results are shown in Table 5.4, and Fig. 5.2 presents the subtest and factor loadings. Fit is excellent, and the loadings are similar to those from the analysis of all primary and secondary subtests. These results support the effectiveness of this five-factor model in accounting for the subtest intercorrelations (Wechsler 2014, p. 84).

Furthermore, the isomorphism observed between fluid reasoning and *g* (i.e., .99) suggests that the model may be overfactored as evinced in numerous WISC-IV and WISC-V studies. In sum, the limited amount of evidentiary support furnished by the test publisher and available in the extant literature suggests that our understanding of the theoretical structure of the 10 primary subtest WISC-V battery remains incomplete.

To address this evidentiary lacuna, the purpose of this study was to investigate the theoretical structure of the WISC-V 10 primary subtest battery using a large clinical sample and three different exploratory factor analytic (EFA) modeling techniques: (a) a direct hierarchical solution via exploratory bifactor analysis (EBFA; Jennrich & Bentler, 2011); (b) a higher-order solution via common factor analysis (Gorsuch, 1983; Thurstone, 1947); and (c) a transformation of the higher-order solution via the Schmid-Leiman (SL) Orthogonalization procedure (Schmid & Leiman, 1957).

Multiple exploratory techniques were utilized in the present study given the remaining questions about the theoretical structure of the 10-subtest battery and the bifactor vs. higher-order model debate in the assessment literature (i.e., Beaujean, 2015b; Dombrowski, McGill & Morgan, 2019; Dombrowski, McGill et al., 2020). It is beyond the scope of the present discussion to fully adjudicate this debate; instead, our primary purpose was to attempt to ascertain whether the publisher-preferred five-factor CHC model was able to be replicated in the 10 primary subtest battery using a clinical sample and recommended EFA methods. It is important to acknowledge that when multiple methods of factor analysis converge then we can be more confident in a derived factor structure (Gorsuch, 1983). The results of this study portend to shed additional insight into the theoretical structure of the WISC-V 10 primary subtest battery via the use of several exploratory modeling techniques and a clinical sample that is more than double the size of the standardization sample. These results are not just important for the theoretical musings of researchers in the academy. They have direct relevance for how the WISC-V should be scored and interpreted both directly and via various methods of interpretation such as cross battery assessment and patterns of strengths and weaknesses analysis required in many states.

Method

Participants

A total of 5359 children between the ages of 6 and 16 years were administered the WISC-V as part of clinical assessments through a large, outpatient pediatric psychology/neuropsychology clinic within a children's specialty hospital. Deidentified data were retrieved from the electronic medical record for participants whose assessments included the 10 primary WISC-V subtests. The study was approved by the hospital's Institutional Review Board. Table 1 presents demographic characteristics of the total clinical sample. As shown, the sample was primarily composed of White/Caucasian and Black/African American youths. The participants' ages ranged from 6.0 to 16.93 years and averaged 10.69 years (SD = 2.74 years). Table 2 shows that three diagnostic groups (ADHD, 47.6%; anxiety, 11.6%; and encephalopathy, e.g., non-traumatic diffuse brain dysfunction; 10.1%) comprised just over two-thirds of the sample. Table 3 shows that the sample, compared with the US standardization sample, was slightly below average in subtest and composite scores as is typical in clinical samples. All subtests and composite scores showed univariate normal distributions with no appreciable skewness or kurtosis.

Table 1 Demographic characteristics of the clinical sample

			Sex		
Race/ethnicity	*N*	Percent	Female	Male	Unknown
White	2865	53.50%	885	1764	216
Black	1513	28.20%	443	940	130
Multi-racial	376	7.00%	117	257	2
Unknown/other	209	3.90%	40	100	69
Asian	191	3.60%	54	108	29
Hispanic/Spanish origin	190	3.50%	57	132	1
American Indian/Alaskan native	12	0.20%	5	7	0
Native Hawaiian or Pacific Islander	3	0.10%	1	0	1
Total	5359		1602	3308	448
Percent		100.00%	29.90%	61.70%	8.40%

Procedure and Analyses

All analyses were conducted using either SPSS Version 25 or the **R** Statistical programming Environment (R Development Core Team, 2020) featuring the *psych* (Revelle, 2012) package for factor extraction and the *GPArotation* (Bernaards & Jennrich, 2005) package for rotation. First, the target correlation matrix created from the raw data was examined for its adequacy for factor analysis using Bartlett's test of sphericity (Bartlett, 1954) and the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) criteria. Second, the number of factors to extract was examined by conducting the minimum average partial (MAP; Velicer, 1976) test, parallel analysis (PA; Horn, 1965) using both the mean and 95th percentile criteria, and the Bayesian Information Criteria (BIC). Third, the factors were extracted using principal axis factoring. For exploratory bifactor analysis, an orthogonal bigeomin rotation (Jennrich & Bentler, 2011) was applied to the extracted solution. This was conducted in **R** using the FindBifactor.orth syntax furnished by Loehlin and Beaujean (2016). The higher-order solution was obtained using SPSS through principal axis factoring with a promax rotation. For the SL solution, the principal axis promax-rotated solution was subsequently transformed via the Schmid-Leiman orthogonalization procedure using the SPSS syntax provided by Wolff and Preising (2005). Watkins' *Omega* (2013) program was used to provide estimates for various indices of interpretive relevance (e.g., H, ω) and percentage of uncontaminated correlations.

Table 2 Diagnostic categories of the clinical sample

ICD-10 diagnosis	*N*	Percent
ADHD/ADD	2552	47.6%
Encephalopathy	620	11.6%
Anxiety	539	10.1%
Adjustment disorder	213	4.0%
Behavior disorder	213	4.0%
Epilepsy	140	2.6%
Mood disorder	133	2.5%
Congenital anomaly	112	2.1%
Genetic condition	112	2.1%
Frontal lobe deficit	100	1.9%
Disorder of the nervous system	76	1.4%
Major depression	69	1.3%
Brain/spine injury	43	0.8%
Neoplasm/tumor	40	0.7%
Hearing loss	38	0.7%
Leukemia	38	0.7%
Unknown	37	0.7%
Other depressive disorder	30	0.6%
Autism spectrum disorder	25	0.5%
Cancer (not brain/nervous system)	23	0.4%
Emotional disturbance	16	0.3%
Expressive/receptive language disorder	13	0.2%
Fetal alcohol syndrome	13	0.2%
Bipolar disorder	11	0.2%
Other mental/psychological disorder	11	0.2%
Reading/learning disability	11	0.2%
Tic/Tourette's disorder	11	0.2%
MISC medical/psychiatric conditions	120	2.2%
Total	5359	100.0%

Note. *ICD*, International Classification of Diseases, tenth edition; *ADD*, attention-deficit disorder; *ADHD*, attention-deficit/hyperactivity disorder

Results

WISC-V Exploratory Factor Analysis

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy of .911 far exceeded the minimum standard of .60 (Kaiser, 1974; Kline, 1994; Tabachnick & Fidell, 2007) and Bartlett's Test of

Table 3 Descriptive statistics for the WISC-V clinical sample

Subtest/ Composite	Total Sample (N = 5,359)			
	M	*SD*	Skewness	Kurtosis
Subtests				
Block Design	8.69	3.33	0.12	-0.17
Similarities	9.14	3.30	-0.02	-0.07
Matrix Reasoning	9.02	3.38	0.08	-0.10
Digit Span	7.96	3.10	0.11	0.07
Coding	7.57	3.33	-0.03	-0.37
Vocabulary	8.99	3.56	0.06	-0.53
Figure Weights	9.52	3.12	-0.03	-0.28
Visual Puzzles	9.58	3.29	-0.04	-0.43
Picture Span	8.55	3.12	0.14	-0.16
Symbol Search	8.21	3.23	0.00	0.03
Composites				
VCI	94.96	17.41	-0.04	-0.18
VSI	95.15	17.23	0.08	-0.06
FRI	95.80	16.77	0.02	-0.37
WMI	89.94	15.82	0.14	-0.13
PSI	88.10	17.10	-0.15	0.03
FSIQ	91.03	17.27	-0.00	0.04

Sphericity ($\mathcal{X}^2 = 29{,}505.12, p < .000$; Bartlett, 1954) indicated that the WISC-V correlation matrix was not random. Therefore, the correlation matrix was deemed appropriate for factor analysis. The correlation matrix was extracted and examined using three different modeling procedures. For comparison purposes, all three models (i.e., higher-order, SL, and bifactor) used principal axis factor extraction and then applied either a promax or bigeomin (i.e., bifactor) rotation. The Schmid-Leiman orthogonalization procedure was applied to principal axis/promax-rotated factors.

Factor Extraction Criteria The number of factors to extract and rotate was determined by both empirical and theoretical considerations. The MAP criterion indicated extracting a single factor, while the BIC and PA (50th & 95th percentile) suggested four group factors. Although no index supported extracting five group factors, the results from such a solution were examined since that is the number of group factors indicated by the WISC-V publisher, likely of most interest to users of the instrument in clinical practice, and it was suggested that it is better to overextract than underextract (Wood, Tataryn, & Gorsuch, 1996) to examine performance of small factors.

Exploratory Bifactor Analysis Principal axis factoring followed by an orthogonal bigeomin rotation was applied to the WISC-V clinical sample (Table 4). Four and five factors were extracted, rotated, and examined. Extracting five factors revealed four interpretable group factors (*Gc*, *Gv*, *Gwm*, and *Gs*) and a fifth factor containing no salient loadings. This suggested over-factoring. Figure weights (FW) and matrix reasoning (MR) did not saliently load on any group factor and only loaded on the general factor. Extracting four factors revealed the same four interpretable group factors and loading patterns that were theoretically consistent except for the two hypothesized *Gf* subtests (MR and FW).

Schmid-Leiman Orthogonalization Principal axis factoring with a promax rotation followed by the SL orthogonalization procedure was the second method used to examine the theoretical structure of the WISC-V clinical sample (Table 4). Four and five factors were extracted, rotated, and examined. Extracting five factors revealed four interpretable group factors (PR, *Gc*, *Gwm*, and *Gs*) and one singlet, non-viable, fifth factor that was loaded by matrix reasoning. MR also cross-loaded on PR. Extracting four factors revealed four well-defined group factors (PR, *Gc*, *Gwm*, and *Gs*) with the desired simple structure that resulted in the *Gf* and *Gv* subtests coalescing into a PR group factor.

Higher-Order Principal Axis Factoring Four and five factors were extracted and examined using principal axis factoring followed by an oblique promax rotation (Table 4). Extracting five factors yielded four interpretable factors (PR, *Gc*, *Gwm*, and *Gs*) and one single subtest factor (MR cross-loaded on a fifth untenable factor and the PR factor). Such observations are symptoms of overextraction (Gorsuch, 2003). On the contrary, extracting four factors revealed four viable factors (PR, *Gc*, *Gwm*, and *Gs*) and a subtest loading pattern similar to the SL orthogonalization.

Sources of Variance and Indices of Interpretive Relevance Examination of explained common variance (ECV) and explained total variance (ETV) across the four-factor EBFA and SL solutions revealed the dominance of the general factor (ECV = .73 and .71 for EBFA and SL, respectively; ETV = .49 and .46, respectively; see Table 4). This exceeded the variance apportioned to group factors (ECV range = .05–.11 and .06–.09 for EBFA and SL, respectively; ETV range = .03–.07 and .04–.06, respectively) by multiples in excess of 10 in most cases. Model-based reliability estimates (ω_{H} and ω_{HS}) and construct reliability or construct replicability coefficients (*H*) indicated that while the broad *g* factor permits confident individual interpretation (EBFA $\omega_{\mathrm{H}} = .87$, SL $\omega_{\mathrm{H}} = .85$), the ω_{HS} and *H* estimates for the four WISC-V group factors were low (see Table 3) and limited for standalone clinical interpretation purposes (Brunner et al., 2012; Hancock & Mueller, 2001; Reise, 2012; Rodriguez et al., 2016). The PUC (.80) across both models was also suggestive of the dominance of the general factor and indicated that primary

Table 4 EBFA, Schmid-Leiman, and Higher Order Loading Patterns

Four Factor Extraction

EBFA Bigeomin (Orthogonal)

	g	*Gv*	*Gc*	*Gs*	*Gwm*	h^2
VP	.77	.45	-.01	-.01	-.02	.80
BD	.74	.35	-.07	.02	-.03	.68
FW	.74	.17	.01	-.03	-.03	.58
MR	.80	.04	-.21	-.04	-.06	.68
VC	.74	.01	.45	-.02	.08	.76
SI	.74	-.05	.41	-.04	-.02	.72
CD	.53	-.03	-.03	.60	.04	.63
SS	.57	.04	.00	.58	.02	.65
PS	.60	-.02	.02	.15	.48	.61
DS	.69	-.11	.09	.09	.24	.56
ECV	.73	.06	.06	.11	.05	
ETV	.49	.04	.04	.07	.03	
ω_H / ω_{HS}	.87	.09	.21	.42	.17	
H	.91	.30	.31	.52	.27	
PUC	.80					

Schmid-Leiman Orthogonalization

	g	PR	*Gc*	*Gs*	*Gwm*	h^2
VP	.76	.44	.03	.00	-.06	.77
BD	.72	.43	-.02	.02	-.04	.70
MR	.67	.33	-.04	.00	.14	.57
FW	.69	.31	.07	-.01	.04	.58
VC	.78	-.01	.52	.01	-.02	.87
SI	.70	.07	.35	-.01	.06	.62
CD	.63	-.02	-.01	.48	.02	.62
SS	.67	.03	.01	.46	-.02	.67
DS	.61	.00	.02	-.01	.60	.73
PS	.57	.03	.04	.10	.31	.43
ECV	.71	.09	.06	.07	.07	
ETV	.46	.06	.04	.04	.05	
ω_H / ω_{HS}	.85	.19	.22	.27	.27	
H	.90	.41	.34	.36	.40	
PUC	.80					

Higher-Order Factor Factor Analysis

	*g**	PR	*Gc*	*Gs*	*Gwm*	h^2
VP	.80	.89	.06	.00	-.09	.75
BD	.76	.86	-.03	.04	-.05	.69
MR	.73	.68	-.07	.00	.19	.59
FW	.74	.63	.13	-.02	.05	.58
VC	.78	-.03	.96	.02	-.02	.87
SI	.73	.13	.64	-.02	.09	.64
CD	.61	-.03	-.02	.81	.03	.64
SS	.65	.05	.02	.78	-.03	.65
DS	.73	-.01	.03	-.02	.85	.74
PS	.63	.07	.08	.17	.43	.44

Group factor loadings on *g*: PR = .87; *Gc* = .84; *Gs* = .81; *Gwm* = .71

Five Factor Extraction

EBFA Bigeomin (Orthogonal)

	g	*Gv*	*Gc*	*Gs*	*Gwm*	?	h^2
VP	.79	.38	.00	-.01	-.03	.04	.76
BD	.76	.36	-.07	.01	-.03	-.05	.71
FW	.75	.12	.01	-.04	-.04	.20	.63
MR	.78	.06	-.16	-.03	-.04	.02	.64
VC	.73	.00	.47	.00	.10	.03	.76
SI	.73	-.07	.43	-.04	-.02	-.03	.73
CD	.53	-.04	-.03	.60	.04	.02	.64
SS	.57	.03	.00	.57	.02	-.04	.65
PS	.60	-.03	.02	.15	.47	.01	.60
DS	.70	-.15	.08	.09	.25	-.05	.60
ECV	.73	.05	.07	.11	.04	.01	
ETV	.49	.03	.04	.07	.03	.01	
ω_H / ω_{HS}	.87	.07	.23	.42	.17	--	
H	.91	.25	.34	.51	.26	--	
PUC	.80						

Schmid-Leiman Orthogonalization

	g	*PR*	*Gc*	*Gs*	*Gwm*	?	h^2
VP	.79	.42	.02	-.01	.01	-.04	.79
BD	.75	.36	-.02	.03	.01	.04	.69
FW	.72	.24	.10	.00	.00	.12	.60
MR	.74	.24	-.03	.00	.01	.34	.72
SI	.71	.01	.49	.01	-.06	.03	.76
VC	.71	.02	.46	-.01	.06	-.08	.73
SS	.66	.03	.02	.49	-.01	-.02	.68
CD	.61	-.01	-.01	.49	.02	.01	.61
PS	.55	.02	-.02	.02	.53	-.03	.58
DS	.64	.00	.12	.03	.33	.09	.54
ECV	.71	.06	.07	.07	.06	.02	
ETV	.48	.04	.05	.05	.04	.01	
ω_H / ω_{HS}	.86	.13	.26	.29	.25	--	
H	.91	.33	.37	.39	.34	--	
PUC	.80						

Higher-Order Factor Factor Analysis

	*g**	PR	*Gc*	*Gs*	*Gwm*	?	h^2
VP	.80	.90	.03	-.01	.02	-.06	.80
BD	.76	.78	-.03	.04	.01	.06	.68
FW	.74	.52	.19	.00	.00	.16	.58
MR	.74	.52	-.05	.00	.01	.46	.68
SI	.74	.02	.87	.01	-.09	.04	.73
VC	.77	.05	.82	-.01	.09	-.10	.76
SS	.65	.05	.03	.80	-.02	-.03	.67
CD	.60	-.03	-.02	.79	.03	.02	.62
PS	.64	.04	-.04	.02	.77	-.05	.57
DS	.71	-.01	.22	.04	.48	.13	.59

Group factor loadings on *g*: PR = .89; *Gc* = .83; *Gs* = .78; *Gwm* = .72; ? = .67

Note. WISC–V Subtests: *SI*, Similarities; *VO*, Vocabulary; *BD*, Block Design; *VP*, Visual Puzzles; *MR*, Matrix Reasoning; *FW*, Figure Weights; *DS*, Digit Span; *PS*, Picture Span; *CD*, Coding; *SS*, Symbol Search; *PR*, Perceptual Reasoning; *Gc*, Crystallized Ability; *Gs*, Processing Speed; *Gwm*, Working Memory; *Gv*, Visual-Spatial; *h2*, Communality; *ECV*, Explained Common Variance; *ETV*, Explained Total Variance; *ωH /ωHS*, Omega-Hierarchical/Omega-Hiearchical Subscale; *H*, Construct reliability or replicability index; *PUC*, Percentage of Uncontaminated Correlations; *EBFA*, Exploratory Bifactor Analysis; *g**, reflects the first unrotated factor loadings for the principal axis factoring results

interpretive evidence should be placed upon the WISC-V general factor.

Discussion

Throughout its history, the theoretical/factor structure of the WISC has vacillated from edition to edition. Given its popularity within clinical practice, this lack of consistency has important implications for use of the WISC for high-stakes clinical applications such as identification of specific learning disorders and intellectual disability (Dombrowski, 2015, 2020a). For the WISC-V, the publisher proposed a five-factor higher-order theoretical structure (e.g., *Gc*, *Gf*, *Gv*, *Gwm*, & *Gs*) linked to CHC theory. This linkage to a theory of cognitive ability is laudable; however, the accuracy of its linkage with theory may be readily tested via factor analysis. Although the WISC-V 10 subtest primarily battery is arguably the most frequently administered battery the world over, its theoretical/factor structure has not been fully investigated by the test publisher and within the literature. An inability to replicate posited findings suggests concern with attempting to apply the scoring structure implied by that measurement model (Lilienfeld, 2018).

The results of this large clinical replication study suggest that the publisher-proposed five-factor theoretical alignment with CHC theory is not entirely supported. This is one of the few consistently replicated findings in the assessment literature featuring the WISC-V 16 subtest primary and secondary battery. This study suggests that this finding may be extended to the 10 subtest primary battery as well. Instead, the results suggest an alternative four-factor structure with a slightly different factor composition depending upon the model used (i.e., higher-order and SL vs. EBFA). The use of the SL procedure and a higher-order factor analysis indicated that the WISC-V contains a factor structure similar to the prior WISC-IV (e.g., *Gc*, *Gs*, and *Gwm* with the *Gf* and *Gv* subtests fused into a fourth PR factor). The similarity in results between the SL and the higher-order model is not surprising because the SL is derived from the higher-order model (Carroll, 1993) and uses a few additional quantitative steps to partition the variance independently attributable to *g* and group factors (Schmid & Leiman, 1957; Wolff & Preising, 2005). Put simply, it represents a statistical transformation of the higher-order model for the purpose of variance partitioning.[1] On the other hand, EBFA, a *true* exploratory bifactor model (Jennrich & Bentler, 2011), demonstrated that the WISC-V is a four-factor instrument comprising four of the five factors proposed by the publisher (*Gc*, *Gv*, *Gwm*, and *Gs*). There was one notable difference between EBFA and the SL models. With EBFA, both of the *Gf* subtests saliently loaded only on the general factor but not on any group factor. Among extant WISC-V factor analytic studies, this finding is particularly unique. Research studies on other cognitive ability instruments (e.g., Dombrowski et al., 2018) and other editions of the WISC (e.g., WISC-IV; Golay et al., 2013) have similarly found that the *Gf* subtests are isomorphic, and possibly empirically redundant, with the general factor.

Regardless of whether the fourth factor is a combined *Gf*/*Gv* (i.e., PR) or *Gv* factor, whether one ascribes to a bifactor or higher-order measurement conceptualization of intelligence, whether the 16-subtest full battery or 10 primary subtest battery is investigated, or whether an extant study has used an exploratory or confirmatory factor analytic technique, nearly all independent research studies using the WISC-V standardization and clinical samples have suggested that the WISC-V is a four-factor instrument with a factor structure similar to the WISC-IV. A recent simulation that modeled the structure of the WISC-V across 1000 replications (Dombrowski, McGill & Morgan, 2019) also supported this result. Even so, the fact that new plausible measurement models continue to be identified for the WISC-V almost 6 years after its publication illustrate well the danger in relying on the publisher's promoted model or any one particular structural validity study as it relates to how an instrument should be interpreted (Dombrowski, 2020b). As noted by Beaujean (2015a), it may very well take several years and many studies before there is any consensus as to what a test *actually* measures.

From a practical, interpretive perspective, it is also important to consider other metrics of interpretability. Inspection of omega estimates, explained common variance (ECV), explained total variance (ETV), and *H* suggest that there is insufficient group factor variance to overlook interpretation of the general factor in favor of group factors. This is a consistent finding regardless of whether a bifactor model (e.g., EBFA) or the hierarchical SL model was used to extract and examine the underlying factors.

If one were to insist upon interpreting group factors as proposed by the test publisher, then there appears to be theoretical alignment of subtests on the *Gc*, *Gwm*, and *Gs* factors, and perhaps the *Gv* factor (i.e., via EBFA), but not the *Gf* factor for the 10 primary subtest battery. As previously mentioned, identification issues have been noted for the *Gf* construct since it was specified post hoc on the WISC-IV (i.e., Weiss et al., 2013). In only two extant independent studies do we find evidence for subtest alignment of matrix reasoning and figure weights on the *Gf* factor (e.g., Dombrowski, Canivez et al., 2017; Reynolds & Keith, 2017). However, Dombrowski, Canivez et al. (2017) found this alignment for only one of the four age ranges (i.e., age 15 to 16) but not for the remaining age ranges (i.e., 6 to 8; 9 to 11; 12 to 14) when looking at the 10 primary subtest battery. It is unknown why this difference in factor structure occurred only at this age range. In a CFA investigation looking at the WISC-V full-test battery, Reynolds and Keith engaged in an extensive number of model adjustments to arrive at their final, validated model. For instance, this entailed having the Arithmetic subtest load on both the working memory factor and the second-order general factor. Additionally, Reynolds and Keith correlated the disturbance between the *Gv* and *Gf* group factors. Although circumspect post hoc model adjusting may be supported in some cases (e.g., Byrne, 2005), it has been questioned by other researchers (e.g., Brown, 2015; Cucina & Howardson, 2017; Dombrowski, McGill et al., 2020; Kerr, 1998; Kline, 2016) who caution against the use of CFA for exploratory model fitting which may produce final structures

[1] The variance in a higher-order model may be residualized to disclose the apportionment of variance between the group factors and the general factor. The SL orthogonalization approach is considered a more elegant technique for this purpose and would likely result in approximately the same variance estimates if the oblique/higher-order model were residualized. Thus, variance apportionment is not reported for the higher-order model in this study.

that capitalize on chance or may not replicate under more stringent measurement conditions (MacCallum et al., 1992; Meehl, 1978). We are not necessarily criticizing the use of this parameter constraint (i.e., correlating the residual terms between *Gf* and *Gv*). However, given the lineage of the PRI, and recent independent results which suggest a fusion of the Gv and Gf subtests onto a single factor, this permits questioning as to whether the two constructs can be empirically separated on the WISC.

In a subsequent CFA analysis of the Canadian version of the WISC-V by Watkins et al. (2018) and in a simulation study conducted by Dombrowski, McGill & Morgan (2019), the five-factor higher-order model posited by Reynolds and Keith (2017) was found to be inferior to the four-factor bifactor model, or at best to provide equivalent fit to more parsimonious models. In totality, we suggest placing emphasis on the preponderance of the research evidence when making a decision about the WISC-V theoretical structure and how the instrument should be interpreted. This study adds to that literature base suggesting that the WISC-V 10 primary subtest battery contains four theoretical factors but with a slightly different factor composition depending upon the method of exploratory modeling used (i.e., EBFA vs. SL and higher-order). Based on results from ω_H, ω_{HS}, and H indices, the WISC-V appears to be essentially a measure that should receive primary interpretive emphasis at the Full-Scale Intelligence Quotient (FSIQ) level with only secondary, yet cautious, emphasis at the group factor level, but only when there is factor analytic evidence for the existence of group factors (Hancock & Mueller, 2001; Rodriguez et al., 2016).

The finding of an alternative four-factor structure on the WISC-V has implications for PSW and XBA. If an index does not measure what it is intended to measure, then this makes it difficult to use these techniques on indices such as *Gf* as the index does not appear to measure its intended CHC construct in the way proposed by the test publisher. A lack of structural validity precludes any further interpretive action such as PSW/XBA or even direct WISC-V index level interpretation when these indices that lack structural validity support.

Limitations

One of the unique aspects of this study—a large clinical sample—may also pose a limitation. Can the conclusions drawn from this study be extended to the general population as this sample's demographic characteristics do not match those of the US population in terms of gender, ethnicity, and socioeconomic status? However, it does appear that the factor analytic results are consistent across both the standardization sample and this large clinical sample, which is comprised primarily of ADHD, anxiety, and brain injury participants. In fact, given that assessment of cognitive functioning in referred youth is one of the primary purposes of intelligence tests such as the WISC-V, the findings of this study may be directly extended to those populations. Additionally, one of the strengths of factor analytic studies of an exploratory nature—it offers less room for post hoc adjusting and the potential inclusion of biases in scientific decision-making—is also a weakness: it does not have fully evolved statistical fit indices to potentially aid in model comparison. In totality, no modeling method is perfect, and each has its own strengths and limitations making the use of both exploratory and confirmatory factor analytic techniques useful. This suggests that any final, adopted model produced by the test publisher or proposed by an independent study should be viewed as tentative and subject to further replication using different methods and samples. Ultimately, the structure of any test must be evaluated against external criteria.

Conclusion and Implications for Practice

Given the mercurial nature of the WISC theoretical structure over the past 75 years, researchers and school psychologists must be cautious about accepting a model based upon what is proposed by the test publisher or even within a single factor analytic study. After all, what has *really* changed over the WISC's history of clinical use (Frazier & Youngstrom, 2007)? Many of the subtests continue to trace their lineage to the original Army Alpha and Beta tests developed in the early 1900s (Warne, 2019). Yet, with each iteration, researchers are discovering new factors and labels for previously discovered factors. Instead, school psychology researchers and practitioners should adopt a "preponderance of the evidence" standard looking at the agglomeration of available factor analytic research evidence when deciding whether and how to interpret the instrument.

The results of this study are valuable given that the study utilized a rarely available clinical sample, focused on the 10 primary subtest battery which has been scarcely investigated but predominantly utilized by school psychologists, and used the largest sample to date—more than double the size of the WISC-V standardization sample. Findings suggest that the WISC-V 10 primary subtest battery is a four-factor instrument (e.g., *Gc*, *Gwm*, and *Gs*) containing either PR (via higher-order & SL) or *Gv* (via EBFA) as the fourth factor. Regardless of the composition of the fourth factor (*Gv* or PR) or the method of modeling used, metrics of interpretability and sources of variance suggest that primary, and perhaps, exclusive interpretive emphasis should be placed on the general factor (*g*; i.e., FSIQ) with secondary emphasis placed on the interpretation of the group factors.

This latter interpretive statement is frequently offered by some in the research community regardless of the cognitive ability instrument investigated: WJ-IV full-test battery (Dombrowski et al., 2018a) and WJ-IV Cognitive

(Dombrowski, McGill et al., 2017, 2018b); WJ-III full-test battery, WJ-III Cognitive (Dombrowski, 2013, 2014a, 2014b, 2015b; Dombrowski & Watkins, 2013); WISC-V (Canivez et al., 2016, 2017; Dombrowski, Canivez et al., 2017; Dombrowski et al., 2015); Stanford-Binet, Fifth Edition (Canivez, 2008; DiStefano & Dombrowski, 2006); DAS-II (Canivez & McGill, 2016; Dombrowski, Canivez, McGill & Dombrowski, 2020; Dombrowski, Golay et al., 2018; Dombrowski, McGill, Canivez Dombrowski et al., 2019); KABC–2 (McGill & Dombrowski, 2018; McGill & Spurgin, 2017); and RIAS (Canivez, 2008; Dombrowski et al., 2009). The accumulating structural validity literature suggests that, although the lower-order group factors may be present, they are of less significance in comparison with the general factor (Dombrowski, McGill et al. 2020). Thus, interpretation of group factor indices (e.g., VCI, WMI, PSI) should only be undertaken cautiously and contingent upon structural and external validity support. Primary interpretive emphasis should be placed upon the FSIQ. In other words, school psychologists should note that this index score is the most reliable and valid score that will replicate consistently from administration to administration with the least fluctuation of all the scores presented in the WISC-V. The VCI, WMI, and PSI scores have structural validity support but contain a meager proportion of variance relative to the FSIQ. Consequently, little interpretive emphasis should be placed upon the WISC-V index level scores. This study did not locate a fluid reasoning factor as both of its subtests either fused with the visual spatial factor or loaded only on the general factor. Thus, this study indicates that the fluid reasoning index should not be interpreted even though the publisher produced a score for it.

References

Bartlett, M. S. (1954). A further note on the multiplying factors for various chi square approximations in factor analysis. *Journal of the Royal Statistical Society Series A (General), 16*, 296–298.

Beaujean, A. A. (2015a). Adopting a new test edition: Psychometric and practical considerations. *Research and Practice in the Schools, 3*(*1*), 51–57.

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

Beaujean, A. A. (2016). Reproducing the Wechsler Intelligence Scale for Children-Fifth Edition: Factor model results. *Journal of Psychoeducational Assessment, 34*, 404–408. https://doi.org/10.1177/0734282916642679.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology, 72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004.

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*, 676–696. https://doi.org/10.1177/0013164404272507.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

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

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*, 17–32. https://doi.org/10.1207/s15327752jpa8501_02.

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. https://doi.org/10.1037/a0012884.

Canivez, G. L., & Kush, J. C. (2013). WAIS-IV and WISC-IV structural validity: Alternate methods, alternate results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). *Journal of Psychoeducational Assessment, 31*, 157–169. https://doi.org/10.1177/0734282913478036.

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. https://doi.org/10.1037/pas0000279.

Canivez, G. L., & Watkins, M. W. (2016). Review of the Wechsler Intelligence Scale for Children–Fifth Edition: Critique, commentary, and independent analyses. In Kaufman, A. S, Raiford, S. E. & Coalson D. L. (Eds.), *Intelligent testing with the WISC–V* (pp. 683–702). Wiley.

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. 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*, 458–472. https://doi.org/10.1037/pas0000358.

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology, 89*, 195–224. https://doi.org/10.1111/bjep.12230.

Canivez, G. L., McGill, R. J., Dombrowski, S. C. (2020) Factor Structure of the Differential Ability Scales–Second Edition Core Subtests: Standardization Sample Confirmatory Factor Analyses. *Journal of Psychoeducational Assessment 38*, (7):791–815

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*, 274–296. https://doi.org/10.1177/1073191118811609.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge University Press.

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

Cucina, J. M., & Howardson, G. N. (2017). Woodcock-Johnson-III, Kaufman Adolescent and Adult Intelligence Test (KAIT), Kaufman Assessment Battery for Children (KABC), and Differential Ability Scales (DAS) support Carroll but not Cattell-Horn. *Psychological Assessment, 29*, 1001–1015. https://doi.org/10.1037/pas0000389.

R Development Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet, Fifth Edition. *Journal of Psychoeducational Assessment, 24*, 123–136. https://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. 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*, 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*, 483–494. https://doi.org/10.1177/0734282914530838.

Dombrowski, S. C. (2015). *Psychoeducational assessment and report writing*. New York: Springer Science.

Dombrowski, S. C. (2020a). *Psychoeducational assessment and report writing* (2nd ed.) Springer Nature.

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

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

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. 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*, 279–286. https://doi.org/10.1177/0734282909333179.

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*. https://doi.org/10.1016/j.intell.2015.10.009.

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

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

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018b). 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*, 1–13. https://doi.org/10.1037/arc0000039.

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*, 847–860. https://doi.org/10.1080/10705511.2019.1622421.

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*, 91–104. https://doi.org/10.1177/0734282918760724.

Dombrowski, S. C., McGill, R. J., & Morgan, G. W. (2019). Monte Carlo modeling of contemporary intelligence test (IQ) factor structure: Implications for IQ assessment, interpretation and theory. *Assessment.* https://doi.org/10.1177/1073191119869828.

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), 28–28. https://doi.org/10.1177/0734282920961952.

Dombrowski, S. C., Watkins, M. W., McGill, R. J., Canivez, G. L., Holingue, C., Pritchard, A. E., & Jacobson, L. A. (2021). Measurement invariance of the Wechsler Intelligence Scale for Children, Fifth Edition 10-subtest primary battery: Can index scores be compared across age, sex, and diagnostic groups? *Journal of Psychoeducational Assessment, 39*(1), 89–99. https://doi.org/10.1177/0734282920954583.

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. https://doi.org/10.1016/j.intell.2006.07.002.

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. https://doi.org/10.1037/a0030676.

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

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology (pp. 143-164)*. Wiley.

Graves, S. L., Smith, L. V., & Nichols, K. D. (2020). Is the WISC-V a fair test for black children: Factor structure in an urban public school sample. *Contemporary School Psychology*, 1–13. https://doi.org/10.1007/s40688-020-00306-9.

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)* Scientific Software International.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185. https://doi.org/10.1007/BF02289447.

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

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 6*, 537–549. https://doi.org/10.1007/s11336-011-9218-4.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36. https://doi.org/10.1007/BF02291575.

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the cognitive assessment system does and does not measure. *School Psychology Review, 28*, 303–321. https://doi.org/10.1080/02796015.1999.12085967.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*, 196–217. https://doi.org/10.1207/s15327957pspr0203_4.

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

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

Lecerf, T., & Canivez, G. L. (2017). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment, 30*, 793–808. https://doi.org/10.1037/pas0000526.

Lilienfeld, S. O. (2018). Direct replication and clinical psychological science. *Behavioral and Brain Sciences, 41*, e140. https://doi.org/10.1017/S0140525X18000754.

Loehlin, J. C., & Beaujean, A. A. (2016). *Syntax companion for Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). Baylor Psychometric Laboratory. Retrieved from https://sites.baylor.edu/lvm5/

MacCallum, R. C., Roznowski, M., & Neocowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490–504. https://doi.org/10.1037/0033-2909.111.3.490.

McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC II: Exploratory factor analyses with the 16 core and supplemental subtests. *Contemporary School Psychology, 22*, 279–293. https://doi.org/10.1007/s40688-017-0152-z.

McGill, R. J., & Dombrowski, S. C. (2019). Critically reflecting on the origins, evolution, and impact of the Cattell-Horn-Carroll (CHC) Model. *Applied Measurement in Education, 32*, 216–231. https://doi.org/10.1080/08957347.2019.1619561.

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(4), 540–552. https://doi.org/10.1177/1073191115614081

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834. https://doi.org/10.1037/0022-006X.46.4.806.

Oakland, T., Douglas, S., & Kane, H. (2016). Top ten standardized tests used internationally with children and youth by school psychologists in 64 countries: a 24-year follow-up study. *Journal of Psychoeducational Assessment, 34*, 166–176. https://doi.org/10.1177/0734282915595303.

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

Revelle, W. (2012). *psych: Procedures for psychological, psychometric, and personality research* (version 1.2.4) [computer software]. Northwestern University.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children–Fifth Edition: What does it measure? *Intelligence, 62*, 31–47. https://doi.org/10.1016/j.intell.2017.02.005.

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

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 theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues (4th ed., pp. 73-163)* Guilford.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Inc: Pearson Education.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of factors of mind*. University of Chicago Press.

Velicer, W. F. (1976). Determining the number of components form the matrix of partial correlations. *Psychometrika, 31*, 321–327. https://doi.org/10.1007/BF02293557.

Warne, R. T. (2019). First publication of subtests in the Standord-Binet 5, WAIS-IV, WISC-V, and WPPSI-IV. *Intelligence, 75*, 9–18. https://doi.org/10.1016/j.intell.2019.02.005.

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

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*, 252–265. https://doi.org/10.1080/21683603.2017.1342580.

Wechsler, D. (1946). *Wechsler-Bellevue Intelligence Scale*. Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. Psychological Corporation.

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

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31*, 114–131. https://doi.org/10.1177/0734282913478032.

Wolff, H.-G., & Preising, K. (2005). Exploring item and higher order factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods, 37*, 48–58. https://doi.org/10.3758/BF03206397.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*, 254–265. https://doi.org/10.1037/1082-989X.1.4.354.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Stefan C. Dombrowski, Ph.D. is a professor, researcher, and licensed psychologist. He is director of the graduate program in school psychology at Rider University in Lawrenceville, NJ. For more information about Dr. Dombrowski's background and scholarship please visit www.stefandombrowski.com.

Ryan J. McGill, Ph.D., BCBA-D, NCSP is Associate Professor and Chair of the Department of School Psychology and Counselor Education at the William & Mary School of Education. His scholarly and practice interests include psychological assessment and measurement, judgement and decision-making in school psychology, and assessment and identification of specific learning disability.

Marley W. Watkins earned a Ph.D. in Educational Psychology and Measurements with a specialty in School Psychology from the University of Nebraska-Lincoln. He is currently Research Professor in the Department of Educational Psychology at Baylor University and has authored more than 200 articles, books, and chapters and presented more than 150 papers at professional conferences.

Gary L. Canivez, Ph.D., is a Professor of Psychology at Eastern Illinois University principally involved in the School Psychology training program. The author of over 100 peer reviewed research and professional publications and presenter of over 200 professional presentations and continuing education workshops, Dr. Canivez has research interests in applied psychometrics in evaluating the psychometric fitness of psychological and educational tests (including international applications); and

his research has been supported by the National Institutes of Health/National Institute of Mental Health.

Dr. Alison Pritchard is a licensed and board-certified child and adolescent clinical psychologist who serves as the Co-Director of and research scientist in the Department of Neuropsychology at the Kennedy Krieger Institute. She also holds an appointment as assistant professor of Psychiatry and Behavioral Sciences at the Johns Hopkins University School of Medicine. Dr. Pritchard sees patients for psychological assessment in the Targeted Assessment Clinic and in the Consultation Clinic within the Neuropsychology Department's outpatient program.

Dr. Lisa Jacobson is the Director of Research, as well as a licensed psychologist and pediatric neuropsychologist, in the Department of Neuropsychology, Kennedy Krieger Institute. She is also the co-Director of the Center for Innovation and Leadership in Special Education (CILSE). Additionally, she holds an appointment as an associate professor in Psychiatry and Behavioral Sciences at The Johns Hopkins University School of Medicine.