RUNNING HEAD: Construct Validity of the WISC–V

Please use the following citation when referencing this work:

Canivez G. L., McGill, R. J., Dombrowksi, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2018). Construct validity of the WISC-V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*. Advance online publication. doi: 10.1177/1073191118811609

Construct Validity of the WISC–V in Clinical Cases: Exploratory and Confirmatory Factor Analyses of the 10 Primary Subtests

Gary L. Canivez
Eastern Illinois University

Ryan J. McGill
William & Mary

Stefan C. Dombrowski
Rider University

Marley W. Watkins
Baylor University

Alison E. Pritchard
Kennedy Krieger Institute, Johns Hopkins University School of Medicine

Lisa A. Jacobson
Kennedy Krieger Institute, Johns Hopkins University School of Medicine

Author Note

Preliminary results were presented at the 2018 annual conventions of the National Association of School Psychologists and the American Psychological Association.

Corresponding Author: 600 Lincoln Avenue, Charleston, IL 61920, glcanivez@eiu.edu

Abstract

Independent exploratory (EFA) and confirmatory (CFA) factor analytic research with the Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V; Wechsler, 2014a) standardization sample has failed to provide support for the five group factors proposed by the publisher (Canivez, Watkins, & Dombrowski, 2016; Canivez, Dombrowski, & Watkins, 2017; Dombrowski, Canivez, & Watkins, 2017; Dombrowski, Canivez, Watkins, & Beaujean (2015), but there have been no independent examinations of the WISC–V structure among clinical samples. The present study examined the latent structure of the 10 WISC–V primary subtests with a large (N = 2,512), bifurcated clinical sample (EFA n = 1,256, CFA n = 1,256). EFA did not support five factors as there were no salient subtest factor pattern coefficients on the fifth extracted factor. EFA indicated a four-factor model resembling the WISC–IV with a dominant general factor. A bifactor model with four group factors was supported by CFA as suggested by EFA. Variance estimates from both EFA and CFA found that the general intelligence factor dominated subtest variance and omega-hierarchical coefficients supported interpretation of the general intelligence factor. In both EFA and CFA, group factors explained small portions of common variance and produced low omega-hierarchical subscale coefficients, indicating that the group factors were of poor interpretive value.

Keywords: WISC–V; exploratory factor analysis; confirmatory factor analysis; bifactor; intelligence

Construct Validity of the WISC–V in Clinical Cases: Exploratory and Confirmatory Factor Analyses of the 10 Primary Subtests

The Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V; Wechsler, 2014a) is a major test of cognitive abilities for children ages 6-16 years. Its development and construction was influenced by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Blankson, 2005; Horn & Cattell, 1966), often referred to as Cattell-Horn-Carroll (CHC) theory (Schneider & McGrew, 2012), and neuropsychological constructs (Wechsler, 2014b). The Wechsler Intelligence Scale for Children-Fourth Edition (WISC–IV; Wechsler, 2003) Word Reasoning and Picture Completion subtests were deleted and, to better measure purported CHC broad abilities, three new subtests were added. Specifically, Picture Span was adapted from the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI–IV; Wechsler, 2012) to measure visual working memory; while Visual Puzzles and Figure Weights were adapted from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV; Wechsler, 2008) to better measure visual spatial and fluid reasoning, respectively. The addition of Visual Puzzles and Figure Weights was made to facilitate splitting the former Perceptual Reasoning (PR) factor into distinct Visual Spatial (VS) and Fluid Reasoning (FR) factors in an attempt to make the WISC–V more consistent with CHC theory.

The WISC–V measurement model preferred by the publisher is illustrated in Figure 1. The structural validation procedures and analyses reported in the WISC–V *Technical and Interpretive Manual* (Wechsler, 2014b) that were provided in support of this preferred model and upon which scores and interpretations were created have been criticized as problematic (Beaujean, 2016; Canivez & Watkins, 2016; Canivez, Watkins, & Dombrowski, 2016, 2017). Specifically, problems include (a) use of weighted least squares (WLS) estimation without explicit justification rather than maximum likelihood (ML) estimation (Kline, 2011); (b) failure to fully disclose details of confirmatory factor analytic (CFA) methods; (c) preference for a complex measurement model (cross-loading Arithmetic on three group factors) thereby abandoning parsimony of simple structure (Thurstone, 1947); (d) retention of a model with a standardized path coefficient of 1.0 between general intelligence and the FR factor indicating that FR and *g* are empirically redundant; (e) failure to consider rival bifactor models (Beaujean, 2015); (f) omission of decomposed variance estimates; and (g) absence of model based reliability estimates (Watkins, 2017). These problems call into question the publisher's preferred WISC–V measurement model.

A number of these concerns are not new and were previously identified and discussed with other Wechsler scales (Canivez, 2010, 2014a; Canivez & Kush, 2013; Gignac & Watkins, 2013), but they were not addressed in the WISC–V *Technical and Interpretive Manual* thereby continuing a tendency by the publisher to ignore "contradictory findings available in the literature" (Braden & Niebling, 2012, p. 744). For example, the publisher referenced Carroll's (1993) three stratum theory as a foundation for the WISC–V but decomposed variance estimates provided by the Schmid and Leiman (SL; 1957) transformation were not provided even though Carroll (1995) *insisted* on use of the SL transformation of EFA loadings to allow subtest variance apportionment among the first- and higher-order dimensions. Additionally, Beaujean (2015a) noted that Carroll's (1993) model was ostensibly a bifactor model but no examination of an alternative bifactor structure for the WISC–V was reported (Wechsler, 2014b).

Higher-order representations of Wechsler scales (and other intelligence tests) specify general intelligence (*g*) as a superordinate (second-order) factor that is fully mediated by the first-order group factors which have direct influences (paths) on the subtest indicators (Gignac, 2008). Thus, *g* has indirect influences on subtest indicators, which may obfuscate the role of *g*. The bifactor model initially conceptualized by Holzinger and Swineford (1937) does not include a hierarchy of *g* and the first-order group factors. Rather, bifactor models specify *g* as a breadth factor with direct influences (paths) on subtest indicators, and group factors also have direct influences on subtest indicators (Gignac, and 2005, 2006, 2008). Because the bifactor model includes *g* and group factors at the same level of inference and includes simultaneous influence on subtest indicators the bifactor model can be considered a more conceptually parsimonious model (Gignac, 2006) and also more consistent with Spearman (1927). According to Beaujean (2015a), Carroll (1993) favored the bifactor model where all subtests load directly on *g* and on one (or more) of the first-order group factors. For further discussion of bifactor models see Canivez (2016) or Reise (2012).

Because EFA was not reported in the WISC–V *Technical and Interpretive Manual,* Canivez et al. (2016) conducted independent EFA with the 16 WISC–V primary and secondary subtests and did not find support for five-factors with the total WISC–V standardization sample. The fifth factor consisted of only one salient subtest pattern coefficient. When the standardization sample was divided into four age groups (6-8, 9-11, 12-14, 15-16), only one salient subtest factor loading was found for the fifth factor for all but the 15-16 year old age group (Dombrowski, Canivez, & Watkins, 2017). Both studies found support for four first-order WISC–V factors resembling the traditional WISC–IV structure (i.e., Verbal Comprehension [VC], PR, Working Memory [WM], Processing Speed [PS]).

Schmid and Leiman (1957) orthogonalization of the second-order EFA with the total WISC–V standardization sample and the four age groups yielded substantial portions of variance apportioned to the general factor (*g*) and considerably smaller portions of variance uniquely apportioned to the group factors (Dombrowski et al., 2017). Omega-hierarchical (ω_H) coefficients (Reise, 2012; Rodriguez, Reise, & Haviland, 2016) for the general factor ranged from .817 (Canivez et al., 2016) to .847 (Dombrowski et al., 2017) and exceeded the preferred level (.75) for clinical interpretation (Reise, 2012; Reise, Bonifay, & Haviland, 2013; Rodriguez et al., 2016). Omega-hierarchical subscale (ω_{HS}) coefficients (Reise, 2012) for the four WISC–V group factors ranged from .131 to .530. The ω_{HS} coefficients for VC, PR, and WM group factor scores failed to approach or exceed the minimum criterion (.50) desired for clinical interpretation (Reise, 2012; Reise et al., 2013), but ω_{HS} coefficients for PS scores approached or exceeded the .50 criterion that might allow clinical interpretation.

Dombrowski, Canivez, Watkins, and Beaujean (2015), using exploratory bifactor analysis (i.e., EFA with a bifactor rotation; Jennrich & Bentler, 2011), also failed to identify five WISC–V factors within the WISC–V standardization sample. The failure to find a verbal comprehension factor by Dombrowski et al. (2015) is inconsistent with the long-standing body of structural validity evidence for the Wechsler scales where every other study located a distinct verbal ability dimension. It is unknown why this anomalous result was produced. Dombrowski et al. speculated that it could be a function of the WISC-V simply having verbal subtests that are predominantly g loaded. Unlike the Schmid-Leiman procedure, an approximate bifactor solution, Jennrich and Bentler's (2011) EBFA procedure is a true exploratory bifactor analysis procedure that may produce different results. Thus, it could be possible that the WISC-V verbal subtests "collapsed" onto the general factor following simultaneous extraction of general and specific factors. In other words, following the bifactor rotation it is plausible that most of the variance could have been apportioned to the general factor leaving nominal variance to the specific verbal factor producing the results evident in the Dombrowski et al. study. This speculation is supported by recent simulation research that found these exploratory bifactor routines to be prone to group factor collapse onto the general factor and to local minima problems, especially with variables that are either poorly or complexly related to one another (Mansolf & Reise, 2016).

Lecerf and Canivez (2018) similarly assessed the French WISC–V standardization sample (French WISC–V; Wechsler, 2016a) with hierarchical EFA and also found support for four first-order factors (not five), the dominant general intelligence factor, and little unique reliable measurement of the four group factors. Assessment of the WISC-V^{UK} (Wechsler, 2016b) using hierarchical EFA also failed to identify five WISC-V factors and like the French WISC-V and US versions contained too little unique variance among the four group factors for confident interpretation (Canivez, Watkins, & McGill, 2018).

In a follow-up study, Canivez et al. (2017) examined the latent factor structure of the 16 WISC–V primary and secondary subtests using CFA with ML estimation and found that all higher-order models that included five group factors (including the final publisher-preferred WISC–V model presented in the WISC–V *Technical and Interpretative Manual*) produced improper solutions (i.e., negative variance estimates for the FR factor) potentially caused by misspecification of the models. An acceptable solution for a bifactor model that included five group factors fit the standardization sample data well based on global fit, but examination of local fit identified problems where Matrix Reasoning, Figure Weights, and Picture Concepts did *not* have statistically significant FR group factor loadings, rendering this model inadequate. Consistent with the Canivez et al. (2016) WISC–V EFA results, the WISC–V bifactor model with four group factors (VC, PR, WM, PS) appeared to be the most acceptable solution based on a combination of statistical fit and Wechsler theory. As with the EFA analyses, a dominant general intelligence dimension but weak group factors with limited unique measurement beyond *g* was found. Similar CFA findings were also found with the WISC–V^{Spain} (Wechsler, 2015) in an independent study of standardization sample data (Fenollar-Cortés & Watkins, 2018) as well as with the French WISC–V (Lecerf & Canivez, 2018) and the WISC–V^{UK} (Canivez et al., 2018).

Chen, Zhang, Raiford, Zhu, and Weiss (2015) reported invariance of the final publisher preferred WISC–V higher-order model with five group factors across gender, but invariance for rival higher-order or bifactor models was not examined. Reynolds and Keith (2017) also investigated the measurement invariance of the WISC–V across age groups with CFA, but only examined an *oblique* five-factor model, which did not include a general intelligence dimension. As noted by Hayduk (2016), if the number of factors are not accurately specified then "asking about invariance between groups is asking whether the groups agree in their misrepresentation of the connections between the indicators and the underlying latent variables" (p. 2).

Reynolds and Keith (2017) also explored numerous (perhaps post-hoc) model modifications for five-factor first-order models and then for both higher-order and bifactor models including five group factors to better understand WISC–V measurement. Based on these alternate models (modifications), Reynolds and Keith suggested a model different from the publisher preferred model that allowed a direct loading from general intelligence to Arithmetic, a cross-loading of Arithmetic on Working Memory, and correlated disturbances of the Visual Spatial and Fluid Reasoning group factors. Even with these modifications the model still produced a general intelligence to Fluid Reasoning standardized path coefficient of .97, suggesting that these dimensions may be empirically redundant. However, post-hoc modifications capitalize on chance and "such changes often lead the model away from the population model, not towards it" (Gorsuch, 2003, p. 151). Of note, when that same VS-FR factor covariance was allowed in a structural model for the Canadian WISC–V standardization sample (WISC–V^{CDN}; Wechsler, 2014c), it was *not* superior to a bifactor model with four group factors (Watkins, Dombrowski, & Canivez, 2017).

Understanding the structural validity of tests is essential for evaluating the interpretability of scores and score comparisons (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Accordingly, test users must select technically sound instruments with demonstrated validity for the population under evaluation (Evers, Hagemeister, Høstmaelingen, Lindley, Muñiz, & Sjöberg, 2013; International Test Commission, 2001; Public Law [P.L.] 108-446, 2004). Presently, studies of the latent factor structure of the WISC–V have been restricted to analyses of data from the standardization sample. Although such studies are informative, the results provided by such investigations may not generalize to clinical samples (Strauss, Sherman, & Spreen, 2006). Additionally, independent analyses of the WISC–V standardization data have contested the structure preferred by its publisher (Beaujean, 2016; Canivez et al., 2016, 2017; Dombrowski et al., 2017; Dombrowski et al., 2015; Reynolds & Keith, 2017). Whereas these investigations have produced several plausible alternative models, it remains unclear which should be preferred. To provide additional insight on these matters, the present study examined the latent factor structure of the 10 WISC–V primary subtests with a large clinical sample and: (a) followed best practices in EFA and CFA, (b) compared bifactor models to higher-order models as rival explanations, (c) examined decomposed factor variance sources in EFA and CFA, and (d) estimated model-based reliabilities. Results from these analyses are essential for users of the WISC–V to determine the value of the various scores and score comparisons provided in the WISC–V and interpretive guidelines emphasized by the publisher.

Method

Participants and Selection

A total of 2,512 children (65% male) between the ages of 6 and 16 years were administered the WISC–V as part of assessments conducted in a large outpatient neuropsychology clinic between October 2014 and February 2017. All test data are routinely entered into the department's clinical database via the electronic medical record and securely maintained by the hospital's Information Systems Department. Following approval from the hospital's Institutional Review Board, the clinical database was queried and a limited, de-identified data set was constructed of patients for whom subtest scores from all 10 WISC–V primary subtests were available. With regard to the referred nature of the sample, billing diagnosis codes were queried to provide descriptive information regarding presenting concerns. Approximately 20% of cases were seen for primarily medical concerns (e.g., 21.2% epilepsy, 19.2% encephalopathy, 10.6% pediatric cancer diagnoses, 49% other congenital or acquired conditions). Among the remaining 80% of cases seen for mental health concerns, 58.9% were diagnosed with ADHD, 14.0% with anxiety or depression, 7.2% with an adjustment disorder, and 19.9% other.

The sample was randomly bifurcated into EFA and CFA samples by sex. Table 1 presents demographic characteristics of the EFA (n = 1,256) and CFA (n = 1,256) samples with equal distributions of male and female participants. The sample was primarily composed of White/Caucasian and Black/African American youths. The ages of participants were similar in EFA (M = 10.63, SD = 2.74) and CFA (M = 10.46, SD = 2.68) samples. Table 2 illustrates the distribution of Race/Ethnicity across the 11 age groups of WISC–V. Given the clinical nature of the sample, these data do not represent the general public.

WISC–V descriptive statistics for the EFA and CFA samples are presented in Table 3 and show that average subtest and composite scores were slightly below average, but within one standard deviation of population means, as is typical in clinical samples. All subtests and composite scores showed univariate normal distributions with no appreciable skewness or kurtosis. However, Mardia's (1970) multivariate kurtosis estimates for the EFA sample (χ^2= 123.7) and the CFA sample (χ^2= 128.5) indicated significant ($p < .05$) multivariate non-normality for both samples (Cain, Zhang, & Yuan, 2017). There were no statistically significant subtest or composite score mean differences between the EFA and CFA samples.

Instrument

The WISC–V (Wechsler, 2014a), is a test of general intelligence composed of 16 subtests expressed as scaled scores (*M* = 10, *SD* = 3). There are seven primary subtests (Similarities [SI], Vocabulary [VO], Block Design [BD], Matrix Reasoning [MR], Figure Weights [FW], Digit Span [DS], and Coding [CD]) that produce the FSIQ and three additional primary subtests (Visual Puzzles [VP], Picture Span [PS], and Symbol Search [SS]) used to produce the five factor index scores (two subtests each for Verbal Comprehension [VCI], Visual Spatial [VSI], Fluid Reasoning [FRI], Working Memory [WMI], and Processing Speed [PSI]).

In addition, there are six secondary subtests (Information [IN], Comprehension [CO], Picture Concepts [PC], Arithmetic [AR], Letter-Number Sequencing [LN], and Cancellation [CN]) that are used either for substitution in FSIQ estimation (when one primary subtest is spoiled) or in estimating the General Ability Index and Cognitive Proficiency Index and three newly created Ancillary Index Scores (Quantitative Reasoning, Auditory Working Memory, Nonverbal). Ancillary Index Scores (pseudofactors) are not, however, factorially derived and, thus, were not examined in the present investigation. The FSIQ and Index scores are expressed as standard scores (M = 100, SD = 15). Five new subtests (Naming Speed Literacy, Naming Speed Quality, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation) combine to measure three Complementary Index scales (Naming Speed, Symbol Translation, and Storage and Retrieval); but are not intelligence subtests so may not be substituted for any of the primary or secondary subtests.

Analyses

Exploratory factor analyses (EFA). Multiple criteria were used to determine the number of factors to extract and retain: eigenvalues > 1 (Kaiser, 1960), the scree test (Cattell, 1966), standard error of scree (SE_{scree}; Zoski & Jurs, 1996), parallel analysis (PA; Horn, 1965), Glorfeld's (1995) modified PA, and minimum average partials (MAP, Velicer, 1976; Frazier & Youngstrom, 2007). Simulation studies have found that HPA and MAP are useful *a priori* empirical criteria with scree sometimes a helpful adjunct (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). Some criteria were estimated using SPSS 24 for Macintosh while others were computed with open source software. The SE_{scree} program (Watkins, 2007) was used in scree analysis and *Monte Carlo PCA for Parallel Analysis* software (Watkins, 2000) produced random eigenvalues for PA using 100 iterations to provide stable estimates. Glorfeld's (1995) modified PA criterion utilized eigenvalues at the 95% confidence interval using the *CIeigenvalue* program (Watkins, 2011). Typically, PA suggests retaining too few factors when there is a strong general factor (Crawford et al., 2010); therefore, the publisher's theory was also considered.

Principal axis extraction was employed to assess the WISC–V factor structure using SPSS 24 for Macintosh followed by Promax rotation (k = 4; Gorsuch, 1983). Following Canivez and Watkins (2010a, 2010b), iterations in first-order principal axis factor extraction were limited to two in estimating final communality estimates (Gorsuch, 2003).

Factors were required to have at least two salient loading subtests (≥ .30; Child, 2006) to be considered viable. Variance apportionment of first- and second-order factors was accomplished with the Schmid and Leiman procedure (SL; Schmid & Leiman, 1957), which has been recommended by Carroll (1993) and Gignac (2005) and has been used in numerous Wechsler scale EFA studies: WISC–IV (Watkins, 2006; Watkins et al., 2006), WISC–V (Canivez et al., 2016; Dombrowski et al., 2017; Dombrowski et al., 2015); WISC–IV Spanish (McGill & Canivez, 2016), French WAIS–III (Golay & Lecerf, 2011), French WISC–IV (Lecerf et al., 2011), and the French WISC–V (Lecerf & Canivez, 2018). The SL procedure derives a hierarchical factor model from higher-order models and decomposes the variance of subtest scores first to the general factor and then to the first-order factors and is labeled SL bifactor (Reise, 2012) for convenience. The first-order factors are orthogonal to each other and also to the general factor (Gignac, 2006; Gorsuch, 1983). The SL procedure is an approximate bifactor model (and labeled SL Bifactor for convenience) and was produced using the *MacOrtho* program (Watkins, 2004).

Confirmatory factor analyses (CFA). EQS 6.3 (Bentler & Wu, 2016) was used to conduct confirmatory factor analysis (CFA) using maximum likelihood estimation. In the WISC–V, each of the five latent factors (VC, VS, FR, WM, PS) have only two observed indicators and thus are underidentified. Consequently, those subtests were constrained to equality in bifactor CFA models to ensure identification (Little, Lindenberger, & Nesselroade, 1999). Given the significant multivariate kurtosis of the scores, robust maximum likelihood estimation with the Satorra and Bentler (S-B; 2001) corrected chi-square was applied. Byrne (2006, p. 138) indicated "the S-B χ^2 has been shown to be the most reliable test statistic for evaluating mean and covariance structure models under various distributions and sample sizes."

The structural models with the 10 WISC–V primary subtests previously examined by Canivez et al. (2017) were investigated (both higher-order and bifactor models) with the present CFA clinical sample. Model 1 is a unidimensional *g* factor model with all 10 primary subtests loading only on *g*. Table 4 illustrates the subtest associations within the various models. Models with more than one group factor included a higher-order *g* factor and models with four- and five-group factors included higher-order and bifactor variants, including that suggested by EFA.

Given that the large sample size may unduly influence the χ^2 value (Kline, 2016), approximate fit indices were used to aid model evaluation and selection. While universally accepted criterion values for approximate fit indices do not exist (McDonald, 2010), the comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) were used to evaluate overall global model fit. Higher values indicate better fit for the CFI and TLI whereas lower values indicate better fit for the RMSEA. Hu and Bentler's (1999) combinatorial heuristics were applied where CFI and TLI $\geq .90$ along with RMSEA $\leq .08$ were criteria for adequate model fit; whereas CFI and TLI $\geq .95$ and RMSEA $\leq .06$ were criteria for good model fit. The Akaike Information Criterion (AIC) was also considered, but because AIC does not have a meaningful scale, the model with the smallest AIC value was preferred as most likely to replicate (Kline, 2016). Superior models required adequate to good overall fit *and* indication of meaningfully better fit (ΔCFI > .01, ΔRMSEA > .015, ΔAIC > 10) than alternative models (Burnham & Anderson, 2004; Cheung & Rensvold, 2002; Chen, 2007). Local fit was also considered in addition to global fit as models should never be retained “solely on global fit testing” (Kline, 2016, p. 461). The large sample size allowed for sufficient statistical power to detect even small differences as well as more precise estimates of model parameters.

Coefficients omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}) were estimated as model-based reliabilities and provide estimates of reliability of unit-weighted scores produced by the indicators (Reise, 2012; Rodriguez et al., 2016; Watkins, 2017). The ω_H coefficient is the general intelligence factor reliability estimate with variability from the group factors removed, whereas the ω_{HS} coefficient is the group factor reliability estimate with variability from all other group *and* general factors removed (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012). Omega estimates (ω_H and ω_{HS}) are calculated from CFA bifactor solutions or decomposed variance estimates from higher-order models and were obtained using the *Omega* program (Watkins, 2013), which is based on the Brunner et al. (2012) tutorial and the works of Zinbarg, Revelle, Yovel, and Li (2005) and Zinbarg, Yovel, Revelle, and McDonald (2006). ω_H and ω_{HS} coefficients should exceed .50, but .75 might be preferred (Reise, 2012; Reise et al., 2013). Omega coefficients were supplemented with Hancock and Mueller's (2001) construct reliability or construct replicability coefficient (*H*), which estimates the adequacy of the latent construct represented by the indicators, with a criterion value of .70 (Hancock & Mueller, 2001; Rodriguez et al., 2016). *H* coefficients were produced by the *Omega* program (Watkins, 2013).

Results

WISC–V Exploratory Factor Analyses

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy of .902 far exceeded the minimum standard of .60 (Kaiser, 1974) and Bartlett's Test of Sphericity (Bartlett, 1954), χ^2 = 6,372.06, $p < .0001$; indicated that the WISC–V correlation matrix was not random. Initial communality estimates ranged from .377 to .648. Therefore, the correlation matrix was deemed appropriate for factor analysis.

Factor Extraction Criteria

Scree, *SE*scree, PA, Glorfeld's modified PA, and MAP criteria all suggested only one factor while the eigenvalues > 1 criterion suggested 2 factors. The publisher of the WISC–V, however, claims five factors and the traditional Wechsler structure suggests four factors. Because Wood, Tataryn, and Gorsuch (1996) noted that it is better to overextract than underextract, EFA began by extracting five factors to examine subtest associations with latent factors based on the publisher's promoted WISC–V structure. This permitted the assessment of smaller factors and subtest alignment. Models with four, three, and two factors were then sequentially examined for adequacy.

Exploratory Factor Analyses Models

Five–Factor model. When five WISC–V factors were extracted followed by promax rotation, a fifth factor with no salient factor pattern coefficients resulted (see Table 5). The BD, VP, MR, and FW subtests had salient pattern coefficients on a common factor, but MR and FW did *not* share sufficient common variance separate from BD and VP to constitute separate Fluid Reasoning and Visual Spatial dimensions. Given that no salient fifth factor emerged, the five-factor model was judged inadequate.

Four–Factor model. Table 6 presents the results from extraction of four WISC–V factors followed by promax rotation. The *g* loadings ranged from .567 (CD) to .796 (VP) and all were within the fair to good range based on Kaufman's (1994) criteria (≥ .70 = good, .50 –.69 = fair, < .50 = poor). Table 6 illustrates strong, well defined Verbal Comprehension (SI, VO), Perceptual Reasoning (BD, VP, MR, FW), Working Memory (DS, PS), and Processing Speed (CD, SS) factors with theoretically consistent subtest associations resembling the traditional WISC–IV structure. None of the subtests had salient factor pattern coefficients on more than one factor, thereby achieving desired simple structure. The factor intercorrelations (.531 to .755) were moderate to high and suggested the presence of a general intelligence factor that should be further explicated (Gorsuch, 1983).

Two- and three–factor models. Results from the two and three WISC–V factor extractions with promax rotation are presented in Table 7. For the three-factor model, the Perceptual Reasoning factor remained intact as the first factor but the second factor was a merging of Verbal Comprehension and Working Memory factors. The Processing Speed factor emerged as the third factor. When extracting only three factors the PS subtest cross-loaded on PR and PS factors. In the two-factor model, Factor 1 included all subtests (except MR and SS that had salient factor pattern coefficients on the second factor along with CD). Coding also cross-loaded on Factor 1. Thus, the two– and three–factor models clearly displayed fusion of theoretically meaningful constructs, subtest migration to alternate factors that would not be expected, and cross-loadings. This appears to be due to underextraction, thereby rendering them unacceptable (Gorsuch, 1983; Wood et al., 1996).

Hierarchical EFA: SL Bifactor Model

The EFA results indicated that the four-factor solution was the most appropriate and was accordingly subjected to higher-order EFA and transformed with the SL orthogonalization procedure (see Table 8). Following SL transformation, all subtests were properly associated with their theoretically proposed factors resembling the WISC–IV (Wechsler model). The hierarchical *g* factor accounted for 42.4% of the total variance and 70.2% of the common variance. The general factor also accounted for between 28.6% (CD) and 51.0% (SI and VO) of individual subtest variability.

The PR group factor accounted for an additional 7.1% and 11.8%, VC an additional 3.6% and 5.9%, PS an additional 5.7% and 9.5%, and WM an additional 1.5% and 2.6% of the total and common variance, respectively. The general and group factors combined to measure 60.3% of the common variance in WISC–V scores, leaving 39.7% unique variance (a combination of specific and error variance).

Based on SL results in Table 8, omega–hierarchical (ω_H) and omega–hierarchical subscale (ω_{HS}) coefficients were estimated. The general intelligence ω_H coefficient (.821) was high and indicated that a unit-weighted composite score based on the indicators would be sufficient for scale interpretation; however, the group factor (PR, VC, PS, WM) ω_{HS} coefficients were considerably lower (.083-.351). This suggests that unit-weighted composite scores based on the four WISC–V group factors' indicators would likely contain too little true score variance for clinical interpretation (Reise, 2012; Reise et al., 2013). Table 8 also presents *H* coefficients which reflect the correlation between the latent factor and optimally weighted composite scores (Rodriguez et al., 2016). The *H* coefficient for the general factor[1] (.883) signaled that the general factor was well defined by the 10 WISC–V primary subtest indicators and was a good indicator of construct reliability or replicability (Rodriguez et al.); but the *H* coefficients for the four group factors ranged from .116 to .505 and suggested that the four group factors were inadequately defined by their subtest indicators.

Table 9 presents decomposed variance estimates from the SL bifactor solution of the second-order EFA with the forced five factor extraction. Like the first-order EFA, subtests purported to measure fluid reasoning (MR and FW) had their largest portions of residual variance apportioned to the PR factor along with BD and VP subtests. The MR and FW subtests also had small amounts of residual variance apportioned to the fifth factor (5.2% and 2.5%, respectively). These portions of unique residual variance appear to be the result of diverting small amounts of variance from the general intelligence factor. Another indication of the extremely poor measurement of the fifth factor is the ω_{HS} coefficient of .052 which indicates that a unit-weighted composite score based on MR and FW subtests would account for a meager 5.2% true score variance.

Confirmatory Factor Analyses

Results of CFA for the 10 WISC–V primary subtests with the CFA clinical sample are presented in Table 10. The combinatorial heuristics of Hu and Bentler (1999) revealed that Model 1 (*g*) and Model 2 (V, P) were inadequate due to low CFI and TLI and high RMSEA values. Model 3 (V, P, PS) was inadequate due to high RMSEA values. Both models with four group factors reflecting traditional Wechsler (VC, PR, WM, PS) configurations, 4a Higher-Order (see Figure 2) and 4b Bifactor (see Figure 3), were well fitting models to these data. Both models with five group factors reflecting CHC (VC, VS, FR, WM, PS) configurations, 5a Higher-Order (see Figure 2) and 5b Bifactor (see Figure 3), were also adequate fitting models to these data.

Assessment of local fit for all models with four and five group factors indicated statistically significant standardized path coefficients and there were no problems identified with impermissible parameter estimates. Model 4a Higher-Order and Model 4b Bifactor were not meaningfully different based on global fit statistics, but the bifactor model had the lower AIC index, which exceeded the $\Delta AIC > 10$ criterion (Burnham & Anderson, 2004). Because CHC based WISC–V models with 10 primary subtests are underidentified, Model 5a Higher-Order and Model 5b Bifactor were mathematically equivalent (see Table 10). Based on the $\Delta AIC > 10$ criterion (Burnham & Anderson, 2004), the Wechsler Higher-Order model (Model 4a) was superior to the CHC Higher-Order model (Model 5a) and the Wechsler Bifactor model (Model 4b) was superior to the CHC Bifactor model (Model 5b) and thus more likely to replicate.

According to the ΔAIC > 10 criterion, the best fitting model was the Wechsler based Model 4b Bifactor, which was also consistent with the present EFA results. Table 11 presents sources of variance for Model 4b Bifactor from the 10 WISC–V primary subtests. The general intelligence dimension accounted for most of the subtest variance and substantially smaller portions of subtest variance were uniquely associated with the four WISC–V group factors (except for CD and SS). Omega-hierarchical and omega-hierarchical subscale coefficients estimated using bifactor results from Table 11 found the ω_H coefficient for general intelligence (.836) was high and indicated a unit-weighted composite score based on the 10 subtest indicators would produce 83.6% true score variance. The ω_{HS} coefficients for the four WISC–V factors (VC, PR, WM, PS) were considerably lower ranging from .100 (WM) to .397 (PS). Thus, unit-weighted composite scores for the four WISC–V first-order factors possess too little true score variance to recommend clinical interpretation (Reise, 2012; Reise et al., 2013). Table 11 also presents *H* coefficients that reflect correlations between the latent factors and optimally weighted composite scores (Rodriguez et al., 2016). The *H* coefficient for the general factor[1] (.895) indicated the general factor was well defined by the 10 WISC–V subtest indicators, but the *H* coefficients for the four group factors ranged from .144 to .484 and, as with the EFA sample, indicated that the four group factors were not adequately defined by their subtest indicators.

Discussion

The present WISC–V EFA and CFA results with a large clinical sample bifurcated into EFA and CFA samples provided replication of independent WISC–V EFA and CFA results previously reported with the standardization sample (Canivez et al., 2016, 2017; Dombrowski et al., 2017; Dombrowski et al., 2015). EFA results with the present clinical sample did not identify the five latent WISC–V factors specified by the publisher because the VS and FR factors did not emerge as separate and distinct dimensions. Subtests thought to measure distinct VS and FR factors shared variance associated with a single PR dimension similar to the former WISC–IV. Further, hierarchical EFA and Schmid and Leiman (1957) orthogonalization replicated the dominance of the general intelligence factor and the limited unique measurement of the four group factors; the general factor accounted for more than 5.9 times as much common subtest variance as any individual WISC–V group factor and about 2.4 times as much common subtest variance as all four WISC–V group factors combined. Despite publisher claims of five group factors as well as scoring and interpretive guidelines for five factors, independent EFA of the WISC–V standardization sample and the present clinical sample supports only four factors. These results are also consistent with an independent EFA examinations of the French WISC–V (Wechsler, 2016a) standardization sample (Lecerf & Canivez, 2018) and WISC–V^{UK} (Wechsler, 2016b) standardization sample (Canivez et al., 2018).

CFA results with the present clinical sample generally paralleled those of previous independent CFA of the WISC–V standardization sample (Canivez et al. 2017), although in the present clinical sample, models with five group factors did not produce model specification errors and improper parameter estimates. Consistent with the present EFA results, the best fitting CFA measurement model was the traditional four-factor Wechsler model in a bifactor structure. While a CHC based bifactor model provided adequate fit, standardized coefficients for MR and FW were *higher* with the Perceptual Reasoning factor (Wechsler model) than they were with the Fluid Reasoning factor (CHC model) where they were weak (see Figure 4). Like the EFA results, the assessment of variance sources from the Wechsler-based bifactor model (Model 4b) showed the dominance of the general intelligence factor and the limited unique measurement of the four group factors. The subtest variance apportions indicated that the general factor accounted for more than 6.75 times as much common subtest variance as any individual WISC–V group factor and about 2.4 times as much common subtest variance as all four WISC–V group factors combined. The present CFA results are consistent with independent CFAs of standardization samples from the Canadian WISC–V (WISC–V^{CDN}; Wechsler, 2014c), WISC–V^{Spain} (Wechsler, 2015), French WISC–V, and WISC–V^{UK} (Canivez et al., 2018; Fenollar-Cortés & Watkins, 2018; Lecerf & Canivez, 2018; Watkins et al., 2017).

Model-based reliability estimates (ω_H and ω_{HS}) and construct reliability or construct replicability coefficients (*H*) from both EFA and CFA results of the bifactor models indicated that while the broad *g* factor would allow confident individual interpretation (EFA ω_H = .811, CFA ω_H = .829, EFA *H* = .883, CFA *H* = .895), the ω_{HS} and *H* estimates for the four WISC–V group factors were unacceptably low (see Tables 8 and 11), and thus extremely limited for measuring unique cognitive constructs (Brunner et al, 2012; Hancock & Mueller, 2001; Reise, 2012; Rodriguez et al., 2016).

Similar EFA and CFA results have also been observed in studies of the WISC–IV (Bodin et al., 2009; Canivez, 2014b; Keith, 2005; Watkins, 2006, 2010; Watkins, Wilson, Kotz, Carbone, & Babula, 2006) and with other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Canivez, Watkins, Good, James, & James, 2017; Canivez et al., 2018; Fenollar-Cortés & Watkins, 2018; Golay & Lecerf, 2011; Golay et al., 2013; Gignac, 2005, 2006; Lecerf & Canivez, 2018; McGill & Canivez, 2016, 2017; Watkins & Beaujean, 2014; Watkins et al., 2017; Watkins et al., 2013), so these results are not unique to the WISC–V. While some of these studies were of standardization samples, some EFA and CFA studies were of clinical samples (Bodin et al., 2009; Canivez, 2014a; Canivez, Watkins, Good, James, & James, 2017; Watkins, 2010; Watkins et al., 2013; Watkins et al., 2006). Further, similar results have been reported with the DAS (Cucina & Howardson, 2017); DAS–II (Canivez & McGill, 2016; Dombrowski, Golay, McGill, & Canivez, 2018; Dombrowski, McGill, Canivez, & Peterson, 2018), KAIT (Cucina & Howardson, 2017), KABC (Cucina & Howardson, 2017), KABC-2 (McGill & Dombrowski, 2018), SB5 (Canivez, 2008; DiStefano & Dombrowski, 2006), WASI and WRIT (Canivez et al., 2009), RIAS (Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson et al., 2007), CAS (Canivez, 2011), WJ III (Cucina & Howardson, 2017; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Strickland, Watkins, & Caterino, 2015), and the WJ IV Cognitive and full battery (Dombrowski, McGill, & Canivez, 2017a, 2017b), so results of domination of general intelligence and limited unique measurement of group factors are not unique to Wechsler scales. These results and the advantages of bifactor modeling for understanding test structure (Canivez, 2016; Cucina & Byle, 2017; Reise, 2012; Gignac, 2008) indicate that comparisons of bifactor models to the higher-order models are needed.

Within CFA models, a higher-order representation of intelligence test structure is an indirect hierarchical model (Gignac, 2005, 2006, 2008) and the first-order factors fully mediate the subtest influences of the *g* factor to influence subtests *indirectly* (Yung et al., 1999). The higher-order model conceives of *g* as a *superordinate* factor and as Thompson (2004) noted, *g* would be an abstraction from abstractions. While higher-order models have been most commonly applied to assess "construct-relevant psychometric multidimensionality" (Morin, Arens, & Marsh, 2016, p. 117) of intelligence tests, the alternative bifactor model was originally specified by Holzinger and Swineford (1937) and has been referred to as a direct hierarchical (Gignac, 2005, 2006, 2008) or nested factors model (Gustafsson, & Balke, 1993). In bifactor models, *g* is conceptualized as a *breadth factor* (Gignac, 2008) because both the general (*g*) *and* the group factors *directly* influence the subtests and are at the same level of inference. Both *g* *and* first-order group factors are simultaneous abstractions derived from the observed subtest indicators and therefore should be considered a more parsimonious and less complicated conceptual model (Canivez, 2016; Cucina & Byle, 2017; Gignac, 2008). In bifactor models, the general factor direct subtest indicator influences are easy to interpret, both general *and* specific subtest influences can be simultaneously examined, and the psychometric properties necessary for determining scoring and interpretation of subscales can be directly examined (Canivez, 2016; Reise, 2012).

Bifactor and higher-order representations of intelligence have generated scholarly debate and varying perspectives. Some have questioned the appropriateness of bifactor models of intelligence on theoretical grounds. Reynolds & Keith (2013) stated that "we believe that higher-order models are theoretically more defensible, more consistent with relevant intelligence theory (e.g., Jensen, 1998), than are less constrained hierarchical [bifactor] models" (p. 66). In contrast, Gignac (2006, 2008) argued that general intelligence is the most substantial factor of a battery of tests and subtest influences should be *directly* modeled and it is the higher-order model that demands explicit theoretical justification of the full mediation of general intelligence by the group factor. Carroll (1993, 1995) pointed out that subtest scores reflect variation on both a general and a more specific group factor, so while subtest scores may appear reliable, the reliability is primarily a function of the general factor, *not* the specific group factor. Other researchers have indicated that the bifactor model better represents Spearman’s (1927) and Carroll’s (1993) conceptualizations of intelligence (Beaujean, 2015a; Frisby & Beaujean, 2015; Brunner et al., 2012; Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993). Beaujean (2015a) elaborated that Spearman's conception of general intelligence was of a factor "that was directly involved in all cognitive performances, not indirectly involved through, or mediated by, other factors" (p. 130) and also pointed out that "Carroll was explicit in noting that a bi-factor model best represents his theory" (p. 130). The present results (both EFA and CFA) seem to support Carroll's theory due to the large contributions of *g* in WISC–V measurement and further support previous commentary by Cucina and Howardson (2017) who also concluded that their analyses supported Carroll but not Horn-Cattell.

Murray & Johnson (2013) suggested that bifactor models might better account for unmodeled complexity when compared to higher-order models and thus benefit from statistical bias in favor of the bifactor model. Morgan, Hodge, Wells, and Watkins (2015) found that both bifactor and higher-order models produced good model fit in simulations regardless of the true test structure. Mansolf and Reise (2017) distinguished higher-order and bifactor models in terms of tetrad constraints, indicating that while all models impose *rank constraints*, higher-order models contain unique tetrad constraints not present in a bifactor model. Mansolf and Reise noted that when tetrad constraints are violated, goodness-of-fit statistics are biased in favor of the bifactor model but a technical solution does not appear to be available. Systematic bias favoring the bifactor model *was not* found by Canivez, Watkins, Good, James, and James (2017) in their investigation of the WISC–IV^{UK}.

Some have argued (e.g., Reynolds & Keith, 2017) that the bifactor model may not be appropriate for cognitive data that might deviate from desired simple structure as bifactor models assume factor orthogonality and subtest indicator loadings on only one group factor. Subtest cross-loadings, intermediate factors, and correlated disturbance and/or error terms are frequently added to CFA models produced by researchers preferring a higher-order structure for Wechsler scales. However, such parameters are rarely specified *a priori* and unmodeled complexities are later added iteratively in the form of post-hoc model modifications designed to improve model fit or remedy local fit problems[2] (e.g., Heywood cases). Specification of these parameters may be problematic due to lack of conceptual grounding in previous theoretical work, lack of consideration of earlier EFA, and dangers of hypothesizing after results are known (HARKing; Cucina & Byle, 2017). These CFA methodological concerns were also noted by Horn (1989):

> "At the present juncture of history in the study of human abilities, it is probably overly idealistic to expect to fit confirmatory models to data that well represent the complexities of human cognitive functioning: too much is unknown. Even when we can, a priori, specify a multiple-variable model that fits data in a general way—with chi-square three or four times as large as the number of degrees of freedom (*df*)—we cannot anticipate all the small loadings that must be in a model for a particular sampling of variables and subjects if the model is to 'truly' fit data" (p. 39). Horn continued, "The statistical demands of structure equation theory are stringent. If there is tinkering with results to get a model to fit, the statistical theory, and thus the basis for strong inference, goes out the window" (p. 39).

Horn (1989, p. 40) also noted that if there was overuse of post hoc model modifications then "…one should not give any greater credence to results from modeling analyses than one can give to results from comparably executed factor analytic studies of the older variety" (e.g., EFA). Previous post-hoc attempts with the WAIS–IV (Weiss, Keith, Zhu, & Chen, 2013a) and the WISC–IV (Weiss, Keith, Zhu, & Chen, 2013b) were reported, but numerous psychometric difficulties with the proposed higher-order models including five group factors in both the WAIS–IV and WISC–IV were pointed out by Canivez and Kush (2013).

Although there is debate regarding which model (bifactor or higher-order) is the "correct" model to represent intelligence, Murray and Johnson (2013) concluded that if there is an attempt to estimate or account for domain-specific abilities, the “bifactor model factor scores should be preferred” (Murray & Johnson, 2013, p. 420). By providing factor index scores, comparisons between factor index scores, and suggestions of interpretation of meaning of these scores and comparisons, the WISC–V publisher emphasizes such domain-specific abilities. Thus, the bifactor model is critical in evaluation of the WISC–V construct validity because of publisher claims of what factor index scores measure as well as the numerous factor index score comparisons and inferences derived from such comparisons. Researchers and clinicians must consider empirical *evidence* of how well WISC–V group factor scores (domain-specific) uniquely measure the represented construct independent of the general intelligence (*g*) factor score (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Chen, West, & Sousa, 2006). A bifactor model, which contains a general factor but permits multidimensionality, is better than the higher-order model for determining the relative contribution of group factors independent of the general intelligence factor (Reise, Moore, & Haviland, 2010).

A final note regarding the poor unique contributions to measurement by the four broad WISC–V factors is that there are implications for clinical application. Use of ipsative or pairwise comparisons of WISC–V factor index scores as reflections of processing strengths or weaknesses (PSWs) within CHC or other interpretation schemes *does not* consider the fact that such index scores conflate general intelligence with group factor variance and in most instances *g* is the dominant contributor of reliable variance and little unique true score variance is provided by broad factor. Longitudinal stability of such PSWs (see Watkins & Canivez, 2004) or diagnostic and treatment utility of such WISC–V PSWs has yet to be demonstrated, but given the limited portions of unique measurement factor index scores provide, such evidence may be elusive.

Limitations

The present study examined EFA and CFA of the WISC–V with heterogeneous clinical samples but it is possible that specific clinical groups (ADHD, SLD, etc.) might produce somewhat different results. Further, specific clinical groups at different ages might also show varied EFA and CFA so examination of structural invariance across age within specific clinical groups would also be useful. Other demographic variables where invariance should be examined include sex/gender, race/ethnicity, and socioeconomic status; which is the next step in examining these data. Chen et al. (2015) examined structural invariance across gender with the WISC–V, but bifactor models and models with fewer than five group factors were not examined so invariance of alternative models should also be examined across demographic groups among clinical samples. Finally, the results of the present study only pertain to the latent factor structure and do not answer other WISC–V construct validity questions. Latent class analysis or latent profile analysis might be useful to identify if the WISC–V is able to identify various clinical groups that might differ from normative samples. Further, examinations of WISC–V relations to external criteria such as incremental predictive validity (Canivez, 2013a; Canivez, Watkins, James, James, & Good, 2014; Glutting, Watkins, Konold, & McDermott, 2006) should be conducted to determine if reliable achievement variance is incrementally accounted for by the WISC–V factor index scores beyond that accounted for by the FSIQ (or through latent factor scores [see Kranzler, Benson, & Floyd, 2015]). Diagnostic utility (see Canivez, 2013b) studies should also be examined because of the use of the WISC–V in clinical decision making. The small portions of true score variance uniquely contributed by the group factors in the WISC–V standardization sample (Canivez et al., 2016, 2017) and in the present clinical sample might make it unlikely that the WISC–V factor index scores would provide meaningful value.

Conclusion

Based on the present results with a large clinical sample, the WISC–V appears to be overfactored when extracting five factors and the strong replication of previous EFA and CFA findings with the WISC–V (Canivez et al., 2016, 2017; Dombrowski et al., 2015), WISC–V^{CDN} (Watkins et al., 2017), WISC–V^{UK} (Canivez et al., 2018), WISC–V^{Spain} (Fenollar-Cortés & Watkins, 2018), and French WISC–V (Lecerf & Canivez, 2018) further reinforces the need for extreme caution in WISC–V interpretation beyond the FSIQ. The attempt to divide the Perceptual Reasoning factor into separate and distinct Visual Spatial and Fluid Reasoning factors was again unsuccessful and further suggests that standard scores and comparisons for FR and VS are potentially misleading. Better measurement of FR as distinct from *g* may require creation and inclusion of more or better indicators. Given the insubstantial amounts of unique true score variance captured by the WISC-V group factors in both EFA and CFA, and lack of evidence for incremental validity or diagnostic utility, it seems prudent to recommend more efficient methods of estimating general intelligence in clinical assessment through the use of more cost and time effective tests to estimate general intelligence (Kranzler & Floyd, 2013). Clinicians interpreting WISC–V scores beyond the FSIQ risk engaging in misinterpretation or over-interpretation of scores because the factor index scores conflate general intelligence and group factor variance. Consideration of these and other independent WISC–V studies allow users to "know what their tests can do and act accordingly” (Weiner, 1989, p. 829).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bartlett, M. S. (1954). A further note on the multiplying factors for various χ^2 approximations in factor analysis. *Journal of the Royal Statistical Society Series A (General), 16*, 296 –298.

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3,* 121-136. doi: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. doi:0.1177/0734282916642679.

Bentler, P. M., & Wu, E. J. C. (2016). *EQS for Windows.* Encino CA: Multivariate Software, Inc.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC–IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417-424. doi:10. .1080/09297040802603661

Braden, J. P., & Niebling, B. C. (2012). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 739-757). New York, NY: Guilford.

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

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33,* 261-304.

doi:10.1177/0049124104268644

Byrne, B. M. (2006) *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods, 49,* 1716-1735. doi:10.3758/s13428-016-0814-1

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. doi:10.1037/a0012884

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

Canivez, G. L. (2010). Review of the Wechsler Adult Intelligence Test-Fourth Edition. In R. A. Spies, J. F. Carlson, and K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 684-688). Lincoln, NE: Buros Institute of Mental Measurements.

Canivez, G. L. (2013a). Incremental validity of WAIS-IV factor index scores: Relationships with WIAT–II and WIAT–III subtest and composite scores. *Psychological Assessment, 25,* 484-495. doi:10.1037/a0032092

Canivez, G. L. (2013b). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean, (Eds.), *The Oxford handbook of child psychological assessments* (pp. 84-112). New York: Oxford University Press.

Canivez, G. L. (2014a). Review of the Wechsler Preschool and Primary Scale of Intelligence-

Fourth Edition. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook* (pp. 732-737). Lincoln, NE: Buros Center for Testing.

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

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

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

Canivez, G. L., & Kush, J. C. (2013). WISC–IV and WAIS–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. doi: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. doi:10.1037/pas0000279

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

Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher-order factor analyses of the

Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV) adolescent subsample. *School Psychology Quarterly, 25,* 223-235. doi:10.1037/a0022046

Canivez, G. L., & Watkins, M. W. (2016). Review of the Wechsler Intelligence Scale for Children–Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Authors), *Intelligent testing with the WISC–V* (pp. 683–702). Hoboken, NJ: 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. doi: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. doi:10.1037/pas0000358

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

Canivez, G. L., Watkins, M. W., James, T., James, K., & Good, R. (2014). Incremental validity of WISC–IV^{UK} factor index scores with a referred Irish sample: Predicting performance on the WIAT–II^{UK}. *British Journal of Educational Psychology, 84,* 667-684. doi:10.1111/bjep.12056

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2018). 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*. Advance online publication. http://dx.doi.org/10.1111/bjep.12230

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

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429-452. doi: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). New York, NY: Pergamon Press.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276. doi:10.1207/s15327906mbr0102_10

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139-164. doi:10.1111/j.1745-3984.1978.tb00065.x

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

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80,* 219 -251. doi:10.1111/j.1467-6494.2011.00739.x

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189-225. doi:10.1207/s15327906mbr4102_5

Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children-Fifth Edition. *Personality and Individual Differences, 86,* 1-5. doi:10.1016/j.paid.2015.05.020

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

Child, D. (2006). *The essentials of factor analysis* (3rd. Ed.). New York, NY: Continuum.

Crawford, A. V., Green, S. B., Levy, R., Lo, W.–J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70,* 885–901. doi:10.1177/0013164410379332

Cucina, J. M., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence,5,* 27-48. doi:10.3390/jintelligence5030027

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. doi:10.1037/pas0000389.

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

Dombrowski, S. C. (2013). Investigating the structure of the WJ–III Cognitive at school age. *School Psychology Quarterly, 28,* 154-169. doi: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. doi: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. doi:10.1177/0734282914530838

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.* Advance online publication. doi:10.1007/s40688-017-0125-2

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

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,* 190-207. doi:10.1002/pits.22096

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

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017b). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology Quarterly*. Advance online publication. doi:10.1037/spq0000221

Dombrowski, S. C., McGill, R. J., Canivez, G. L., & Peterson, C. H. (2018, March). Investigating the theoretical structure of the Differential Ability Scales—Second Edition through hierarchical exploratory factor analysis. *Journal of Psychoeducational Assessment*. Advance online publication. doi:10.1177/0734282918760724

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. doi:10.1037/a0031335

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

Evers, A., Hagemeister, C., Høstmaelingen, A., Lindley, P. Muñiz, J., & Sjöberg, A. (2013). *EFPA review model for the description and evaluation of psychological and educational tests*. Brussels, Belgium: European Federation of Psychologists' Associations.

Fenollar-Cortés, J. & Watkins, M. W. (2018): Construct validity of the Spanish Version of the Wechsler Intelligence Scale for Children Fifth Edition (WISC-V[Spain]). *International Journal of School & Educational Psychology*. doi: 10.1080/21683603.2017.1414006

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182. http://dx.doi.org/10 .1016/j.intell.2006.07.002

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS–IV/WMS–IV standardization data. *Intelligence, 51,* 79-97. doi:10.1016/j.intell.2015.04.007

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

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

Gignac, G. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science Quarterly, 50,* 21-43.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS–IV. *Multivariate Behavioral Research, 48,* 639-662. doi:10.1080/00273171.2013.804398

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377-393. doi: https://doi.org/10.1177/0013164495055003002.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC–IV in estimating reading and math achievement on the WIAI–II. *Journal of Special Education, 40,* 103-114. doi:10.1177/00224669060400020101

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

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

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

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

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407-434. doi:10.1207/s15327906mbr2804_2

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

Hayduk, L. A. (2016). Improving measurement-invariance assessments: Correcting entrenched testing deficiencies. *BMC Medical Research Methodology, 16*(130), 1-10. doi:10.1186/s12874-016-0230-3

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41-54. doi:10.1007/BF02287965

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185. doi: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: Urbana, IL.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson technical manual* (Rev. ed., pp. 197-232). Itasca, IL: Riverside.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan and P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd Ed.) (pp. 41-68). New York: Guilford.

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

International Test Commission. International guidelines for test use. *International Journal of Testing, 1,* 93-114. doi:10.1207/S15327574IJT0102_1

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi–factor analysis. *Psychometrika, 76,* 537-549. doi: 10.1007/s11336–011–9218–4

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

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. doi:10.1177/001316446002000116

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

Kaufman, A. S. (1994). Intelligent testing with the WISC-III. New York, NY: Wiley.

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

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

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

Kranzler, J. H., Benson, N., & Floyd, R. G. (2015). Using estimated factor scores from a bifactor analysis to examine the unique effects of the latent variables measured by the WAIS–IV on academic achievement. *Psychological Assessment, 27,* 1402-1416. doi:10.1037/pas0000119

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. New York, NY: Guilford.

Lecerf, T., Golay, P., Reverte, I., Senn, D., Favez, N., & Rossier, J. (2011, July). *Orthogonal higher-order structure and confirmatory factor analysis of the French Wechsler Children Intelligence Scale – Fourth Edition (WISC-IV).* Paper presented at the 12th European Congress of Psychology, Istanbul, Turquie.

Lecerf, T., & Canivez, G. L. (2018). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment, 30,* 793–808. doi:10.1037/pas0000526

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

Mansolf, M., & Reise, S. P. (2016). Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. *Multivariate Behavioral Research, 51,* 698-717. doi:10.1080.00273171.2016.1215898

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence, 61*, 120-129. doi:10.1016/j.intell.2017.01.012

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

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

McGill, R. J., & Canivez, G. L. (2016). Orthogonal higher order structure of the WISC–IV Spanish using hierarchical exploratory factor analytic procedures. *Journal of Psychoeducational Assessment, 36,* 600-606. doi:10.1177/0734282915624293

McGill, R. J., & Canivez, G. L. (2017, October). Confirmatory factor analyses of the WISC–IV Spanish core and supplemental subtests: Validation evidence of the Wechsler and CHC models. *International Journal of School and Educational Psychology*. Advance online publication. doi:10.1080/21683603.2017.1327831

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. doi:10.1007/s40688-017-0152-z

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3,* 2-20. doi:10.3390/jintelligence3010002

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23,* 116-139. doi:10.1080/10705511.2014.961800

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41,* 407–422. doi:10.1016/j.intell.2013.06.004

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

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

Public Law (P.L.) 108-446. *Individuals with Disabilities Education Improvement Act of 2004* (IDEIA). (20 U.S.C. 1400 et seq.). 34 CFR Parts 300 and 301. Assistance to States for the education of children with disabilities and preschool grants for children with disabilities; Final Rule. *Federal Register, 71 (156),* 46540-46845.

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

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

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment, 92, 544-559. doi:10.1080/00223891.2010.496477.

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Saklofske, V. L. Schwean, & C. R. Reynolds (Eds.), *Oxford handbook of child psychological assessment* (pp. 48-83). New York, NY: Oxford University Press.

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. doi: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. doi:10.1037/met0000045

Satorra, A., & Bentler, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. doi: 10.1007/BF02296192

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

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York: Guilford.

Spearman, C. (1927). *The abilities of man*. New York: Cambridge.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York, NY: Oxford University Press.

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

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago, IL: University of Chicago Press.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327, doi:10.1007/BF02293557.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, & E. Helms (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41−71). Norwell, MA: Springer.

Watkins, M. W. (2000). *Monte Carlo PCA for Parallel Analysis* [Computer Software]. State College, PA: Author.

Watkins, M. W. (2004). *MacOrtho.* [Computer Software]. State College, PA: Ed & Psych Associates.

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

Watkins, M. W. (2007). *SEscree* (Computer software). State College, PA: Ed & Psych Associates.

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

Watkins, M. W. (2011). *CIeigenvalue* (Computer software). Phoenix, AZ: Ed & Psych Associates.

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

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*. doi:10.1080/13854046.2017.1317364

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

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC–III subtest composite strengths and weaknesses. *Psychological Assessment, 16,* 133–138. doi:10.1037/1040–3590.16.2.133

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

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2017, October). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–Fifth Edition. *International Journal of School and Educational Psychology*. doi:10.1080/21683603.2017.1342580

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

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition.* San Antonio, TX: NCS Pearson.

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

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

Wechsler, D. (2014c). *Wechsler Intelligence Scale for Children–Fifth Edition: Canadian*. Toronto, Canada: Pearson Canada Assessment.

Wechsler, D. (2015). *Escala de inteligencia de Wechsler para ninos-V. Manual tecnico y de interpretacion.* Madrid, Spain: Pearson Educacion.

Wechsler, D. (2016a). WISC–V. *Echelle d'intelligence de Wechsler pour enfants-5e édition.* Paris, France: Pearson France-ECPA.

Wechsler, D. (2016b). *Wechsler Intelligence Scale for Children-Fifth UK Edition.* London, UK: Harcourt Assessment.

Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of*

Personality Assessment, 53, 827-831. doi:10.1207/s15327752jpa5304_18

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psycheducational Assessment, 31,* 94-113. doi:10.1177/0734282913478030

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

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under– and over–extraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354–365. doi:10.1037/1082–989X.1.4.354

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64,* 113-128. doi:10.1007/BF02294531

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123-133. doi:10.1007/s11336-003-0974-7

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement, 30,* 121-144. doi:10.1177/0146621605278814

Zoski, K. W., & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement, 56,* 443-451. doi: 10.1177/0013164496056003006

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432−442. doi: 10.1037/0033-2909.99.3.432

Footnote

[1]The actual scoring structure of the WISC–V produces the FSIQ score from only 7 subtests so omega hierarchical and *H* estimates based on 10 subtests is theoretical.

[2] It is also important for clinicians to bear in mind that the standardized scores that have been developed for the WISC–V, do not account for these complexities.

Table 1

Demographic Characteristics of the Clinical EFA and CFA Samples

	EFA Sample (n = 1,256)		CFA Sample (n = 1,256)	
	N	%	*N*	%
Sex				
Male	816	65.0	816	65.0
Female	440	35.0	440	35.0
Race/Ethnicity				
White/Caucasian	687	54.7	710	56.5
Black/African American	369	29.4	348	27.7
Asian American	41	3.3	36	2.9
Hispanic/Latino	28	2.2	56	4.5
Native American	3	0.2	2	0.2
Multiracial	94	7.5	75	6.0
Native Hawaiian/Pacific Islander	1	0.1	0	0.0
Other	2	0.2	8	0.6
Unknown	31	2.5	21	1.7

Table 2
Sample Sizes of Race/Ethnicity by Age Group in the EFA and CFA Samples

	Age Group										
EFA Sample (*n* = 1,256)	6	7	8	9	10	11	12	13	14	15	16
White/Caucasian	68	97	86	91	63	70	61	63	43	39	6
Black/African American	23	40	37	47	37	36	40	41	32	28	8
Asian American	3	6	3	6	5	9	2	2	1	3	1
Hispanic/Latino	1	4	5	6	5	2	1	1	1	0	2
Native American	0	0	0	0	1	0	0	2	0	0	0
Multiracial	8	14	15	13	13	10	7	5	6	3	0
Native Hawaiian/Pacific Islander	0	0	0	0	0	0	0	0	1	0	0
Other	0	0	0	1	1	0	0	0	0	0	0
Unknown	0	4	3	8	3	2	5	1	4	1	0
CFA Sample (*n* = 1,256)											
White/Caucasian	77	95	104	94	83	63	62	46	49	37	0
Black/African American	30	35	47	42	47	48	33	21	28	17	0
Asian American	4	6	5	3	5	3	2	4	2	2	0
Hispanic/Latino	4	8	12	11	6	5	4	2	2	2	0
Native American	0	0	0	1	0	1	0	0	0	0	0
Multiracial	7	11	8	13	10	5	4	6	8	3	0
Native Hawaiian/Pacific Islander	0	0	0	0	0	0	0	0	0	0	0
Other	0	0	2	1	3	0	1	0	1	0	0
Unknown	0	0	2	1	5	2	4	1	2	4	0

Table 3

Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Descriptive Statistics for the Clinical EFA and CFA Samples

	EFA Sample (n = 1,256)				CFA Sample (n = 1,256)			
Subtest/Composite	*M*	*SD*	Skewness	Kurtosis	*M*	*SD*	Skewness	Kurtosis
Subtests								
Block Design	8.77	3.30	0.11	-0.21	8.67	3.17	0.02	-0.12
Similarities	8.93	3.25	-0.05	-0.07	9.07	3.29	-0.04	-0.05
Matrix Reasoning	9.14	3.39	0.07	-0.04	8.97	3.37	0.00	-0.24
Digit Span	8.05	3.04	0.13	0.20	7.90	3.09	0.11	0.02
Coding	7.74	3.25	-0.06	-0.43	7.73	3.27	0.00	-0.15
Vocabulary	8.87	3.53	0.06	-0.42	8.89	3.49	0.03	-0.51
Figure Weights	9.45	3.15	-0.04	-0.31	9.51	3.14	-0.03	-0.29
Visual Puzzles	9.51	3.29	-0.04	-0.52	9.54	3.30	-0.01	-0.46
Picture Span	8.59	3.14	0.17	-0.16	8.61	3.03	0.06	-0.02
Symbol Search	8.19	3.20	0.01	0.06	8.21	3.18	-0.07	0.05
Composites								
VCI	94.09	17.21	-0.05	0.02	94.44	17.16	-0.05	-0.22
VSI	95.23	17.18	0.09	-0.15	94.96	16.70	0.00	0.03
FRI	95.93	16.73	0.05	-0.48	95.61	16.77	0.01	-0.43
WMI	90.26	15.44	0.21	0.09	89.89	15.40	0.09	-0.16
PSI	88.45	16.72	-0.18	-0.04	88.46	16.60	-0.22	0.22
FSIQ	91.09	16.90	-0.01	-0.24	90.91	16.90	-0.02	-0.29

Note. VCI = Verbal Comprehension Index, VSI = Visual Spatial Index, FRI = Fluid Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full Scale IQ. Mardia's (1970) multivariate kurtosis estimate (EQS 6.3) was 4.23 for the EFA sample and 9.71 for the CFA sample. Independent t-tests for mean differences of WISC–V subtests and composite scores between the EFA and CFA samples indicated no statistically significant differences with t values ranging from -1.07 to 1.23 (p > .20).

Table 4
Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Primary Subtest Assignment to Theoretical First-Order Group Factors for CFA Model Testing

2 Factor Model		3 Factor Model			Wechsler 4 Factor Model				Cattell-Horn-Carroll (CHC) 5 Factor Model				
V	P	V	P	PS	VC	PR	WM	PS	VC	VS	FR	WM	PS
SI	BD	SI	BD	CD	SI	BD	DS	CD	SI	BD	MR	DS	CD
VO	VP	VO	VP	SS	VO	VP	PS	SS	VO	VP	FW	PS	SS
DS	MR	DS	MR			MR							
	FW		FW			FW							
	PS		PS										
	CD												
	SS												

Note. Factors: V = Verbal, P = Performance, PS = Processing Speed, VC = Verbal Comprehension, WM = Working Memory. 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.

Table 5

Exploratory Factor Analysis of the 10 Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Primary Subtests: Five Oblique Factor Solution with Promax Rotation (k = 4) for the Clinical EFA Sample (N = 1,256)

	General	F1: PR		F2: VC		F3: PS		F4: WM		F5		
WISC–V Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
SI	.749	.049	.619	**.778**	.826	.048	.476	-.036	.626	.031	.376	.685
VO	.746	.054	.624	**.773**	.825	-.033	.457	.080	.646	-.067	.307	.687
BD	.760	**.816**	.825	-.028	.580	.061	.528	-.011	.551	-.001	.200	.683
VP	.796	**.854**	.865	.042	.637	-.034	.503	.001	.576	.002	.230	.750
MR	.719	**.597**	.713	-.031	.585	.029	.479	.087	.578	.249	.426	.577
FW	.705	**.582**	.708	.158	.619	-.028	.424	-.022	.532	.174	.375	.552
DS	.673	.019	.526	.160	.632	.019	.508	**.529**	.722	.121	.406	.552
PS	.610	.032	.490	.068	.532	.092	.524	**.556**	.670	-.058	.216	.460
CD	.567	-.019	.439	-.043	.392	**.752**	.755	.047	.536	.023	.162	.572
SS	.618	.037	.500	.060	.453	**.745**	.772	-.034	.549	-.016	.148	.600
Eigenvalue		5.28		1.06		0.82		0.60		0.52		
% Variance		48.72		6.19		4.27		1.39		0.60		
Factor Correlations		F1: PR		F2: VC		F3: PS		F4: WM		F5		
F1: PR		–										
F2: VC		.716		–								
F3: PS		.600		.536		–						
F4: WM		.663		.750		.698		–				
F5		.252		.434		.191		.393		–		

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, VC = Verbal Comprehension, PS = Processing Speed, WM = Working Memory. *S* = Structure Coefficient, *P* = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (*g* loadings). Salient pattern coefficients (≥ .30) presented in bold.

Table 6

Exploratory Factor Analysis of the 10 Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Primary Subtests: Four Oblique Factor Solution with Promax Rotation (k = 4) for the Clinical EFA Sample (N = 1,256)

	General	F1: Perceptual Reasoning		F2: Verbal Comprehension		F3: Processing Speed		F4: Working Memory		
WISC–V Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
Similarities	.749	.055	.639	**.768**	.825	.028	.469	.002	.638	.683
Vocabulary	.746	.051	.636	**.762**	.826	-.010	.453	.042	.646	.684
Block Design	.760	**.834**	.819	-.029	.582	.095	.526	-.073	.538	.677
Visual Puzzles	.796	**.873**	.861	.039	.638	.002	.501	-.062	.566	.744
Matrix Reasoning	.719	**.631**	.736	-.030	.579	-.027	.470	.209	.599	.560
Figure Weights	.705	**.611**	.726	.156	.615	-.068	.417	.059	.549	.543
Digit Span	.673	.027	.552	.158	.628	.012	.501	**.588**	.733	.551
Picture Span	.610	.025	.495	.067	.532	.151	.523	**.485**	.653	.444
Coding	.567	-.016	.440	-.041	.393	**.739**	.754	.071	.519	.571
Symbol Search	.618	.038	.498	.060	.455	**.741**	.772	-.035	.528	.600
Eigenvalue		5.28		1.06		0.82		0.60		
% Variance		48.72		6.19		4.27		1.39		
Promax Based Factor Correlations		F1: PR		F2: VC		F3: PS		F4: WM		
F1: Perceptual Reasoning (PR)		–								
F2: Verbal Comprehension (VC)		.738		–						
F3: Processing Speed (PS)		.594		.531		–				
F4: Working Memory (WM)		.683		.755		.663		–		

Note. S = Structure Coefficient, *P* = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (g loadings). Salient pattern coefficients (≥ .30) presented in bold.

Table 7

Exploratory Factor Analysis of the 10 Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Primary Subtests: Two and Three Oblique Factor Solutions for the Clinical EFA Sample (N = 1,256)

	Two Oblique Factors				Three Oblique Factors				
WISC-V Subtest	*g*[1]	F1: *g*	F2: PS	h^2	*g*[1]	F1: PR	F2: VC/WM	F3: PS	h^2
SI	.754	**.744** (.765)	.031 (.528)	.586	.748	.079 (.635)	**.781** (.809)	-.052 (.473)	.658
VO	.739	**.702** (.745)	.065 (.533)	.558	.745	.070 (.631)	**.809** (.814)	-.077 (.460)	.667
BD	.719	**.712** (.730)	.028 (.503)	.534	.761	**.828** (.820)	-.074 (.596)	.080 (.530)	.676
VP	.668	**.466** (.641)	.263 (.574)	.450	.797	**.866** (.862)	.009 (.649)	-.018 (.506)	.743
MR	.569	-.070 (.458)	**.791** (.745)	.557	.718	**.601** (.732)	.135 (.617)	.048 (.491)	.547
FW	.735	**.713** (.744)	.047 (.522)	.555	.705	**.599** (.725)	.217 (.629)	-.063 (.428)	.544
DS	.707	**.786** (.735)	-.076 (.447)	.543	.670	.000 (.544)	**.577** (.690)	.185 (.538)	.498
PS	.792	**.858** (.819)	-.058 (.514)	.673	.608	.001 (.489)	**.411** (.595)	**.300** (.552)	.410
CD	.609	**.324** (.565)	**.362** (.578)	.392	.568	-.010 (.436)	-.022 (.444)	**.774** (.754)	.569
SS	.620	.020 (.516)	**.744** (.758)	.574	.618	.054 (.495)	.007 (.493)	**.727** (.764)	.585
Eigenvalue		5.28	1.06			5.28	1.06	0.82	
% Variance		48.25	5.97			48.65	6.15	4.19	
Factor Correlations		F1	F2			F1	F2	F3	
	F1	–			F1	–			
	F2	.667	–		F2	.751	–		
					F3	.598	.612	–	

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, *g* = general intelligence, PS = Processing Speed, WM = Working Memory, h^2 = Communality. [1]General structure coefficients based on first unrotated factor coefficients (*g*–loadings). Factor pattern coefficients (structure coefficients) based on principal factors extraction with promax rotation (k = 4). Coefficient, P = Pattern Coefficient, h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (g loadings). Salient pattern coefficients presented in bold (pattern coefficient ≥ .30)

Table 8

Sources of Variance in the Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) 10 Primary Subtests for the Clinical EFA Sample (N = 1,256) According to an Exploratory Bifactor Model (Orthogonalized Higher-Order Factor Model) with Four First–Order Factors

	General		F1: Perceptual Reasoning		F2: Verbal Comprehension		F3: Processing Speed		F4: Working Memory			
WISC–V Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
Similarities	.714	.510	.031	.001	**.413**	**.171**	.020	.000	.001	.000	.682	.318
Vocabulary	.714	.510	.029	.001	**.410**	**.168**	-.007	.000	.021	.000	.679	.321
Block Design	.667	.445	**.471**	**.222**	-.016	.000	.067	.004	-.036	.001	.673	.327
Visual Puzzles	.700	.490	**.493**	**.243**	.021	.000	.001	.000	-.030	.001	.734	.266
Matrix Reasoning	.658	.433	**.357**	**.127**	-.016	.000	-.019	.000	.102	.010	.571	.429
Figure Weights	.639	.408	**.345**	**.119**	.084	.007	-.048	.002	.029	.001	.538	.462
Digit Span	.677	.458	.015	.000	.085	.007	.009	.000	**.288**	**.083**	.549	.451
Picture Span	.606	.367	.014	.000	.036	.001	.107	.011	**.237**	**.056**	.436	.564
Coding	.535	.286	-.009	.000	-.022	.000	**.524**	**.275**	.035	.001	.563	.437
Symbol Search	.574	.329	.021	.000	.032	.001	**.526**	**.277**	-.017	.000	.608	.392
Total Variance		.424		.071		.036		.057		.015	.603	.397
Explained Common Variance		.702		.118		.059		.095		.026		
ω		.921		.867		.811		.738		.655		
ω_H / ω_{HS}		.821		.270		.194		.351		.083		
Relative ω		.891		.311		.238		.476		.127		
H		.883		.505		.280		.435		.116		
PUC		.800										

Note. *b* = loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness, ω = Omega, ω_H = Omega-hierarchical (general factor), ω_{HS} = Omega-hierarchical subscale (group factors), *H* = construct reliability or replicability index, PUC = percentage of uncontaminated correlations. Bold type indicates highest coefficients and variance estimates and consistent with the theoretically proposed factor.

Table 9

Sources of Variance in the 10 Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) Primary Subtests for the Clinical EFA Sample (N = 1,256) According to an Exploratory SL Bifactor Model (Orthogonalized Higher-Order Factor Model) with Five First–Order Factors

	General		F1: PR		F2: VC		F3: PS		F4: WM		F5			
WISC–V Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
SI	.718	.516	.030	.001	**.405**	**.164**	.034	.001	-.016	.000	.028	.001	.683	.317
VO	.724	.524	.033	.001	**.402**	**.162**	-.023	.001	.037	.001	-.061	.004	.692	.308
BD	.653	.426	**.501**	**.251**	-.015	.000	.043	.002	-.005	.000	-.001	.000	.680	.320
VP	.687	.472	**.525**	**.276**	.022	.000	-.024	.001	.000	.000	.002	.000	.749	.251
MR	.642	.412	**.367**	**.135**	-.016	.000	.020	.000	.040	.002	.228	.052	.601	.399
FW	.624	.389	**.358**	**.128**	.082	.007	-.020	.000	-.010	.000	.159	.025	.550	.450
DS	.684	.468	.012	.000	.083	.007	.013	.000	**.242**	**.059**	.111	.012	.546	.454
PS	.620	.384	.020	.000	.035	.001	.065	.004	**.255**	**.065**	-.053	.003	.458	.542
CD	.533	.284	-.012	.000	-.022	.000	**.530**	**.281**	.022	.000	.021	.000	.567	.433
SS	.573	.328	.023	.001	.031	.001	**.525**	**.276**	-.016	.000	-.015	.000	.606	.394
Total S^2		.420		.079		.034		.057		.013		.010	.613	.387
ECV		.686		.129		.056		.092		.021		.016		
ω_H / ω_{HS} [1]		.821		.270		.194		.351		.083				
ω_H / ω_{HS} [2]		.849		.308		.194		.351		.083		.052		

Note. 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, VC = Verbal Comprehension, PS = Processing Speed, WM = Working Memory, ECV = Explained Common Variance. *b* = loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness. Bold type indicates highest coefficients and variance estimates. [1]Matrix Reasoning and Figure Weights included on Factor 1 (Perceptual Reasoning). [2]Matrix Reasoning and Figure Weights included on Factor 5 (supposedly Fluid Reasoning).

Table 10

Robust Maximum Likelihood CFA Fit Statistics for 10 WISC–V Primary Subtests for the Clinical CFA Sample (N = 1,256)

Model[1]	S-Bχ^2	*df*	CFI	TLI	RMSEA	RMSEA 90% CI	AIC
1 (*g*)	898.33	35	.839	.792	.140	[.132, .148]	59,650.94
2[2] (V, P)	594.04	33	.895	.857	.116	[.108, .125]	59,321.48
3 (V, P, PS)	361.42	32	.938	.913	.091	[.082, .099]	59,037.53
4a Higher-Order (VC, PR, WM, PS)	170.66	31	.974	.962	.060	[.051, .069]	58,831.45
4b Bifactor[3] (VC, PR, WM, PS)	**144.20**	**28**	**.978**	**.965**	**.058**	**[.048, .067]**	**58,813.56**
5a Higher-Order (VC, VS, FR, WM, PS)	216.84	30	.965	.948	.070	[.062, .079]	58,886.17
5b Bifactor[4] (VC, VS, FR, WM, PS)	216.84	30	.965	.948	.070	[.062, .079]	58,886.17

Note. Mardia's multivariate kurtosis estimate was 9.71 indicating multivariate non-normality and need for robust estimation. All models were statistically significant ($p < .001$). S-B = Satorra-Bentler, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, AIC = Akaike's Information Criterion, *g* = general intelligence, V = Verbal, P = Performance, PS = Processing Speed, VC = Verbal Comprehension, PR = Perceptual Reasoning, WM = Working Memory, VS = Visual Spatial, FR = Fluid Reasoning. Bold text illustrates best fitting model. [1]Model numbers correspond to those reported in the WISC-V *Technical and Interpretive Manual* and are higher-order models (unless otherwise specified) when more than one first-order factor was specified. [2]EQS condition code indicated Factor 2 (Performance) and the higher-order factor (g) were linearly dependent on other parameters so variance estimate set to zero for model estimation and loss of 1 *df*. [3]VC, WM, and PS factor subtest loadings were constrained to equality to identify the bifactor version of Model 4b due to under-identified latent factors (VC, WM, PS). [4]VC, VS, FR, WM, and PS factor subtest loadings were constrained to equality to identify the bifactor version of Model 4b due to under-identified latent factors (VC, VS, FR, WM, PS). Due to constraining each factor's loadings to equality because of under-identified latent factors (VC, VS, FR, WM, PS), bifactor Model 5b is mathematically equivalent to higher-order Model 5a.

Table 11

Sources of Variance in the Wechsler Intelligence Scale for Children-Fifth Edition (WISC–V) 10 Primary Subtests for the Clinical CFA Sample (N = 1,256) According to a Bifactor Model with Four Group Factors

	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed			
WISC–V Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
Similarities	.711	.506	.472	.223							.728	.272
Vocabulary	.735	.540	.445	.198							.738	.262
Block Design	.637	.406			.499	.249					.655	.345
Visual Puzzles	.711	.506			.477	.228					.733	.267
Matrix Reasoning	.679	.461			.320	.102					.563	.437
Figure Weights	.692	.479			.287	.082					.561	.439
Digit Span	.761	.579					.276	.076			.655	.345
Picture Span	.632	.399					.281	.079			.478	.522
Coding	.521	.271							.557	.310	.582	.418
Symbol Search	.553	.306							.573	.328	.634	.366
Total Variance		.445		.042		.066		.016		.064	.633	.367
Explained Common Variance		.704		.066		.104		.025		.101		
ω		.930		.846		.869		.722		.756		
ω_H / ω_{HS}		.836		.243		.220		.100		.397		
Relative ω		.899		.287		.253		.138		.525		
H		.895		.348		.454		.144		.484		
PUC		.800										

Note. *b* = loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness, ω = Omega, ω_H = Omega-hierarchical (general factor), ω_{HS} = Omega-hierarchical subscale (group factors), *H* = construct reliability or replicability index, PUC = percentage of uncontaminated correlations.

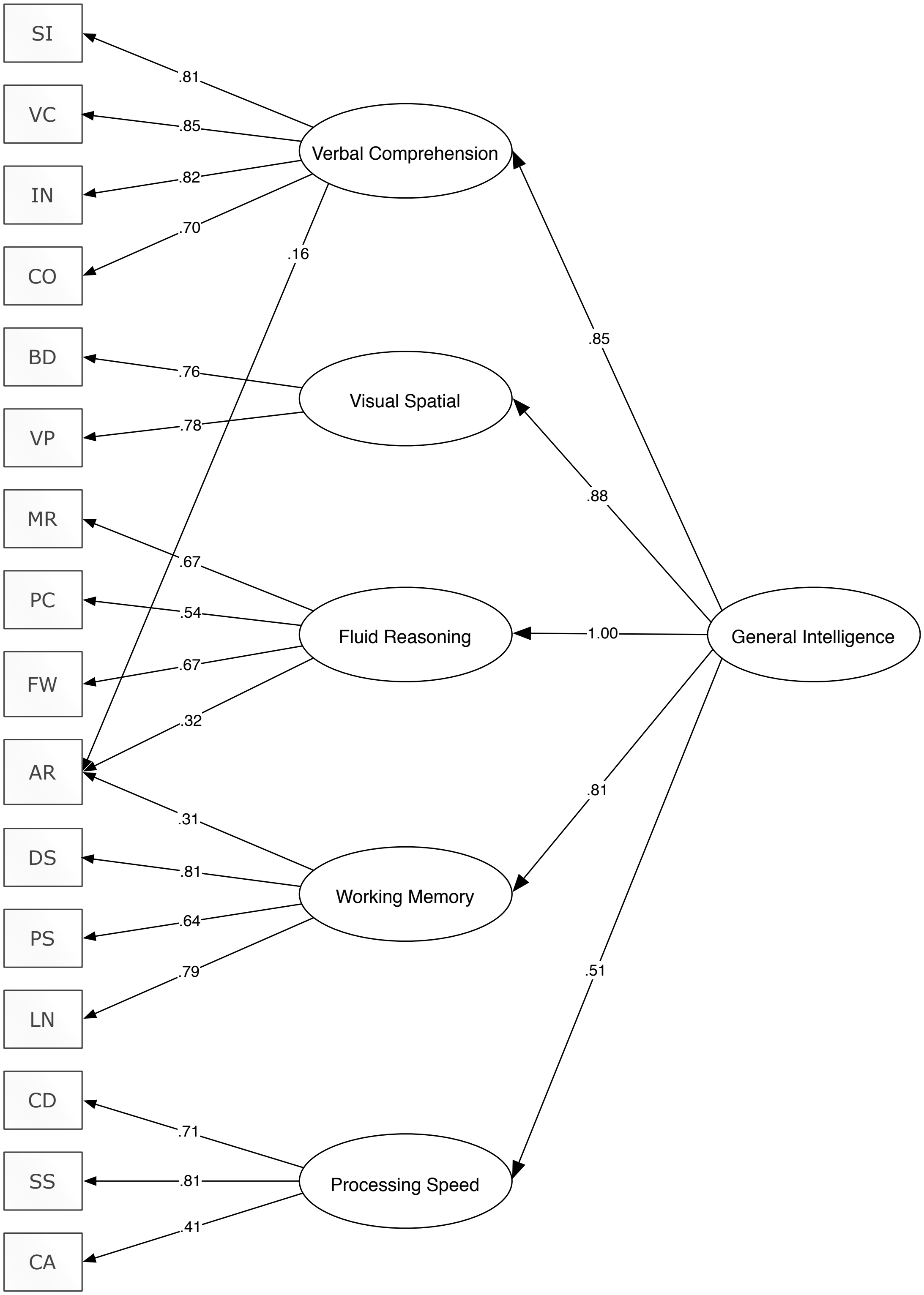
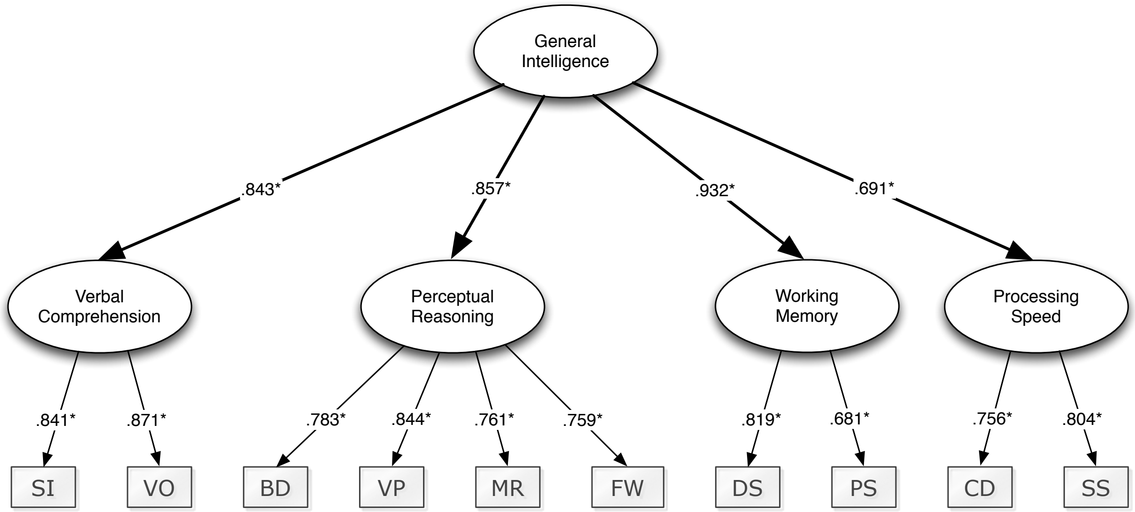


Figure 1. Higher-order measurement model with standardized coefficients (adapted from Figure 5.1 [Wechsler, 2014b]), for WISC–V standardization sample (*N* = 2,200) 16 Subtests. SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, PC = Picture Concepts, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LN = Letter-Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation.

Model 4a Wechsler Higher-Order



Model 5a CHC Higher-Order

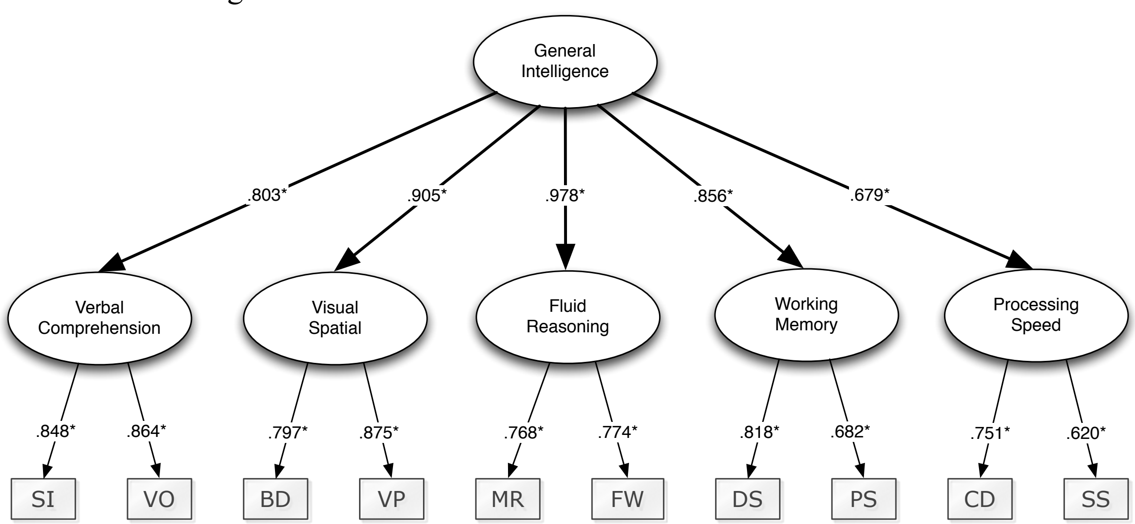
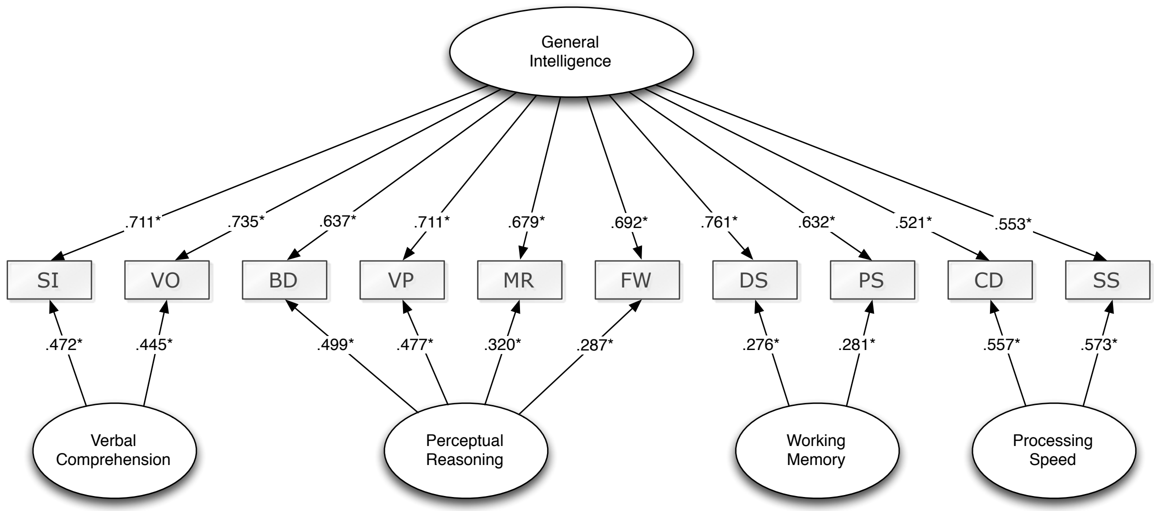


Figure 2. Higher-order measurement models (4a [Wechsler Model] and 5a [CHC Model]), with standardized coefficients, for the 10 WISC–V primary subtests with the clinical CFA sample (N = 1,256). SI = Similarities, VO = Vocabulary, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, DS = Digit Span, PS = Picture Span, CD = Coding, SS = Symbol Search. *p < .05.

Model 4b Wechsler Bifactor



Model 5b CHC Bifactor

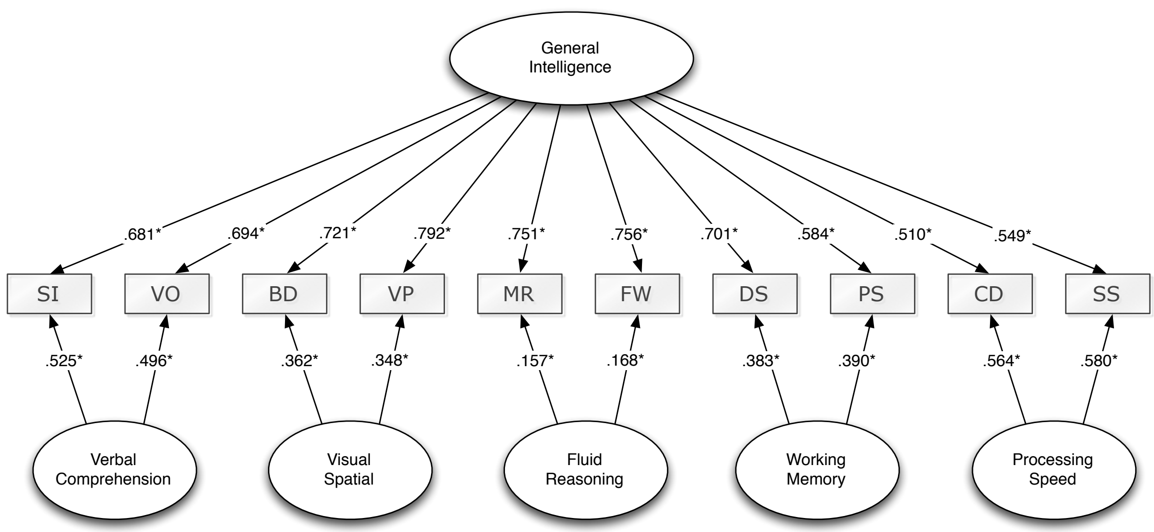


Figure 3. Bifactor measurement models (4b Bifactor [Wechsler Model] and 5b Bifactor [CHC Model]), with standardized coefficients, for the 10 WISC–V primary subtests with the clinical CFA sample (N = 1,256). SI = Similarities, VO = Vocabulary, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, DS = Digit Span, PS = Picture Span, CD = Coding, SS = Symbol Search. *p < .05.