Article

Assessment
2021, Vol. 28(2) 327–352
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191120936330
journals.sagepub.com/home/asm
SAGE

Construct Validity of the German Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory and Confirmatory Factor Analyses of the 15 Primary and Secondary Subtests

Gary L. Canivez[1], Silvia Grieder[2], and Anette Buenger[2]

Abstract

The latent factor structure of the German Wechsler Intelligence Scale for Children–Fifth edition (German WISC-V) was examined using complementary hierarchical exploratory factor analyses (EFAs) with Schmid and Leiman transformation and confirmatory factor analyses (CFAs) for all reported models from the German WISC-V *Technical Manual* and rival bifactor models using the standardization sample (N = 1,087) correlation matrix of the 15 primary and secondary subtests. EFA results did not support a fifth factor (Fluid Reasoning). A four-factor model with the dominant general intelligence (g) factor resembling the WISC-IV was supported by EFA. CFA results indicated the best representation was a bifactor model with four group factors, complementing EFA results. Present EFA and CFA results replicated other independent assessments of standardization and clinical samples of the United States and international versions of the WISC-V and indicated primary, if not exclusive, interpretation of the Full Scale IQ as an estimate of g.

Keywords

German WISC-V, exploratory factor analysis, confirmatory factor analysis, bifactor model, hierarchical CFA, intelligence

Worldwide popularity of Wechsler scales has resulted in numerous translations, adaptations, and norms for many different countries, languages, and cultures (Georgas et al., 2003; Oakland et al., 2016); and H. Chen et al. (2010) reported latent factor structure invariance of the Wechsler Intelligence Scale for Children–Fourth edition (WISC-IV) across cultures. The Wechsler Intelligence Scale for Children–Fifth edition (WISC-V; Wechsler, 2014a) is the most recent version and purports to measure five first-order factors (Verbal Comprehension [VC], Visual Spatial [VS], Fluid Reasoning [FR], Working Memory [WM], Processing Speed [PS]) and a higher-order general intelligence (g) factor. This is consistent with contemporary conceptualizations of intelligence influenced by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Blankson, 2005; Horn & Cattell, 1966), often referred to as the so-called Cattell–Horn–Carroll (CHC) theory (Schneider & McGrew, 2012, 2018), and was also influenced by neuropsychological constructs (Wechsler, 2014c). A major revision goal in constructing the WISC-V was to separate subtests from the former Perceptual Reasoning (PR) factor into distinct VS and FR factors for better match to CHC. Similar attempts were previously made with the WAIS-IV (Weiss et al., 2013a) and WISC-IV (Weiss et al., 2013b), but Canivez and Kush (2013) highlighted numerous psychometric problems with the proposed higher-order models that included five group factors in both the WAIS-IV *and* the WISC-IV. Among the problems noted by Canivez and Kush (2013) were selective reporting of extant literature, creating intermediary factors that make models appear statistically better, post hoc model modifications, neglecting rival bifactor models, and lack of disclosure of decomposed variance estimates. WISC-V adaptations and norms are available for Canada, Spain, France, the United Kingdom, and Germany; and a version for Japan is forthcoming.

Canivez and Watkins (2016) criticized the publisher's claimed supportive evidence for the preferred higher-order measurement model (Model 5e) presented in the U.S. WISC-V *Technical and Interpretive Manual* (Wechsler, 2014c) that included numerous methodological and

[1]Eastern Illinois University, Charleston, IL, USA
[2]University of Basel, Basel, Switzerland

Corresponding Author:
Gary L. Canivez, Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920-3099, USA.
Email: glcanivez@eiu.edu

statistical problems including failure to report results of exploratory factor analyses (EFAs), use of weighted least squares (WLS) estimation in confirmatory factor analyses (CFA) without explicit justification (Kline, 2016), failure to fully disclose details of CFA, abandoning parsimony of simple structure (Thurstone, 1947) by cross-loading Arithmetic [AR] on *three* group factors, empirical redundancy of FR and *g* due to the standardized path coefficient of 1.0 between *g* and the FR factor, no consideration or testing of rival bifactor models, omission of decomposed variance sources between the higher-order *g* and lower order group factors, and absence of model-based reliability/validity estimates for *g* (omega-hierarchical [ω_H]) and the lower-order group (omega-hierarchical subscale [ω_{HS}]) factors (Watkins, 2017). Furthermore, degrees of freedom often do not add up to what is expected based on freely estimated parameters of stated models that suggests undisclosed fixing parameters to not go beyond permissible bounds. These problems cast substantial doubt for the viability of the publisher preferred "confirmatory" model (Beaujean, 2016; Canivez & Watkins, 2016).

German WISC-V

The German adaptation of the U.S. Wechsler Intelligence Scale for Children–Fifth edition (German WISC-V; Wechsler, 2017a), was reported to follow "contemporary intelligence theories, factor analytic studies, and clinical research" (Wechsler, 2017b, p. 15). However, while the U.S. WISC-V explicitly noted CHC theory (Wechsler, 2014c) the German WISC-V does not. Instead, a hierarchical two- or three-stratum intelligence structure (with or without narrow abilities) is assumed without references to specific intelligence theories or models and devoid of bifactor consideration. References to reviews in Flanagan and Harrison (2012) and Sattler (2008a, 2008b) are provided for the German WISC-V for different interpretation models because detailing descriptions of all intelligence theories was reported not to be within the scope of the chapter (Wechsler, 2014c). While a detailed U.S. WISC-V review (Canivez & Watkins, 2016) and several published independent analyses of the U.S. WISC-V were available prior to publication of the German WISC-V (Canivez et al., 2016; Canivez & Watkins, 2016; Dombrowski et al., 2015) none were referenced and it is unknown if they were reviewed or considered by the publisher in developing the German WISC-V. The German WISC-V includes all primary and secondary subtests from the U.S. version, except Picture Concepts, which also was not included in the versions for France and Spain (but *was* included in the Canadian and U.K. versions).

German WISC-V subtests are composed of items retained from the German WISC-IV (Petermann & Petermann, 2011), items adapted and modified from the U.S. WISC-V, and newly developed items. The German WISC-V *Technical Manual* does not provide a rationale for this mixture of kept, adapted, and newly developed items and there is no presentation of equivalence with subtests from the U.S. version. Specific guidelines or standards on which the adaptation and translation process was based were not provided; however, reference to standards applied for the standardization program of the U.S. version, namely the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; Wechsler, 2017b, p. 57) was noted; although a more recent version of the *Standards* was published in 2014 (AERA, APA, & NCME, 2014). Psychometric properties and understanding of instructions were empirically tested in a pilot study prior to the standardization process, but only for the verbal subtests (Similarities [SI], Vocabulary [VC], Information [IN], Comprehension [CO], and AR). Internal consistency (split-half reliability) coefficients based on the standardization sample were high for all German WISC-V subtests and index scores (ranging from .80 for Cancellation [CA] to .96 for the Full Scale IQ [FSIQ]), and short-term test–retest reliabilities with a mean retest interval of 26 days (SD = 19, range: 7-116 days) for a subsample of 94 individuals ranged from .72 (Picture Span [PS]) to .90 (IN), with a stability coefficient of .89 for the FSIQ. Validity evidence reported in the German WISC-V *Technical Manual* includes factorial validity (described in detail below), convergent and discriminant validity, and distinct group differences validity. The subtests and indexes showed medium to strong relationships with corresponding subtests and indexes from the German adaptations of the WISC-IV, WPPSI-III (Wechsler Preschool and Primary Scale of Intelligence–Third edition; Petermann et al., 2014), WAIS-IV (Petermann, 2012), and Kaufman Assessment Battery for Children–Second edition (KABC-II; Kaufman & Kaufman, 2015). Furthermore, a subsample of gifted individuals obtained significantly higher scores on all subtests, except for PS and some WM subtests, and had higher mean index scores compared with matched controls; while a subsample of intellectually disabled individuals obtained significantly lower scores on all subtests and had lower mean index scores and FSIQ compared with matched controls.

Figure 1 presents the publisher preferred structural measurement model for the German WISC-V as a basis for creation of standardized factor scores and interpretation. This is the identical model proffered for the United States, Canadian, the United Kingdom, French, and Spanish versions. The publisher claimed the German WISC-V "enables an estimation of general intelligence which is represented by five cognitive domains" (Wechsler, 2017b, p. 102). This

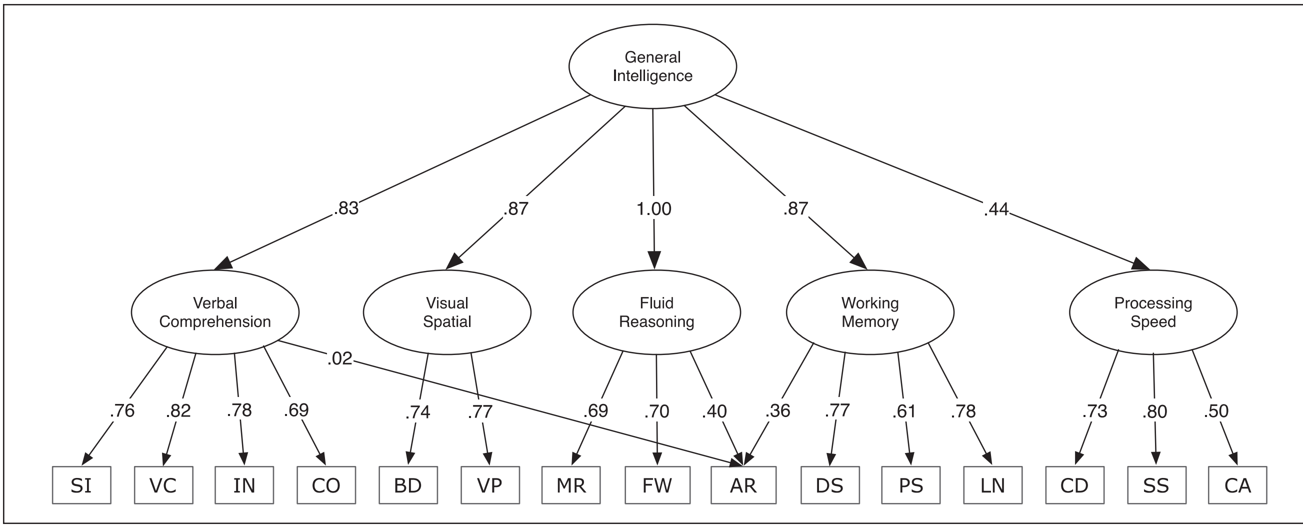

Figure 1. Publisher preferred higher-order measurement Model 5e with standardized coefficients.
Note. SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation.
Source. Adapted from Figure 5.1 (Wechsler, 2017b, p. 107) for the German WISC-V standardization sample (*N* = 1,087).

five-factor model was preferred over the four-factor model based on reported better global fit, but like the U.S. WISC-V, this preferred model includes problems of the standardized path coefficient of 1.0 from the higher-order *g* factor to FR and includes three cross-loadings for the AR subtest (VC [.02], FR [.40], WM [.36]). Pauls et al. (2020) used multi-group confirmatory factor analysis (MGCFA) to examine the latent factor structure invariance of the publisher preferred German WISC-V measurement model across gender [sic] and reported configural, first-order and second-order metric invariance. Unlike the German WISC-V *Technical Manual*, Pauls et al. explicitly used maximum likelihood estimation with the reported normally distributed subtests scores. Oddly, the reported *df* for baseline models was one *higher* than would be expected based on the tested measurement model. Full scalar invariance was not supported, but partial scalar invariance showed subtest intercepts for IN, Figure Weights (FW), Coding (CD), and CA were not invariant across gender [sic], but invariance was observed for the other German WISC-V subtests. Error variances were also reported to be invariant. In contrast to the German WISC-V *Technical Manual*, Pauls et al. (2020) reported decomposed sources of variance in the German WISC-V according to a Schmid and Leiman (1957) orthogonalized higher-order model and found that while *g* had ample unique variance as reflected by high ω_H (.798) and construct replicability index (H = .896) values, the five group factors did not contain minimally acceptable unique variance. This led Pauls et al. (2020) to conclude primary interpretation of the FSIQ as an estimate of *g* and cautious interpretation of factor index scores, if at all.

German WISC-V Concerns

The same major concerns and shortcomings observed in the U.S. WISC-V reported by Canivez and Watkins (2016) were also observed in the German WISC-V *Technical Manual* (Wechsler, 2017b). EFA were not reported, opting for exclusive use of CFA despite the complementary nature of EFA and CFA (Brown, 2015; Gorsuch, 1983; Kline, 2016). Also, given the adaptations and modifications of subtests, creation of new item content, proposed change in factor structure (separation of VS and FR), and a new standardization sample; it cannot be assumed that the factor structure would be unchanged and thus appropriate and necessary to use EFA to inform plausible CFA models to test. The method of estimation was not disclosed nor was the method for setting scales. While maximum likelihood estimation would customarily be used with tests like the German WISC-V given continuous variables and reasonably normally distributed data, WLS estimation was used with other WISC-V versions (Wechsler, 2014b, 2014c, 2015a, 2015b, 2016a, 2016c) but without specific justification (see Canivez & Watkins, 2016). There was no report of the estimator used in German WISC-V CFA and model comparisons relied solely on the χ^2 difference despite reporting Akaike information criterion (AIC) and Bayesian information criterion (BIC) estimates. It was stated in the German WISC-V *Technical Manual* that nested models can be compared using the χ^2 difference test, but "when models are not nested, change in fit is assessed through subjective evaluation rather than statistical comparisons in model fit" (p. 105). More troubling is the continued observation that there are fewer degrees of freedom reported than expected, that is, they are not consistent with the number of freely estimated parameters suggested by specified models. This suggests that parameters may have been fixed (without disclosure) to allow model estimation by not allowing parameters to go beyond permissible bounds as was apparently done with the U.S. WISC-V (Beaujean, 2016). This calls into question reported global model fit indexes as "supportive." Additionally, bifactor measurement models were apparently disregarded and variance estimates for contributions of first- and second-order factors remain absent. Bifactor models have several advantages: (a) direct influences of the general factor are more easily interpretable, (b) influences of both general *and* specific factors on indicators (subtests) are simultaneously examined, and (c) the psychometric properties necessary for determining scoring and interpretation of subscales (i.e., ω_H and ω_{HS} estimations) are directly examined (Canivez, 2016; Cucina & Byle, 2017; Reise, 2012). Gignac (2006) also noted that the direct hierarchical (i.e., bifactor) model can be considered more parsimonious because it specifies a unidimensional general factor. Furthermore, a major local fit problem—a standardized path coefficient of 1.0 between higher-order *g* and the FR group factor—was dismissed as a common finding in current studies on intelligence tests, without further discussion. Another German WISC-V local fit problem included the standardized path coefficient of .02 from VC to AR, which was not addressed at all in the manual. Finally, no statistical significances are reported in the German WISC-V *Technical Manual* for any of the parameters from the final model, thus hampering the examination of local fit problems.

WISC-V Research

Exploratory Factor Analyses. While EFAs were not reported in the WISC-V *Technical and Interpretive Manual* (Wechsler, 2014c), best practices (Watkins, 2018) applied in independent EFA of the U.S. WISC-V did not support the existence of five group factors in the total WISC-V standardization sample (Canivez et al., 2016) or in four age groups (6-8, 9-11, 12-14, and 15-16 years) within the WISC-V standardization sample (Canivez, Dombrowski, et al., 2018; Dombrowski, Canivez, et al., 2018), as the fifth extracted factor included only one salient subtest loading. Instead, a four-factor solution consistent with the WISC-IV was found to best represent the standardization data. Schmid and Leiman (1957) orthogonalization of the second-order EFA for the total U.S. WISC-V standardization sample and the four age groups found substantial portions of variance apportioned to *g* and substantially smaller portions of variance apportioned to the group factors (VC, PR, WM, and PS). ω_H coefficients for *g* (Reise, 2012; Rodriguez et al., 2016) ranged from .817 (Canivez et al., 2016) to .847 (Dombrowski, Canivez, et al., 2018) and exceeded the preferred level (.75) for clinical interpretation (Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). ω_{HS} coefficients for the four U.S. WISC-V group factors (Reise, 2012) ranged from .131 to .530, but no ω_{HS} coefficients for VC, PR, or WM approached or exceeded the minimum criterion (.50) for clinical interpretation (Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). ω_{HS} coefficients for PS, however, approached or exceeded the .50 criterion for possible clinical interpretation. Dombrowski et al. (2015) also failed to find support for five-factors in the total U.S. WISC-V standardization sample using exploratory bifactor analysis through the bifactor rotation criterion (Jennrich & Bentler, 2011). Furthermore, EFA did not support five group factors with a large U.S. clinical sample (Canivez, McGill, et al., 2018). Recent independent research with the French WISC-V (Wechsler, 2016a) and the WISC-V U.K. edition (WISC-V^{UK}; Wechsler, 2016b) found identical EFA results supporting four first-order factors (not five), dominant general intelligence, and poor unique measurement of the four group factors (Canivez et al., 2019; Lecerf & Canivez, 2018).

Confirmatory Factor Analyses. Independent CFA conducted with the 16 U.S. WISC-V primary and secondary subtests (Canivez, Watkins, & Dombrowski, 2017) found all five of the higher-order models that included five first-order group factors (including the final U.S. WISC-V measurement model presented in the U.S. WISC-V *Technical and Interpretative Manual* as the preferred model) resulted in inadmissible solutions (i.e., negative variance estimates for the FR factor) potentially caused by misspecification of the models. A bifactor model that included five first-order factors produced an admissible solution and fit the standardization data well, but examination of local fit indicated problems where the Matrix Reasoning (MR), FW, and Picture Concepts subtests did not have statistically significant loadings on the FR group factor. The bifactor model with four group factors (VC, PR, WM, and PS) was selected as best based on the combination of statistical fit, local fit, and theory. This was consistent with previous EFA results (Canivez et al., 2016) showing a dominant general intelligence dimension and weak group factors with limited unique measurement beyond *g*. One study, however, (H. Chen et al., 2015) reported factorial invariance of the final publisher preferred WISC-V higher-order model with five group factors across gender [sic], although they did not examine invariance for alternative rival higher-order or bifactor models.

Reynolds and Keith (2017) suggested U.S. WISC-V invariance across age groups, but the model they examined for invariance was an oblique five-factor model, which ignores general intelligence altogether. Then they used CFA to explore numerous post hoc model modifications for first-order models with five-factors and then for both higher-order and bifactor models with five group factors in an attempt to better understand U.S. WISC-V measurement. While such explorations are possible, they may capitalize on chance and sample size. The final best fitting U.S. WISC-V higher-order model produced by Reynolds and Keith was different from the publisher preferred model in that AR was given a direct loading from *g and* a "cross-loading" on WM, and they also added correlated disturbances for the VS and FR group factors (.77) to represent an intermediate nonverbal general reasoning factor between the broad abilities and *g*. Yet the model still produced a standardized path coefficient of .97 from *g* to FR suggesting inadequate discriminant validity. Another concern was reliance on statistically significant χ^2 difference tests for model improvement despite the large sample and multiple comparisons but no meaningful changes in global fit. Researchers preferring higher-order Wechsler scale structures often introduce *post hoc* cross-loadings and correlated disturbance and error terms in altered CFA models; however, such procedures may capitalize on chance and sample size (MacCallum et al., 1992; Schreiber et al., 2006; Ullman, 2001) and it is rare when such parameters are specified *a priori*. Typically, previously unmodelled complexities are *post hoc* model adjustments iteratively added to improve model fit and/or remedy local fit problems, but specification of such parameters may be problematic due to lack of conceptual grounding in previous theoretical work, may be unlikely to replicate, and increase the dangers of hypothesizing after results are known (HARKing) as noted by Cucina and Byle (2017). Preregistration would help address this potential problem. Furthermore, decomposed variance estimates of the Reynolds and Keith higher-order model showed the U.S. WISC-V subtests primarily reflected general intelligence variance with small portions of unique group factor variance (except for the PS subtests). Their best U.S. WISC-V bifactor model included a covariance estimate between VS and FR (.62), which appears necessary in order to salvage five group factors that EFA (Canivez et al., 2016) failed to locate. Watkins et al. (2018) also tested a similar bifactor model with the Canadian WISC-V (WISC-V^{CDN}), but this bifactor model with five group factors and VS–FR covariance estimate was *not* superior to the bifactor model with four group factors, so the Wechsler-based bifactor model with four group factors (VC, PR, WM, and PS) was preferred and Reynolds and Keith's findings failed replication. Independent research regarding the factor structure of international versions of the WISC-V replicated both EFA and CFA findings yielded by independent assessments of the U.S. WISC-V version (cf. Canivez et al., 2016; Canivez, Watkins, & Dombrowski, 2017), all failing to support five group factors (Canivez et al., 2019; Fenollar-Cortés & Watkins, 2019; Lecerf & Canivez, 2018; Watkins et al., 2018).

Higher-Order Versus Bifactor Models

Publisher references to Carroll's (1993) three stratum theory are provided in WISC-V technical manuals, but repeatedly fail to report EFA findings and decomposed variance estimates using the Schmid and Leiman transformation (SLT; Schmid & Leiman, 1957) which Carroll (1995) insisted on; or use more recently developed exploratory bifactor analysis (Jennrich & Bentler, 2011, 2012). SLT (sometimes referred to as an approximate bifactor solution; Reise, 2012) of EFA loadings apportion subtest variance to the first-order *and* higher-order dimensions because intelligence test subtests are influenced by both first-order factors *and* the higher-order *g* factor in a higher-order model. Interpretation of higher-order models requires this partitioning of variance in EFA, as well as CFA, so the relative influence of the first-order factors in comparison with the higher-order factor(s) may be determined. However, the SLT is just a reparameterization of the higher-order (second-order) model and may not be equivalent to a bifactor model (Beaujean, 2015b).

Higher-order representation of intelligence test structure is an *indirect* hierarchical model (Gignac, 2005, 2006, 2008) where the *g* factor influences subtests indirectly through full mediation through the first-order factors (Yung et al., 1999). The higher-order model conceptualizes *g* as a *superordinate* factor and an abstraction from abstractions (Thompson, 2004). While higher-order models have been commonly applied to assess intelligence test "construct-relevant psychometric multidimensionality" (Morin, Arens, & Marsh, 2016, p. 117), the bifactor model predated widespread use of higher-order models and was originally specified by Holzinger and Swineford (1937) and 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 *g* and the group factors directly and independently influence the subtest indicators. This means that both *g* and first-order group factors are at the same level of inference constituting a less complicated (more parsimonious) conceptual model (Cucina & Byle, 2017; Gignac, 2008). Carroll (1993) and his three stratum theory appear to reflect bifactor intelligence structure (Beaujean, 2015b) and there are major theoretical differences between higher-order and bifactor models. In the higher-order model, *g* is what the broad first-order factors have in common, whereas in the bifactor model, *g* is what is common among a diverse set of tasks or indicators which is how Spearman and Carroll thought of *g* (Beaujean, 2019). Cucina and Byle (2017) illustrated superiority of bifactor representations among a variety of cognitive tests and, given such results and the advantages of bifactor modeling for understanding test structure (Canivez, 2016; Cucina & Byle, 2017; Gignac, 2008; Reise, 2012), bifactor model comparisons should be routinely examined in addition to higher-order models for structural validation of cognitive tests.

Purpose

Understanding the structural validity of tests is crucial for evaluating interpretability of provided scores (AERA, APA, & NCME, 2014) and the German WISC-V *Technical Manual* lacks sufficient and necessary information regarding evidence of the German WISC-V structure to properly interpret test results. Numerous unanswered questions and incomplete information regarding the German WISC-V structure prohibits users of the German WISC-V to exercise good judgment about which scores have acceptable evidence of construct validity. Beaujean (2015a) indicated that a revised test should be treated like a new test as it cannot be assumed that scores from the revision would be directly comparable to the previous version without supporting evidence. Given the absence of EFA, questionable CFA methods identified in the German WISC-V *Technical Manual* (Wechsler, 2017b), and the lack of details regarding structural validity evidence, the present study (a) used best practices in EFA (Watkins, 2018) to examine the German WISC-V factor structure suggested by the 15 primary and secondary subtest relationships; (b) examined the German WISC-V factor structure using CFA with customary maximum likelihood estimation; (c) compared bifactor models with higher-order models as rival explanations; (d) decomposed factor variance sources in EFA and CFA; and (e) examined model-based reliability/validity (Watkins, 2017). Answers to these questions are essential for users of the German WISC-V to determine the interpretive value of the plethora of scores and score comparisons provided in the German WISC-V and interpretive guidelines promulgated by the publisher.

Method

Participants

To conduct independent EFA and CFA with the German WISC-V, standardization sample raw data were requested from the publisher (NCS Pearson, Inc.) but access was denied without rationale. Absent raw data, the present analyses required use of summary statistics (correlations, means, and standard deviations) provided in the German WISC-V *Technical Manual* (Wechsler, 2017b, Table 5.1, pp. 96-97). The published correlation matrix includes correlations rounded to only 2 decimals, but Carroll (1993) stated, “Little precision is lost by using two-decimal values” (p. 82). These correlations were reportedly produced by participants who were members of the full German WISC-V standardization sample (N = 1,087) who ranged in age from 6 to 16 years. The sample was stratified according to the key variables indicated by the Federal Statistical Office of the Federal Republic of Germany (2014): age, sex, migration background, parental education (age groups 6-9 years, four education levels), and children’s school level (age groups 10-16 years, five school levels). Institutional review board review and approval of methods were obtained by the first author although no data were directly collected in this study.

Instrument

The German WISC-V (Wechsler, 2017a) is a general intelligence test composed of 15 subtests with scaled scores (M = 10, SD = 3). Like the United States and other versions there are 10 primary subtests (SI, VC, Block Design [BD], Visual Puzzles [VP], MR, FW, Digit Span [DS], PS, CD, Symbol Search [SS]) that are used for the measurement of five factor-based Primary Index scales: Verbal Comprehension Index (VCI), Visual Spatial Index (VSI), Fluid Reasoning Index (FRI), Working Memory Index (WMI), and Processing Speed Index (PSI). Seven of the primary subtests are used to produce the FSIQ. Ancillary index scales (pseudo-composites) are provided and include Quantitative Reasoning Index (QRI), Auditory Working Memory Index (AWMI), Nonverbal Index (NVI), General Ability Index (GAI), and Cognitive Proficiency Index (CPI), but are *not* factorially derived. Index scores and FSIQ are standard score (M = 100, SD = 15) metrics. Secondary subtests (IN, CO, AR, Letter–Number Sequencing [LN], CA) are used for substitution in FSIQ estimation when one subtest is spoiled or for use in estimating newly created (QR, AWM, and NV) or previously existing (General Ability and Cognitive Proficiency) Ancillary index scores. Ancillary index scores are not factorially derived composite scores, but logically or theoretically constructed. Picture Concepts, a subtest present in the U.S. WISC-V (and Canadian and U.K. versions) was not included in the German WISC-V.

Analyses

Exploratory Factor Analyses. The 15 German WISC-V primary and secondary subtest correlation matrix was used to conduct EFAs. Several criteria were examined and compared for their recommendation of the number of factors that might be extracted and retained (Gorsuch, 1983) and included 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 (see Figure 2), and minimum average partials (MAP; Frazier & Youngstrom, 2007; Velicer, 1976). Statistics were estimated with SPSS 24 for Macintosh or with specific software where noted. The Watkins (2007) SE_{scree} program was used as SE_{scree} reportedly is the most accurate objective scree method (Nasser et al., 2002).

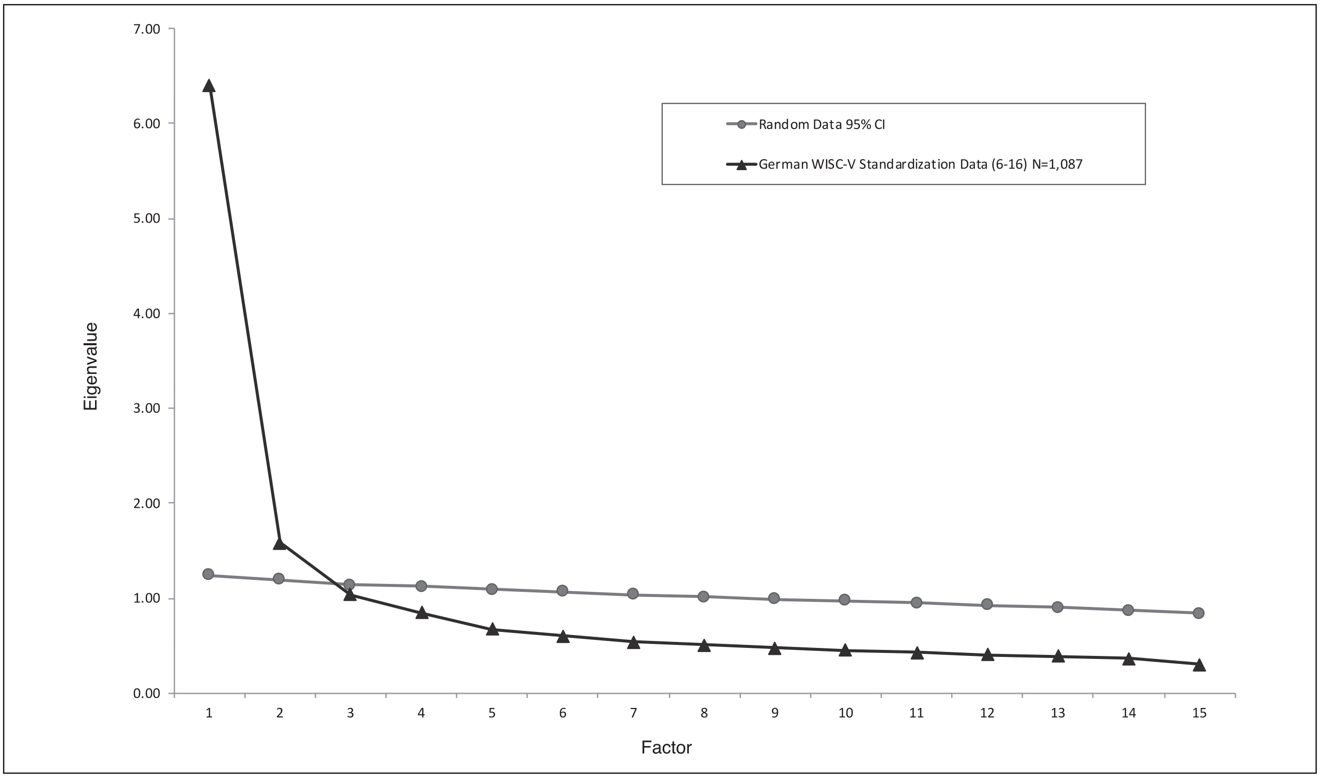

Figure 2. Scree plots for Horn's parallel analysis for the German WISC-V standardization sample (N = 1,087).

Random data and resulting eigenvalues for PA using both mean and 95% confidence intervals (Glorfeld, 1995) were produced using the O'Connor (2000) SPSS syntax with 100 replications to provide stable eigenvalue estimates. PA frequently suggests retaining too few factors (underextraction) in the presence of a strong general factor (Crawford et al., 2010) so it was not the exclusive criterion. MAP was also conducted using the O'Connor (2000) SPSS syntax.

Principal axis EFAs were conducted to analyze the factor structure of the German WISC-V using SPSS 24 for Macintosh. Retained factors were obliquely rotated with promax ($k = 4$; Gorsuch, 1983) and viable factors required a minimum of two subtests with salient factor pattern coefficients ($\geq$.30; Child, 2006). Because the German WISC-V explicitly adopted a higher-order structure, the SLT (Schmid & Leiman, 1957) procedure was applied to disentangle the contribution of 1st and 2nd order factors, as advocated by Carroll (1993) and Gignac (2005). The SLT has been used in numerous EFA studies with the WISC-IV (Watkins, 2006; Watkins et al., 2006), WISC-V (Canivez, Dombrowski, et al., 2018; Canivez et al., 2016; Dombrowski, Canivez, et al., 2018); RIAS (Dombrowski et al., 2009; Nelson et al., 2007), Wechsler Abbreviated Scale of Intelligence (WASI) and Wide Range Intelligence Test (WRIT; Canivez et al., 2009), SB5 (Canivez, 2008), WISC-IV Spanish (McGill & Canivez, 2016), French WAIS-III (Golay & Lecerf, 2011), French WISC-IV (Lecerf et al., 2010), French WISC-V (Lecerf & Canivez, 2018), and WISC-V^{UK} (Canivez et al., 2019). The SLT allows for deriving a hierarchical factor model from higher-order models and decomposes the variance of each subtest score into general factor variance first and then first-order factor variance. The first-order factors are modeled orthogonally to each other and to the general factor (Gignac, 2006; Gorsuch, 1983). The SLT was produced using the *MacOrtho* program (Watkins, 2004). This procedure disentangles the common variance explained by the general factor and the residual common variance explained by the first-order factors.

Confirmatory Factor Analyses. CFA with maximum likelihood estimation was conducted using EQS 6.3 (Bentler & Wu, 2016). Covariance matrices were reproduced for CFA using the correlation matrix, means, and standard deviations obtained from the German WISC-V standardization sample. As with other similar studies (e.g., Canivez, Watkins, & Dombrowski, 2017; Watkins et al., 2018) identification of latent variable scales set a reference indicator to 1.0 in higher-order models and in bifactor models, by setting the variance of latent variables to 1.0 (Brown, 2015; Byrne, 2006). As with other versions of the WISC-V, the VS factor and FR factor are underidentified in some of the five-factor models because they are measured by only two subtests (BD and VP, MR, and FW). Thus, in specifying the VS factor and FR factor (in some five-factor models) in

Table 1. German WISC–V Primary and Secondary Subtest Configuration for CFA Models With 1 to 4 Factors.

	Model 1	Model 2		Model 3			Model 4a				Model 4a Bi-factor					Model 4b				Model 4c				Model 4d			
Subtest	g	F1	F2	F1	F2	F3	F1	F2	F3	F4	g	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
SI	•	•		•			•				•	•				•				•				•			
VC	•	•		•			•				•	•				•				•				•			
IN	•	•		•			•				•	•				•				•				•			
CO	•	•		•			•				•	•				•				•				•			
BD	•		•		•			•			•		•				•				•				•		
VP	•		•		•			•			•		•				•				•				•		
MR	•		•		•			•			•		•					•			•				•		
FW	•		•		•			•			•		•					•			•				•		
AR	•	•		•					•		•			•				•			•	•		•	•	•	
DS	•	•		•					•		•			•				•				•				•	
PS	•		•		•				•		•			•				•				•				•	
LN	•	•		•					•		•			•				•				•				•	
CD	•		•			•				•	•				•				•				•				•
SS	•		•			•				•	•				•				•				•				•
CA	•		•			•				•	•				•				•				•				•

Note. All models include a higher-order general factor except for the bifactor model. WISC-V = Wechsler Intelligence Scale for Children–Fifth edition; CFA = confirmatory factor analysis; SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation.

CFA bifactor models, the two subtests' path coefficients on their group factors were constrained to equality prior to estimation to ensure identification (Little et al., 1999).

The structural models specified in Table 5.2 of the German WISC-V *Technical Manual* (Wechsler, 2017b) were also examined in present CFA analyses and are reproduced in Table 1 and Table 2 with the addition of alternative bifactor models that were not included in analyses reported in the German WISC-V *Technical Manual*. Although there are no universally accepted cutoff values for approximate fit indices (McDonald, 2010), overall global model fit was evaluated using the comparative fit index (CFI), Tucker–Lewis index (TLI), standardized root mean squared residual (SRMR), and the root mean square error of approximation (RMSEA). Higher CFI and TLI values indicate better fit whereas lower SRMR and RMSEA values indicate better fit. Hu and Bentler (1999) combinatorial heuristics indicated adequate model fit with CFI and TLI $\geq .90$ along with SRMR $\leq .09$ and RMSEA $\leq .08$. Good model fit required CFI and TLI ≥ 0.95 with SRMR and RMSEA ≤ 0.06 (Hu & Bentler, 1999). Additionally, the AIC was considered. AIC does not have a meaningful scale; the model with the smallest AIC value is most likely to replicate (Kline, 2016) and would be preferred. Superior model fit required adequate to good overall fit *and* display of meaningfully better fit. Meaningful differences between well-fitting models were assessed using ΔCFI $> .01$ and ΔRMSEA $> .015$ (F. F. Chen, 2007; Cheung & Rensvold, 2002) and ΔAIC $>$ 10 (Burnham & Anderson, 2004). In addition to assessing global fit, local fit assessment was conducted as models should never be retained "solely on global fit testing" (Kline, 2016, p. 461).

Model-based reliabilities/validities were estimated with coefficients ω_H and ω_{HS}, which estimate reliability of unit-weighted scores produced by the indicators (Reise, 2012; Rodriguez et al., 2016). ω_H is the model-based reliability estimate for the general intelligence factor with variability of group factors removed. ω_{HS} is the model-based reliability estimate of a group factor with all other group *and* general factors removed (Brunner et al., 2012; Reise, 2012). Omega estimates (ω_H and ω_{HS}) may be obtained from CFA bifactor solutions or decomposed variance estimates from higher-order models and were produced using the *Omega* program (Watkins, 2013), which is based on the tutorial by Brunner et al. (2012) and the work of Zinbarg et al. (2005) and Zinbarg et al. (2006). Omega coefficients should at a minimum exceed .50, but .75 is preferred (Reise, 2012; Reise et al., 2013).

Omega coefficients were supplemented with the *H* coefficient (Hancock & Mueller, 2001), a construct reliability or construct replicability coefficient, and the correlation between a factor and an optimally weighted composite score. *H* represents how well the latent factor is represented by the indicators and a criterion value of .70 (Hancock & Mueller, 2001; Rodriguez et al., 2016) was used. *H* coefficients were produced by the *Omega* program (Watkins, 2013).

Table 2. German WISC-V Primary and Secondary Subtest Configurations for CFA Models With Five Factors.

	Model 5a					Model 5a Bifactor						Model 5b					Model 5c					Model 5d					Model 5e				
Subtest	F1	F2	F3	F4	F5	g	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
SI	•					•	•					•					•					•					•				
VC	•					•	•					•					•					•					•				
IN	•					•	•					•					•					•					•				
CO	•					•	•					•					•					•					•				
BD		•				•		•					•					•					•					•			
VP		•				•		•					•					•					•					•			
MR			•			•			•					•					•					•					•		
FW			•			•			•					•					•					•					•		
AR				•		•				•				•					•	•		•			•		•		•	•	
DS				•		•				•					•					•					•					•	
PS				•		•				•					•					•					•					•	
LN				•		•				•					•					•					•					•	
CD					•	•					•					•					•					•					•
SS					•	•					•					•					•					•					•
CA					•	•					•					•					•					•					•

Note. All models include a higher-order general factor except for the bifactor model. WISC-V = Wechsler Intelligence Scale for Children–Fifth edition; CFA = confirmatory factor analysis; SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation.

Results

Exploratory Factor Analyses

The Kaiser–Meyer–Olkin Measure of Sampling Adequacy of .931 far exceeded the .60 minimum standard (Kaiser, 1974; Tabachnick & Fidell, 2007) and Bartlett's Test of Sphericity (Bartlett, 1954), $\chi^2 = 6{,}987.57$, $p < .0001$; indicated that the German WISC-V correlation matrix was not random. The correlation matrix was thus deemed appropriate for factor analysis. Without standardization sample raw data, it was not possible to estimate univariate subtest skewness or kurtosis or multivariate normality, but principal axis extraction does not require normality. While univariate and multivariate skewness and kurtosis were not reported in the German WISC-V *Technical Manual* (Wechsler, 2017b), Pauls et al. (2020) reported reasonably normally distributed subtest scores for the 15 German WISC-V subtests within the two gender [sic] groups based on univariate estimates (Male sample skewness ranged −.39 to .12, kurtosis ranged −.35 to .60, Female sample skewness ranged −.34 to .09 and kurtosis ranged −.26 to .71); however, multivariate estimates were not provided.

Figure 2 illustrates the scree plots from Horn's parallel analysis for the German WISC-V total standardization sample. Scree, PA, and Glorfeld's modified PA criteria suggested two factors, while eigenvalues > 1 and SE_{scree} criteria suggested 3 factors. The MAP criterion suggested only one factor. In contrast, the German WISC-V publisher desired and claimed five latent factors. EFA began by extracting five factors to examine subtest associations based on the publisher's desired and promoted structure to allow examination of the performance of smaller factors because Wood et al. (1996) noted that it is better to overextract than underextract. Models with four, three, and two factors were subsequently examined for adequacy.

Results of a five-factor extraction with promax rotation presented in Table 3 include a fifth factor with only one salient factor pattern coefficient (SI). This extraction and rotation also produced Factor 1 (WM) that included salient pattern coefficients for theoretically related subtests (AR, DS, PS, LN) but also included salient pattern coefficients for MR and FW. Factor 2 (VC) included salient pattern coefficients for VC, IN, and CO. Factor 3 (VS [formerly PR]) included salient pattern coefficients for BD, VP, and MR. However, MR also cross-loaded on Factor 1 (WM) which indicated a lack of simple structure. Factor 4 (PS) included salient subtest pattern coefficients by the theoretically consistent subtests (CD, SS, and CA). Thus, MR and FW did not share sufficient common variance to constitute a FR dimension as specified by the publisher. This pattern of psychometrically unsatisfactory results is indicative of overextraction (Gorsuch, 1983; Wood et al., 1996) and the five-factor model was judged inadequate.

Table 4 presents the results of extracting four factors with promax rotation. The *g* loadings (first unrotated factor structure coefficients) ranged from .287 (CA) to .744 (VC) and—except CD, SS, and CA—were within the fair to good range based on Kaufman's (1994) criteria (≥.70 = good,

Table 3. German Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) Exploratory Factor Analysis: Five Oblique Factor Solution for the Total Standardization Sample (N = 1,087).

German WISC-V	General	F1: Working Memory		F2: Verbal Comprehension		F3: Visual Spatial		F4: Processing Speed		F5: Inadequate		
Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
SI	.749	.020	.611	.173	.708	.013	.633	−.004	.275	**.736**	.877	.787
VC	.741	−.030	.578	**.821**	.856	.026	.581	−.015	.276	.063	.630	.736
IN	.716	.056	.592	**.526**	.739	.186	.622	−.057	.239	.103	.614	.587
CO	.633	.069	.508	**.822**	.752	−.137	.434	.098	.327	−.084	.469	.586
BD	.645	−.009	.543	−.079	.462	**.713**	.732	.073	.325	.076	.514	.544
VP	.668	.018	.569	.022	.502	**.780**	.764	.013	.293	−.072	.481	.586
MR	.640	*.341*	.622	−.006	.471	**.455**	.651	−.036	.247	−.059	.448	.471
FW	.666	**.323**	.631	.158	.559	.277	.619	−.050	.239	.025	.512	.464
AR	.719	**.532**	.724	.065	.559	.157	.625	.045	.341	.022	.522	.546
DS	.688	**.832**	.777	−.067	.489	−.005	.562	−.037	.272	.013	.479	.607
PS	.579	**.507**	.601	.029	.447	.024	.476	.031	.270	.070	.436	.368
LN	.698	**.834**	.779	.073	.536	−.082	.538	−.006	.300	−.065	.458	.614
CD	.400	.123	.338	−.077	.243	−.097	.263	**.727**	.737	.080	.245	.553
SS	.438	−.014	.335	.001	.285	.085	.337	**.772**	.789	−.031	.236	.626
CA	.287	−.141	.183	.143	.235	.082	.221	**.509**	.516	−.054	.155	.280
Eigenvalue		6.41		1.58		1.04		.85		.67		
% Variance		39.91		7.37		4.33		2.71		1.37		
Factor correlations		F1: WM		F2: VC		F3: VS		F4: PS		F5		
Working Memory (WM)		—										
Verbal Comprehension (VC)		.676		—								
Visual Spatial (VS)		.740		.658		—						
Processing Speed (PS)		.396		.336		.366		—				
F5		.633		.698		.670		.283		—		

Note. General structure coefficients are based on the first unrotated factor coefficients (g loadings). Salient pattern coefficients (≥.30) presented in bold. German WISC-V Subtests: SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LN = Letter–Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation; *S* = Structure Coefficient; *P* = Pattern Coefficient; h^2 = Communality.

.50-.69 = fair, <.50 = poor). Table 4 illustrates robust VC (SI, VC, IN, and CO) and PS (CD, SS, and CA) factors with theoretically consistent subtest associations. The WM factor included the four theoretically related subtests (AR, DS, PS, and LN) but also included salient pattern coefficients of MR and FW. The VS (formerly PR) factor included salient pattern coefficients of BD, VP, and MR, but FW did not have a salient loading on this factor. Overall the four-factor model resembled WISC-IV structure but was not a perfect match. MR had primary loading on the VS factor but cross-loaded on WM. The moderate to high factor correlations presented in Table 4 (.341 to .747) suggested the presence of a general intelligence factor (Gorsuch, 1983) requiring explication.

Table 5 presents results from the three- and two-factor extractions with promax rotation. For the three-factor model, the VS/PR and WM factors merged, leaving fairly distinct VC and PS factors. Oddly, SI cross-loaded on the VS/PR/WM factor. The two-factor model showed merging of VC, VS/PR, and WM factors, leaving only the separate PS factor. No subtest cross-loadings were observed in the two-factor model. The two- and three-factor models clearly displayed fusion of potentially theoretically meaningful constructs that is symptomatic of underextraction, thereby rendering them unsatisfactory (Gorsuch, 1983; Wood et al., 1996).

Because the four-factor EFA solution appeared to be the most reasonable it was subsequently subjected to second-order EFA and results transformed with the SLT procedure (see Table 6). Following SLT, all German WISC-V subtests were properly associated with their theoretically proposed factors (Wechsler model), except for FW, which had residual variance approximately evenly split between the WM factor and VS factor. The hierarchical g factor accounted for 35.1% of the total variance and 65.1% of the common variance. The general factor also accounted for between 6.0% (CA) and 47.1% (AR) of individual subtest variability.

At the group factor level, WM accounted for an additional 3.4%, VC for an additional 5.1%, VS for an additional 2.8%, and PS for an additional 7.5% of the total variance. Of the common variance, WM accounted for an additional

Table 4. German Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) Exploratory Factor Analysis: Four Oblique Factor Solution for the Total Standardization Sample (N = 1,087).

	General	F1: Working Memory		F2: Verbal Comprehension		F3: Visual Spatial		F4: Processing Speed		
German WISC-V Subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
Similarities	.723	.090	.609	**.522**	.729	.215	.639	−.018	.273	.568
Vocabulary	.744	−.051	.575	**.919**	.868	−.018	.579	−.011	.276	.756
Information	.718	.041	.590	**.621**	.758	.184	.625	−.057	.238	.596
Comprehension	.630	.065	.506	**.778**	.725	−.196	.433	.108	.327	.547
Block Design	.648	−.029	.542	−.041	.495	**.768**	.742	.068	.322	.556
Visual Puzzles	.668	.017	.570	−.007	.523	**.738**	.752	.017	.290	.566
Matrix Reasoning	.641	*.332*	.622	−.034	.491	**.438**	.650	−.033	.245	.466
Figure Weights	.668	**.314**	.630	.186	.580	.276	.621	−.050	.237	.465
Arithmetic	.720	**.526**	.724	.085	.580	.161	.629	.045	.339	.547
Digit Span	.690	**.832**	.777	−.057	.512	.000	.568	−.038	.269	.607
Picture Span	.580	**.506**	.601	.074	.468	.043	.483	.029	.268	.367
Letter–Number Sequencing	.699	**.826**	.778	.045	.549	−.106	.541	−.002	.299	.609
Coding	.400	.126	.337	−.036	.254	−.060	.270	**.715**	.731	.540
Symbol Search	.439	−.014	.335	−.020	.289	.078	.337	**.776**	.792	.631
Cancellation	.287	−.141	.183	.114	.230	.058	.218	**.511**	.516	.277
Eigenvalue		6.41		1.58		1.04		.85		
% Variance		39.77		7.33		4.23		2.66		
Promax-based factor correlations		F1: WM		F2: VC		F3: VS		F4: PS		
F1: Working Memory (WM)		—								
F2: Verbal Comprehension (VC)		.700		—						
F3: Visual Spatial (VS)		.747		.695		—				
F4: Processing Speed (PS)		.392		.341		.364		—		

Note. General structure coefficients are based on the first unrotated factor coefficients (g loadings). Salient pattern coefficients (≥.30) presented in bold. Italic type denotes salient cross-loading. S = Structure Coefficient; P = Pattern Coefficient; h^2 = Communality.

6.3%, VC for an additional 9.5%, VS for an additional 5.2%, and PS for an additional 13.9%. The general and group factors combined to measure 54.0% of the total variance in German WISC-V scores, leaving 46.0% unique variance (combination of specific and error variance).

ω_H and ω_{HS} coefficients were estimated based on the SLT results and presented in Table 6, assigning FW to the VS factor. The ω_H coefficient for general intelligence (.823) was high and sufficient for scale interpretation of a unit-weighted composite score based on the indicators; however, the ω_{HS} coefficients for the four German WISC-V group factors (WM, VC, VS, and PS) were considerably lower (.135-.562). Thus, unit-weighted composite scores based on all subtest indicators of the four German WISC-V group factors, likely possess too little true score variance for confident clinical interpretation, with the possible exception of PS (Reise, 2012; Reise et al., 2013). ω_H and ω_{HS} were also estimated with FW assigned to the WM factor (see Table 6) and coefficients showed a slight decrease in ω_{HS} for WM but a slight increase for VS, but still well below the .50 criterion. H indexes indicated an optimally weighted composite score for g accounted for 90.7% of g variance but WM, VC, and VS group factors were not well defined by their optimally weighted indicators (Hs < .70). The H index of .749 for PS indicated that it was well defined by optimal weighting of its three indicators.

Confirmatory Factor Analyses

Global Fit. Results from CFAs for the 15 German WISC-V primary and secondary subtests are presented in Table 7. Combinatorial heuristics of Hu and Bentler (1999) indicated that Models 1 (g) and 2 (Verbal and Performance) were inadequate due to too low CFI and TLI and too high SRMR and RMSEA values. Model 3 was adequate but all models (higher-order and bifactor) that included four or five group factors produced global fit statistics that indicated good model fit to these data, better than one-, two-, or three-factor models. Bifactor versions of models with four and five group factors where AR was not cross-loaded were meaningfully better than higher-order versions in CFI and AIC, but meaningful differences in RMSEA were observed only for Model 4a bifactor and the EFA suggested bifactor compared with the higher-order version. All bifactor models were superior to higher-order versions (ΔAIC > 10) and thus more likely to replicate.

Table 5. German Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) Exploratory Factor Analysis: Two and Three Oblique Factor Solutions for the Total Standardization Sample (N = 1,087).

German WISC-V	Two oblique factors				Three oblique factors				
Subtest	g^a	F1: g	F2: PS	h^2	g^a	F1: PR/WM	F2: VC	F3: PS	h^2
SI	.725	**.751** (.735)	−.036 (.295)	.541	.725	**.304** (.667)	**.509** (.724)	−.018 (.288)	.566
VC	.722	**.748** (.732)	−.035 (.294)	.537	.748	−.035 (.620)	**.904** (.875)	−.009 (.289)	.766
IN	.716	**.765** (.731)	−.077 (.260)	.538	.720	.233 (.650)	**.604** (.755)	−.056 (.252)	.593
CO	.616	**.583** (.613)	.069 (.326)	.380	.629	−.051 (.515)	**.712** (.713)	.108 (.333)	.517
BD	.636	**.595** (.632)	.083 (.346)	.405	.634	**.584** (.644)	.041 (.491)	.071 (.335)	.420
VP	.658	**.643** (.659)	.037 (.320)	.435	.656	**.613** (.669)	.062 (.518)	.025 (.308)	.449
MR	.640	**.653** (.647)	−.012 (.275)	.419	.642	**.720** (.679)	−.036 (.477)	−.035 (.260)	.463
FW	.673	**.706** (.684)	−.049 (.262)	.470	.670	**.577** (.675)	.166 (.568)	−.053 (.251)	.470
AR	.723	**.697** (.722)	.057 (.364)	.524	.722	**.686** (.738)	.047 (.562)	.042 (.351)	.547
DS	.679	**.687** (.685)	−.005 (.298)	.469	.683	**.809** (.731)	−.086 (.492)	−.035 (.281)	.539
PS	.582	**.565** (.582)	.039 (.288)	.340	.580	**.557** (.594)	.035 (.451)	.026 (.277)	.355
LN	.689	**.683** (.693)	.021 (.322)	.480	.690	**.713** (.717)	.006 (.526)	.001 (.308)	.514
CD	.403	.004 (.322)	**.723** (.724)	.525	.400	.062 (.333)	−.047 (.245)	**.713** (.723)	.525
SS	.445	.001 (.355)	**.802** (.803)	.644	.441	.033 (.360)	−.015 (.282)	**.790** (.799)	.639
CA	.289	.010 (.233)	**.505** (.509)	.260	.288	−.098 (.212)	.121 (.229)	**.516** (.517)	.274
Eigenvalue		6.41	1.58			6.41	1.58	1.04	
% Variance		39.20	7.26			39.55	7.31	4.06	
Factor correlations		F1	F2			F1	F2	F3	
	F1	—			F1	—			
	F2	.441	—		F2	.730	—		
					F3	.427	.346	—	

Note. Factor pattern coefficients (structure coefficients) based on principal factors extraction with promax rotation (k = 4). Salient pattern coefficients (≥.30) presented in bold. German WISC-V Subtests: SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LN = Letter–Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation; g = general intelligence; PS = Processing Speed; PR = Perceptual Reasoning; WM = Working Memory; VC = Verbal Comprehension; h^2 = Communality.
[a]General structure coefficients based on first unrotated factor coefficients (g loadings).

Local Fit. While all models with four or five group factors achieved good *global fit*, assessment of local fit identified numerous problems. Table 8 presents each of the models that contained local fit problems (i.e., nonstatistically significant standardized path coefficients, negative standardized path coefficients, and standardized path coefficients of 1.0). Most of these models were thus considered inadequate.

Model Selection. According to the ΔAIC > 10 criterion, the models most likely to generalize were Models 4a bifactor and the EFA suggested bifactor. These were also identified best by ΔCFI and ΔRMSEA criteria. However, local fit difficulties with Models 4a bifactor and EFA suggested bifactor (see Table 8) weighed against their selection without modification. Thus, Model 4a bifactor (Figure 3) and EFA suggested bifactor (Figure 4) and their modifications show remarkable similarity. Differences between these models are with which group factor FW is placed, and in both instances, FW had a negative and not-statistically significant standardized path coefficient with the assigned group factor. Figures 3 and 4 also illustrate modification where the FW group factor path was dropped and the model reestimated, which resulted in an identical model with theoretical alignment of all subtests but FW having only a path from g.

Variance and Reliability: Modified Model 4a Bifactor. Table 9 presents sources of variance for the modified Model 4a bifactor (see Figure 3) from the 15 German WISC-V primary and secondary subtests where the group factor path for FW was dropped. This model is identical to the EFA suggested bifactor model with the group factor path for FW dropped (see Figure 4). Most subtest variance was associated with the general intelligence dimension and substantially smaller portions of variance were uniquely associated with the four German WISC-V group factors. ω_H and ω_{HS} coefficients were estimated based on the bifactor results from Table 9 and the ω_H coefficient for general intelligence (.836) was high and sufficient for confident scale interpretation. The ω_{HS} coefficients for the four German WISC-V factors (VC, VS, WM, and PS), however, were considerably lower, ranging from .086 (VS) to .575

Table 6. Sources of Variance in the German Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) for the Total Standardization Sample (N = 1,087) According to the Schmid–Leiman Orthogonalized Higher-Order EFA Model With Four First-Order Factors.

German WISC-V	General		F1: Working Memory		F2: Verbal Comprehension		F3: Visual Spatial		F4: Processing Speed			
Subtest	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2
Similarities	.675	.456	.044	.002	**.310**	**.096**	.111	.012	−.016	.000	.566	.434
Vocabulary	.674	.454	−.025	.001	**.546**	**.298**	−.009	.000	−.010	.000	.753	.247
Information	.667	.445	.020	.000	**.369**	**.136**	.095	.009	−.051	.003	.593	.407
Comprehension	.562	.316	.031	.001	**.463**	**.214**	−.101	.010	.097	.009	.551	.449
Block Design	.629	.396	−.014	.000	−.024	.001	**.397**	**.158**	.061	.004	.558	.442
Visual Puzzles	.649	.421	.008	.000	−.004	.000	**.382**	**.146**	.015	.000	.567	.433
Matrix Reasoning	.623	.388	.161	.026	−.020	.000	**.226**	**.051**	−.030	.001	.466	.534
Figure Weights	.638	.407	*.152*	*.023*	.111	.012	**.143**	**.020**	−.045	.002	.465	.535
Arithmetic	.686	.471	**.255**	**.065**	.051	.003	.083	.007	.040	.002	.547	.453
Digit Span	.665	.442	**.403**	**.162**	−.034	.001	.000	.000	−.034	.001	.607	.393
Picture Span	.552	.305	**.245**	**.060**	.044	.002	.022	.000	.026	.001	.368	.632
Letter–Number Sequencing	.667	.445	**.400**	**.160**	.027	.001	−.055	.003	−.002	.000	.609	.391
Coding	.347	.120	.061	.004	−.021	.000	−.031	.001	**.641**	**.411**	.536	.464
Symbol Search	.382	.146	−.007	.000	−.012	.000	.040	.002	**.696**	**.484**	.632	.368
Cancellation	.244	.060	−.068	.005	.068	.005	.030	.001	**.458**	**.210**	.279	.721
Total Variance		.351		.034		.051		.028		.075	.540	.460
Explained Common Variance		.651		.063		.095		.052		.139		
ω		.923		.815		.857		.794		.727		
ω_H/ω_{HS}		.823		.167		.257		.135		.562		
Relative ω		.892		.204		.300		.170		.773		
H		.907		.408		.507		.367		.749		
PUC		.800										
ω_H/ω_{HS} Figure Weights on WM		.822		.142		.257		.167		.562		

Note. Bold type indicates coefficients and variance estimates consistent with the theoretically proposed factor. Italic type indicates coefficients and variance estimates associated with an alternate factor (where residual cross-loading b was larger than for the theoretically assigned factor).
EFA = exploratory factor analysis; b = loading of subtest on factor; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ω_H = Omega-hierarchical; ω_{HS} = Omega-hierarchical subscale; H = construct reliability or replicability index; WM = Working Memory; PUC = percentage of uncontaminated correlations.

(PS). Thus, three German WISC-V group factors (VC, VS, and WM) likely possess too little unique true score variance in a unit-weighted composite score to support confident clinical interpretation (Reise, 2012; Reise et al., 2013); however, the PS group factor exceeded the minimum criterion for possible interpretation. H indexes indicated an optimally weighted composite score for g accounted for 90.7% of g variance, but the four group factors were not well defined by their optimally weighted indicators (Hs < .70). For comparison purposes, Table A1 (see online supplement) presents sources of variance for Model 4a bifactor (see Figure 3) from the 15 German WISC-V primary and secondary subtests including FW group factor path and results of explained variances, ω_H and ω_{HS}, and H indexes were virtually identical to the modified Model 4a bifactor.

Discussion

Results from the present independent EFA and CFA substantially challenge the German WISC-V structure promoted in the German WISC-V *Technical Manual* on which standard scores and interpretive guidelines are provided. EFA results failed to support a five-factor model as only the SI subtest had a salient loading on the fifth factor which was inadequate and indicative of overfactoring. As with other versions of the WISC-V, there appears to be no separate FR factor and EFA results from both five- and four-factor models show the

Table 7. Maximum Likelihood CFA Fit Statistics for the 15 German WISC-V Primary and Secondary Subtests for the Standardization Sample (N = 1,087).

Model[a]	χ^2	*df*	CFI	ΔCFI	TLI	SRMR	RMSEA	ΔRMSEA	RMSEA 90% CI	AIC	ΔAIC
1. General intelligence	1,242.69	90	.833	−.157	.806	.072	.109	.080	[.103, .114]	76177.08	1069.43
2. Higher-order[b]	1,155.06	88	.846	−.144	.816	.071	.106	.077	[.100, .111]	76093.45	985.80
3. Higher-order[c]	651.33	87	.918	−.072	.902	.044	.077	.048	[.072, .083]	75591.72	484.07
4a. Higher-order[d]	276.83	86	.972	−.018	.966	.030	.045	.016	[.039, .051]	75219.21	111.56
4a. Bifactor[e]	**143.26**	**75**	**.990**	**.000**	**.986**	**.023**	**.029**	**.000**	**[.022, .036]**	**75107.65**	**0.00**
4a. Bifactor (no FW–VS path)*	**143.27**	**76**	**.990**	**.000**	**.987**	**.023**	**.029**	**.000**	**[.021, .036]**	**75105.66**	**−1.99**
4b. Higher-order[f]	279.23	86	.972	−.018	.966	.030	.045	.016	[.040, .051]	75221.62	113.97
4c. Higher-order[g]	259.36	85	.975	−.015	.969	.030	.043	.014	[.037, .049]	75203.75	96.10
4d. Higher-order[h]	257.51	84	.975	−.015	.969	.030	.044	.015	[.038, .050]	75203.90	96.25
EFA suggested bifactor[i]	**143.25**	**75**	**.990**	**.000**	**.986**	**.023**	**.029**	**.000**	**[.022, .036]**	**75107.63**	**−0.02**
5a. Higher-order[j]	237.98	85	.978	−.012	.973	.029	.041	.012	[.035, .047]	75182.36	74.71
5a. Bifactor[k]	153.60	77	.989	−.001	.985	.024	.030	.001	[.023, .037]	75117.98	10.33
5b. Higher-order[l]	236.19	85	.978	−.012	.973	.029	.040	.011	[.034, .047]	75180.57	72.92
5c. Higher-order[m]	217.47	84	.981	−.009	.976	.028	.038	.009	[.032, .044]	75163.86	56.21
5d. Higher-order[n]	228.20	84	.979	−.011	.974	.029	.040	.011	[.034, .046]	75174.59	66.94
5e. Higher-order[o]	217.25	83	.981	−.009	.975	.028	.039	.010	[.032, .045]	75165.63	57.98

Note. Bold text illustrates best fitting models. WISC-V = Wechsler Intelligence Scale for Children–Fifth edition; CFA = confirmatory factor analysis; CFI = comparative fit index; TLI = Tucker–Lewis index (nonnormed fit index); SRMR = standardized root mean square; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike's information criterion; FW = Figure Weights; VS = Visual Spatial.
[a]Model numbers correspond to those reported in the German WISC-V *Technical Manual* Table 5.2 and are higher-order models (unless otherwise specified) when more than one first-order factor was specified. Subtest assignments to latent factors are specified in Tables 1 and 2. [b-o]Models with local fit problems specified in Table 8.
*Best model.

AR subtest to only saliently load on the WM factor. Given this result, including AR in the pseudocomposite QRI appears misguided. Four first-order factors better represented the German WISC-V structure, but FW saliently loaded on the WM and not VS (formerly PR). The SLT of the four-factor oblique solution showed the primary subtest contribution was related mostly to *g* rather than to the first-order group factors (except for the PS subtests that poorly measure *g*). The present results replicate the outcomes of two WISC-V EFA studies with international WISC-V versions, the French WISC-V (Lecerf & Canivez, 2018) and the WISC-V[UK] (Canivez et al., 2019); two EFA studies with the full U.S. WISC-V standardization sample data (Canivez et al., 2016; Dombrowski et al., 2015), and two EFA studies examining the U.S. WISC-V standardization sample normative data partitioned into four age-groups (Canivez, Dombrowski, et al., 2018; Dombrowski, Canivez, et al., 2018). All found a lack of empirical support for five first-order factors and in all these studies the *g* factor accounted for substantially greater common variance and there was strong support for interpretation of composite score estimates of *g*. Also, these studies all showed inadequate portions of unique group factor variance apart from *g* necessary for confident interpretation of factor index scores, except, perhaps, for PS. These results were also observed in a large U.S. clinical sample (Canivez, McGill, et al., 2018).

Present CFA results also failed to support the publisher's preferred measurement model (Model 5e) and instead better supported a bifactor representation of German WISC-V structure with four group factors similar to the present EFA results. When modeling five first-order factors and one higher-order factor with all 15 primary and secondary subtests as promoted by the publisher (including Model 5e), approximate fit statistics *appeared* to support the models, unlike CFA results of five group-factor higher-order models with the U.S. WISC-V that produced model specification errors with negative FR variance (see Canivez, Watkins, & Dombrowski, 2017). However, assessment of local fit identified numerous problems of nonstatistically significant standardized path coefficients, negative standardized path coefficients, and standardized *g* to FR paths of 1.0 or approaching 1.0. The publisher preferred German WISC-V model (Model 5e) included three cross-loadings of AR on VC, FR, and WM identical to the U.S. WISC-V, but present results found the standardized path coefficient of VC to AR (.029) was not statistically significant and the standardized path coefficient from *g* to FR was 1.0 indicating empirical redundancy, thereby indicating Model 5e was not the best

Table 8. Local Fit Problems Identified Within Specified Models.

CFA model[a]	Local fit problem
2. Higher-order[b]	V factor and higher-order *g* factor linearly dependent on other parameters, *g* factor standardized path coefficient with V factor = 1.0
3. Higher-order[c]	*g* factor standardized path coefficients with V factor (.943) and *P* factor (.964) were high
4a. Higher-order[d]	*g* factor standardized path coefficients with VS factor (.946) and WM factor (.919) were high
4a. Bifactor[e]	FW standardized path coefficient with VS factor (−.005) was not statistically significant; and the MR standardized path coefficient with VS factor (.136), PS standardized path coefficient with WM (.200), and AR standardized path coefficient with WM (.169) were statistically significant but low
4b. Higher-order[f]	*g* factor standardized path coefficients with FR and WM factor (.945) were high
4c. Higher-order[g]	*g* factor standardized path coefficients with VS+AR factor (.947) was high
4d. Higher-order[h]	*g* factor standardized path coefficients with VS factor (.947) was high, AR standardized path coefficient with VC (.069) not statistically significant, AR standardized path coefficient on VS (.262) was low, removing AR path from VC produces Model 4c
EFA suggested bifactor[i]	FW standardized path coefficient with WM factor (−.007) was not statistically significant; MR standardized path coefficient with VS (.138), PS standardized path coefficient with WM (.198), and AR standardized path coefficient with WM (.167) were statistically significant but low; removal of WM–FW path produces same model as 4a Bifactor (without VS–FW path)
5a. Higher-order[j]	FR standardized path coefficient with *g* (.995) extremely high
5a. Bifactor[k]	MR (.090) and FW (.090) had low standardized path coefficients with FR and not statistically significant, removal of MR and FW group factor paths eliminates the FR factor
5b. Higher-order[l]	FR standardized path coefficient from *g* = 1.0
5c. Higher-order[m]	FR standardized path coefficient from *g* = 1.0
5d. Higher-order[n]	FR standardized path coefficient from *g* = 1.0, AR standardized path coefficient with VC (.151) was low
5e. Higher-order[o]	FR standardized path coefficient from *g* = 1.0, AR standardized path coefficient (.029) with VC not statistically significant, removal of AR loading with VC produces Model 5d

Note. Model number indicates the number of group factors included in the model and model number and letter correspond to those reported in the German WISC-V *Technical Manual*. Bifactor models were added for comparison. Subtest assignments to latent factors are specified in Tables 1 and 2. CFA = confirmatory factor analysis; *g* = general intelligence; V = Verbal; P = Performance; VC = Verbal Comprehension; WM = Working Memory; VS = Visual Spatial; FR = Fluid Reasoning; FW = Figure Weights; MR = Matrix Reasoning; PS = Picture Span; AR = Arithmetic.
[a]CFA model corresponding to Table 7. [b-o]Superscripts correspond model superscript designating local fit problem from Table 7.

model when one considers local fit. A similar result was observed with the French WISC-V where the AR subtest also failed to yield a statistically significant standardized path coefficient from VC, and thus the publisher preferred Model 5e was also not the best model with the French WISC-V (Lecerf & Canivez, 2018). A bifactor representation of the German WISC-V with *g* and *five* group factors (Model 5a bifactor) produced admissible global fit results, but MR and FW did not have statistically significant standardized path coefficients on the FR group factor, thereby challenging FR viability. Removal of nonstatistically significant MR and FW group factor paths eliminated the FR group factor. Thus, in both the higher-order and bifactor representations of the German WISC-V, FR is empirically indistinguishable from psychometric *g*.

These German WISC-V results are not unique and quite similar to EFA and CFA results observed in studies of the WISC-IV (Bodin et al., 2009; Canivez, 2014; Keith, 2005; Styck & Watkins, 2016; Watkins, 2006, 2010; Watkins et al., 2006) and with other Wechsler scale versions (Canivez & Watkins, 2010a, 2010b; Canivez, Watkins, Good, et al., 2017; Gignac, 2005, 2006; Golay et al., 2013; Golay & Lecerf, 2011; Lecerf & Canivez, 2018; McGill & Canivez, 2016, 2018; Nelson et al., 2013; Watkins & Beaujean, 2014; Watkins et al., 2013). The present results showing dominance of *g* variance and small portions of group factor variance are also not unique to Wechsler scales as similar results have also been observed with the Woodcock–Johnson III (Cucina & Howardson, 2016; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Strickland et al., 2015), the Woodcock–Johnson IV Cognitive and full battery (Dombrowski et al., 2017; Dombrowski, McGill, et al., 2018a, 2018b), the Differential Ability Scale (DAS; Cucina & Howardson, 2016), the DAS–II (Canivez et al., 2020; Canivez & McGill, 2016; Dombrowski et al., 2019), the Kaufman Adolescent and Adult Intelligence Test (Cucina & Howardson, 2016), the KABC (Cucina & Howardson, 2016), the SB5 (Canivez, 2008), the WASI and WRIT (Canivez et al., 2009), and the RIAS (Dombrowski et al., 2009; Nelson & Canivez, 2012, Nelson et al., 2007).

Practical Considerations

The present results have major practical implications in clinical assessment where the FRI is provided yet is not empirically supported by German standardization sample

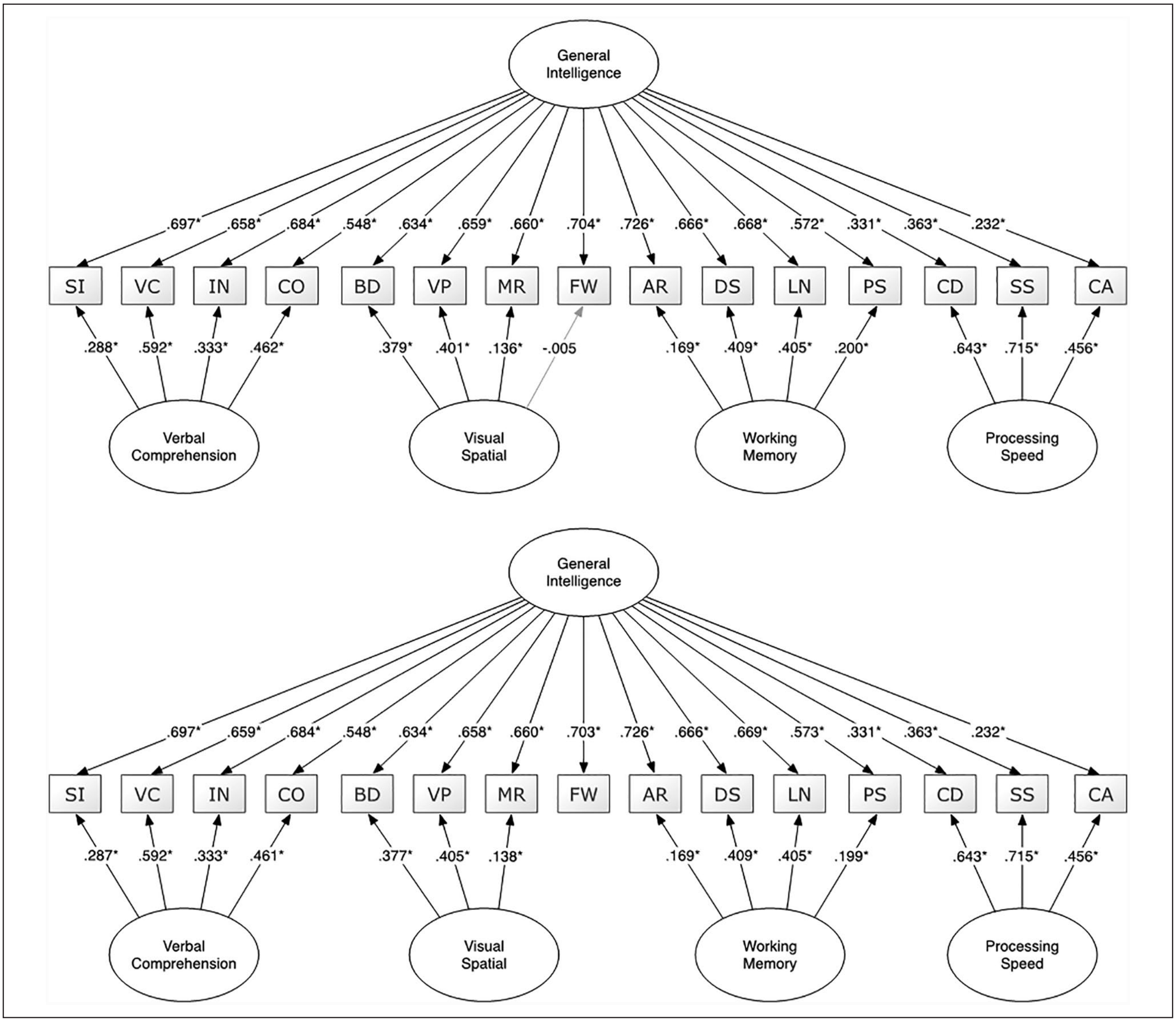

Figure 3. Bifactor measurement model (4a bifactor), with standardized coefficients, for the German WISC-V standardization sample (N = 1,087) 15 subtests, with and without the VS–FW path.

Note. WISC-V = Wechsler Intelligence Scale for Children–Fifth edition; SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation.

$^*p < .05$.

data in either EFA or CFA. This was also observed in the other WISC-V versions (Canivez et al., 2016; Canivez, Watkins, & Dombrowski, 2017; Canivez et al., 2019; Fenollar-Cortés & Watkins, 2019; Lecerf & Canivez, 2018; Watkins et al., 2018). The FR variance is essentially psychometric *g* variance, but this is obfuscated in higher-order models unless variance sources are decomposed (something the publisher has never provided in any WISC-V version) and thus, interpretation of a FR score most likely results in faulty inferences. Furthermore, VCI, VSI, and WMI are scores based on subtests that measure more *g* variance than group factor variance and the unique portions of true score variance provided by VC, VS, and WM are also seemingly inadequate for confident interpretation of scores provided by either unit-weighted or optimally weighted indexes as indicated by low ω_{HS} and *H* coefficients, respectively (Brunner et al., 2012; Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). Thus, "much of the reliable variance of the subscale scores can be attributable to the general factor, and not what is unique to the group factors" (Rodriguez et al., 2016, p. 225). Factor index scores, as provided by the publisher, conflate *g* variance and group factor

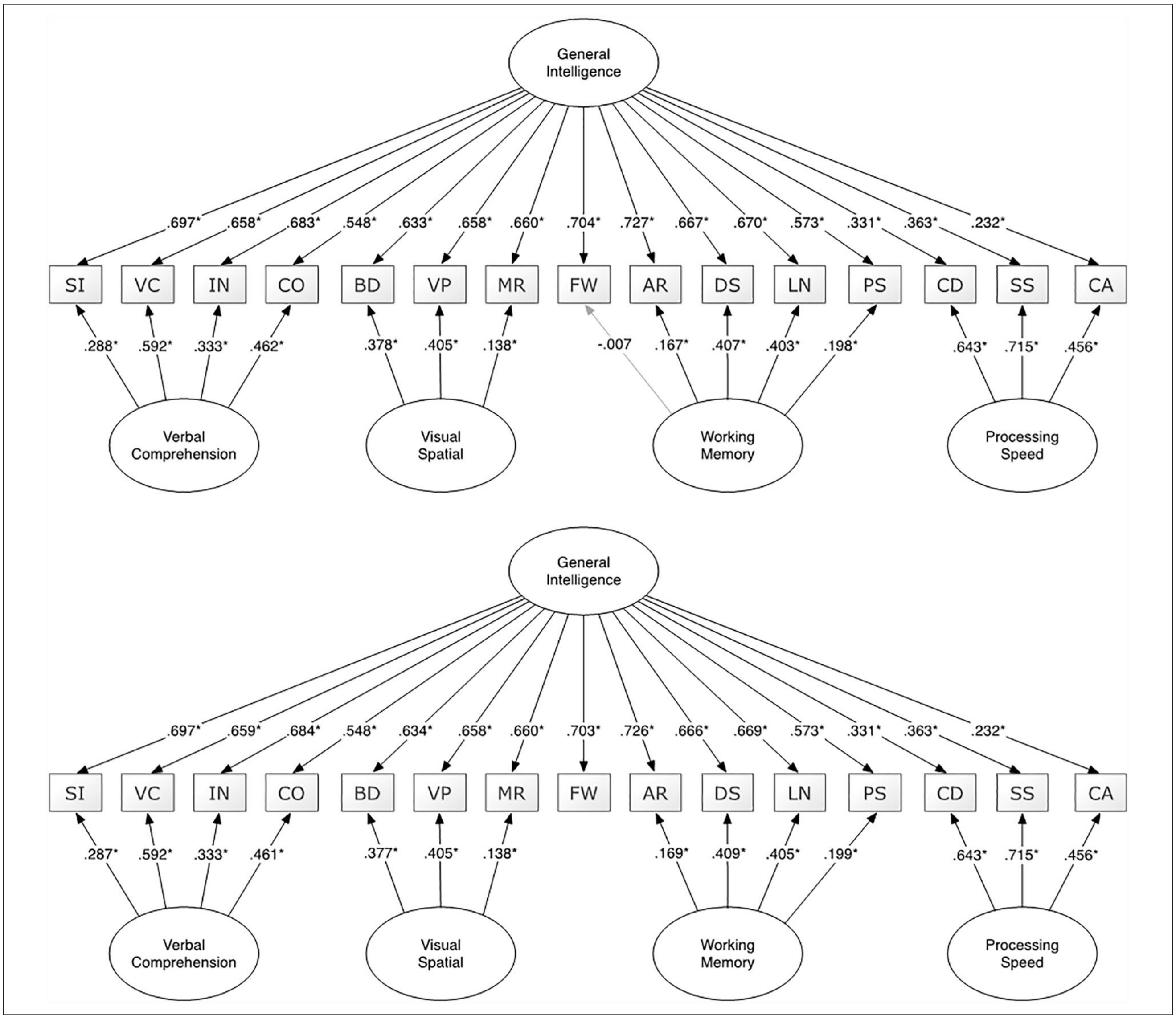

Figure 4. Bifactor measurement model (EFA Suggested Bifactor), with standardized coefficients, for the German WISC-V standardization sample (N = 1,087) 15 Subtests, with and without the VS–FW path.
Note. WISC-V = Wechsler Intelligence Scale for Children–Fifth edition; EFA = exploratory factor analysis; SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation.
$^*p < .05$.

variance which cannot be disentangled at the individual level. This too was observed in other WISC-V versions (Canivez et al., 2016; Canivez et al., 2019; Canivez, Watkins, & Dombrowski, 2017; Fenollar-Cortés & Watkins, 2019; Lecerf & Canivez, 2018; Watkins et al., 2018). Users of the German WISC-V can be confident in their individual clinical inferences regarding FSIQ results, but inferences from index scores beyond the FSIQ are likely overinterpretations or misinterpretations as also noted by Pauls et al. (2020). If it is important to generate scores for constructs represented by the group factors, and distinction between VS and FR, then it appears there is much work to be done to create tasks that accomplish this (if that is even possible). As Beaujean and Benson (2019) argue based on the work of Luecht et al. (2006), to achieve this, “publishers should not attempt to create instruments that concurrently measure some unitary attribute (e.g., a general attribute) and then try to spread out the same information across multiple scores of more specific attributes” (p. 130). Thus, it might be necessary to refrain from creating multidimensional measures of intelligence altogether and instead trying to develop multiple unidimensional tests, each designed to

Table 9. Sources of Variance in the 15 German Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) Primary and Secondary Subtests for the Standardization Sample (N = 1,087) According to a Bifactor Model With Four Group Factors With Visual Spatial to Figure Weights Path Removed.

	General		Verbal Comprehension		Visual Spatial		Working Memory		Processing Speed				
German WISC-V Subtest	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2	ECV
Similarities	.697	.486	.287	.082							.568	.432	.855
Vocabulary	.659	.434	.592	.350							.785	.215	.553
Information	.684	.468	.333	.111							.579	.421	.808
Comprehension	.548	.300	.461	.213							.513	.487	.586
Block Design	.634	.402			.377	.142					.544	.456	.739
Visual Puzzles	.658	.433			.405	.164					.597	.403	.725
Matrix Reasoning	.660	.436			.138	.019					.455	.545	.958
Figure Weights	.703	.494									.494	.506	.999
Arithmetic	.726	.527					.169	.029			.556	.444	.949
Digit Span	.666	.444					.409	.167			.611	.389	.726
Picture Span	.573	.328					.199	.040			.368	.632	.892
Letter–Number Sequencing	.669	.448					.405	.164			.612	.388	.732
Coding	.331	.110							.643	.413	.523	.477	.209
Symbol Search	.363	.132							.715	.511	.643	.357	.205
Cancellation	.232	.054							.456	.208	.262	.738	.206
Total Variance		.366		.050		.022		.027		.076	.541	.459	
Explained Common Variance		.678		.093		.040		.049		.140			
ω		.926		.859		.805		.818		.725			
ω_H/ω_{HS}		.836		.253		.086		.137		.575			
Relative ω		.903		.295		.107		.168		.793			
Factor correlation		.914		.503		.294		.370		.758			
H		.907		.506		.276		.319		.668			
PUC		.800											

Note. b = loading of subtest on factor; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ECV = explained common variance; ω_H = Omega-hierarchical (general factor); ω_{HS} = Omega-hierarchical subscale (group factors); H = construct reliability or replicability index; PUC = percentage of uncontaminated correlations. Illustrated in Figure 3.

measure a single, theoretically well-defined attribute. Furthermore, because the German WISC-V appears to only measure *g* well and provides group factor scores with inadequate interpretive value beyond *g*, it may be time and cost effective to use a measure like the German version of the Reynolds Intellectual Assessment Scales (RIAS; Hagmann-von Arx & Grob, 2014) as a more efficient assessment of *g*. Because the German RIAS includes only four (two verbal, two nonverbal) intelligence subtests representing verbal and nonverbal group factors there are fewer scores and comparisons that might be misused.

Theoretical Considerations

In addition to practical implications there are also theoretical implications for the present results. The superiority of the bifactor model observed with the German WISC-V which allows the general intelligence dimension to directly influence subtest indicators, while simultaneously allowing group factor influences on subtests, is consistent with Spearman's (1927) conceptualization of intelligence as well as Carroll's (1993; Beaujean, 2015b; Brunner et al., 2012; Frisby & Beaujean, 2015; Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993). Beaujean (2015b) noted Spearman's conceptualization 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 he also opined that "Carroll was explicit in noting that a bi-factor model best represents his theory" (p. 130). This conceptualizes *g* as a breadth factor that permits multidimensionality by determining how broad abilities perform independent of the *g* factor and was also preferred by Gignac (2008). Bifactor representation of *g* is less complicated and can be considered more parsimonious (Cucina & Byle, 2017; Gignac, 2008) with *g* and group factors at the same level of inference (see also Canivez, 2013b; Thompson, 2004). This is in contrast to the superordinate conceptualization of *g* represented by the publisher preferred higher-order model where the influence of psychometric *g* is fully mediated by the first-order group factors.

The theoretical appropriateness of bifactor models of intelligence was questioned by Reynolds and Keith (2013) who argued “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). Gignac (2006, 2008) alternatively suggested that because *g* was the most substantial factor it should be directly modeled and that full mediation of *g* in the higher-order model was what required explicit theoretical justification. Carroll (1993, 1995) pointed out that subtest scores reflect variation of both a general and more specific group factor but because they generally contain larger portions of *g* variance the subtest scores reliability is primarily a function of the general factor, not the specific group factor. Other researchers have also argued that Spearman’s (1927) and Carroll’s (1993) conceptualizations of intelligence are better represented by the bifactor model and not the higher-order model (Beaujean, 2015b; Brunner et al., 2012; Frisby & Beaujean, 2015; Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993).

Murray and Johnson (2013), Gignac (2016), and Mansolf and Reise (2017) determined that bifactor models might be found superior in fit due to unmodeled complexities such as small cross-loadings of indicators on multiple factors, proportionality constraint, or tetrad constraints; so the bifactor model may not be statistically better. Analyses of simulations of bifactor and higher-order models by Morgan et al. (2015) confirmed that regardless of the true structure, both types of models exhibited good model fit. Mansolf and Reise (2017) admitted that presently there is no technical solution to resolve the problem that bifactor and higher-order models cannot be distinguished by fit indices. Given this problem, Watkins et al. (2018) suggested requiring “a parsimonious, substantively meaningful model that fits observed data adequately well” (MacCallum & Austin, 2000, p. 218) and that fulfills the purpose of measurement; while Murray and Johnson (2013) concluded that when estimating or accounting for domain-specific abilities, the “bifactor model factor scores should be preferred” (p. 420). In the case of the German WISC-V, and all Wechsler scales, factor index scores and the numerous factor index score comparisons (ipsative and pairwise) and inferences made from such comparisons beyond the FSIQ is focusing on domain-specific abilities, so a bifactor model is necessary. Researchers and clinicians must know how well German WISC-V group factor scores perform independent of the *g* factor score (F. F. Chen et al., 2006; F. F. Chen et al., 2012).

Reise et al. (2010) also concluded that a bifactor model, which contains a general factor but permits multidimensionality, is better than the higher-order model so that relative contribution of group factors independent of the general factor (i.e., general intelligence) may be determined. This has also been recommended by others (Brunner et al., 2012; DeMars, 2013; Morin et al., 2016; Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). Given the absence of the FR factor and poor ω_{HS} and *H* coefficients for VC, VS, and WM, interpretation of these German WISC-V index scores “as representing the precise measurement of some latent variable that is unique or different from the general factor, clearly, is misguided” (Rodriguez et al., 2016, p. 225).

A final theoretical implication of present German WISC-V results relates to the so-called CHC theory (McGrew, 2009; Schneider, & McGrew, 2018). While several group factors (broad abilities) could be located, but not FR, the dominance of the *g* factor in explaining common variance in the German WISC-V is consistent with Carroll’s three stratum theory and not with the Cattell–Horn extended *Gf-Gc* theory. Cucina and Howardson (2017) offered the same conclusion in their analyses. Given the volume of evidence regarding preeminence of *g* variance in Wechsler scales and other intelligence tests, an annulment of the unhappy arranged marriage of the theories of Cattell–Horn and Carroll appears warranted (Canivez & Youngstrom, 2019; Wasserman, 2019).

Limitations

The present study examined EFA and CFA for the full German WISC-V standardization sample but it is possible that different age groups within the German WISC-V standardization sample might produce somewhat different results. EFA and CFA with different age subgroups should be conducted to examine structural invariance across age. Other demographic variables where invariance should be examined include sex and socioeconomic status. While Pauls et al. (2020) reported factor structure invariance for the publisher preferred German WISC-V measurement model (Model 5e) across gender [sic] and reported configural, first-order and second-order metric invariance, this only shows that the inadequate measurement model did not vary between groups. Invariance of the better represented bifactor model with four group factors identified in the present study should be examined. Structural invariance across gender [sic] was also reported for the U.S. WISC-V (H. Chen et al., 2015) but bifactor models and models with fewer group factors were also not examined. Because the publisher denied access to the German WISC-V standardization sample raw data, we are unable to independently conduct such analyses.

The present analyses were of the standardization sample and results may not generalize to other populations such as clinical groups or other independent samples of nonclinical groups, participants of different races/ethnicities, immigration status, or language minorities. Finally, the results of the present study only consider the latent factor structure and cannot fully test the construct validity of the German WISC-V. Examinations of German WISC-V relationships with external criteria (e.g., scholastic achievement) are needed. Examinations of incremental predictive validity (Canivez, 2013a; Canivez et al., 2014; Glutting et al., 2006) to determine if reliable achievement variance is incrementally accounted for by the German WISC-V factor index scores beyond that accounted for by the FSIQ (or through use of latent factor scores, see Kranzler et al., 2015) and diagnostic utility (see Canivez, 2013b) studies should also be conducted. However, the small portions of true score variance uniquely contributed by the four group factors identified here with the German WISC-V standardization sample makes it unlikely that German WISC-V factor index scores would provide meaningful additive interpretive value. Finally, while the present findings show dominance of general intelligence this does not mean that psychometric *g* is a single psychological attribute and perhaps, as indicated by Kovacs and Conway (2016) and Kan et al. (2019), the *g* factor may be a formative variable rather than a reflective variable; although Gottfredson (2016) argued the Kovacs and Conway Process Overlap Theory actually can be considered support for *g*. The present results (and Pauls et al., 2020) suggest clinicians should interpret with caution the factor index scores, if at all, due to low amounts of unique contributions of the broad abilities. However, that does not mean that broad abilities do not exist, they just may not be adequately measured by the German WISC-V or other intelligence tests.

Conclusion

Based on the present results, the German WISC-V as presented in the German WISC-V *Technical Manual* appears to be overfactored and the strong replication of previous EFA and CFA findings with the U.S. WISC-V and other international versions further indicates primary, if not exclusive, focus of interpretation on the German WISC-V FSIQ. The attempt to divide the PR factor into separate VS and FR factors was again unsuccessful by not producing a viable FR factor. Therefore, generating standard scores and comparisons for FR is potentially misleading and users likely misinterpreting scores. If FR cannot be located and does not make a unique contribution then the publisher should provide normative scores for four (VC, VS, WM, and PS) rather than *five* first-order factors, but the small portions of unique variance contribution by VC, VS, and WM likely render them of little utility; and the poor measurement of *g* by PS subtests might call for elimination from a test of general intelligence. The present results will help users of the German WISC-V make informed decisions about whether, when, and how to use the German WISC-V and which scores have adequate psychometric support for confident interpretation. Researchers and clinicians must rely on more than the test technical manuals to appropriately use test scores and their comparisons because test users bear "the ultimate responsibility for appropriate test use and interpretation" (AERA, APA, & NCME, 2014, p. 141). This will also allow professionals ethically using the German WISC-V to "know what their tests can do and act accordingly" (Weiner, 1989, p. 829).

Author's Note

Preliminary results were presented at the 41st Annual Conference of the International School Psychology Association, Basel, Switzerland and the 2019 Annual Convention of the American Psychological Association, Chicago, IL.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by a 2019 Summer Research Grant from the Council on Faculty Research, Eastern Illinois University to the first author.

ORCID iDs

Gary L. Canivez https://orcid.org/0000-0002-5347-6534
Silvia Grieder https://orcid.org/0000-0002-0118-7722

Supplemental Material

Supplemental material for this article is available online.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (1999). *Standards for educational and psychological testing*. American Educational Research Association.

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

Bartlett, M. S. (1954). A further note on the multiplying factors for various chi-square approximations in factor analysis. *Journal of the Royal Statistical Society: Series B*, *16*(2), 296-298. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x

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

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

Beaujean, A. A. (2016). Reproducing the Wechsler Intelligence Scale for Children-Fifth Edition: Factor model results. *Journal*

of Psychoeducational Assessment, 34(4), 404-408. http://doi.org/10.1177/0734282916642679
Beaujean, A. A. (2019). General and specific intelligence attributes in the two-factor theory: A historical review. In D. J. McFarland (Ed.), *General and specific mental abilities* (pp. 25-58). Cambridge Scholars.
Beaujean, A. A., & Benson, N. F. (2019). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*, *23*(2), 126-137. http://doi.org/10.1007/s40688-018-0182-1
Bentler, P. M., & Wu, E. J. C. (2016). *EQS for Windows*. Multivariate Software.
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*(5), 417-424. http://doi.org/10.1080/09297040802603661
Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796-846. http://doi.org/10.1111/j.1467-6494.2011.00749.x
Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261-304. http://doi.org/10.1177/0049124104268644
Byrne, B. M. (2006). *Structural equation modeling with EQS* (2nd ed.). Lawrence Erlbaum.
Canivez, G. L. (2008). Orthogonal higher-order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. *School Psychology Quarterly*, *23*(4), 533-541. http://doi.org/10.1037/a0012884
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*(2), 484-495. http://doi.org/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.), *Oxford handbook of child psychological assessments* (pp. 84-112). Oxford University Press.
Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*, *29*(1), 38-51. http://doi.org/10.1037/spq0000032
Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer, & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Hogrefe.
Canivez, G. L., Dombrowski, S. C., & Watkins, M. W. (2018). Factor structure of the WISC-V for four standardization age groups: Exploratory and hierarchical factor analyses with the 16 primary and secondary subtests. *Psychology in the Schools*, *55*(7), 741-769. http://doi.org/10.1002/pits.22138
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*(4), 252-265. http://doi.org/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*(2), 157-169. http://doi.org/10.1177/0734282913478036
Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales-Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment*, *28*(11), 1475-1488. http://doi.org/10.1037/pas0000279
Canivez, G. L., McGill, R. J., & Dombrowski, S. C. (2020). Factor structure of the Differential Ability Scales-Second Edition core subtests: Standardization sample confirmatory factor analyses. Advance online publication. *Journal of Psychoeducational Assessment*, *38*(7), 791-815. https://doi.org/10.1177/0734282920914792
Canivez, G. L., McGill, R. J., Dombrowski, 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*, *27*(2), 274-296. https://doi.org/10.1177/1073191118811609
Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV) adolescent subsample. *School Psychology Quarterly*, *25*(4), 223-235. http://doi.org/10.1037/a0022046
Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*(4), 827-836. http://doi.org/10.1037/a0020429
Canivez, G. L., & Watkins, M. W. (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 (Eds.), *Intelligent testing with the WISC-V* (pp. 683-702). Wiley.
Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *28*(8), 975-986. http://doi.org/10.1037/pas0000238
Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *29*(4), 458-472. http://doi.org/10.1037/pas0000358
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*(3), 383-407. http://doi.org/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*(4), 667-684. http://doi.org/10.1111/bjep.12056

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*(2), 195-224. http://doi.org/10.1111/bjep.12230

Canivez, G. L., & Youngstrom, E. A. (2019). Challenges to the Cattell–Horn–Carroll theory: Empirical, clinical, and policy implications. *Applied Measurement in Education*, *32*(3), 232-248. https://doi.org/10.1080/08957347.2019.1619562

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

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

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

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. http://doi.org/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*(3), 139-164. https://doi.org/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*(3), 464-504. http://doi.org/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*(1), 219-251. http://doi.org/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*(2), 189-225. http://doi.org/10.1207/s15327906mbr4102_5

Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC–IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, *49*(7), 677-682. http://doi.org/10.1016/j.paid.2010.06.004

Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between gender on the Wechsler Intelligence Scale for Children-Fifth Edition. *Personality and Individual Differences, 86*(November), 1-5. http://doi.org/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*(2), 233-255. http://doi.org/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*(6), 885-901. http://doi.org/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), 1-21. http://doi.org/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*(8), 1001-1015. http://doi.org/10.1037/pas0000389

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, *13*(4), 354-378. http://doi.org/10.1080/15305058.2013.799067

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly*, *28*(2), 154-169. http://doi.org/10.1037/spq0000010

Dombrowski, S. C. (2014a). Exploratory bifactor analysis of the WJ-III Cognitive in adulthood via the Schmid–Leiman procedure. *Journal of Psychoeducational Assessment*, *32*(4), 330-341. http://doi.org/10.1177/0734282913508243

Dombrowski, S. C. (2014b). Investigating the structure of the WJ-III cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment*, *32*(6), 483-494. http://doi.org/10.1177/0734282914530838

Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2018). Factor structure of the 10 WISC-V primary subtests across four standardization age groups. *Contemporary School Psychology*, *22*(1), 90-104. http://doi.org/10.1007/s40688-017-0125-2

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

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Cognitive at school age. *Psychological Assessment*, *29*(4), 394-407. http://doi.org/10.1037/pas0000350

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

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

Dombrowski, S. C., McGill, R. J., Canivez, G. L., & Peterson, C. H. (2019). Investigating the theoretical structure of the Differential Ability Scales—Second Edition through hierarchical exploratory factor analysis. *Journal of Psychoeducational Assessment*, *37*(1), 94-104. http://doi.org/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*(2), 442-455. http://doi.org/10.1037/a0031335

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, *27*(6), 494-507. http://doi.org/10.1177/0734282909333179

Federal Statistical Office of the Federal Republic of Germany. (Eds.). (2014). *Statistisches Jahrbuch Deutschland und Internationales* [Statistical yearbook]. Statistisches Bundesamt.

Fenollar-Cortés, J., & Watkins, M. W. (2019). Construct validity of the Spanish version of the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V^{Spain}). *International Journal of School & Educational Psychology*, *7*(3), 150-164. http://doi.org/10.1080/21683603.2017.1414006

Flanagan, D. P., & Harrison, D. (Eds.). (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). Guilford Press.

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*(2), 169-182. http://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*(July), 79-97. http://doi.org/10.1016/j.intell.2015.04.007

Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC–III. In J. Georgas, L. G. Weiss, & F. J. R. van de Vijver (Eds.), *Culture and children's intelligence* (pp. 277-313). Academic Press. http://doi.org/10.1016/B978-012280055-9/50021-7

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment*, *12*(3), 320-329. http://doi.org/10.1177/1073191105278118

Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, *27*(2), 73-86. http://doi.org/10.1027/1614-0001.27.2.73

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science Quarterly*, *50*(1), 21-43.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*(March-April), 57-68. http://doi.org/10.1016/j.intell.2016.01.006

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*(5), 639-662. http://doi.org/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*(3), 377-393. http://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*(2), 103-114. http://doi.org/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*(1), 143-152. http://doi.org/10.1037/a0021230

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling (BSEM). *Psychological Assessment*, *25*(2), 496-508. http://doi.org/10.1037/a0030676

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

Gottfredson, L. S. (2016). A *g* theorist on why Kovacs and Conway's process overlap theory amplifies, not opposes, *g* theory. *Psychological Inquiry*, *27*(3), 210-217. http://doi.org/10.1080/1047840X.2016.1203232

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407-434. http://doi.org/10.1207/s15327906mbr2804_2

Hagmann-von Arx, P., & Grob, A. (2014). *Reynolds Intellectual Assessment Scales and Screening (RIAS)™: German adaptation of the Reynolds Intellectual Assessment Scales (RIAS) TM & the Reynolds Intellectual Screening Test (RIST)™ from Cecil R. Reynolds and Randy W. Kamphaus*. Hans Huber.

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

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*(1), 41-54. http://doi.org/10.1007/BF02287965

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica*, *30*(2), 179-185. http://doi.org/10.1007/BF02289447

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). Riverside.

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

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology*, *57*(5), 253-270. https://doi.org/10.1037/h0023816

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

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

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

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*(3), 442-454. http://doi.org/10.1007/s11336-012-9269-1

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

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36. http://doi.org/10.1007/BF02291575

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, *73*(March-April), 52-62. http://doi.org/10.1016/j.intell.2018.12.004

Kaufman, A. S. (1994). *Intelligent testing with the WISC–III*. Wiley.

Kaufman, A. S., & Kaufman, N. L. (2015). *Kaufman Assessment Battery for Children–II* (German Version of P. Melchers & M. Melchers). Pearson Assessment.

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). Guilford Press.

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

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151-177. http://doi.org/10.1080/1047840X.2016.1153946

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*(4), 1402-1416. http://doi.org/10.1037/pas0000119

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*(6), 793-808. http://doi.org/10.1037/pas0000526

Lecerf, T., Rossier, J., Favez, N., Reverte, I., & Coleaux, L. (2010). The four- vs. alternative six-factor structure of the French WISC–IV. *Swiss Journal of Psychology*, *69*(4), 221-232. http://doi.org/10.1024/1421-0185/a000026

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*(2), 192-211. http://doi.org/10.1037/1082-989X.4.2.192

Luecht, R. M., Gierl, M. J., Tan, X., & Huff, K. (2006, April). *Scalability and the development of useful diagnostic scales* [Paper presentation]. Annual meeting of the National Council on Measurement in Education, San Francisco, CA.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226. https://doi.org/10.1146/annurev.psych.51.1.201

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

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

McDonald, R. P. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science*, *5*(6), 675-686. http://doi.org/10.1177/1745691610388766

McGill, R. J., & Canivez, G. L. (2016). Orthogonal higher order structure of the WISC-IV Spanish using hierarchical exploratory factor analytic procedures. *Journal of Psychoeducational Assessment*, *34*(6), 600-606. http://doi.org/10.1177/0734282915624293

McGill, R. J., & Canivez, G. L. (2018). 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*, *6*(4), 239-351. http://doi.org/10.1080/21683603.2017.1327831

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1-10. http://dx.doi.org/10.1016/j.intell.2008.08.004

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*(1), 2-20. http://doi.org/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. http://doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, *25*(4), 277-288. http://doi.org/10.1002/mpr.1485

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407-422. http://doi.org/10.1016/j.intell.2013.06.004

Nasser, F., Benson, J., & Wisenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurement*, *62*(3), 397-419. http://doi.org/10.1177/00164402062003001

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, *24*(1), 129-140. http://doi.org/10.1037/a0024878

Nelson, J. M, Canivez, G. L, Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology*, *45*(4), 439-456. http://doi.org/10.1016/j.jsp.2007.03.003

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale–Fourth edition (WAIS-IV) with a clinical sample. *Psychological Assessment*, *25*(2), 618-630. http://doi.org/10.1037/a0032086

Oakland, T., Douglas, S., & Kane, H. (2016). Top ten standardized tests used internationally with children and youth by school psychologists in 64 countries: A 24-year follow-up study. *Journal of Psychoeducational Assessment*, *34*(2), 166-176. http://doi.org/10.1177/0734282915595303

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402. http://doi.org/10.3758/BF03200807

Pauls, F., Daseking, M., & Petermann, F. (2020). Measurement invariance across gender on the second-order five-factor model of the German Wechsler Intelligence Scale for Children–Fifth edition. *Assessment*, *27*(8), 1836-1852. https://doi.org/10.1177/1073191119847762

Petermann, F. (Ed.). (2012). *Wechsler Adult Intelligence Scale–Fourth edition (WAIS-IV): German Adaptation*. Pearson Assessment.

Petermann, F., & Petermann, U. (Eds.). (2011). *Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV)*. Pearson Assessment.

Petermann, F., Ricken, G., Fritz, A., Schuck, K. D., & Preuß, U. (Eds.). (2014). *Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III): German Adaptation*. Pearson Assessment.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. http://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129-140. http://doi.org/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*(6), 544-559. http://doi.org/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). 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*(May), 31-47. http://doi.org/10.1016/j.intell.2017.02.005

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237. http://doi.org/10.1080/00223891.2015.1089249

Sattler, J. (2008a). *Assessment of children: Cognitive foundations* (5th ed.). Author.

Sattler, J. (2008b). *Resource guide to accompany assessment of children* (5th ed.). Author.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 53-61. http://doi.org/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). Guilford Press.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan, & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues: Fourth edition* (pp. 73-163). Guilford Press.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323-337. https://doi.org/10.3200/JOER.99.6.323-338

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

Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment*, *27*(2), 689-697. http://doi.org/10.1037/pas0000052

Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities*, *49*(2), 216-224. http://doi.org/10.1177/0022219414539565

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

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

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

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick, & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed.). Allyn & Bacon.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327. http://doi.org/10.1007/BF02293557

Wasserman, J. D. (2019). Deconstructing CHC. *Applied Measurement in Education*, *32*, 249-268. https://doi.org/10.1080/08957347.2019.1619563

Watkins, M. W. (2004). *MacOrtho*. [Computer software]. Ed & Psych Associates.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children–Fourth edition. *Psychological Assessment*, *18*(1), 123-125. http://doi.org/10.1037/1040-3590.18.1.123

Watkins, M. W. (2007). *SEscree* [Computer software]. 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*(4), 782-787. http://doi.org/10.1037/a0020043

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

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*, *31*(6-7), 1113-1126. http://doi.org/10.1080/13854046.2017.1317364

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219-246. http://doi.org/10.1177/0095798418771807

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence–Fourth edition. *School Psychology Quarterly*, *29*(1), 52-63. http://doi.org/10.1037/spq0000038

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 & Educational Psychology*, *1*(2), 102-111. http://doi.org/10.1080/21683603.2013.794439

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

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred students. *Educational and Psychological Measurement*, *66*(6), 975-983. http://doi.org/10.1177/0013164406288168

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children-Fifth edition*. NCS Pearson.

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children-Fifth edition: Canadian manual*. Pearson Canada Assessment.

Wechsler, D. (2014c). *Wechsler Intelligence Scale for Children-Fifth edition: Technical and interpretive manual*. NCS Pearson.

Wechsler, D. (2015a). *Escala de inteligencia de Wechsler para niños-V*. Pearson Educación.

Wechsler, D. (2015b). *Escala de inteligencia de Wechsler para niños-V: Manual técnico y de interpretación*. Pearson Educación.

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

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

Wechsler, D. (2016c). *Wechsler Intelligence Scale for Children-Fifth UK edition: Administration and scoring manual*. Harcourt Assessment.

Wechsler, D. (2017a). *Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V). Durchführungs- und Auswertungsmanual* (German version of F. Petermann). Pearson.

Wechsler, D. (2017b). *Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V): Technisches Manual*. (German version of F. Petermann). Pearson.

Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment*, *53*(4), 827-831. https://doi.org/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 Psychoeducational Assessment*, *31*(2), 94-113. http://doi.org/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 Psychoeducational Assessment*, *31*(2), 114-131. http://doi.org/10.1177/0734282913478032

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*(4), 254-265. http://doi.org/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*(2), 113-128. http://doi.org/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*(1), 123-133. http://doi.org/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*(2), 121-144. http://doi.org/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*(3), 443-451. http://doi.org/10.1177/0013164496056003006