British Journal of Educational Psychology (2019), 89, 195–224

The British Psychological Society

www.wileyonlinelibrary.com

Construct validity of the Wechsler Intelligence Scale For Children – Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests

Gary L. Canivez[1]*, Marley W. Watkins[2] and Ryan J. McGill[3]

[1]Eastern Illinois University, Charleston, Illinois
[2]Baylor University, Waco, Texas
[3]William & Mary, Williamsburg, Virginia

Background. There is inadequate information regarding the factor structure of the Wechsler Intelligence Scale for Children – Fifth UK Edition (WISC-V^{UK}; Wechsler, 2016a, Wechsler Intelligence Scale for Children-Fifth UK Edition, Harcourt Assessment, London, UK) to guide interpretation.

Aims and methods. The WISC-V^{UK} was examined using complementary exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) for all models proposed by Wechsler (2016b, Wechsler Intelligence Scale for Children-Fifth UK Edition: Administration and scoring manual, Harcourt Assessment, London, UK) as well as rival bifactor models.

Sample. The WISC-V^{UK} standardization sample ($N = 415$) correlation matrix was used in analyses due to denial of standardization sample raw data.

Results. EFA did not support a theoretically posited fifth factor because only one subtest (Matrix Reasoning) had a salient pattern coefficient on the fifth factor. A model with four group factors and a general intelligence factor resembling the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003, Wechsler Intelligence Scale for Children-Fourth Edition, Psychological Corporation, San Antonio, TX, USA) was supported by both EFA and CFA. General intelligence (*g*) was the dominant source of subtest variance and large omega-hierarchical coefficients supported interpretation of the Full Scale IQ (FSIQ) score. In contrast, the four group factors accounted for small portions of subtest variance and low omega-hierarchical subscale coefficients indicated that the four-factor index scores were of questionable interpretive value independent of *g*. Present results replicated independent assessments of the Canadian, Spanish, French, and US versions of the WISC-V (Canivez, Watkins, & Dombrowski, 2016, *Psychological Assessment,* 28, 975; 2017, *Psychological Assessment*, 29, 458; Fennollar-Cortés & Watkins, 2018, *International Journal of School & Educational Psychology*; Lecerf & Canivez, 2018, *Psychological Assessment*; Watkins, Dombrowski, & Canivez, 2018, *International Journal of School and Educational Psychology*).

Conclusion. Primary interpretation of the WISC-V^{UK} should be of the FSIQ as an estimate of general intelligence.

*Correspondence should be addressed to Gary L. Canivez, Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920-3099 (email: glcanivez@eiu.edu).

DOI:10.1111/bjep.12230

The Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V; Wechsler, 2014a) is the latest edition of one of the most popular intelligence tests in applied psychological practice and likely to be extensively used throughout the world (Oakland, Douglas, & Kane, 2016). Based on neuropsychological research and Cattell–Horn–Carroll (CHC) theory (CHC; Schneider & McGrew, 2012), which is an amalgam of the work of Carroll, Cattell, and Horn (Carroll, 1993; Horn, 1989; Horn & Cattell, 1966), two Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003) subtests were deleted and three new subtests were added. In addition, all 13 subtests retained from the WISC-IV included new and modified items (Wechsler, 2014b).

A major goal in revising the WISC-V was to separate subtests from the Perceptual Reasoning factor (PR) into distinct Visual Spatial (VS) and Fluid Reasoning (FR) factors making the instrument more consistent with CHC theory (Wechsler, 2014b). Accordingly, Visual Puzzles (VP) and Figure Weights (FW), both adapted from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008), were added to be better measure VS and FR factors, respectively. Picture Span (PSpan), which was adapted from the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV; Wechsler, 2012), was also added to the WISC-V to enhance measurement of the Working Memory (WM) factor.

WISC-V^{UK}

The WISC-V was anglicized and adapted for the United Kingdom (WISC-V^{UK}; Wechsler, 2016a) with few changes reportedly required in items, language, or spelling (Wechsler, 2016b). It was reported that substantial changes in item difficulty were not observed when comparing the WISC-V^{UK} to the US version so item order for the subtests was retained. The resulting WISC-V^{UK} subtests were then standardized and normed on a sample of 415 children between the ages of 6–0 and 16–11 years who were reported to be representative of the UK population stratified by geographic region, sex, race/ethnicity, and parent education level. This represents a substantial reduction in normative sample size from prior versions in the United Kingdom that may have affected sampling error (Bridges & Holler, 2007).

Unlike the WISC-IV^{UK} (Wechsler, 2004), *some* reliability and validity data based on the WISC-V^{UK} standardization sample were included in the WISC-V^{UK} *Administration and Scoring Manual* (Appendix D; Wechsler, 2016b). However, there was no separate technical manual presenting detailed descriptions of WISC-V^{UK} psychometric analyses. Additionally, the 16 intelligence subtests, Full Scale IQ (FSIQ), factor index scores, and ancillary index scores for the WISC-V^{UK} were identical to the US WISC-V.

Structural validity evidence

Structural validity evidence for intelligence tests is mainly derived from factor analytic methods. Both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are based on the common factor model, but EFA evaluates the correlational data to suggest a satisfactory model to describe those data, whereas CFA tests the hypothesis that a model could generate the observed data (Carroll, 1997). Wechsler (2014b) opined that CFA 'is preferred to exploratory factor analysis when an explicit theory of the factor structure is present or when there are competing models in the research literature' (p. 77). However, CFA methods may be vulnerable to confirmation bias or 'unwitting selectivity in the acquisition and use of evidence' (Nickerson, 1998, p. 175) by more readily allowing researchers to disregard plausible alternative models and confirm that their preferred 'theory-based' model fits the data (DiStefano & Hess, 2005). For example, Table D.10 in Wechsler (2016b) clearly shows that many of the tested models exhibited almost identical global fit (e.g., eight separate models exhibited root mean square error of approximation values of .04) and that Model 5d actually exhibited better fit (according to the Akaike information criterion) than the publisher preferred Model 5e (see Figure 1).

Other researchers have noted that CFA fit indices may be biased when there are signs of local misfit (Ropovik, 2015) or the model has been misspecified (Kline, 2016). While global fit refers to the overall model fit, local fit relates to individual parameter estimates, standard errors, or *z* values. Over-reliance on global fit indices can lead to weak factor structures that are unlikely to replicate (Ferrando & Navarro-González, 2018) and 'may account for uninterestingly small proportions of variance' (DeVellis, 2017, p. 197). Additionally, the statistical tests in CFA may be misleading when evaluating the discriminant validity of factors, leading to a proliferation of empirically indistinct constructs (Shaffer, DeGeest, & Li, 2016).

Rather than preferring one method over another, EFA and CFA should be considered complementary rather than competing methods that can be valuable when used together (Carroll, 1997; Haig, 2014; Keith, 2005; Tukey, 1980). For example, one complementary advantage of EFA methods is that they do not require advanced specification of models and thus are unbiased with respect to such prior specification (Carroll, 1985). Additionally, CFA results can be strengthened when supported by prior EFA that have identified the correct number of factors and indicator–factor relationships (Brown & Moore, 2012; Carroll, 1998). Given the relative strengths and weaknesses of EFA and CFA methods, Carroll (1995) recommended that both be employed when analysing cognitive data. Horn (1989) also suggested that CFA methods alone might be insufficient for analysing cognitive data. Given their influence in developing the CHC theory upon which the WISC-V was reportedly based, it seems apposite that the recommendations of Carroll and Horn be honoured in analyses of the WISC-V.

Problems with the publisher's factor analyses of the WISC-V

Contrary to the recommendations of Carroll (1995) and Horn (1989), the publisher relied exclusively on CFA for investigation of the internal structure of the WISC-V^{UK}. Users of the WISC-V^{UK} were directed to the US WISC-V *Technical and Interpretive Manual* (Wechsler, 2014b) for an 'overview of confirmatory factor analysis procedures and full details of the models tested' (Wechsler, 2016b; p. 371), as these were identically applied to the WISC-V^{UK} standardization sample. Table D.10 in the WISC-V^{UK} *Administration and Scoring Manual* (Appendix D) presented CFA fit statistics for the tested models paralleling the US WISC-V and claimed that CFA results 'support the allocation of the subtests to the respective indexes as in the US analyses' (Wechsler, 2016b, p. 371).

Figure 1 presents the publisher preferred measurement model for the US WISC-V, which was reportedly the model (Model 5e) that was also preferred with the WISC-V^{UK}. This higher-order model places *g* as a second-order factor being loaded by five first-order factors (Verbal Comprehension [VC], VS, FR, WM, and Processing Speed [PS]). Although CFA global fit statistics were presented for the WISC-V^{UK} standardization sample data, standardized path coefficients and the structural measurement model were not presented so it is not possible to assess local fit for the WISC-V^{UK} final preferred model.

The same substantive problems identified by Canivez and Watkins (2016); Canivez *et al.* (2016); Canivez, Watkins, and Dombrowski (2017); and Beaujean (2016) with the

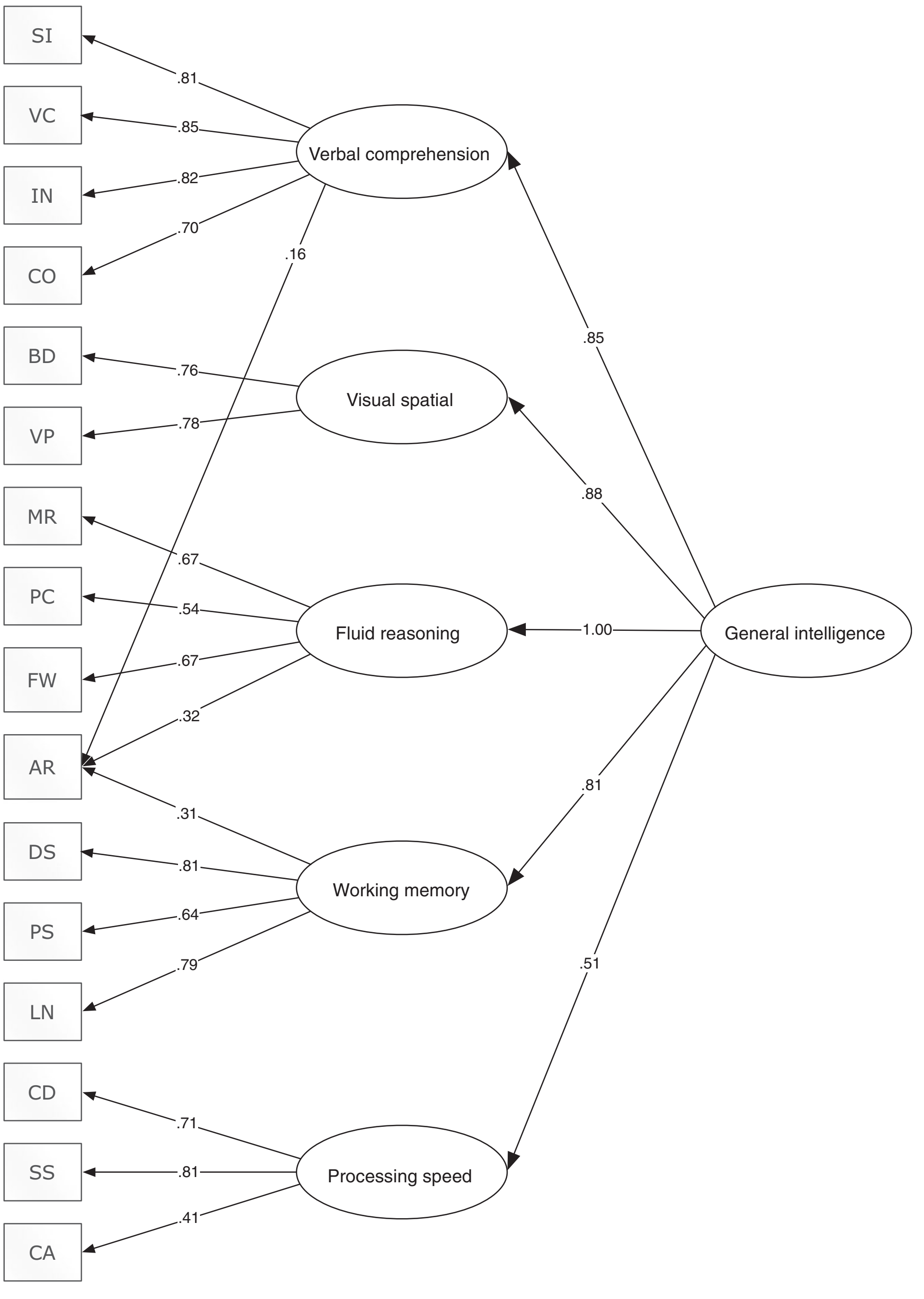

Figure 1. Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) higher-order measurement model with standardized coefficients (adapted from Figure 5.1 [Wechsler, 2014b]), for the standardization sample (N = 2,200). SI, Similarities; VC, Vocabulary; IN, Information; CO, Comprehension; BD, Block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

CFA methods employed by the publisher with the US WISC-V also apply to the WISC-V^{UK}. Among the noted problems was use of unweighted least squares estimation without explicit justification rather than maximum-likelihood estimation as well as failure to fully disclose details of CFA (Kline, 2016). Second, a complex CFA measurement model (cross-loading Arithmetic on three group factors) was retained, thereby abandoning parsimony of simple structure (Thurstone, 1947). Third, the standardized path coefficient of 1.0 between general intelligence (*g*) and the new FR factor is a threat to discriminant validity and indicates that FR and *g* may be empirically redundant (Kline, 2016; Le, Schmidt, Harter, & Lauver, 2010). Additionally, other areas of local fit may have been compromised. In fact, inspection of the degrees of freedom presented in Table D.10 (Wechsler, 2016b) indicates that there are fewer degrees of freedom than would be expected based on the number of indicators and the number of parameters that should be freely estimated. This suggests that some undisclosed parameters were fixed in some of the models prior to estimation (see Beaujean, 2016). Fourth, decomposed sources of variance between the higher-order *g* factor and lower-order group factors that are important for accurate interpretation of common factors were not reported (Brown, 2015). Fifth, model-based reliability estimates for factor scores were not provided (Watkins, 2017).

Finally, there was no consideration or testing of rival models as alternatives to the higher-order models examined by Wechsler (2014b, 2016b). A 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, Thissen, & McLeod, 1999). This model is illustrated in Figure 1. The higher-order model conceptualizes *g* as a *superordinate* factor and is thus an abstraction from abstractions (Gorsuch, 1983; Thompson, 2004). Wechsler (2014b, 2016b) exclusively relied on a higher-order structural representation for analyses of the WISC-V and WISC-V^{UK}.

Bifactor model

While higher-order models have been commonly applied to assess intelligence test structure, the bifactor model is an alternative conceptualization (illustrated in Figure 4). Originally specified by Holzinger and Swineford (1937), bifactor models have also been called direct hierarchical (Gignac, 2005, 2006, 2008) or nested factors models (Gustafsson & Balke, 1993). In bifactor models, *g* is conceptualized as a *breadth* factor (Gignac, 2008) because both the general and group factors *directly* influence the subtests. This means that both *g* and first-order group factors are simultaneous abstractions derived from the observed subtest indicators and therefore a less complicated (more parsimonious) conceptual model (Canivez, 2016; Cucina & Byle, 2017; Gignac, 2006, 2008).

Bifactor models have been found to fit data as well or better than higher-order models in more than 90% of published comparisons (Cucina & Byle, 2017). Additionally, bifactor models have several advantages, including the direct influences of the general factor are easy to interpret, both general and specific influences on indicators (subtests) can be examined simultaneously, and the psychometric properties necessary for determining scoring and interpretation of subscales can be directly examined (Canivez, 2016; Reise, 2012). Accordingly, Rodriguez, Reise, and Haviland (2016) concluded that 'the bifactor model is consistent with any measure found to have correlated factors or a second-order structure and, thus, it has quite broad generalizability' (p. 234) and Morin, Arens, Tran, and Caci (2016) argued that 'bifactor models provide a more flexible, realistic, and meaningful representation of the data whenever these dimensions are assumed to reflect a global underlying construct' (p. 9).

However, Keith (2005) questioned the theoretical appropriateness of bifactor models of intelligence, stating that they are 'not consistent with any modern theoretical orientation' (p. 594). Other researchers have disagreed with that conclusion. For example, Gignac (2006, 2008) contended that the most substantial factor of a battery of tests (i.e., *g*) should be directly modelled, whereas its full mediation in the higher-order model demands explicit theoretical justification; that is, a rationale is needed for why general intelligence should directly influence group factors but not subtests. Other researchers have argued that a bifactor model better represents Spearman's (1927) and Carroll's (1993) conceptualizations of intelligence than the higher-order model (Beaujean, 2015; Beaujean, Parkin, & Parker, 2014; Brunner, Nagy, & Wilhelm, 2012; Frisby & Beaujean, 2015; Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993). Beaujean (2015) elaborated that Spearman's conception of general intelligence was of a factor 'that was directly involved in all cognitive performances, not indirectly involved through, or mediated by, other factors' (p. 130) and noted that 'Carroll was explicit in noting that a bi-factor model best represents his theory' (p. 130). In fact, Jensen and Weng (1994) suggested a bifactor model as the first step in their strategy for identifying general intelligence (Jensen & Weng, 1994).

Many of these problems were previously identified and discussed with other Wechsler versions (Canivez, 2010, 2014a; Canivez & Kush, 2013; Gignac & Watkins, 2013), but were not addressed in the WISC-V *Technical and Interpretive Manual* nor in the WISC-V^{UK} *Administration and Scoring Manual.* These problems substantially challenge the preferred measurement model promulgated by the publisher of the WISC-V and WISC-V^{UK}, and it remains unclear whether the final measurement model presented by the publisher is viable for the WISC-V^{UK}.

Independent EFA of the WISC-V

Although EFA was not reported in the WISC-V *Technical and Interpretive Manual,* independent EFA of the WISC-V has not supported the existence of five factors in the total US WISC-V standardization sample (Canivez *et al.*, 2016; Dombrowski, Canivez, Watkins, & Beaujean, 2015), in four age groups (6–8, 9–11, 12–14, 15–16) with the 16 WISC-V primary and secondary subtests (Canivez, Dombrowski, & Watkins, 2018), nor in three of the four age groups (6–8, 9–11, and 12–14 years) with the 10 WISC-V primary subtests in the US standardization sample (Dombrowski, Canivez, & Watkins, 2018). In these cases, the fifth extracted factor included only one salient subtest loading. Recent EFA research with the French WISC-V (Wechsler, 2016c) also failed to find evidence for five factors (Lecerf & Canivez, 2018).

These EFAs of the US WISC-V standardization sample found substantial portions of variance apportioned to the general factor but substantially smaller portions of variance apportioned to the group factors (VC, PR, WM, PS). Omega-hierarchical (ω_H) coefficients (McDonald, 1999) ranged from .817 (Canivez *et al.*, 2016, 2018) to .847 (Canivez *et al.*, 2018; Dombrowski, Canivez, *et al.*, 2018) for the general factor but omega-hierarchical subscale (ω_{HS}) coefficients for the four WISC-V group factors ranged from .131 to .530. Similar reliability estimates were found with the French WISC-V (Lecerf & Canivez, 2018). Thus, independent EFA results have suggested that a four-factor solution appears to be the best measurement model for the WISC-V.

Independent CFA of the WISC-V

Independent CFA conducted with the 16 WISC-V primary and secondary subtests from the total US WISC-V standardization sample (Canivez, Watkins, *et al.*, 2017) found all five of the higher-order models that included five first-order factors (including the final WISC-V model presented in the WISC-V *Technical and Interpretative Manual*) resulted in statistically 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 local fit problems were identified where the Matrix Reasoning, Figure Weights, and Picture Concepts subtests did not evince statistically significant loadings on the FR factor. Consequently, the bifactor model with four group factors (VC, PR, WM, PS) was preferred based on the combination of statistical fit and Wechsler theory and provided complementary results to previous WISC-V EFA results (Canivez *et al.*, 2016) with a dominant general intelligence dimension and weak group factors with limited reliable measurement beyond *g*.

However, one study (H. Chen, Zhang, Raiford, Zhu, & Weiss, 2015) reported factorial invariance of the final publisher preferred WISC-V higher-order model with five-group factors across gender, although it did not examine invariance for rival higher-order or bifactor models. Likewise, Reynolds and Keith (2017) reported WISC-V invariance across age groups, but the model they examined for invariance was an oblique five-factor model, which ignores general intelligence altogether.

Reynolds and Keith (2017) also explored numerous post hoc 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 WISC-V measurement. Based on these explorations, their best-fitting WISC-V higher-order model was different from the publisher preferred model, yet it still produced a standardized path coefficient of .97 from *g* to Fluid Reasoning, suggesting that these dimensions may be isomorphic. In agreement with prior independent CFA, decomposed variance estimates from this higher-order model showed that the WISC-V subtests primarily reflected variance from *g* with small portions of variance unique to the group factors. An alternative bifactor model added a covariance estimate between VS and FR factors that 'recognizes the nonverbal related nature of these two factors' (p. 38). However, there was no justification for why the non-verbal PS factor was not also recognized. A similar bifactor model with correlated FR and VS factors was tested with the Canadian and Spanish WISC-V standardization samples (Wechsler, 2014c, 2015). It was *not* superior to the bifactor model with four group factors in the Canadian sample (Watkins *et al.*, 2018) but statistically equivalent to the four-factor solution with the Spanish sample, albeit with low discriminant validity and concomitant interpretational confounding (Fennollar-Cortés & Watkins, 2018).

Post hoc cross-loadings and correlated disturbance and error terms are frequently invoked in CFA models produced by researchers that prefer a higher-order structure for Wechsler scales. However, such explorations may capitalize on chance and sample size (MacCallum, Roznowski, & Necowitz, 1992). Additionally, it is rare for such parameters to be specified *a priori*. Instead, these previously unmodelled complexities are later added iteratively in the form of post hoc model adjustments designed to improve model fit or remedy issues encountered with local fit. However, Cucina and Byle (2017) suggested that specification of these parameters may be problematic due to lack of conceptual grounding in previous theoretical work and dangers of hypothesizing after results are known (HARKing).

In summary, the factorial structure of WISC-V standardization samples has been investigated by several independent researchers via CFA and results have been inconsistent. Some researchers favoured a traditional Wechsler four-factor model, while others preferred a CHC-based five-factor model. However, all studies have found a dominant general intelligence dimension and weak group factors with limited reliable measurement beyond *g*.

Independent CFA of the WISC-IV^UK

To date, there are no extant studies, technical supplements, or technical manuals providing EFA or CFA information with the WISC-IV^{UK} or WISC-V^{UK} standardization samples (Wechsler, 2004, 2016a). Only two studies have examined the latent factor structure of the WISC-IV^{UK}, and both applied CFA to data from Irish children referred for evaluation of learning difficulties (Canivez, Watkins, Good, James, & James, 2017; Watkins, Canivez, James, Good, & James, 2013). In the first study, Watkins *et al.* (2013) analysed the 10 core subtests and found a four-factor structure (VC, PR, WM, PS). In the second study (Canivez, Watkins, Good, *et al.*, 2017), all 15 WISC-IV^{UK} subtests were analysed to allow a comparison of CHC-based models with five factors to Wechsler-based models with four factors. Meaningful differences in fit were not observed between the CHC and Wechsler representations, leading the researchers to favour the more parsimonious Wechsler model. Both studies found that *g* accounted for the largest proportion of explained variance, and the group factors accounted for small to miniscule portions of explained variance. Both studies also found that FSIQ scores were relatively reliable ($\omega_H \cong .85$), while the group factor index scores were not reliable after removing the stabilizing influence of *g* ($\omega_{HS} \cong .14$ to $.43$).

Research aims

Understanding the structural validity of tests is essential for evaluating interpretability of test scores (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), and detailed information regarding evidence of the WISC-V^{UK} structure is necessary to properly interpret score results according to the *Code of Good Practice for Psychological Testing* of the British Psychological Society (2007, 2016) as well as the *Guidelines for Test Use* of the International Test Commission (2013). Given the absence of EFA, questionable CFA methods identified in the WISC-V *Technical and Interpretive Manual* (Wechsler, 2014b) that were also used with the WISC-V^{UK}, and lack of details regarding validity evidence for the WISC-V^{UK} provided in the *Administration and Scoring Manual* (Wechsler, 2016b); the present study: (1) used best practices in EFA (Watkins, 2018) to examine the WISC-V^{UK} factor structure suggested by the 16 primary and secondary subtest relationships, (2) examined the WISC-V^{UK} factor structure using CFA with customary maximum-likelihood estimation, (3) compared alternative bifactor models to higher-order models as rival explanations, (4) decomposed factor variance sources in EFA and CFA, and (5) estimated model-based reliabilities. The information afforded by these analyses is essential for users of the WISC-V^{UK} to determine the value of the scores and score comparisons provided in the WISC-V^{UK} and interpretive guidelines promoted by the publisher (Beaujean & Benson, 2019).

Method

Participants

The request for WISC-V^{UK} standardization sample raw data to conduct these independent analyses was denied without rationale by NCS Pearson, Inc. Absent raw data, the summary statistics (correlations and descriptive statistics) provided in Table D.9, Appendix D, in the WISC-V^{UK} *Administration and Scoring Manual* (Wechsler, 2016b) were used in the present analyses. These correlations were reportedly produced by participants who were members of the full WISC-V^{UK} standardization sample (N = 415) of children that ranged in age from 6 to 16 years. Demographic characteristics provided by Wechsler (2016b) illustrate the demographic representation of the UK standardization sample obtained using stratified proportional sampling across variables of age, sex, race/ethnicity, parental education level, and geographic region.

Instrument

The WISC-V^{UK} (Wechsler, 2016a) is an individually administered general intelligence test composed of 16 subtests expressed as scaled scores (M = 10, SD = 3). It includes seven 'Primary' subtests (Similarities [SI], Vocabulary [VC], Block Design [BD], Matrix Reasoning [MR], Figure Weights [FW], Digit Span [DS], and Coding [CD]) that produce the FSIQ score and three additional 'Primary' subtests (Visual Puzzles [VP], Picture Span [PSpan], and Symbol Search [SS]) that combine with the seven FSIQ subtests to produce the five-factor index scores (two subtests each for Verbal Comprehension [VCI], Visual Spatial [VSI], Fluid Reasoning [FRI], Working Memory [WMI], and Processing Speed [PSI]). There are six 'Secondary' subtests (Information [IN], Comprehension [CO], Picture Concepts [PC], Arithmetic [AR], Letter-Number Sequencing [LN], and Cancellation [CN]) that are used either for substitution in FSIQ estimation or in estimating the General Ability Index and Cognitive Proficiency Index scores. Index scores and FSIQ scores are expressed as standard scores (M = 100, SD = 15).

Analyses

Exploratory factor analysis

The 16 WISC-V^{UK} primary and secondary subtest correlation matrix included in Table D.9 of Wechsler (2016b, p. 370) was used to conduct EFAs. Although the published matrix includes correlations rounded to only two decimals, Carroll (1993) found that, 'little precision is lost by using two-decimal values' (p. 82).

The scree test (Cattell, 1966), standard error of scree (SE_{scree}; Zoski & Jurs, 1996), parallel analysis (PA; Horn, 1965), and minimum average partials (MAP; Velicer, 1976) criteria were considered when determining the number of factors to extract. Previous research and publisher theory suggested that four and five factors, respectively, should also be considered (Canivez *et al.*, 2016; Lecerf & Canivez, 2018; Wechsler, 2016b).

Principal axis extraction and promax rotation were accomplished with SPSS 24 for Macintosh. Other analyses were completed with open source software (Watkins, 2000, 2004, 2007). For a factor to be considered viable at least two subtests required salient loadings ($\geq$.30; McDonald, 1999). Then, to disentangle the contribution of first- and second-order factors, the Schmid and Leiman procedure was applied (SL; Schmid & Leiman, 1957). Carroll (1995) insisted on use of the SL transformation of EFA loadings to 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. Adhering to Carroll's (1995) directive, the SL procedure has been successfully applied in numerous studies of cognitive ability tests (e.g., Canivez, 2008; Canivez *et al.*, 2016; Dombrowski, Watkins, & Brogan, 2009; Golay & Lecerf, 2011; Lecerf & Canivez, 2018; Watkins, 2006).

Confirmatory factor analysis

EQS 6.3 (Bentler & Wu, 2016) was used to conduct CFA using maximum-likelihood estimation. Because of the absence of standardization sample raw data, covariance matrices were reproduced for CFA using the correlation matrix, means, and standard deviations from the total WISC-V^{UK} standardization sample presented by Wechsler (Table D.9, Appendix D, Wechsler, 2016b).

The structural models specified in Table 5.3 of the WISC-V *Technical and Interpretative Manual* (Wechsler, 2014b) were also examined in CFA with the WISC-V^{UK} (Table D.10; Wechsler, 2016b) and are reproduced in Figures 2 and 3 with the addition of alternative bifactor models that were not included in analyses reported by Wechsler (2014b, 2016b). Model 1 is a unidimensional *g* factor model loaded by all 16 subtests. Bifactor models were examined for all models that did not include cross-loadings on multiple factors. Because the VS factor was measured by only two subtests, those two loadings were constrained to equality when estimating bifactor models to ensure identification (Little, Lindenberger, & Nesselroade, 1999).

Although there are no universally accepted cut-off values for approximate fit indices (McDonald, 2010), overall global model fit was evaluated using the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Higher values indicate better fit for the CFI, whereas lower values indicate better fit for the RMSEA. Applying the

	Model 2		Model 3			Model 4a				Model 4a bifactor					Model 4b				Model 4b bifactor					Model 4c				Model 4d				Model 4e				Model 4e bifactor				
Subtest	F1	F2	F1	F2	F3	F1	F2	F3	F4	g	F1	F2	F3	F4	F1	F2	F3	F4	g	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4	g	F1	F2	F3	F4
SI	■		■			■				■	■				■				■	■				■				■				■				■	■			
VC	■		■			■				■	■				■				■	■				■				■				■				■	■			
IN	■		■			■				■	■				■				■	■				■				■				■				■	■			
CO	■		■			■				■	■				■				■	■				■				■				■				■	■			
BD		■		■			■			■		■				■			■		■				■				■				■			■		■		
VP		■		■			■			■		■				■			■		■				■				■				■			■		■		
MR		■		■			■			■		■					■		■			■			■				■				■			■		■		
FW		■		■			■			■		■					■		■			■			■				■				■			■		■		
PC		■		■			■			■		■					■		■			■			■				■				■			■		■		
AR	■		■					■		■			■				■		■			■			■	■		■	■	■			■			■		■		
DS	■		■					■		■			■				■		■			■				■				■				■		■			■	
PS		■		■				■		■			■				■		■			■				■				■				■		■			■	
LN	■		■					■		■			■				■		■			■				■				■				■		■			■	
CD		■			■				■	■				■				■	■				■				■				■				■	■				■
SS		■			■				■	■				■				■	■				■				■				■				■	■				■
CA		■			■				■	■				■				■	■				■				■				■				■	■				■

Figure 2. WISC-V^{UK} Primary and Secondary Subtest configuration for confirmatory factor analysis (CFA) models with two to four factors. SI, Similarities; VC, Vocabulary; IN, Information; CO, Comprehension; BD, Block Design; VP, Visual Puzzles; MR, Matrix Reasoning; FW, Figure Weights; PC, Picture Concepts; AR, Arithmetic; DS, Digit Span; PS, Picture Span; LN, Letter-Number Sequencing; CD, Coding; SS, Symbol Search; CA, Cancellation. All models include a higher-order general factor except for the bifactor models.

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

Figure 3. WISC-V[UK] Primary and Secondary Subtest configuration for confirmatory factor analysis (CFA) models with five factors. SI, Similarities; VC, Vocabulary; IN, Information; CO, Comprehension; BD, Block Design; VP, Visual Puzzles; MR, Matrix Reasoning; FW, Figure Weights; PC, Picture Concepts; AR, Arithmetic; DS, Digit Span; PS, Picture Span; LN, Letter-Number Sequencing; CD, Coding; SS, Symbol Search; CA, Cancellation. All models include a higher-order general factor except for the bifactor models.

Hu and Bentler (1999) combinatorial heuristics, criteria for adequate model fit were CFI $\geq$ .90 along with RMSEA $\leq$ .08. Good model fit required CFI $\geq$ .95 with RMSEA $\leq$ .06. Additionally, the Akaike Information Criterion (AIC) was considered. AIC does not have a meaningful scale but the model with the smallest AIC value is most likely to replicate (Kline, 2016). For a model to be considered superior, it had to exhibit good overall fit *and* display meaningfully better fit (ΔCFI > .01, ΔRMSEA < .015, and ΔAIC < 10) than alternative models (Burnham & Anderson, 2004; F. Chen, 2007; Cheung & Rensvold, 2002). All models were examined for presence of local fit problems (e.g., negative, too high, or too low standardized path coefficients, coefficients exceeding limits [-1, 1], negative variance estimates) as models should never be retained 'solely on global fit testing' (Kline, 2016, p. 461).

Model-based reliabilities

Model-based reliabilities were estimated with omega coefficients (Reise, 2012; Reise, Bonifay, & Haviland, 2013; Rodriguez *et al.*, 2016). McDonald (1999) described several omega coefficient variants based on decomposing total test variance into common and unique components: (1) omega (ω) that is similar to coefficient alpha in that it indexes the proportion of variance in a unit-weighted score attributable to all sources of common variance; (2) omega-hierarchical (ω_H) that estimates the reliability of a unit-weighted total score (i.e., FSIQ) after removing the influence of the group factors; and (3) omega-hierarchical subscale (ω_{HS}) that estimates the reliability of a unit-weighted group factor score (i.e., VCI, PRI) after removing the influence of all other factors. Omega coefficients make fewer and more realistic statistical assumptions than coefficient alpha and have been recommended for use with multidimensional tests like the WISC-V[UK] (Watkins, 2017). Omega estimates 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). Omega coefficients should at a minimum exceed .50, but .75 is preferred (Reise, 2012; Reise *et al.*, 2013; Rodriguez *et al.*, 2016).

Results

Exploratory factor analysis

The Kaiser–Meyer–Olkin measure of sampling adequacy of .924 far exceeded the .60 minimum standard (Kaiser, 1974) and Bartlett's test of sphericity (Bartlett, 1954), $\chi^2 = 2{,}560.45$, $p < .0001$; indicated that the WISC-V^{UK} correlation matrix was not random. Without standardization sample raw data, it was not possible to estimate skewness or kurtosis or determine whether multivariate normality existed, but principal axis extraction does not assume normality. Therefore, the correlation matrix was deemed appropriate for EFA.

Regarding the number of factors to extract, Scree, PA (see Figure S1), and MAP criteria suggested two, *SE*scree indicated three, prior research with the WISC-V indicated that four would suffice, and the WISC-V^{UK} publisher claimed five factors. Wood, Tataryn, and Gorsuch (1996) opined that it is better to overextract than underextract, so EFA began by extracting five factors and then sequentially examined the adequacy of models with four, three, and two factors.

Extracting five WISC-V^{UK} factors (see Table S1) produced a fifth factor with only one salient factor pattern coefficient (MR). Thus, MR and FW did not share sufficient common variance to constitute the FR dimension posited by the publisher. Furthermore, PC failed to achieve a salient pattern coefficient on any factor. This pattern of results is emblematic of overextraction (Gorsuch, 1983; Wood *et al.*, 1996), and the five-factor model was judged inadequate.

Table 1 presents the results of extracting four WISC-V^{UK} factors and reveals four robust factors with theoretically consistent subtest associations resembling the traditional Wechsler structure. None of the subtests loaded saliently on more than one factor and the moderate-to-high factor correlations (.357 to .699) signalled the presence of a general intelligence factor (Gorsuch, 1983).

For the three-factor model, the PR and WM factors merged, leaving distinct VC and PS factors, but no subtest cross-loadings were observed. The two-factor model showed merging of VC, PR, and WM factors, leaving only the separate PS factor. The two- and three-factor models (see Table S2) clearly displayed fusion of theoretically meaningful constructs that is symptomatic of underextraction, thereby rendering them unsatisfactory (Gorsuch, 1983; Wood *et al.*, 1996).

Given these results, the four-factor EFA solution appeared to be the most appropriate and was accordingly subjected to second-order EFA that was transformed with the SL procedure (see Table 2). Following SL transformation, all WISC-V^{UK} subtests were properly associated with their theoretically proposed factors (Wechsler model). The hierarchical *g* factor accounted for 31.7% of the total variance and 65.3% of the common variance. The general factor also accounted for between 5.3% (CA) and 45.3% (IN) of individual subtest variability. For comparison, results of SL transformation of five-factor EFA solution are presented in Table S3 and illustrate how little unique variance the fifth factor provides (3.4% total variance, 6.4% common variance).

Omega coefficients were estimated based on the SL results in Table 2. The ω_H coefficient for a unit-weighted FSIQ score based on all indicators (.811) was high;

Table 1. Wechsler Intelligence Scale for Children – Fifth UK Edition (WISC-V^{UK}) exploratory factor analysis: Oblique four-factor solution for the total standardization sample (N = 415)

	General	Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed		
WISC-V^{UK} subtest	*S*	*P*	*S*	*P*	*S*	*P*	*S*	*P*	*S*	h^2
Similarities	.716	**.704**	.776	.052	.585	.037	.549	.031	.329	.608
Vocabulary	.770	**.815**	.859	.029	.622	.046	.586	−.017	.314	.740
Information	.757	**.604**	.774	.178	.654	.045	.588	.047	.375	.628
Comprehension	.594	**.827**	.726	−.113	.440	−.032	.425	−.003	.219	.536
Block Design	.663	.043	.527	**.695**	.728	−.021	.513	.035	.378	.531
Visual Puzzles	.603	−.059	.457	**.846**	.724	−.113	.437	−.005	.323	.534
Matrix Reasoning	.620	.008	.484	**.529**	.647	.176	.538	−.023	.338	.433
Figure Weights	.573	.192	.513	**.438**	.583	.053	.453	−.054	.258	.364
Picture Concepts	.473	.016	.369	**.352**	.478	.165	.421	−.001	.271	.244
Arithmetic	.652	.176	.544	.076	.542	**.458**	.652	.047	.407	.455
Digit Span	.664	.003	.502	.014	.534	**.795**	.764	−.074	.385	.587
Picture Span	.504	−.055	.355	.258	.475	**.310**	.500	.081	.363	.288
Letter-Number Sequencing	.630	.043	.488	−.046	.490	**.769**	.729	−.064	.366	.536
Coding	.451	−.149	.214	.033	.365	.150	.462	**.679**	.727	.543
Symbol Search	.532	.035	.339	−.068	.401	.190	.529	**.640**	.727	.550
Cancellation	.273	.119	.176	.013	.217	−.291	.173	**.666**	.549	.339
Eigenvalue		6.32		1.52		1.05		0.98		
% Variance		36.55		6.50		3.38		3.03		
Promax-based factor correlations		VC		PR		WM		PS		
Verbal Comprehension (VC)		–								
Perceptual Reasoning (PR)		.698		–						
Working Memory (WM)		.650		.699		–				
Processing Speed (PS)		.357		.488		.568		–		

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

Table 2. Sources of variance in the Wechsler intelligence scale for children – Fifth UK Edition (WISC-V^{UK}) for the total standardization sample (N = 415) according to an exploratory SL model (Orthogonalized higher-order factor model) with four first-order factors

	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed			
WISC-V^{UK} subtest	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2
Similarities	.625	.391	**.463**	**.214**	.027	.001	.018	.000	.025	.001	.607	.393
Vocabulary	.669	.448	**.536**	**.287**	.015	.000	.023	.001	−.014	.000	.736	.264
Information	.673	.453	**.397**	**.158**	.092	.008	.022	.000	.038	.001	.621	.379
Comprehension	.497	.247	**.544**	**.296**	−.058	.003	−.016	.000	−.002	.000	.547	.453
Block Design	.629	.396	.028	.001	**.359**	**.129**	−.010	.000	.029	.001	.526	.474
Visual Puzzles	.579	.335	−.039	.002	**.437**	**.191**	−.056	.003	−.004	.000	.531	.469
Matrix Reasoning	.598	.358	.005	.000	**.273**	**.075**	.088	.008	−.019	.000	.440	.560
Figure Weights	.534	.285	.126	.016	**.226**	**.051**	.026	.001	−.044	.002	.355	.645
Picture Concepts	.456	.208	.011	.000	**.182**	**.033**	.082	.007	−.001	.000	.248	.752
Arithmetic	.622	.387	.116	.013	.039	.002	**.228**	**.052**	.038	.001	.455	.545
Digit Span	.661	.437	.002	.000	.007	.000	**.396**	**.157**	−.061	.004	.598	.402
Picture Span	.495	.245	−.036	.001	.133	.018	**.154**	**.024**	.066	.004	.292	.708
Letter-Number Sequencing	.623	.388	.028	.001	−.024	.001	**.383**	**.147**	−.052	.003	.539	.461
Coding	.437	.191	−.098	.010	.017	.000	.075	.006	**.556**	**.309**	.516	.484
Symbol Search	.501	.251	.023	.001	−.035	.001	.095	.009	**.524**	**.275**	.536	.464
Cancellation	.231	.053	.078	.006	.007	.000	−.145	.021	**.545**	**.297**	.378	.622
Total variance		.317		.060		.030		.024		.055	.495	.505
ECV		.653		.123		.062		.049		.113		
ω		.914		.867		.773		.767		.712		
ω_H/ω_{HS}		.811		.332		.169		.145		.469		

Note. b = standardized loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness (specificity plus error), ECV = explained common variance, ω = omega, ω_H = omega-hierarchical (general factor), and ω_{HS} = omega-hierarchical subscale (group factors). Bold type indicates largest coefficients and variance estimates consistent with the theoretically proposed factor. The highest subtest loading with the specific group factor was used in omega subscale estimates.

however, the ω_{HS} coefficients for four unit-weighted WISC-V^{UK} factor index scores (VCI, WMI, PRI, PSI) based on all indicators were considerably lower (.145–.469).

Confirmatory factor analysis

Global fit

Results from CFAs for the 16 WISC-V^{UK} primary and secondary subtests are presented in Table 3. Models 1 and 2 were inadequate due to low CFI and too high RMSEA values. Model 3 was adequate, but all models (both higher-order and bifactor) that included four- or five-group factors produced global fit statistics that indicated good fit to these data. Bifactor models where AR was not cross-loaded were often meaningfully better than their higher-order versions when considering ΔCFA values, but meaningful differences in RMSEA were only observed for Model 4b bifactor and Model 4e bifactor compared to their higher-order versions. In contrast, all bifactor models were meaningfully superior to their higher-order versions when considering ΔAIC and therefore more likely to replicate.

Local fit

Although several models achieved good global fit, assessment of local fit identified numerous problems. Table 4 presents each of the models that exhibited local fit problems (i.e., non-statistically significant standardized path coefficients, negative standardized path coefficients, and negative variance estimates) or issues with either very low or very high standardized path coefficients (DiStefano & Hess, 2005). Many of these models were thus considered inadequate. For example, the publisher's preferred model (5e higher-order) produced good global fit to these data (CFI = .979, RMSEA = .036), but the standardized path coefficient (.063) of AR on FR was not statistically significant, the standardized path coefficient (.192) of AR on VC was statistically significant but low, and the removal of the non-statistically significant AR loading on FR produces Model 5d.

Model selection

Model 4a bifactor displayed the best fit according to CFI, RMSEA, and AIC indices, but it was not meaningfully superior to bifactor Models 4b, 4e, 5a, and 5b. However, local fit problems with those alternative models (see Table 4) weighed against their selection. Thus, Model 4a bifactor (Figure 4) appears the best model to represent WISC-V^{UK} measurement despite the weak standardized path coefficients of PC on PR and PSpan on WM. Model 4a bifactor did not manifest any negative standardized path coefficients or negative variance estimates and was consistent with CFA results from the WISC-IV (Canivez, Watkins, Good, *et al.*, 2017; Watkins *et al.*, 2013) as well as the current EFA results from the WISC-V^{UK}.

Variance and reliability

Table 5 presents sources of variance for Model 4a bifactor from the 16 WISC-V^{UK} primary and secondary subtests. Most subtest variance was associated with the general intelligence dimension, and substantially smaller portions of variance were uniquely associated with the four WISC-V^{UK} group factors. The ω_H coefficient of .829 for a unit-weighted FSIQ score with all indicators was robust, but the ω_{HS} coefficients for four

Table 3. Confirmatory factor analysis (CFA) fit statistics for WISC-V^{UK} 16 subtests for the total standardization sample (N = 415) of the Wechsler Intelligence Scale for Children – Fifth UK Edition

Model[a]	χ^2	*df*	CFI	ΔCFI	RMSEA	90% CI RMSEA	ΔRMSEA	AIC	ΔAIC
1	499.62	104	.840	.150	.096	[.087, .104]	.069	31346.37	353.60
2 Higher-Order[2]	411.10	102	.875	.115	.086	[.077, .094]	.059	31261.85	269.08
3 Higher-Order	273.95	101	.930	.060	.064	[.055, .073]	.037	31126.71	133.94
4a Higher-Order	163.29	100	.974	.016	.039	[.028, .050]	.012	31018.05	25.28
4a Bifactor[3]	**114.02**	**88**	**.990**	**.000**	**.027**	**[.008, .040]**	**.000**	**30992.77**	**0.00**
4b Higher-Order[4]	187.18	100	.965	.025	.046	[.036, .056]	.019	31041.94	49.17
4b Bifactor[5]	115.79	89	.989	.001	.027	[.009, .040]	.000	30994.55	1.78
4c Higher-Order[6]	162.33	99	.974	.016	.039	[.028, .050]	.012	31019.08	26.31
4d Higher-Order[7]	157.54	98	.976	.014	.038	[.027, .049]	.011	31016.30	23.53
4e Higher-Order[8]	188.24	100	.964	.026	.046	[.036, .046]	.019	31043.00	50.23
4e Bifactor[9]	114.80	88	.989	.001	.027	[.009, .040]	.000	30993.56	0.79
5a Higher-Order[10]	157.99	99	.976	.014	.038	[.026, .049]	.011	31014.75	21.98
5a Bifactor[11]	118.38	89	.988	.002	.028	[.012, .041]	.001	30997.14	4.37
5b Higher-Order[12]	164.47	99	.974	.016	.040	[.029, .050]	.013	31021.22	28.45
5b Bifactor[13]	123.70	89	.986	.004	.031	[.016, .043]	.004	31002.45	9.68
5c Higher-Order[14]	153.76	98	.978	.012	.037	[.025, .048]	.010	31012.52	19.75
5d Higher-Order[15]	150.11	98	.979	.011	.036	[.024, .047]	.009	31008.86	16.09
5e Higher-Order[16]	149.95	97	.979	.011	.036	[.024, .047]	.009	31010.70	17.93

Notes. CFI, Comparative Fit Index; RMSEA, root mean square error of approximation; AIC, Akaike's information criterion. Bold text illustrates best-fitting model. [a]Model numbers (number of group factors) and letters correspond to those reported in the WISC-V *Technical and Interpretive Manual* and the WISC-V^{UK} *Administration and Scoring Manual Appendix D* (except 4e, which was added for comparison to Canivez, Watkins, *et al.*, 2017). Subtest assignments to latent factors are specified in Figures 2 and 3. [2-16]Models with local fit problems specified in Table 4.

Table 4. Local fit problems identified within specified models

CFA model	Local fit problem
2 Higher-Order[a]	Factor 1 (Verbal) and Factor 3 (*g*) linearly dependent on other parameters so variance estimate set to zero for model estimation and loss of 1 *df*
4a Bifactor	PC standardized path coefficient on PR (.100) and PS standardized path coefficient on WM (.127) not statistically significant
4b Higher-Order	Factor 3 (PR/WM) standardized path coefficient with *g* (.997) very high
4b Bifactor	MR (−.069), PC (.017), and PS (.124) standardized path coefficients on Factor 2 (PR/WM) not statistically significant, FW standardized path coefficient on Factor 2 (PR/WM) (−.159) statistically significant
4c Higher-Order	AR standardized path coefficient on PR (.125) not statistically significant
4d Higher-Order	AR standardized path coefficient on PR (.038) not statistically significant, AR standardized path coefficient on VC (.184) not practically significant
4e Higher-Order	Model 4e placed AR only on PR group factor with no cross-loadings (see Canivez, Watkins, Good, *et al.*, 2017), PR standardized path coefficient with *g* (.977) very high
4e Bifactor	PC standardized path coefficient on PR (.115) and PS standardized path coefficient on WM (.099) not statistically significant, and negative AR standardized path coefficient on PR (−.161)
5a Higher-Order	FR standardized path coefficient with *g* (.973) very high
5a Bifactor	MR and FW had negative loadings (−.029 and −.795, respectively) on FR and PC standardized path coefficient on FR not statistically significant
5b Higher-Order[a]	Negative variance estimate, FR standardized path coefficient with $g > 1.0$
5b Bifactor	MR standardized path coefficient on FR (.193), FW standardized path coefficient on FR (.125), PC standardized path coefficient on FR (.107), AR standardized path coefficient on FR (−.374), and PS standardized path coefficient on WM (.112) not statistically significant
5c Higher-Order[a]	Negative variance estimate, FR standardized path coefficient with $g > 1.0$
5d Higher-Order	AR standardized path coefficient on VC (.215) was weak; FR standardized path coefficient with *g* (.980) was very high
5e Higher-Order	AR standardized path coefficient (.063) on FR not statistically significant, AR standardized path coefficient (.192) on VC was low but statistically significant, removal of AR loading on FR produces Model 5d

Notes. Model number indicates the number of group factors included in the model, and model number and letter correspond to those reported in the WISC-V *Technical and Interpretive Manual* and the WISC-V[UK] *Administration and Scoring Manual Appendix D* (except 4e, which was added for comparison to Canivez, Watkins, *et al.*, 2017). Subtest assignments to latent factors are specified in Figures 2 and 3.
[a]Statistically inadmissible model.

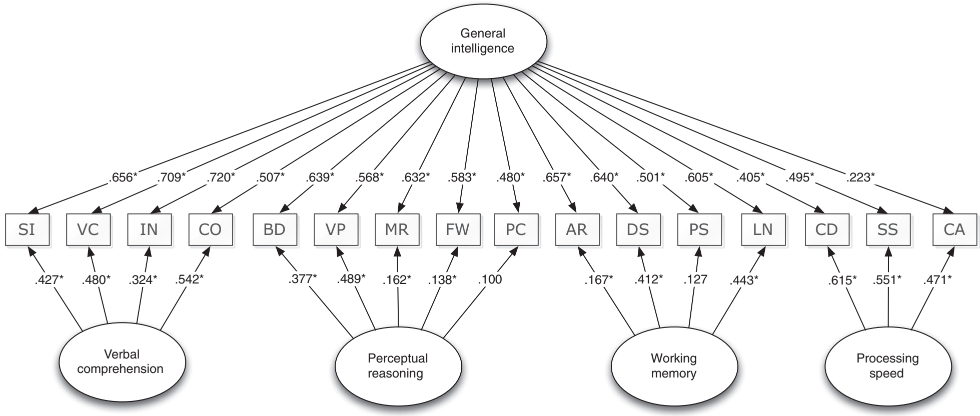

Figure 4. Bifactor measurement model (4a Bifactor), with standardized coefficients, for WISC-V^{UK} standardization sample (N = 415) 16 subtests. SI, Similarities; VC, Vocabulary; IN, Information; CO, Comprehension; BD, Block Design; VP, Visual Puzzles; MR, Matrix Reasoning; FW, Figure Weights; PC, Picture Concepts; AR, Arithmetic; DS, Digit Span; PS, Picture Span; LN, Letter-Number Sequencing; CD, Coding; SS, Symbol Search; CA, Cancellation. *p < .05.

unit-weighted WISC-V^{UK} factor scores (VCI, PRI, WMI, PSI) with all indicators were considerably lower, ranging from .142 (WM) to .452 (PS). For comparison, Table S4 presents variance sources for Model 4a higher-order illustrated in Figure S2. As shown in Table S4, and identical to the bifactor model, only the general intelligence dimension conveyed meaningful portions of true-score variance, while the four group factors conveyed little unique measurement and included low ω_{HS} coefficients.

Discussion

Results from the present EFA and CFA challenge the WISC-V^{UK} structure promoted in the WISC-V^{UK} *Administration and Scoring Manual*. Exploratory factor analysis results failed to support a five-factor model as only the MR subtest had a salient loading on the fifth factor. In contrast, four robust factors with theoretically consistent subtest associations resembling the traditional Wechsler structure emerged from the EFA. The present results replicated the outcomes of EFA studies of the WISC-V in the United States and in other countries in regard to the inadequate fifth factor (Canivez *et al.*, 2016, 2018; Dombrowski, Canivez, Watkins, & Beaujean, 2015; Dombrowski, Canivez, *et al.*, 2018; Lecerf & Canivez, 2018). Of interest, the AR subtest was the sole salient loading on the fifth factor in the French standardization sample but FW, MR, and VP subtests were singlets in the US sample depending on examinee age.

When modelling five first-order factors and one higher-order factor with all 16 primary and secondary subtests as promoted by the publisher, CFA approximate fit statistics appeared to be supportive. The publisher preferred WISC-V^{UK} model (Model 5e higher-order) included three cross-loadings of AR on VC, FR, and WM, but the standardized path coefficient of AR to FR was not statistically significant in the present study, and although the standardized path coefficient of AR to VC was statistically significant, it was low. Additionally, the FR factor loaded at .98 on the g factor, making those factors empirically redundant. These local misfits indicate that Model 5e higher-order (publisher preferred)

Table 5. Sources of variance in the WISC-V^{UK} 16 subtests for the total standardization sample (N = 415) according to confirmatory factor analysis (CFA) model 4a bifactor

	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed				
WISC-V^{UK} subtest	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2	ECV
Similarities	.656	.430	.427	.182							.613	.387	.702
Vocabulary	.709	.503	.480	.230							.733	.267	.686
Information	.720	.518	.324	.105							.623	.377	.832
Comprehension	.507	.257	.542	.294							.551	.449	.467
Block Design	.639	.408			.377	.142					.550	.450	.742
Visual Puzzles	.568	.323			.489	.239					.562	.438	.574
Matrix Reasoning	.632	.399			.162	.026					.426	.574	.938
Figure Weights	.583	.340			.138	.019					.359	.641	.947
Picture Concepts	.480	.230			.100	.010					.240	.760	.958
Arithmetic	.657	.432					.167	.028			.460	.540	.939
Digit Span	.640	.410					.412	.170			.579	.421	.707
Picture Span	.501	.251					.127	.016			.267	.733	.940
Letter-Number Sequencing	.605	.366					.443	.196			.562	.438	.651
Coding	.405	.164							.615	.378	.542	.458	.302
Symbol Search	.495	.245							.551	.304	.549	.451	.447
Cancellation	.223	.050							.471	.222	.613	.387	.702
Total variance		.333		.045		.024		.021		.054	.476	.524	
ECV		.698		.094		.050		.044		.114			
ω		.917		.870		.778		.769		.706			
ω_H / ω_{HS}		.829		.252		.145		.142		.452			

Note. b = standardized loading of subtest on factor, S^2 = variance explained, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω = omega, ω_H = omega-hierarchical (general factor), and ω_{HS} = omega-hierarchical subscale (group factors).

was *not* the best model. In contrast, CFA results supported a bifactor version of the WISC-V^{UK} structure with four group factors akin to the traditional Wechsler representation. That model exhibited no negative standardized path coefficients nor negative variance estimates and was also consistent with results from the WISC-IV^{UK}. However, that model was flawed by weak loadings of the PC and PSpan subtests on their respective factors. Similar results were observed with the Canadian, French, Spanish, and US WISC-V standardization samples where the publisher preferred Model 5e higher-order was not the best-fitting model, the FR and *g* factors were empirically redundant, and a bifactor version of the traditional Wechsler structure was preferred (Canivez, Watkins, *et al.*, 2017; Fennollar-Cortés & Watkins, 2018; Lecerf & Canivez, 2018; Watkins *et al.*, 2018).

Model-based reliability estimates from both WISC-V^{UK} EFA and CFA results indicated that the FSIQ score was sufficiently reliable for individual interpretation (*Md* ω_H = .82). Although the ω coefficients for the WISC-V^{UK} factor index scores were all above .70, the ω_{HS} estimates for those index scores were generally low (*Md* = .21; see Tables 2 and 5). This demonstrates that most of the factor index score reliability could be attributed to the general intelligence factor rather than the group factors. Scores with such low ω_{HS} estimates are extremely limited for measuring unique cognitive constructs (Brunner *et al.*, 2012; Reise, 2012; Reise *et al.*, 2013) and to interpret factor index scores with such low ω_{HS} values '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).

Thus, the WISC-V^{UK} factor index scores likely possess too little reliability beyond the influence of general intelligence to support confident clinical interpretation (Reise, 2012; Reise *et al.*, 2013; Rodriguez *et al.*, 2016). This outcome was predicted by Beaujean and Benson (2019), who contended that a strategy of creating cognitive instruments that measure both a general attribute (i.e., *g*) as well as more specific attributes (i.e., group factors) will result 'in creating less reliable scores of the specific attributes' (p. 5).

These EFA, CFA, and model-based reliability results are not unique to the WISC-V or WISC-V^{UK} nor to national standardization samples. Similar results have been observed in studies of the WISC-IV (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014b; Gomez, Vance, & Watson, 2016; Keith, 2005; Styck & Watkins, 2016; Watkins, 2006, 2010) and with other Wechsler scales (Canivez & Watkins, 2010; Canivez, Watkins, Good, *et al.*, 2017; Gignac, 2005, 2006; Golay & Lecerf, 2011; McGill & Canivez, 2016, 2017; Watkins & Beaujean, 2014; Watkins *et al.*, 2013). Nor are these results unique to Wechsler scales as similar findings have been reported with other cognitive scales (Canivez, 2008, 2011; Canivez, Konold, Collins, & Wilson, 2009; Canivez & McGill, 2016; Cucina & Howardson, 2017; DiStefano & Dombrowski, 2006; Dombrowski, 2013; Dombrowski, McGill, & Canivez, 2017; Dombrowski & Watkins, 2013; Dombrowski *et al.*, 2009; Dombrowski, McGill, *et al.*, 2018; Nelson & Canivez, 2012; Strickland, Watkins, & Caterino, 2015).

Limitations

The present study examined EFA and CFA for the full WISC-V^{UK} standardization sample, but it is possible that different age groups within the WISC-V^{UK} standardization sample might produce somewhat different results. Exploratory factor analysis and CFA with different age groups should be conducted to examine structural invariance across age. Other demographic variables where invariance should be examined include sex/gender and socioeconomic status. However, the WISC-V^{UK} standardization sample is considerably smaller than the WISC-IV^{UK} standardization sample so sampling error may affect such estimates and additional studies with new and much larger samples may be required. Further, the only available correlation matrix for the WISC-V^{UK} standardization sample is for the total sample (no separate matrices by age were provided by the publisher) so standardization sample raw data would be needed, something denied by NCS Pearson, Inc. for the present study.

Also, the present analyses were of the standardization sample and thus may not generalize to other populations such as clinical groups or independent samples of non-clinical groups, participants of different races/ethnicities, or language minorities. While structural invariance across gender has been reported for the US WISC-V (H. Chen *et al.*, 2015), bifactor models and models with fewer group factors were not examined so invariance of alternative models should also be examined across gender.

Of course, the results of the present study only pertain to the latent factor structure and do not fully test the construct validity of the WISC-V^{UK}, which would involve examinations of relations with external criteria (Canivez, 2013a). Examinations of incremental predictive validity (Canivez, 2013b; Canivez, Watkins, James, James, & Good, 2014; Glutting, Watkins, Konold, & McDermott, 2006; Nelson, Canivez, & Watkins, 2013) to determine whether reliable achievement variance is incrementally accounted for by the WISC-V^{UK} factor index scores beyond that accounted for by the FSIQ score (or through latent factor scores [see Kranzler, Benson, & Floyd, 2015]) and diagnostic utility (see Canivez, 2013a) studies should also be examined. Given the small portions of true-score variance uniquely contributed by the four group factors in the WISC-V^{UK} standardization sample, it seems unlikely that WISC-V^{UK} factor index scores will provide meaningful value (DeMars, 2013).

Finally, it has been suggested that fit indices in bifactor models might be statistically biased when compared to higher-order models due to unmodelled complexities (Murray & Johnson, 2013), proportionality constraints (Gignac, 2016), or violation of tetrad constraints (Mansolf & Reise, 2017). However, Morgan, Hodge, Wells, and Watkins (2015) found in their Monte Carlo simulations that the bifactor model 'did not generally produce a better fit when the true underlying structure was not a bi-factor one' (p. 15). There is no satisfactory statistical solution as to whether or why bifactor models might be biased (Mansolf & Reise, 2017). Fortunately, the preferred model (higher-order vs. bifactor) can be selected based on the purpose of measurement. As described by Murray and Johnson (2013), both models will provide a good estimate of *g*, the higher-order model may be more appropriate for testing factor to subtest paths in measurement models, and the bifactor model should be preferred when 'pure' measures of specific factors are desired because factor scores from a higher-order model 'conflate *g* and specific variance, so any associations with these scores will reflect (to possibly a very large extent) *g* rather than just the target specific ability' (p. 420). Given that scores from the WISC-V^{UK} will likely be used by psychologists to provide an estimate of general ability *and* to interpret cognitive strengths and weaknesses operationalized through the factor index scores as recommended by the publisher and popular textbooks (Sattler, Dumond, & Coalson, 2016; Wechsler, 2016b), it has been argued that a bifactor representation of its structure should be preferred (Murray & Johnson, 2013).

Conclusions

The WISC-V^{UK}, as presented in the WISC-V^{UK} *Administration and Scoring Manual,* appears to be overfactored (Beaujean & Benson, 2019; Frazier & Youngstrom, 2007) and the robust replication of previous EFA and CFA findings from the US WISC-V (Canivez *et al.*, 2016; Canivez, Watkins, *et al.*, 2017; Canivez, Watkins, Good, *et al.*, 2017 Canivez *et al.*, 2018; Dombrowski *et al.*, 2015), Canadian WISC-V (Watkins *et al.*, 2018), French WISC-V (Lecerf & Canivez, 2018), and Spanish WISC-V (Fennollar-Cortés & Watkins, 2018) further support that conclusion. The attempt to divide the PR factor into separate VS and FR factors appears to have been unsuccessful and therefore standard scores and comparisons for FRI scores are potentially misleading. If the publisher wishes to measure separate VS and FR factors then subtests that are stronger measures of the VS and FR factors and simultaneously poorer measures of *g* will be required; but, given the dominance of general intelligence in most cognitive subtests, there may still be too little unique variance captured to make such an endeavour fruitful (Rodriguez *et al.*, 2016).

As a result of the current study, psychologists in the United Kingdom and Ireland now have information to properly interpret WISC-V^{UK} scores according to the *Code of Good Practice for Psychological Testing* (British Psychological Society, 2007, 2016) and the *Guidelines on Test Use* (International Test Commission, 2013). Specifically, the WISC-V^{UK} may be best represented by a four-factor structure akin to the prior WISC-IV representation with factor index scores that contribute little reliable information beyond *g* because they conflate the variance from general intelligence and group factors and cannot, therefore, be interpreted as pure measures of Verbal Comprehension, Perceptual Reasoning, Visual Spatial Reasoning, Fluid Reasoning, Working Memory, or Processing Speed. In contrast, the FSIQ exhibited good reliability across factor methods and samples. In agreement with Dombrowski, Canivez, *et al.* (2018), Dombrowski, McGill, *et al.* (2018), we recommend that 'primary interpretive emphasis should be placed upon the FSIQ with only. . .secondary, yet extremely cautious, interpretive emphasis with the WISC-V index scores' (p. 100).

Acknowledgements

This research was partially supported by a 2017 Summer Research Grant from the Council on Faculty Research, Eastern Illinois University, to the first author. Preliminary results were presented at the 39th Annual Conference of the International School Psychology Association, Manchester, UK.

References

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

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

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

Beaujean, A. A. (2016). Reproducing the wechsler intelligence scale for children-fifth edition: Factor model results. *Journal of Psychoeducational Assessment*, *34*, 404–408. https://doi.org/0.1177/0734282916642679

Beaujean, A. A., & Benson, N. F. (2019). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*, *23*, 126–137. https://doi.org/10.1007/s40688-018-0182-1

Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment*, *26*, 789–805. https://doi.org/10.1037/a0036745

Bentler, P. M., & Wu, E. J. C. (2016). *EQS for windows*. Encino, CA: 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*, 417–424. https://doi.org/10. 1080/09297040802603661

Bridges, A. J., & Holler, K. A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology*, *13*, 528–538. https://doi.org/10.1080/09297040701233875

British Psychological Society (2007). *Psychological testing: A user's guide*. Leicester, UK: Author.

British Psychological Society (2016). *Code of good practice for psychological testing*. Leicester, UK: Author.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). New York, NY: Guilford.

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

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

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

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

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

Canivez, G. L. (2013a). 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). New York, NY: Oxford University Press.

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

Canivez, G. L. (2014a). Review of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. In J. F. Carlson, K. F. Geisinger & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook* (pp. 732–737). Lincoln, NE: Buros Center for Testing.

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

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

Canivez, G. L., 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*, 741–769. https://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*, 252–265. https://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*, 157–169. https://doi.org/10.1177/0734282913478036

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

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

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

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

Canivez, G. L., Watkins, M. W., Good, R., James, K., & James, T. (2017). Construct validity of the Wechsler Intelligence Scale for Children-Fourth UK Edition with a referred Irish sample: Wechsler and Cattell-Horn-Carroll model comparisons with 15 subtests. *British Journal of Educational Psychology*, *87*, 383–407. https://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*, 667–684. https://doi.org/10.1111/bjep.12056

Carroll, J. B. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence: Research methodology*, Vol. *1* (pp. 25–58). Norwood, NJ: Ablex.

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

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

Carroll, J. B. (1997). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S. E. Fienberg, D. P. Resnick & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to the Bell Curve* (pp. 125–156). New York, NY: Springer-Verlag. https://doi.org/10.1007/978-1-4612-0669-9

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–23). Mahwah, NJ: Erlbaum.

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

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

Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children-Fifth Edition. *Personality and Individual Differences*, *86*, 1–5. https://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*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

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*, 1–21. https://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*, 1001–1015. https://doi.org/10.1037/pas0000389

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, *13*, 354–378. https://doi.org/10.1080/1 5305058.2013.799067

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: Sage.

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

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, *23*, 225–241. https://doi.org/10.1177/073428290502300303

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

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

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

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

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

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

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

Fennollar-Cortés, J., & Watkins, M. W. (2018). Construct validity of the Spanish version of the Wechsler Intelligence Scale for Children Fifth Edition (WISC-V^{Spain}). *International Journal of School & Educational Psychology*. Advance online publication. https://doi.org/10.1080/21683603.2017.1414006

Ferrando, P. J., & Navarro-González, D. (2018). Assessing the quality and usefulness of factor-analytic applications to personality measures: A study with the statistical anxiety scale. *Personality and Individual Differences*, *123*, 81–86. https://doi.org/10.1016/j.paid.2017.11.014

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

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

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

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

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

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57–68. https://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*, 639–662. https://doi.org/10.1080/00273171.2013.804398

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAI-II. *Journal of Special Education*, *40*, 103–114. https://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*, 143–152. https://doi.org/10.1037/a0021230

Gomez, R., Vance, A., & Watson, S. (2016). Bifactor model of WISC-IV: Applicability and measurement invariance in low and normal IQ groups. *Psychological Assessment*, *29*, 902–912. https://doi.org/10.1037/pas0000369

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

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

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262027366.001.0001

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

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

Horn, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana-Champaign, IL: University of Illinois 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*, 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*, *5*, 1–55. https://doi.org/10.1080/10705519909540118

International Test Commission (2013). *International guidelines for test use (Version 1.2)*. Retrieved from https://www.intestcom.org

Jensen, A. R., & Weng, L.-J. (1994). What is a good *g*? *Intelligence*, *18*, 231–258. https://doi.org/10.1016/0160-2896(94)90029-9

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

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

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

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

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes*, *112*, 112–125. https://doi.org/10.1016/j.obhdp.2010.02.003

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

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

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*, 120–129. https://doi.org/10.1016/j.intell.2017.01.012

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

McDonald, R. P. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science*, *5*, 675–686. https://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*, *36*, 600–606. https://doi.org/10.1177/0734282915624293

McGill, R. J., & Canivez, G. L. (2017). 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*, 239–251. https://doi.org/10.1080/21683603.2017.1327831

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

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*, 277–288. https://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*, 407–422. https://doi.org/10.1016/j.intell.2013.06.004

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

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*, 618–630. https://doi.org/10.1037/a0032086

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. https://doi.org/10.1037/1089-2680.2.2.175

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

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

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

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

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

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, *6*, 1–8. https://doi.org/10.3389/fpsyg.2015.01715

Sattler, J. M., Dumond, R., & Coalson, D. L. (2016). *Assessment of children: WISC-V and WPPSI-IV*. San Diego, CA: Sattler.

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

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

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, *19*, 80–110. https://doi.org/10.1177/1094428115598239

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

Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment*, *27*, 689–697. https://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*, 216–224. https://doi.org/10.1177/0022219414539565

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

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

Tukey, J. W. (1980). We need both exploratory and confirmatory. *American Statistician*, *34*, 23–25. https://doi.org/10.2307/2682991

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

Watkins, M. W. (2000). *Monte Carlo PCA for parallel analysis [Computer software]*. Phoenix, AZ: Ed & Psych Associates.

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

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

Watkins, M. W. (2007). *SEscree [Computer software]*. Phoenix, AZ: Ed & Psych Associates.

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

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

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

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*, 219–246. https://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*, 52–63. https://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 and Educational Psychology*, *1*, 102–111. https://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 and Educational Psychology*, *6*, 252–265. https://doi.org/10.1080/21683603.2017.1342580

Wechsler, D. (2003). *Wechsler intelligence scale for children-fourth edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2004). *Wechsler intelligence scale for children—fourth* (UK ed.). London, UK: Harcourt Assessment.

Wechsler, D. (2008). *Wechsler adult intelligence scale—fourth edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence—fourth edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014a). *Wechsler intelligence scale for children-fifth edition*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2014b). *Wechsler intelligence scale for children-fifth edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2014c). *Wechsler intelligence scale for children-fifth edition: Canadian manual*. Toronto, ON: Pearson Canada Assessment.

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

Wechsler, D. (2016a). *Wechsler intelligence scale for children-fifth* (UK ed.). London, UK: Harcourt Assessment.

Wechsler, D. (2016b). *Wechsler intelligence scale for children-fifth UK edition: Administration and scoring manual*. London, UK: Harcourt Assessment.

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

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

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

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

Received 23 September 2017; revised version received 11 May 2018

Supporting Information

Additional supporting information may be found online in the supporting information section at the end of the article:

Figure S1. Scree plots for Horn's parallel analysis for WISC–V^{UK} standardization sample (N = 415).
Figure S2. Higher-order measurement model (4a), with standardized coefficients, for WISC–V^{UK} standardization sample (N = 415) 16 Subtests.
Table S1. Wechsler Intelligence Scale for Children-Fifth UK Edition (WISC–V^{UK}) exploratory factor analysis: Five oblique factor solution for the total standardization sample (N = 415).
Table S2. Wechsler Intelligence Scale for Children-Fifth UK Edition (WISC–V^{UK}) exploratory factor analysis: Two and three oblique factor solutions for the total standardization sample (N = 415).
Table S3. Sources of variance in the Wechsler Intelligence Scale for Children-Fifth UK Edition (WISC–V^{UK}) for the total standardization sample (N = 415) according to an exploratory SL bifactor model (orthogonalized higher-order factor model) with five first–order factors.
Table S4. Sources of variance in the WISC–V^{UK} 16 subtests for the total standardization sample (N = 415) according to CFA higher-order model 4a.