British Journal of Educational Psychology (2017), 87, 383–407


The British Psychological Society

www.wileyonlinelibrary.com

# 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

Gary L. Canivez[1]*, Marley W. Watkins[2], Rebecca Good[3], Kate James[3], and Trevor James[3]

[1]Eastern Illinois University, Charleston, Illinois, USA
[2]Baylor University, Waco, Texas, USA
[3]Éirim: The National Assessment Agency, Ltd., Dublin, Ireland

**Background.** Irish educational psychologists frequently use the Wechsler Intelligence Scale for Children – Fourth UK Edition (WISC–IV$^{UK}$; Wechsler, 2004, Wechsler Intelligence Scale for Children-Fourth UK Edition, London, UK, Harcourt Assessment) in clinical assessments of children with learning difficulties. Unfortunately, reliability and validity studies of the WISC–IV$^{UK}$ standardization sample have not yet been reported. Watkins *et al.* (2013, *International Journal of School and Educational Psychology*, 1, 102) found support for a bifactor structure with a large sample ($N$ = 794) of Irish children who were administered the 10 WISC–IV$^{UK}$ core subtests in clinical assessments of learning difficulties and dominance of general intelligence. Because only 10 subtests were available, Cattell–Horn–Carroll (CHC; McGrew, 1997, 2005, Contemporary intellectual assessment: Theories, tests, and issues, New York, NY: Guilford; Schneider & McGrew, 2012, Contemporary intellectual assessment: Theories, tests, and issues, New York, NY, Guilford Press) models could not be tested and compared.

**Aim, Sample and Method.** The present study utilized confirmatory factor analyses to test the latent factor structure of the WISC–IV$^{UK}$ with a sample of 245 Irish children administered all 15 WISC–IV$^{UK}$ subtests in evaluations assessing learning difficulties in order to examine CHC- and Wechsler-based models. One through five, oblique first-order factor models and higher order versus bifactor models were examined and compared using CFA.

**Results.** Meaningful differences in fit statistics were not observed between the Wechsler and CHC representations of higher-order or bifactor models. In all four structures, general intelligence accounted for the largest portions of explained common variance, whereas group factors accounted for small to miniscule portions of explained common variance. Omega-hierarchical subscale coefficients indicated that unit-weighted composites that would be generated by WISC–IV$^{UK}$ group factors (Wechsler or CHC) would contain little unique variance and thus be of little value.

*Correspondence should be addressed to Gary L. Canivez, Eastern Illinois University, 600 Lincoln Avenue, Charleston, Illinois 61920, USA (email: glcanivez@eiu.edu).

*Preliminary analyses on a smaller sample were presented at the 2016 Annual Convention of the American Psychological Association and the 10th Conference of the International Test Commission.*

DOI:10.1111/bjep.12155

**Conclusion.** These results were similar to those from other investigations, further demonstrating the replication of the WISC–IV factor structure across cultures and the importance of focusing primary interpretation on the FSIQ.

The popularity of Wechsler scales has led to their translation, adaptation, and norming for use in other countries with different languages and cultures (Georgas, van Vijver, Weiss, & Saklofske, 2003), and factor invariance evidence has been reported across cultures and between standardization and clinical samples (Chen, Keith, Weiss, Zhu, & Li, 2010; Chen & Zhu, 2012; Weiss, Keith, Zhu, & Chen, 2013a,b). Wechsler intelligence scales are also among the most frequently used tests among educational (school) psychologists and clinical psychologists (Alfonso, Oakland, LaRocca, & Spanakos, 2000; Alfonso & Pratt, 1997; Belter & Piotrowski, 2001; Goh, Teslow, & Fuller, 1981; Hutton, Dubes, & Muir, 1992; Kaufman & Lichtenberger, 2000; Oakland & Hu, 1992; Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Stinnett, Havey, & Oehler-Stinnett, 1994; Watkins, Campbell, Nieberding, & Hallmark, 1995).

During the revision of the Wechsler Intelligence Scale for Children – Third Edition (WISC–III; Wechsler, 1991) in the United States, the British version, the WISC–III$^{UK}$, was simultaneously revised and normed for use in the United Kingdom. The Wechsler Intelligence Scale for Children – Fourth Edition (WISC–IV; Wechsler, 2003a) included the addition of new subtests (Picture Concepts, Letter-Number Sequencing, Matrix Reasoning, Cancellation, and Word Reasoning) and deletion of others (Picture Arrangement, Object Assembly, and Mazes). The Full Scale IQ was retained as an estimate of general intelligence, but the Verbal and Performance IQs were deleted. Greater emphasis was placed on interpretation of factor-based index scores (Verbal Comprehension [VC], Perceptual Reasoning [PR], Working Memory [WM], and Processing Speed [PS]) (Wechsler, 2003b; Weiss, Saklofske, & Prifitera, 2005; Williams, Weiss, & Rolfhus, 2003). The WISC–IV revision for use in the United Kingdom with UK norms was published 1 year later as the Wechsler Intelligence Scale for Children – Fourth UK Edition (WISC–IV$^{UK}$; Wechsler, 2004).

The WISC–IV$^{UK}$ *Administration and Scoring Manual* provides a brief description of the standardization project including stratification and detailed information on administration, scoring, and analysis of index score and subtest score comparisons. Comparisons of raw score means and standard deviations between the UK standardization sample and the US standardization sample were provided in that manual, but no further examinations of the UK standardization sample were reported. The US-based WISC–IV *Technical and Interpretive Manual* (Wechsler, 2003b) is provided with the WISC–IV$^{UK}$, but is based on the US standardization sample and supplemental validity samples, *not* the UK standardization sample or UK-based validity samples. There is no mention in the US-based WISC–IV *Technical and Interpretive Manual* of psychometric analyses with the UK sample.

Although the WISC–IV$^{UK}$ *Administration and Scoring Manual* states, ‘confidence in WISC–IV$^{UK}$ score interpretation is based on the extensive US standardization study’ (Wechsler, 2004, p. 284), there are no reports of analyses beyond mean and standard deviation comparisons with the US sample. Raw score means and standard deviations were similar between the UK and US samples (Wechsler, 2004); however, reliability estimates and standard errors of measurement were based on the larger US sample and no validity data whatsoever were presented for the UK standardization sample. Searches of the extant literature found no studies reporting on the psychometric properties of the WISC–IV$^{UK}$ with the UK standardization sample. Understanding the internal structure of tests is essential for evaluating interpretability of provided scores (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Without extensive psychometric examination of the reliability, validity, and diagnostic efficiency/utility with the UK standardization sample, proper interpretation of WISC–IV$^{UK}$ scores remains unknown and thus an ethical challenge for psychologists using the WISC–IV$^{UK}$ based on the *Code of Ethics and Conduct* (BPS, 2009) and the *Code of Good Practice for Psychological Testing* (BPS, 2010).

Due to the absence of a technical manual with detailed psychometric analyses of the UK standardization sample, in 2013 the first author of this study requested from Pearson, United Kingdom, the WISC–IV$^{UK}$ standardization sample raw data to conduct analyses of internal structural validity, but the publisher denied access. Subsequently, the subtest scaled score correlation matrices and descriptive statistics by age were requested, but Pearson, United Kingdom, also refused to provide subtest correlation matrices, which are summary statistics customarily reported in most technical manuals (and available in the US version of the WISC–IV). Thus, in the 12 years since publication of the WISC–IV$^{UK}$ there still are no technical manual, technical reports, or peer-reviewed articles in the professional literature presenting psychometric examinations of reliability or validity on the UK standardization sample.

In the only known structural analysis of the WISC–IV$^{UK}$, Watkins, Canivez, James, Good, and James (2013) examined the latent factor structure with a large sample ($N$ = 794) of Irish children who were administered the 10 WISC–IV$^{UK}$ core subtests in clinical assessments of learning difficulties. One through four-first-order factor models and both higher-order and bifactor hierarchical models were tested with confirmatory factor analytic (CFA) methods and the bifactor model provided the best explanation of WISC–IV$^{UK}$ factor structure.

Gignac (2005, 2006, 2008) has described the higher-order representation of intelligence test structure as an indirect hierarchical model, where the *g* factor influences subtests *indirectly* through full mediation by the first-order factors (Yung, Thissen, & McLeod, 1999). Thus, *g* is conceptualized as a *superordinate* factor, which Thompson (2004) described as an abstraction from abstractions. Higher-order models have been commonly applied to assess the 'construct-relevant psychometric multidimensionality' (Morin, Arens, & Marsh, 2016; p. 117) of intelligence tests, but an alternative conceptualization is the bifactor model (alternatively referred to as a direct hierarchical [Gignac, 2005, 2006, 2008] or nested factors model [Gustafsson & Balke, 1993]) and was originally specified by Holzinger and Swineford (1937). Figure 1 illustrates hypothetical higher-order and bifactor models of the WISC–IV$^{UK}$. Gignac (2008) noted that in bifactor models, both the general (*g*) *and* the group factors *directly* influence the subtest indicators and *g* is conceptualized as a *breadth* factor. 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 (Gignac, 2008). Reise (2012) and Canivez (2016) noted several advantages of bifactor models 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. Gignac (2006) also noted that the bifactor model can be considered more conceptually parsimonious because it specifies a unidimensional

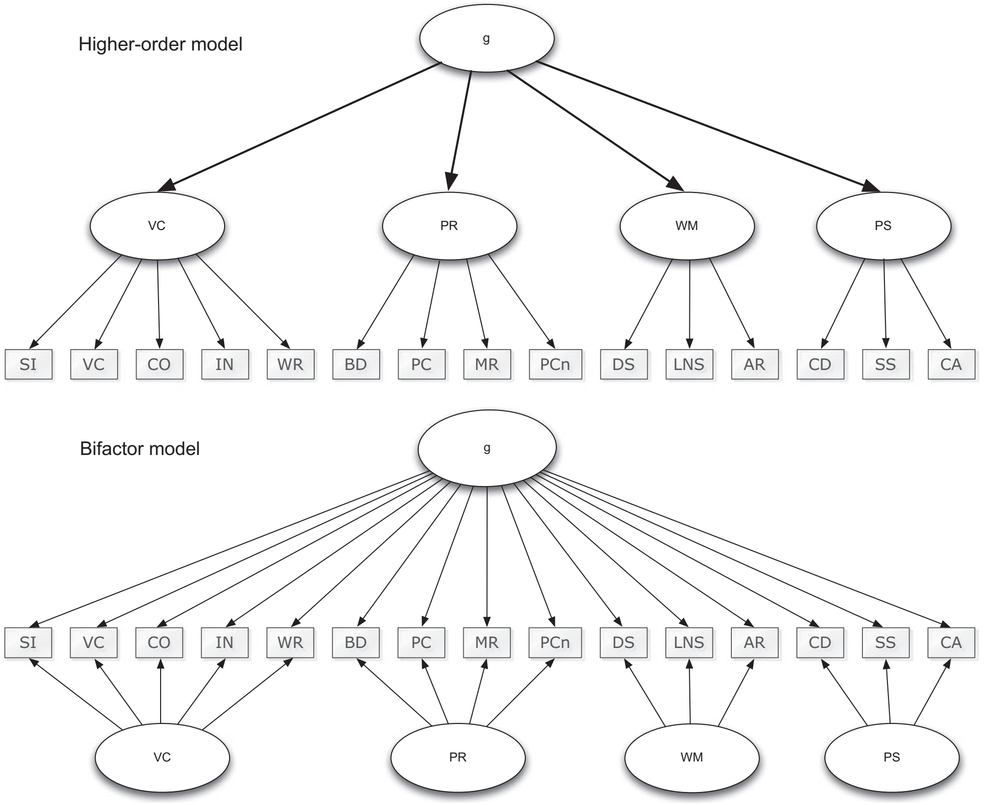


**Figure 1.** Hypothetical Wechsler-based higher-order and bifactor measurement models for the WISC–IV$^{UK}$ 15 Subtests. *g* = General Intelligence; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PC = Picture Completion; MR = Matrix Reasoning; PCn = Picture Concepts; DS = Digit Span; LNS = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation.

general factor and this model seems more consistent with Spearman's (1927) and Carroll's (1993) models of intelligence (Beaujean, 2015).

While Watkins *et al.* (2013) found the general intelligence factor accounted for 63.7% of the explained common variance, the four group factors (VC, PR, WM, PS) each accounted for <10% of the explained common variance. The omega-hierarchical coefficient for the general intelligence factor was .802 and satisfactory for confident interpretation (Reise, 2012; Reise, Bonifay, & Haviland, 2013); however, omega-hierarchical subscale coefficients for the four group factors (VC, PR, WM, PS) ranged from .143 to .376, failing to meet the recommended minimum standard of .50 (Reise, 2012; Reise *et al.*, 2013) for interpretation. It was noted that such results were difficult to place in context as no other studies of the WISC–IV$^{UK}$ factor structure with British (normative or clinical) or Irish samples were available for comparison. Further, because only the 10 WISC–IV$^{UK}$ core subtests were used (archival data), it was not possible to examine rival structural models based on the Cattell–Horn–Carroll framework (CHC; McGrew, 1997, 2005; Schneider & McGrew, 2012).

However, the WISC–IV[UK] factor structure identified by Watkins *et al.* (2013) with the Irish sample was consistent with results from other WISC–IV studies using both EFA and CFA (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Keith, 2005; Nakano & Watkins, 2013; Styck & Watkins, 2016; Watkins, 2006, 2010; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), with other versions of Wechsler scales (Canivez & Watkins, 2010a,b; Dombrowski, McGill, & Canivez, 2017; Gignac, 2005, 2006; Golay & Lecerf, 2011; Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010; McGill & Canivez, 2016; Nelson, Canivez, & Watkins, 2013; Niileksela, Reynolds, & Kaufman, 2013; Watkins & Beaujean, 2014), and intelligence tests in general (Canivez, 2008, 2011; Canivez, Konold, Collins, & Wilson, 2009; Canivez & McGill, 2016; DiStefano & Dombrowski, 2006; Dombrowski, 2013, 2014a,b; Dombrowski & Watkins, 2013; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007) in showing the largest portions of variance were captured by the *g* factor and small portions of variance were associated with group factors. Three recent studies of the WISC–V have also yielded identical results (Canivez, Watkins, & Dombrowski, 2016, 2017; Dombrowski, Canivez, Watkins, & Beaujean, 2015) with general intelligence dominating explained common variance and little unique explained common variance among the group factors. These results suggest that primary interpretation of these Wechsler scales (and other intelligence tests) should focus on the global score because it accounts for the largest portion of common variance. The global score has also been shown to be a powerful predictor of academic achievement and factor index scores provided little incremental validity in predicting academic achievement (Canivez, 2013; Canivez, Watkins, James, James, & Good, 2014; Glutting, Watkins, Konold, & McDermott, 2006; Kranzler, Benson, & Floyd, 2015; Nelson, *et al.*, 2013). The small portions of unique common variance captured by the first-order factors in factor analytic studies may be responsible for the poor incremental predictive validity of Wechsler factor scores.

Because the WISC–IV content and structure reflect conceptualizations of intelligence articulated by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Blankson, 2005; Horn & Cattell, 1966), studies have examined the WISC–IV internal structure with alternate structural models based on the Cattell, Horn, Carroll (CHC; McGrew, 1997, 2005; Schneider & McGrew, 2012) framework. The WISC–IV CHC-based models (Chen, Keith, Chen, & Chang, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Lecerf *et al.*, 2010; Weiss *et al.*, 2013b) retain some of the basic Wechsler structure for subtests and associations with Verbal Comprehension (VC; CHC $G_c$), Working Memory (WM; CHC $G_{sm}$) except Arithmetic, and Processing Speed (PS; CHC $G_s$); but, the WISC–IV Perceptual Reasoning (PR) dimension is split into two CHC factors where Block Design and Picture Completion purportedly measure visual processing ($G_v$) and Matrix Reasoning and Picture Concepts purportedly measure fluid reasoning ($G_f$). Canivez and Kush (2013) pointed out numerous problems with the proposed CHC models for the WAIS–IV and WISC–IV (Weiss *et al.*, 2013a,b). Standardized path coefficients from *g* to $G_f$ were 1.0 with the US WISC–IV standardization sample (Keith, 2005; Keith *et al.*, 2006; Weiss *et al.*, 2013b), .98 with the Taiwan WISC–IV (H.–Y. Chen *et al.*, 2009), and 1.0 with French WISC–IV (Lecerf *et al.*, 2010) basic CHC model patterned after Keith *et al.* (2006), but only .84 with the final modified six-factor CHC model of the French WISC–IV (Lecerf *et al.*, 2010). This suggested that $G_f$ was often isomorphic with the higher-order *g* factor and not supportive of a CHC model. The exception was the modified six-factor CHC model of the French WISC–IV that Lecerf *et al.* (2010) suggested may be due to cultural differences. Isomorphism of $G_f$ with higher-order *g* has also been observed in studies of versions of the Wechsler Adult Intelligence Scale as well (Benson, Hulac, & Kranzler, 2010; Golay & Lecerf, 2011; Weiss *et al.*, 2013a), but a recent study suggested that isomorphism of $G_f$ with higher-order *g* could be an artefact of the CFA statistical method (Golay *et al.*, 2013). It is also possible that the Golay *et al.* results may be unique to the French WISC–IV so further assessment with other Wechsler tests would be useful.

Because the Watkins *et al.* (2013) study of the latent structure of the WISC–IV$^{UK}$ could not assess rival CHC models due to the availability of only the 10 core subtests, this study involved a sample of Irish children referred for evaluations of learning difficulties where all 15 WISC–IV$^{UK}$ subtests were administered. This allowed a comparison of both Wechsler- *and* CHC-based measurement models. In addition, this study, like Watkins *et al.* (2013), examined both higher-order *and* bifactor models to determine best fit to these data.

## Method

### *Participants and procedure*

Participants were 245 children from the Republic of Ireland between the ages of 6–0 and 16–10 who were referred to an educational psychologist for evaluation of learning difficulties. Although some children were referred for evaluation by their parents, the vast majority were referred by their schools to determine eligibility for special education services or accommodations. The largest portion of the sample was male ($n$ = 138, 56.3%) as is typically observed in educational evaluation referrals. The mean age of the sample was 11.05 ($SD$ = 2.72) years. Age distribution showed somewhat larger proportions of referred children were ages 8, 9, 10, and 12. The largest number of assessments were provided for children from Dublin (50.6%) followed by Kildare (7.3%), Laois (6.1%), Wicklow (5.3%), and Tipperary (4.5%). Unfortunately, agency practice and confidentiality standards allowed no other demographic information to be included in this archival data set.

All WISC–IV$^{UK}$ administrations were conducted by one of three registered/licensed/certified educational psychologists according to the standardized procedure. Only children with complete data for all 15 WISC–IV$^{UK}$ subtests were included in analyses. Institutional review board approval was obtained but all data were de-identified and no personal information included.

### *Instrument*

The WISC–IV (Wechsler, 2003a,b) is a general intelligence test that is composed of 15 subtests (*M*s = 10, *SD*s = 3), 10 of which are mandatory and contribute to measurement of four-factor-based index scores: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). Each of the four index scores is expressed as a standard score (*M*s = 100, *SD*s = 15). The FSIQ is composed of 10 core subtests (three Verbal Comprehension, three Perceptual Reasoning, two Working Memory, and two Processing Speed).

The WISC–IV was anglicized and adapted for the United Kingdom (UK) in 2002 through item review and minor changes in items or language, spelling, and order of item difficulty (Wechsler, 2004). The resulting WISC–IV$^{UK}$ was standardized and normed on a sample of 780 children between the ages of 6–0 and 16–11 years who were representative of the UK population stratified by geographic region, sex, race/ethnicity, and parent education level (Wechsler, 2004). Of the 780 children in the standardization sample, 17 (2.2%) were from Northern Ireland. There are no separate norms for children in Ireland generally or the Republic of Ireland specifically. Reliability and validity data based on the WISC–IV$^{UK}$ standardization sample were not provided in the WISC–IV$^{UK}$ manual and standard errors of measurement were taken from the US version of the WISC–IV. There is no separate technical manual presenting psychometric information relating to the factor structure of the WISC–IV$^{UK}$ based on the UK standardization sample.

### *Analyses*

Confirmatory factor analyses (CFA) with maximum-likelihood estimation was conducted with the covariance matrix using EQS 6.2 (Bentler & Wu, 2012). Some CHC-based first-order factors were underidentified because they included only two subtest indicators. In those CFAs, the two subtests were constrained to equality prior to estimating bifactor models to ensure identification (Little, Lindenberger, & Nesselroade, 1999).

The structural models examined were similar to those previously specified (Watkins *et al.*, 2013) but because all 15 WISC–IV$^{UK}$ subtests were administered additional CHC-based models could also be tested. Although there are no universally accepted cut-off values for approximate fit indices (Marsh, Hau, & Wen, 2004; McDonald, 2010), overall model fit was evaluated using the comparative fit index (CFI), standardized root-mean-squared residual (SRMR), 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 SRMR and RMSEA. Applying the combinatorial heuristics of Hu and Bentler (1999), criteria for adequate model fit were CFI $\geq$ .90, SRMR $\leq$ .09, and RMSEA $\leq$ .08. Good model fit required CFI $\geq$ 0.95 with SRMR and RMSEA $\leq$ 0.06 (Hu & Bentler, 1999). Statistical comparisons between models were made using the *ChiSquareDiff* program (Watkins, 2012). For a model to be considered superior, it had to exhibit adequate to good overall fit *and* display meaningfully better fit ($\Delta$CFI > .01 and $\Delta$RMSEA > .015) than alternative models (Cheung & Rensvold, 2002; Chen, 2007). Additionally, the Akaike information criterion (AIC) was considered, but the AIC does not have a meaningful scale so the model with the smallest AIC values was preferred as such models are most likely to replicate (Kline, 2016).

Model-based reliabilities were estimated with coefficients omega-hierarchical ($\omega_H$) and omega-hierarchical subscale ($\omega_{HS}$), which estimate reliability of the unit-weighted composite scores produced by the indicators (Reise, 2012; Rodriguez, Reise, & Haviland, 2016). The $\omega_H$ is a model-based reliability estimate for the general intelligence factor with variability of group factors removed. The $\omega_{HS}$ is a model-based reliability estimate of a group factor with the influence of all other group *and* general factors removed (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012). Omega estimates ($\omega_H$ and $\omega_{HS}$) may be obtained from CFA bifactor solutions or decomposed variance estimates from higher-order models and were produced using the *Omega* program (Watkins, 2013), which is based on the tutorial by Brunner *et al.* (2012) and the work of Zinbarg, Revelle, Yovel, and Li (2005) and Zinbarg, Yovel, Revelle, and McDonald (2006). Omega coefficients should at a minimum exceed .50, but .75 is preferred (Reise, 2012; Reise *et al.*, 2013). The value of $\omega_H$ and $\omega_{HS}$ is that one may determine the relative merit of how much true score variance would be provided by a unit-weighted score based on specified subtest indicators, and if < 50% true score variance was uniquely captured, this would not indicate useful measurement of that construct.

## Results

Descriptive statistics for participants' WISC–IV$^{UK}$ subtest, factor index, and FSIQ scores are presented in Table 1 and illustrate univariate normality with the largest subtest skewness index of .72 (Arithmetic) and the largest subtest kurtosis index of .68 (Arithmetic). Mardia's (1970) standardized multivariate kurtosis estimate for these data was 1.25 and well under the criterion of |5.0| for multivariate normality (Byrne, 2006) and thus appropriate use of maximum-likelihood estimation. WISC–IV$^{UK}$ means for this sample were approximately one standard deviation lower than the normative means and there was somewhat less variability observed among participants. Lower subtest, factor index, and FSIQ scores in referred samples are frequently observed (Canivez & Watkins, 1998; Watkins, 2010; Watkins *et al.*, 2013).

Model fit statistics presented in Table 2 illustrate the increasingly better fit from 1 through 4 oblique factors; however, fit statistics indicated that the one-, two-, and three-factor models were inadequate using combinatorial criteria (Hu & Bentler, 1999) as all had RMSEA $\geq$ .08. The oblique four-factor (VC, PR, WM, PS) Wechsler-based model and oblique five-factor ($G_c$, $G_v$, $G_f$, $G_{sm}$, $G_s$) CHC-based model provided the best fit to these data, but meaningful differences in fit statistics (CFI, RMSEA, SRMR) were not observed. The oblique five-factor CHC model did not produce a statistically significant better fit than the oblique four-factor Wechsler model, $\Delta ML\chi^2 = 5.46$, $\Delta df = 4$, $p = .243$. Although the Wechsler- and CHC-based oblique models fit these data well, the latent factor correlations (Table 3) for both models (Wechsler $r_s$ ranging .532–.819; CHC $r_s$ ranging .532–.949)

**Table 1.** WISC–IV$^{UK}$ descriptive statistics for 245 referred Irish children

| Score | *M* | *SD* | Skewness | Kurtosis |
|---|---|---|---|---|
| Subtest | | | | |
| Block Design | 7.89 | 3.02 | 0.25 | −0.17 |
| Similarities | 9.12 | 2.73 | 0.14 | −0.51 |
| Digit Span | 7.56 | 2.64 | 0.17 | 0.07 |
| Picture Concepts | 8.93 | 3.10 | −0.24 | 0.01 |
| Coding | 8.31 | 2.96 | 0.16 | 0.39 |
| Vocabulary | 7.34 | 2.85 | 0.32 | 0.25 |
| Letter-Number Sequencing | 7.86 | 2.75 | −0.62 | −0.09 |
| Matrix Reasoning | 7.90 | 2.96 | 0.19 | 0.09 |
| Comprehension | 8.21 | 2.48 | −0.06 | −0.12 |
| Symbol Search | 8.88 | 2.77 | −0.54 | 0.37 |
| Picture Completion | 8.77 | 2.68 | 0.14 | 0.17 |
| Cancellation | 11.20 | 3.04 | −0.35 | 0.20 |
| Information | 7.89 | 2.91 | 0.31 | −0.41 |
| Arithmetic | 7.02 | 2.61 | 0.72 | 0.68 |
| Word Reasoning | 9.00 | 2.73 | 0.06 | −0.18 |
| Composite | | | | |
| Verbal Comprehension Index (VCI) | 89.57 | 13.27 | −0.14 | 0.07 |
| Perceptual Reasoning Index (PRI) | 89.00 | 16.40 | −0.50 | 1.49 |
| Working Memory Index (WMI) | 86.40 | 13.24 | −0.28 | −0.20 |
| Perceptual Speed Index (PSI) | 92.45 | 13.85 | −0.19 | 0.24 |
| Full Scale IQ (FSIQ) | 87.07 | 14.17 | −0.20 | −0.02 |

*Note.* Mardia's (1970) standardized multivariate kurtosis estimate = 1.25.

**Table 2.** CFA fit statistics for WISC–IV^UK^ among 245 referred Irish children

| Model | $\chi^2$ | *df* | CFI | RMSEA | 90% CI RMSEA | SRMR | AIC |
|---|---|---|---|---|---|---|---|
| One factor | 409.99 | 90 | .837 | .121 | .109, .132 | .074 | 229.99 |
| Two oblique factors | 279.07 | 89 | .903 | .094 | .081, .106 | .059 | 101.07 |
| Three oblique factors | 243.81 | 87 | .920 | .086 | .073, .099 | .057 | 69.81 |
| Four oblique factors (Wechsler) | 148.80 | 84 | .967 | .056 | .041, .071 | .037 | −19.20 |
| Five oblique factors (CHC) | 143.34 | 80 | .968 | .057 | .041, .072 | .036 | −16.66 |
| Wechsler higher-order | 156.86 | 86 | .964 | .058 | .043, .072 | .040 | −15.14 |
| CHC higher-order | 170.68 | 85 | .956 | .064 | .050, .078 | .045 | 0.68 |
| Wechsler bifactor | 138.02 | 75 | .968 | .059 | .043, .074 | .036 | −11.98 |
| CHC bifactor[a] | 155.83 | 77 | .960 | .065 | .050, .079 | .040 | 1.83 |

*Note.* CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual, AIC = Akaike's information criterion, CHC = Cattell-Horn–Carroll.
[a]Two indicators of the second ($G_v$) and third ($G_f$) factors were constrained to be equal to ensure model identification.

**Table 3.** Latent factor correlations for WISC–IV^UK^ Wechsler- and CHC-based oblique models for referred Irish sample (*N* = 245)

| Wechsler model | VC | PR | WM | PS | |
|---|---|---|---|---|---|
| VC | – | | | | |
| PR | .757 | – | | | |
| WM | .819 | .785 | – | | |
| PS | .532 | .663 | .572 | – | |
| **CHC model** | $G_c$ | $G_v$ | $G_f$ | $G_{sm}$ | $G_s$ |
| $G_c$ | – | | | | |
| $G_v$ | .758 | – | | | |
| $G_f$ | .726 | .949 | – | | |
| $G_{sm}$ | .819 | .745 | .803 | – | |
| $G_s$ | .532 | .669 | .628 | .572 | – |

*Note.* VC = Verbal Comprehension, PR = Perceptual Reasoning, WM = Working Memory, PS = Processing Speed, CHC = Cattell–Horn–Carroll, $G_c$ = Crystallized Intelligence/Comprehension Knowledge, $G_v$ = Visual-Spatial, $G_f$ = Fluid Intelligence/Fluid Reasoning, $G_{sm}$ = Short-term Memory (Working Memory), $G_s$ = Processing Speed.

were moderate to very high and thus these models were deemed inadequate as a general intelligence factor is suggested and required explication (Canivez, 2016; Gorsuch, 1988; Reise, 2012; Thompson, 2004). Further, in the oblique CHC model the $G_v$ and $G_f$ factor correlation of .949 indicated considerable overlap and potential lack of discriminant validity (Kline, 2016).

Combinatorial heuristics (Hu & Bentler, 1999) indicated the CHC-based higher-order and bifactor models exhibited adequate fits to these data (although RMSEA slightly exceeded .06) and the Wechsler-based higher-order and bifactor models were good fits to these data and also produced the lowest AIC values and thus were most likely to replicate (Kline, 2016). The Wechsler bifactor model did not produce a statistically significant better fit than the Wechsler higher-order model, $\Delta ML\chi^2 = 18.84$, $\Delta df = 11$, $p = .064$. The CHC bifactor model did not produce a statistically significant better fit than the CHC higher-order model, $\Delta ML\chi^2 = 14.85$, $\Delta df = 8$, $p = .062$. However, the Wechsler bifactor model produced a statistically significant better fit than the CHC bifactor model, $\Delta ML\chi^2 = 17.81$, $\Delta df = 2$, $p < .00001$; and the Wechsler bifactor model produced the lower AIC. There were no meaningful differences ($\Delta$CFI > .01 and $\Delta$RMSEA > .015) between the Wechsler higher-order (Figure 2), CHC higher-order (Figure 3), Wechsler bifactor (Figure 4), and CHC bifactor (Figure 5) models so all are presented for comparison and illustration.

Tables 4–7 present decomposed variance estimates based on the four different hierarchical models for comparison. Explained common variance (ECV) was dominated by the *g* factor in all four models ranging from .688 to .750. The $\omega_H$ coefficients for the *g* factor in all four models were high, ranging from .854 to .869, and exceeded the .75 criterion for confident interpretation (Reise, 2012; Reise *et al.*, 2013). Explained common variance among the Wechsler (VC, PR, WM, PS)- and CHC ($G_c$, $G_v$, $G_f$, $G_{sm}$, $G_s$)-based group factors was considerably less, ranging from .015 to .117. Particularly low were ECVs for the CHC-based $G_v$ and $G_f$ group factors with ECV coefficients < .02. The $\omega_{HS}$ coefficients for the Wechsler- and CHC-based group factors were also low, ranging from .078 to .429; all falling short of the suggested minimum .50 criterion (Reise, 2012; Reise *et al.*, 2013). Consistent with the ECV estimates, $\omega_{HS}$ coefficients were particularly low for CHC-based $G_v$ and $G_f$ group factors with $\omega_{HS} < .10$, meaning unit-weighted composite scores based on $G_v$ and $G_f$ subtest indicators would account for < 10% unique true score variance.

## Discussion

A previous examination of the WISC–IV$^{UK}$ internal structure (Watkins *et al.*, 2013) could only examine Wechsler-based models because only the 10 core subtests were administered and would not include enough indicators to estimate all CHC latent constructs. In

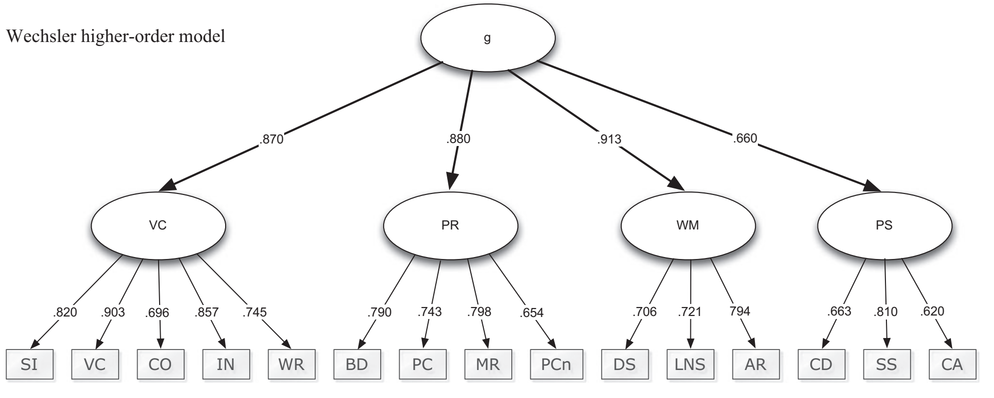


**Figure 2.** Wechsler-based higher-order measurement model, with standardized coefficients, for WISC–IV$^{UK}$ from the Irish referral sample ($N = 245$) 15 Subtests. *g* = General Intelligence; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PC = Picture Completion; MR = Matrix Reasoning; PCn = Picture Concepts; DS = Digit Span; LNS = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation.

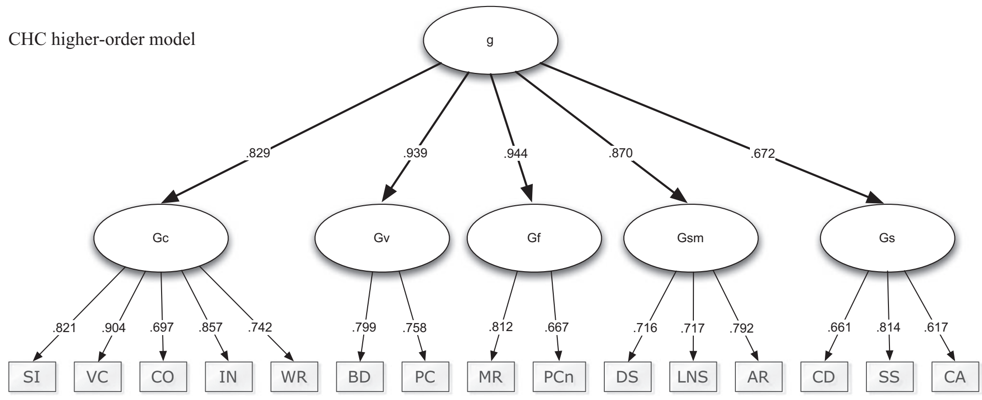


**Figure 3.** CHC-based higher-order measurement model, with standardized coefficients, for WISC–IV[UK] from the Irish referral sample ($N$ = 245) 15 Subtests. $g$ = General Intelligence; $G_c$ = Crystallized Intelligence; $G_v$ = Visual Processing; $G_f$ = Fluid Reasoning; $G_{sm}$ = Short-term Memory; $G_s$ = Processing Speed; SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PC = Picture Completion; MR = Matrix Reasoning; PCn = Picture Concepts; DS = Digit Span; LNS = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation.

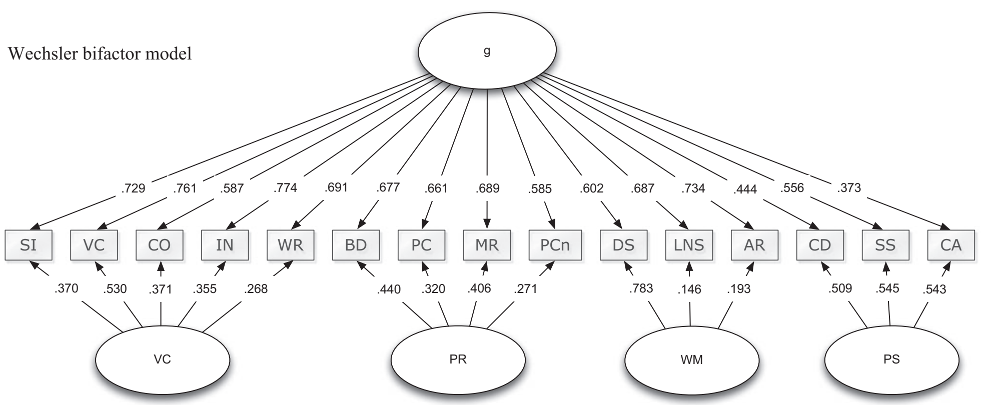


**Figure 4.** Wechsler-based bifactor measurement model, with standardized coefficients, for WISC–IV[UK] from the Irish referral sample ($N$ = 245) 15 Subtests. $g$ = General Intelligence; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PC = Picture Completion; MR = Matrix Reasoning; PCn = Picture Concepts; DS = Digit Span; LNS = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation.

the current study, confirmatory factor analyses were conducted with all 15 core and supplemental WISC–IV[UK] subtests from a sample of Irish children administered the WISC–IV[UK] in clinical evaluations to examine both Wechsler- and CHC-based structures. The CHC higher-order and bifactor models achieved adequate fit to these data, while the Wechsler higher-order and bifactor models achieved good fit to these data. The Wechsler models had lower AIC values compared to rival CHC-based models, but meaningful

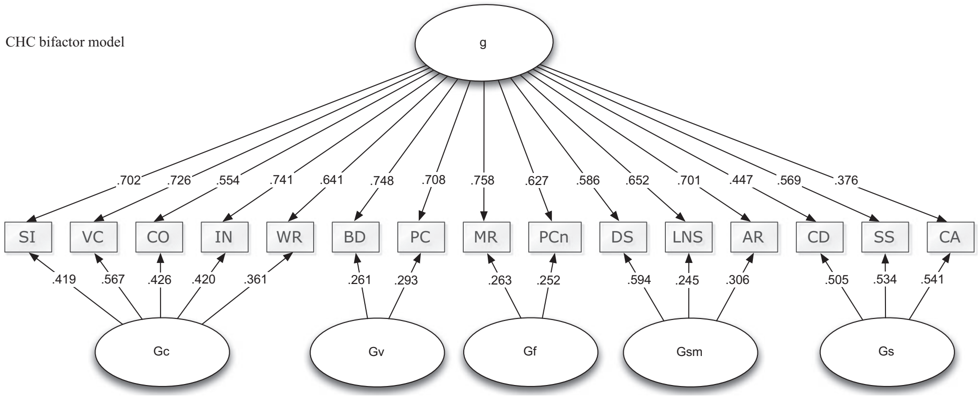


**Figure 5.** CHC-based bifactor measurement model, with standardized coefficients, for WISC–IV[UK] from the Irish referral sample ($N$ = 245) 15 Subtests. g = General Intelligence; $G_c$ = Crystallized Intelligence; $G_v$ = Visual Processing; $G_f$ = Fluid Reasoning; $G_{sm}$ = Short-term Memory; $G_s$ = Processing Speed; SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PC = Picture Completion; MR = Matrix Reasoning; PCn = Picture Concepts; DS = Digit Span; LNS = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation.

differences in fit statistics were not observed between Wechsler and CHC models. Thus, it could be argued that bifactor and higher-order representations of Wechsler and CHC structures explained these data equally well.

Reynolds and Keith (2013) questioned the theoretical appropriateness of bifactor models of intelligence, stating that ‘we believe that higher-order models are theoretically more defensible, more consistent with relevant intelligence theory (e.g., Jensen, 1998), than are less constrained hierarchical [bifactor] models’ (p. 66). Others have challenged this position. Gignac (2006, 2008) argued that the general intelligence factor is the most substantial dimension of a battery of cognitive tests and should be modelled directly, whereas full mediation of general intelligence in the higher-order model demands explicit theoretical justification. A rationale for why general intelligence should directly influence group factors but not subtests seems necessary. Subtest scores reflect variation on both a general and a more specific group factor. As a result, subtest scores may appear reliable, but the reliability estimate is primarily a function of the general factor, *not* the specific group factor. Other researchers maintained that the bifactor model better represents Spearman’s (1904, 1927) and Carroll’s (1993) conceptualizations of intelligence (Beaujean, 2015; Frisby & Beaujean, 2015; Brunner *et al.*, 2012; Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993). Beaujean (2015) noted that Spearman’s conception of general intelligence was of a factor ‘directly involved in all cognitive performances, not indirectly involved through, or mediated by, other factors’ (p. 130), and that ‘Carroll was explicit in noting that a bi-factor model best represents his theory’ (p. 130).

Murray and Johnson (2013) suggested that bifactor models might benefit from statistical bias when compared to higher-order models by better accounting for unmodelled complexity. However, Monte Carlo simulations found that the bifactor model ‘did not generally produce a better fit when the true underlying structure was not a bi-factor one’ (Morgan, Hodge, Wells, & Watkins, 2015; p. 15). Regardless, Murray and

**Table 4.** Decomposed variance sources for the WISC–IV$^{UK}$ 15 subtests for the referred Irish sample ($N$ = 245) according to a Wechsler higher-order model

| WISC–IV$^{UK}$ | General | | Verbal Comprehension | | Perceptual Reasoning | | Working Memory | | Processing Speed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subtest | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $h^2$ | $u^2$ |
| Similarities | .713 | .508 | .404 | .163 | | | | | | | .672 | .328 |
| Vocabulary | .786 | .618 | .445 | .198 | | | | | | | .816 | .184 |
| Comprehension | .606 | .367 | .344 | .118 | | | | | | | .486 | .514 |
| Information | .746 | .557 | .423 | .179 | | | | | | | .735 | .265 |
| Word Reasoning | .648 | .420 | .367 | .135 | | | | | | | .555 | .445 |
| Block Design | .695 | .483 | | | .375 | .141 | | | | | .624 | .376 |
| Picture Completion | .654 | .428 | | | .353 | .125 | | | | | .552 | .448 |
| Matrix Reasoning | .702 | .493 | | | .379 | .144 | | | | | .636 | .364 |
| Picture Concepts | .576 | .332 | | | .311 | .097 | | | | | .428 | .572 |
| Digit Span | .645 | .416 | | | | | .289 | .084 | | | .500 | .500 |
| Letter-Number Sequencing | .658 | .433 | | | | | .294 | .086 | | | .519 | .481 |
| Arithmetic | .725 | .526 | | | | | .325 | .106 | | | .631 | .369 |
| Coding | .438 | .192 | | | | | | | .499 | .249 | .441 | .559 |
| Symbol Search | .535 | .286 | | | | | | | .608 | .370 | .656 | .344 |
| Cancellation | .409 | .167 | | | | | | | .465 | .216 | .384 | .616 |
| Total Variance | | .415 | | .053 | | .034 | | .018 | | .056 | .576 | .424 |
| ECV | | .721 | | .092 | | .059 | | .032 | | .097 | | |
| $\omega_H/\omega_{HS}$ | | .854 | | .220 | | .188 | | .131 | | .419 | | |

*Note.* $b$ = standardized loading of subtest on factor, $S^2$ = variance, $h^2$ = communality, $u^2$ = uniqueness, ECV = explained common variance, $\omega_H$ = omega-hierarchical (general factor), $\omega_{HS}$ = omega-hierarchical subscale (group factors).

**Table 5.** Decomposed variance sources for the WISC–IV$^{UK}$ 15 subtests for the referred Irish sample ($N$ = 245) according to a Wechsler bifactor model

| | General | | Verbal Comprehension | | Perceptual Reasoning | | Working Memory | | Processing Speed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WISC–IV$^{UK}$ Subtest | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $h^2$ | $u^2$ |
| Similarities | .677 | .458 | .370 | .137 | | | | | | | .595 | .405 |
| Vocabulary | .761 | .579 | .530 | .281 | | | | | | | .860 | .140 |
| Comprehension | .587 | .345 | .371 | .138 | | | | | | | .482 | .518 |
| Information | .774 | .599 | .355 | .126 | | | | | | | .725 | .275 |
| Word Reasoning | .691 | .477 | .268 | .072 | | | | | | | .549 | .451 |
| Block Design | .677 | .458 | | | .440 | .194 | | | | | .652 | .348 |
| Picture Completion | .661 | .437 | | | .320 | .102 | | | | | .539 | .461 |
| Matrix Reasoning | .689 | .475 | | | .406 | .165 | | | | | .640 | .360 |
| Picture Concepts | .585 | .342 | | | .271 | .073 | | | | | .416 | .584 |
| Digit Span | .602 | .362 | | | | | .783 | .613 | | | .975 | .025 |
| Letter-Number Sequencing | .687 | .472 | | | | | .146 | .021 | | | .493 | .507 |
| Arithmetic | .734 | .539 | | | | | .193 | .037 | | | .576 | .424 |
| Coding | .444 | .197 | | | | | | | .509 | .259 | .456 | .544 |
| Symbol Search | .556 | .309 | | | | | | | .545 | .297 | .606 | .394 |
| Cancellation | .373 | .139 | | | | | | | .543 | .295 | .434 | .566 |
| Total Variance | | .413 | | .050 | | .036 | | .045 | | .057 | .600 | .400 |
| ECV | | .688 | | .084 | | .059 | | .075 | | .095 | | |
| $\omega_H/\omega_{HS}$ | | .854 | | .204 | | .194 | | .200 | | .429 | | |

*Note.* $b$ = standardized loading of subtest on factor, $S^2$ = variance, $h^2$ = communality, $u^2$ = uniqueness, ECV = explained common variance, $\omega_H$ = omega-hierarchical (general factor), $\omega_{HS}$ = omega-hierarchical subscale (group factors).

**Table 6.** Decomposed variance sources for the WISC–IV^UK 15 subtests for the referred Irish sample ($N$ = 245) according to a CHC higher-order model

| | General | | $G_c$ | | $G_v$ | | $G_f$ | | $G_{sm}$ | | $G_s$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WISC–IV^UK Subtest | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $h^2$ | $u^2$ |
| Similarities | .681 | .464 | .459 | .211 | | | | | | | | | .674 | .326 |
| Vocabulary | .749 | .561 | .504 | .254 | | | | | | | | | .815 | .185 |
| Comprehension | .578 | .334 | .390 | .152 | | | | | | | | | .486 | .514 |
| Information | .710 | .504 | .479 | .229 | | | | | | | | | .734 | .266 |
| Word Reasoning | .615 | .378 | .414 | .171 | | | | | | | | | .550 | .450 |
| Block Design | .750 | .562 | | | .274 | .075 | | | | | | | .638 | .362 |
| Picture Completion | .712 | .507 | | | .262 | .069 | | | | | | | .576 | .424 |
| Matrix Reasoning | .767 | .588 | | | | | .269 | .072 | | | | | .661 | .339 |
| Picture Concepts | .630 | .397 | | | | | .220 | .048 | | | | | .445 | .555 |
| Digit Span | .623 | .388 | | | | | | | .354 | .125 | | | .513 | .487 |
| Letter-Number Sequencing | .624 | .389 | | | | | | | .353 | .125 | | | .514 | .486 |
| Arithmetic | .689 | .475 | | | | | | | .390 | .152 | | | .627 | .373 |
| Coding | .444 | .197 | | | | | | | | | .490 | .240 | .437 | .563 |
| Symbol Search | .708 | .501 | | | | | | | | | .401 | .161 | .662 | .338 |
| Cancellation | .537 | .288 | | | | | | | | | .305 | .093 | .381 | .619 |
| Total Variance | | .436 | | .068 | | .00 | | .008 | | .027 | | .033 | .581 | .419 |
| ECV | | .750 | | .117 | | .016 | | .014 | | .046 | | .057 | | |
| $\omega_H/\omega_{HS}$ | | .869 | | .282 | | .089 | | .078 | | .191 | | .247 | | |

*Note.* $G_c$ = Crystallized Intelligence/Comprehension Knowledge, $G_v$ = Visual-Spatial, $G_f$ = Fluid Intelligence/Fluid Reasoning, $G_{sm}$ = Short-term Memory (Working Memory), $G_s$ = Processing Speed, $b$ = standardized loading of subtest on factor, $S^2$ = variance, $h^2$ = communality, $u^2$ = uniqueness, ECV = explained common variance, $\omega_H$ = omega-hierarchical (general factor), $\omega_{HS}$ = omega-hierarchical subscale (group factors).

**Table 7.** Decomposed variance sources for the WISC–IV$^{UK}$ 15 subtests for the referred Irish sample ($N$ = 245) according to a CHC bifactor model

| | General | | $G_c$ | | $G_v$ | | $G_f$ | | $G_{sm}$ | | $G_s$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WISC–IV$^{UK}$ subtest | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $b$ | $S^2$ | $h^2$ | $u^2$ |
| Similarities | .702 | .493 | .419 | .176 | | | | | | | | | .669 | .331 |
| Vocabulary | .726 | .527 | .567 | .321 | | | | | | | | | .810 | .190 |
| Comprehension | .554 | .307 | .426 | .181 | | | | | | | | | .520 | .480 |
| Information | .741 | .549 | .420 | .176 | | | | | | | | | .699 | .301 |
| Word Reasoning | .641 | .411 | .361 | .130 | | | | | | | | | .571 | .429 |
| Block Design | .748 | .560 | | | .261 | .068 | | | | | | | .627 | .373 |
| Picture Completion | .708 | .501 | | | .293 | .086 | | | | | | | .582 | .418 |
| Matrix Reasoning | .758 | .575 | | | | | .263 | .069 | | | | | .654 | .346 |
| Picture Concepts | .627 | .393 | | | | | .252 | .064 | | | | | .496 | .504 |
| Digit Span | .586 | .343 | | | | | | | .594 | .353 | | | .513 | .487 |
| Letter-Number Sequencing | .652 | .425 | | | | | | | .245 | .060 | | | .561 | .439 |
| Arithmetic | .701 | .491 | | | | | | | .306 | .094 | | | .608 | .392 |
| Coding | .447 | .200 | | | | | | | | | .505 | .255 | .454 | .546 |
| Symbol Search | .569 | .324 | | | | | | | | | .534 | .285 | .649 | .351 |
| Cancellation | .376 | .141 | | | | | | | | | .541 | .293 | .400 | .600 |
| Total Variance | | .416 | | .066 | | .010 | | .009 | | .034 | | .056 | .590 | .410 |
| ECV | | .705 | | .111 | | .017 | | .015 | | .057 | | .094 | | |
| $\omega_H/\omega_{HS}$ | | .856 | | .269 | | .096 | | .086 | | .208 | | .421 | | |

*Note.* $G_c$ = Crystallized Intelligence/Comprehension Knowledge, $G_v$ = Visual-Spatial, $G_f$ = Fluid Intelligence/Fluid Reasoning, $G_{sm}$ = Short-term Memory (Working Memory), $G_s$ = Processing Speed, $b$ = standardized loading of subtest on factor, $S^2$ = variance, $h^2$ = communality, $u^2$ = uniqueness, ECV = explained common variance, $\omega_H$ = omega-hierarchical (general factor), $\omega_{HS}$ = omega-hierarchical subscale (group factors).

Johnson concluded when there is an attempt to estimate or account for domain-specific abilities (something all multifactor intelligence tests attempt to do), the 'bifactor model factor scores should be preferred' (Murray & Johnson, 2013, p. 420). This is critical in evaluation of the WISC–IV$^{UK}$ construct validity based on internal structure because of publisher claims of what factor index scores measure as well as the many comparisons of factor index scores and inferences made from such comparisons. Researchers and clinicians must know how well WISC–IV$^{UK}$ group factors (domain specific) perform independent of the general intelligence (*g*) factor (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Chen, West, & Sousa, 2006). Reise, Moore, and Haviland (2010) concluded that a bifactor model, which contains a general factor, but permits multidimensionality, is better than the higher-order model for determining the relative contribution of group factors independent of the general factor (i.e., general intelligence).

Decomposed variance estimates presented in Tables 4–7 illustrate that the greatest portions of subtest variance were associated with the *g* factor and smaller portions of variance were associated with the four Wechsler group factors (VC, PR, WM, PS) or five CHC ($G_c$, $G_v$, $G_f$, $G_{sm}$, $G_s$) group factors. Numerous studies of Wechsler scales and other intelligence tests have consistently found that the greatest portions of total and common variance are apportioned to or associated with the *g* factor, which is estimated by the Full Scale score, and much smaller portions of total and common variance are apportioned to the first-order or group dimensions, estimated by the respective factor index scores (or CHC-based composites). This has been documented in both EFA and CFA studies of the WISC–IV (Bodin *et al.*, 2009; Canivez, 2014; Keith, 2005; Nakano & Watkins, 2013; Styck & Watkins, 2016; Watkins, 2006, 2010; Watkins *et al.*, 2006) and with other versions of Wechsler scales (Canivez & Watkins, 2010a,b; Canivez *et al.*, 2016, 2017; Dombrowski *et al.*, 2015; Golay & Lecerf, 2011; Golay *et al.*, 2013; Gignac, 2005, 2006; Lecerf *et al.*, 2010; McGill & Canivez, 2016; Watkins & Beaujean, 2014; Watkins *et al.*, 2013). Further, these results are not unique to Wechsler scales as similar results were also observed with the DAS–II (Canivez & McGill, 2016), SB5 (Canivez, 2008), WASI and WRIT (Canivez *et al.*, 2009), RIAS (Dombrowski *et al.*, 2009; Nelson & Canivez, 2012; Nelson *et al.*, 2007), CAS (Canivez, 2011), WJ III (Dombrowski, 2013, 2014a,b; Dombrowski & Watkins, 2013; Strickland, Watkins, & Caterino, 2015), and the WJ IV Cognitive (Dombrowski *et al.*, 2017). The implication of these consistent findings is that primary interpretive weight should be placed on the omnibus FSIQ rather than the first-order group factor-based index scores.

Examination of model-based reliability coefficients indicated that the *g* factor had very strong $\omega_H$ estimates in bifactor and higher-order models of Wechsler and CHC configurations allowing individual interpretation ($\omega_H$ ranging from .854 to .869), but the $\omega_{HS}$ estimates for the four Wechsler or five CHC WISC–IV$^{UK}$ group factors were low ($\omega_{HS}$ ranging from .078 to .429) and extremely limited for measuring unique constructs (Brunner *et al.*, 2012; Reise, 2012), and likely not high enough for individual interpretation (Reise, 2012; Reise *et al.*, 2013). Standardized path coefficients from Watkins (2010) were used to calculate $\omega_H$ and $\omega_{HS}$ estimates for comparison purposes and present results were quite similar. The $\omega_{HS}$ estimates for the four WISC–IV group factors from Watkins (2010) were also very low (.112 to .388). Canivez (2013) also reported very low $\omega_{HS}$ coefficients for the four WISC–IV group factors (.098 to .330) in a sample of referred children demographically similar to Watkins (2010). Explained common variance of the *g* factor in the present study ranged from .688 to .750, and within each Wechsler model and CHC model, the *g* factor accounted for between 6 and 53 times more common variance than the Wechsler or CHC group factors, further illustrating the dominance of general intelligence in WISC–IV$^{UK}$ measurement. In contrast to cross-battery (Flanagan, Alfonso, & Ortiz, 2012) and clinical (Weiss *et al.*, 2005) interpretation approaches, these results further support primary (if not exclusive) interpretation of the FSIQ for the WISC–IV$^{UK}$.

Finally, with respect to CHC conceptualizations about cognitive ability tests, including the WISC–IV$^{UK}$, there remains difficulty reconciling the dramatic and important differences between Horn's perspective and that of Carroll. Horn denied the existence of the *g* factor claiming it to be a statistical artefact, while Carroll repeatedly demonstrated the psychometric ascendance of the *g* factor due to the large portions of subtest variance associated with it (Cucina & Howardson, 2017). The present study results affirm Carroll's perspective (1993, 2003). With the WISC–IV$^{UK}$, the primary difference between the Wechsler and CHC configurations is splitting PR into $G_v$ and $G_f$, and while the CHC models in the present study attained equivalent 'fit' to Wechsler models, they present less parsimonious explanations of intelligence and both $G_v$ and $G_f$ contain miniscule portions of unique variance, rendering them of little to no use for individual clinical application. In fact, none of the group factors in either the Wechsler *or* CHC models attained satisfactory levels of ECV or $\omega_{HS}$ to support individual clinical interpretation. Factor index scores provided for the WISC–IV$^{UK}$ confound general intelligence variance with group factor variance and in most instances the overwhelming portion of index score variance is from general intelligence, thereby misleading the clinician into thinking the important construct is that represented by the group factor. This is further facilitated by interpretation methods prescribed and promoted by test publishers (i.e., Pearson) as well as others promoting use of cross-battery assessment/interpretation techniques (Flanagan *et al.*, 2012).

### *Limitations*
Limitations of the present study are primarily due to the restricted and non-random clinical sample of Irish students referred for evaluations of educational difficulties. Generalization to other populations cannot be recommended despite the identical or very similar results obtained with normative samples or large referred samples outside of Ireland. Because the publisher provided no psychometric studies of the WISC–IV$^{UK}$ internal structure with British (normative or clinical) or Irish samples and has refused to provide WISC–IV$^{UK}$ standardization sample raw data or correlation matrices and descriptive statistics from the WISC–IV$^{UK}$ standardization sample for independent analyses, it is impossible to know how the internal structure based on the present sample compares to the British normative sample or to a normative Irish sample. Clearly there is great need for publication of such crucial psychometric information for the WISC–IV$^{UK}$ normative sample to provide empirical evidence necessary for ethical interpretation of the WISC–IV$^{UK}$ (BPS, 2009, 2010).

### *Conclusion*
Based on the present results and strong replication of previous findings (Watkins *et al.*, 2013), it seems prudent to focus WISC–IV$^{UK}$ interpretation on the FSIQ and if going beyond the FSIQ to interpret factor index scores (Wechsler based or CHC based) doing so with extreme caution so as not to misinterpret or over-interpret scores given the small unique variance provided by the group factors when conflated with general intelligence variance. Further examination of the WISC–IV$^{UK}$ should test relations to external variables or criteria, such as academic achievement, to determine what, if any, reliable achievement

variance is incrementally accounted for by the WISC–IV$^{UK}$ factor index scores (or CHC constructs) beyond that accounted for by the FSIQ (see Canivez *et al.*, 2014). Additionally, diagnostic utility studies should be conducted to ascertain the extent to which various factor indexes of the WISC–IV$^{UK}$ are able to correctly identify individuals from within various diagnostic groups that should hypothetically demonstrate differences in cognitive profiles. Given the low portions of unique variance provided by the Wechsler- or CHC-based WISC–IV$^{UK}$ first-order factors in the present study (and other studies), it is difficult to imagine that they would provide meaningful incremental validity or diagnostic utility beyond general intelligence. In the revision of the WISC–IV$^{UK}$ and forthcoming WISC–V$^{UK}$, it is hoped that Pearson, United Kingdom, publishes a technical manual that includes psychometric details of reliability and validity and similar analyses as presented here to provide evidence for the WISC–V$^{UK}$ structure *based on the UK standardization sample* (including subtest correlation matrices, means and standard deviations) *and* also allow independent analyses of standardization sample raw data. Psychologists in the United Kingdom and Ireland where the WISC–V$^{UK}$ will be used must have such detailed information to properly interpret test results according to the *Code of Ethics and Conduct* (BPS, 2009) and the *Code of Good Practice for Psychological Testing* (BPS 2010).


## Aknowledgement

Gary L. Canivez is Professor of Psychology at Eastern Illinois University. Marley W. Watkins is Nonresident Scholar in the Department of Educational Psychology at Baylor University. Rebecca Good, Kate James, and Trevor James are psychologists at Éirim: The National Assessment Agency, Ltd., Dublin, Ireland. Special thanks to Emma McCarron, also at Éirim: The National Assessment Agency, Ltd., for contributing assessments used in the present study.


## References


Alfonso, V. C., Oakland, T. D., LaRocca, R., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review*, *29*, 52–64.

Alfonso, V. C., & Pratt, S. I. (1997). Issues and suggestions for training professionals in assessing intelligence. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 326–347). New York, NY: Guilford.

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

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

Belter, R. W., & Piotrowski, C. (2001). Current status of doctoral-level training in psychological testing. *Journal of Clinical Psychology*, *57*, 717–726. doi:10.1002/jclp.1044

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*, 121–130. doi:10.1037/a0017767

Bentler, P. M., & Wu, E. J. C. (2012). *EQS for windows*. Encino, CA: Multivariate Software Inc.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC–IV in a clinical neuropsychological sample. *Child Neuropsychology*, *15*, 417–424. doi:10.1080/09297040802603661

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

Byrne, B. M. (2006). *Structural equation modeling with EQS* (2nd ed.). Mahwah, NJ: Erlbaum.

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

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

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

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

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

Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler abbreviated scale of intelligence and wide range intelligence test: Convergent and structural validity. *School Psychology Quarterly*, *24*, 252–265. doi:10.1037/a0018030

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

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

Canivez, G. L., & Watkins, M. W. (1998). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition. *Psychological Assessment*, *10*, 285–291. doi:10.1037/1040-3590.10.3.285

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

Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV) adolescent subsample. *School Psychology Quarterly*, *25*, 223–235. doi:10.1037/a0022046

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

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

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

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

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

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

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

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

Chen, H.–. Y., Keith, T. Z., Chen, Y.–. H., & Chang, B.-S. (2009). What does the WISC-IV measure? Validation of the scoring and CHC-based interpretive approaches *Journal of Research in Education Sciences*, *54*, 85–108.

Chen, H., Keith, T., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, *49*, 677–682. doi:10.1016/j.paid.2011.10.006

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*, 189–225. doi:10.1207/s15327906 mbr4102_5

Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, *52*, 161–166. doi:10.1016/j.paid.2011.10.006

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

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*, Advance online publication. doi:10.1037/pas0000389

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

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

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

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

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

Dombrowski, S. C., 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. doi:10.1037/pas0000350

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

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

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459–483). New York, NY: Guilford.

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

Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, L. G. Weiss & van de Vijver F. J. R. (Eds.), *Culture and children's intelligence* (pp. 277–313). San Diego, CA: Academic Press.

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

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

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

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

Goh, D. S., Teslow, C. J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology*, *12*, 696–706. doi:10.1037/0735-7028.12.6.696

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

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

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.) (pp. 231–258). New York, NY: Plenum.

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

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

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

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

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

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*, 1–55. doi:10.1080/10705519909540118

Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review*, *21*, 271–284.

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

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York, NY: Wiley.

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

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure. *School Psychology Review*, *35*, 108–127.

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

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

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

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

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

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341. doi:10.1207/s15328007sem1103_2

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

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

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York, NY: Guilford.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York, NY: Guilford.

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

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

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

Nakano, S., & Watkins, M. W. (2013). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred Native American students. *Psychology in the Schools*, *50*, 957–968. doi:10.1002/pits.21724

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

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

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. doi:10.1037/a0032086

Niileksela, C. R., Reynolds, M. R., & Kaufman, A. S. (2013). An alternative Cattell–Horn–Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70–90. *Psychological Assessment*, *25*, 391–404. doi:10.1037/a0031175

Oakland, T., & Hu, S. (1992). The top ten tests used with children and youth worldwide. *Bulletin of the International Test Commission*, *19*, 99–120.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385. doi:10.1037/h0088795

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

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

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

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

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

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 Press.

Spearman, C. (1904). "General intelligence": Objectively determined and measured. *American Journal of Psychology*, *15*, 201–293. doi:10.2307/1412107

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

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment*, *12*, 331–350. doi:10.1177/073428299401200403

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

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. doi:10.1177/0022219414539565

The British Psychological Society. (2009). *Code of ethics and conduct*. Leicester, UK: Author.

The British Psychological Society. (2010). *Code of good practice for psychological testing*. Leicester, UK: Author.

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

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

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

Watkins, M. W. (2012). *ChiSquareDiff* [Computer software]. Phoenix, AZ: Ed & Psych Associates.

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

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

Watkins, Jr, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, *26*, 54–60. doi:10.1037/0735-7028.26.1.54

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

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

Wechsler, D. (1991). *Manual for the Wechsler intelligence scale for children-third edition*. San Antonio, TX: The Psychological Corporation.

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

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

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

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

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

Weiss, L. G., Saklofske, D. H., & Prifitera, A. (2005). Interpreting the WISC–IV index scores. In A. Prifitera, D. H. Saklofske & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 71–100). New York, NY: Elsevier Academic Press.

Williams, P. E., Weiss, L. G., & Rolfhus, E. L. (2003). *Clinical validity* (WISC–IV Tech. Rep. No. 3). San Antonio, TX: Psychological Corporation.

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

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

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

*Received 28 November 2016; revised version received 1 March 2017*