Psychological Assessment
2016, Vol. 28, No. 8, 975–986

1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000238

Factor Structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory Factor Analyses With the 16 Primary and Secondary Subtests

Gary L. Canivez
Eastern Illinois University

Marley W. Watkins
Baylor University

Stefan C. Dombrowski
Rider University

The factor structure of the 16 Primary and Secondary subtests of the Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V; Wechsler, 2014a) standardization sample was examined with exploratory factor analytic methods (EFA) not included in the WISC–V Technical and Interpretive Manual (Wechsler, 2014b). Factor extraction criteria suggested 1 to 4 factors and results favored 4 first-order factors. When this structure was transformed with the Schmid and Leiman (1957) orthogonalization procedure, the hierarchical *g*–factor accounted for large portions of total and common variance while the 4 first-order factors accounted for small portions of total and common variance; rendering interpretation at the factor index level less appropriate. Although the publisher favored a 5-factor model where the Perceptual Reasoning factor was split into separate Visual Spatial and Fluid Reasoning dimensions, no evidence for 5 factors was found. It was concluded that the WISC–V provides strong measurement of general intelligence and clinical interpretation should be primarily, if not exclusively, at that level.

Keywords: WISC–V, exploratory factor analysis, factor extraction criteria, Schmid–Leiman higher-order analysis, structural validity

The Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V; Wechsler, 2014a) is the latest version of one of the most frequently used intelligence tests for children. It includes 16 intelligence related subtests; five first-order factor index scores (Verbal Comprehension [VC], Visual Spatial [VS], Fluid Reasoning [FR], Working Memory [WM], and Processing Speed [PS]); and the hierarchically ordered Full Scale score (FSIQ). In addition to eliminating the Word Reasoning and Picture Completion subtests, the WISC–V incorporated a Picture Span subtest (adapted from the Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition [WPPSI–IV; Wechsler, 2012]) to measure visual working memory, and Visual Puzzles and Figure Weights subtests (adapted from the Wechsler Adult Intelligence Scale–Fourth Edition [WAIS–IV; Wechsler, 2008]) to measure visual spatial and fluid reasoning, respectively. A goal for the WISC–V was to split the former Perceptual Reasoning (PR) factor into separate Visual Spatial and Fluid Reasoning factors.

The WISC–V includes seven “Primary” subtests (Similarities [SI], Vocabulary [VC], Block Design [BD], Matrix Reasoning [MR], Figure Weights [FW], Digit Span [DS], and Coding [CD]) that are used in producing the FSIQ and three additional Primary subtests (Visual Puzzles [VP], Picture Span [PS], and Symbol Search [SS]) that are used in producing the five factor index scores (two subtests each). There are six “Secondary” subtests (Information [IN], Comprehension [CO], Picture Concepts [PC], Arithmetic [AR], Letter–Number Sequencing [LN], and Cancellation [CN]) that are used for substitution in FSIQ estimation or in estimating newly created (Quantitative Reasoning, Auditory Working Memory, and Nonverbal) and previously existing (General Ability and Cognitive Proficiency) Ancillary Index Scores. Ancillary Index Scores are not factorially derived composite scores, but logically or theoretically constructed. Complementary subtests (Naming Speed Literacy, Naming Speed Quality, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation) are new to the WISC–V but these subtests are not intelligence subtests and may not be substituted for primary or secondary subtests. The Complementary subtests were included for additional clinical assessment applications (Wechsler, 2014b).

Like other intelligence tests published in the past 15 years (e.g., Wechsler Intelligence Scale for Children–Fourth Edition [WISC–IV; Wechsler, 2003], Stanford–Binet Intelligence Scales–Fifth Edition [SB–5; Roid, 2003a], Kaufman Assessment Battery for Children–Second Edition [KABC–II; Kaufman & Kaufman, 2004]; Reynolds Intellectual Assessment Scales [RIAS; Reynolds & Kamphaus, 2003], Wide Range Intelligence Test [WRIT; Glutting, Adams, & Sheslow, 2000]); the WISC–V attempted to reflect

---

Gary L. Canivez, Department of Psychology, Eastern Illinois University; Marley W. Watkins, Department of Educational Psychology, Baylor University; Stefan C. Dombrowski, School Psychology Program, Rider University.

Correspondence concerning this article should be addressed to Gary L. Canivez, Department of Psychology, 600 Lincoln Avenue, Charleston, IL 61920-3099. E-mail: glcanivez@eiu.edu

current conceptualizations of intellectual measurement articulated by Carroll, Cattell, and Horn (Carroll, 1993, 2003, 2012; Cattell & Horn, 1978; Horn, 1991; Horn & Blankson, 2012; Horn & Cattell, 1966). The WISC–V Technical and Interpretive Manual also notes subtest relations to neuropsychological theories and theories of cognitive psychology. Specifically, the WISC–V includes 16 subtests that provide estimates of narrow abilities, five-factor indexes that provide estimates of broad abilities, and 1 estimate of general intelligence (i.e., FSIQ) consistent with Wechsler's definition of intelligence (i.e., "global capacity;" Wechsler, 1939b, p. 229) and similar to Carroll's (1993, 2003, 2012) intelligence structure.

Evidence of structural validity of the WISC–V was established via confirmatory factor analyses (CFA) reported in the WISC–V Technical and Interpretive Manual (Wechsler, 2014b) that included specification of *higher*-order models with a general intelligence factor *indirectly* influencing subtests via full mediation through two through five first-order factors. CFA models and subtest assignments to latent factors were detailed in Table 5.3, while the publisher preferred five-factor higher-order model was illustrated in Figure 5.10 of the WISC–V Technical and Interpretive Manual (Wechsler, 2014b). This "best fitting" model (see Figure 1) included a higher-order general intelligence dimension with five first-order factors (VC, VS, FR, WM, and PS) and the 16 subtest indicators were uniquely associated with one latent first-order factor except for Arithmetic, which was cross-loaded on VC, FR, and WM. This final model was also reported to fit five different age groupings (6–7, 8–9, 10–11, 12–13, and 14–16) equally well.

There are a number of notable concerns regarding the CFAs reported in the WISC–V Technical and Interpretive Manual. First, there was a lack of detail in describing the CFA methods employed. For example, conspicuously absent are details regarding how the metric of the factors was defined and why weighted least squares (WLS) estimation was selected. By definition, latent constructs (i.e., factors) have no natural scale of measurement, requiring instead specification by the researcher to achieve model identification. The choice of metric can affect unstandardized parameters and may "yield different conclusions regarding the statistical significance of freely estimated parameters" (Brown, 2015, p. 133). Kline (2011) noted that "use of an estimation method other than ML [maximum likelihood] requires explicit justification" (p. 154). WLS is typically applied with categorical or nonnormally distributed data and may not produce χ^2 values nor approximate fit indices equivalent to those generated by ML estimation (Yuan & Chan, 2005); therefore, use of WLS is perplexing and represents a departure from the use of ML estimation most typically observed in CFA of intelligence tests.

Second, the preferred CFA model was complex (because of cross-loadings of the Arithmetic subtest), abandoning the parsimony of simple structure (Thurstone, 1947). Third, the preferred model generated a standardized path coefficient of 1.00 between the higher-order general intelligence factor and the first-order FR factor. Thus, *g* and FR were empirically redundant (Le, Schmidt, Harter, & Lauver, 2010). This constitutes a major threat to discriminant validity and signals that the WISC–V may have been overfactored (Frazier & Youngstrom, 2007). Fourth, after acknowledging the sensitivity of the χ^2 test to trivial differences when analyzing large samples, Wechsler (2014b) subsequently used χ^2 difference tests of nested models to identify the preferred five-factor model. The same sensitivity to large samples is true for χ^2 difference tests (Millsap, 2007), suggesting that the model differences reported in the WISC–V Technical and Interpretive Manual (Wechsler, 2014b) might be significant but trivial. For example, Table 5.4 reveals that the difference between Models 4a and 5a was statistically significant but those two models exhibited identical comparative fit index (CFI) and root mean squared error of approximation (RMSEA) values. Likewise, the preferred five-factor higher-order model was significantly different from other five-factor models but all exhibited identical CFI and RMSEA values (e.g., .98 and .04, respectively). Cheung and Rensvold (2002) demonstrated, in the context of factorial invariance, that practical differences independent of sample size and model complexity could be identified by $\Delta\text{CFI} > .01$.

A fifth concern with the WISC–V CFA is that there was a failure to test rival bifactor models (Holzinger & Swineford, 1937). Bifactor models have several technical benefits over higher-order models (Canivez, in press; Reise, 2012), have been found to fit data from other Wechsler scales (viz., Canivez, 2014b; Gignac & Watkins, 2013; Nelson, Canivez, & Watkins, 2013; Watkins, 2010; Watkins & Beaujean, 2014; Watkins, Canivez, James, James, & Good, 2013), and have been recommended for cognitive tests (Brunner, Nagy, & Wilhelm, 2012; Canivez, in press; Gignac, 2005, 2006). A higher-order structural model of intelligence articulates general intelligence as a superordinate factor fully mediated by the first-order factors to indirectly influence subtest indicators, while the bifactor structural model of intelligence articulates general intelligence as a breadth factor with direct influences on subtest indicators, as do group factors (Canivez, in press; Gignac, 2008). Theoretically, the bifactor model appears to be more consistent with Spearman's (1927) theory and conceptualization of intelligence and a more parsimonious explanation than a higher-order model (Canivez, in press; Gignac, 2006). Canivez (in press) discusses the advantages of the bifactor model. First, the general factor is more easily interpreted because it offers interpretation via a direct influence on subtest indicators rather than an interpretation mediated through an index as is the case with the higher-order model (Chen, Hayes, Carver, Laurenceau & Zhang, 2012; Gorsuch, 1983). Second, a bifactor model permits simultaneous examination of the influence of both general *and* group factors on substests, permitting judgment of general *and* group scale importance (Gorsuch, 1983; Reise, 2012; Reise, Moore, & Haviland, 2010). Third, the bifactor model permits the calculation of model based reliability using omega-hierarchical (ω_h) and omega-subscale (ω_s), which estimate the proportion of variance because of any single factor, general or group, and thereby determine how much interpretive emphasis should be placed upon the general factor and lower-order factor scores (Gignac & Watkins, 2013; Reise, 2012; Zinbarg, Yovel, Revelle, & McDonald, 2006). Finally, bifactor modeling permits an evaluation of the unique contribution of the general factor and group factors in predicting external criteria (Chen et al., 2012; Gignac, 2006; Reise, Moore, & Haviland, 2010).

A sixth notable concern is that the publisher did not provide decomposed variance estimates to illustrate how much subtest variance was because of the general factor and how much was exclusive to the group factors. This is a consequential omission because clinicians and researchers are unable to judge the adequacy of the group factors (VC, VS, FR, WM, and PS) based on how much unique

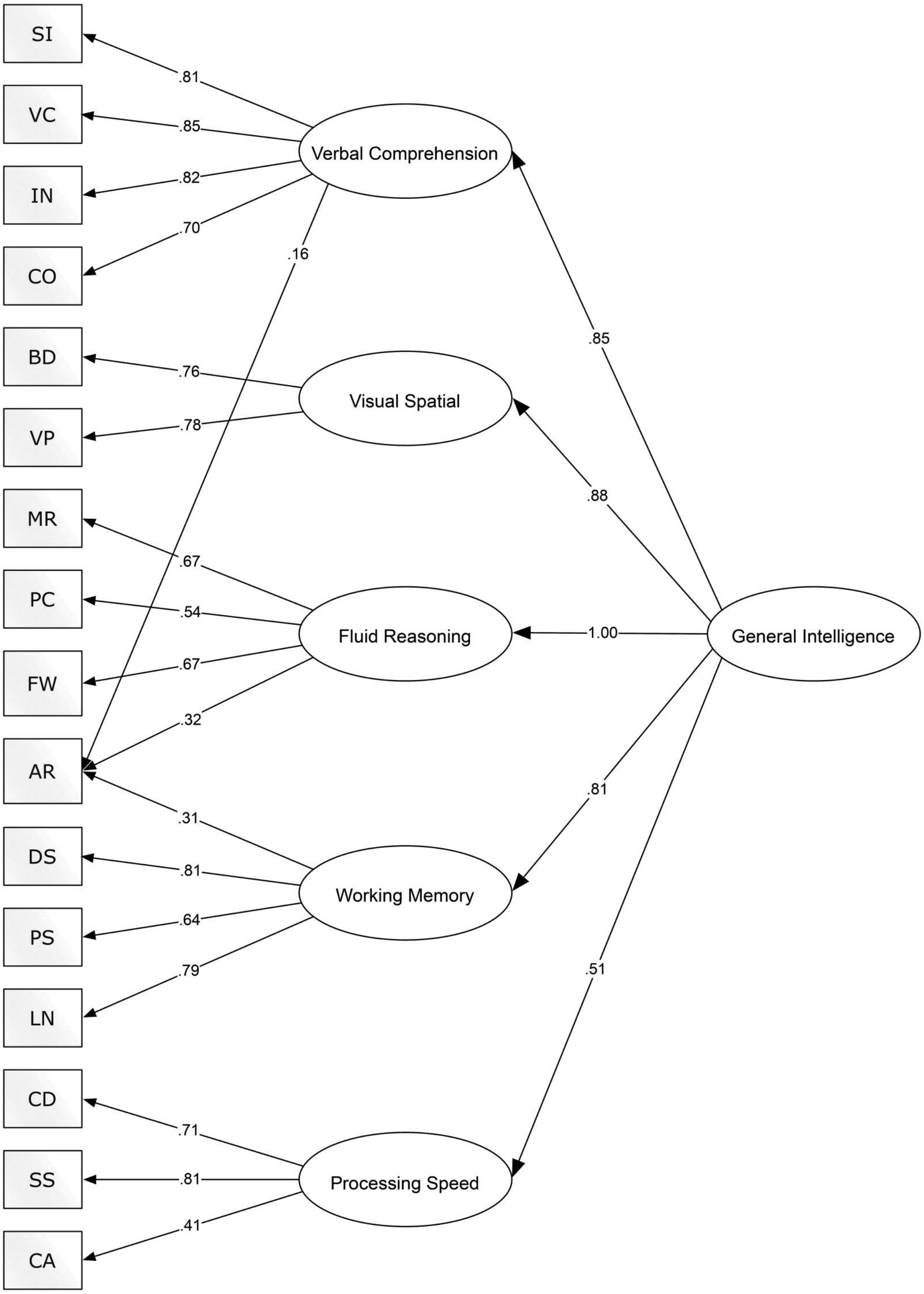

Figure 1. Higher-order measurement model with standardized coefficients (adapted from Figure 5.1 [Wechsler, 2014b]), for Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) standardization sample (N = 2,200) 16 Subtests. SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, PC = Picture Concepts, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LN = Letter–Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation. *Wechsler Intelligence Scale for Children, Fifth Edition* (*WAIS-V*). Copyright 2014 NCS Pearson, Inc. Reproduced with permission. All rights reserved. “*Wechsler Intelligence Scale for Children*” and “*WAIS*” are trademarks, in the United States and/or other countries, of Pearson Education, Inc. or its affiliates(s).

variance they capture when purged of the effects of general intelligence. Relatedly, also missing from the WISC–V Technical and Interpretive Manual (Wechsler, 2014b) were model based reliability estimates. It has long been argued that classical estimates of reliability are biased (Raykov, 1997) and model-based estimates, such as ω_h and ω_s, have been recommended as superior metrics (Gignac & Watkins, 2013). These problems were pointed out in reviews of several Wechsler scales including the WAIS–IV (Canivez, 2010) and WPPSI–IV (Canivez, 2014a) as well as within a commentary on the WISC–IV and WAIS–IV (see Canivez & Kush, 2013). However, ω estimates were conspicuously absent.

Finally, the WISC–V Technical and Interpretive Manual explicitly preferred CFA over EFA rather than taking advantage of both methods. EFA and CFA are complementary procedures, each answering somewhat different questions (Carroll, 1993). Greater confidence in the latent factor structure is achieved when EFA and CFA are in agreement (Gorsuch, 1983). Carroll (1995) and Reise (2012) both noted that EFA procedures are particularly useful in suggesting possible models to be tested in CFA. In fact, Carroll (1998) suggested that "CFA should derive its initial hypotheses from EFA results, rather than starting from scratch or from *a priori* hypotheses . . . [and] CFA analyses should be done to check my EFA analyses" (p. 8). That suggestion was reinforced by Brown (2015), who noted that "in addition to a compelling substantive justification, CFA model specification is usually supported by prior (but less restrictive) exploratory analyses (i.e., EFA) that have established the appropriate number of factors, and pattern of indicator–factor relationships" (p. 141). The deletion of the WISC–IV Word Reasoning and Picture Completion subtests; the addition of Visual Puzzles, Figure Weights, and Picture Span subtests; the reduction in the number of subtests from 10 to 7 to derive a FSIQ; and the inclusion of new or revised items across all WISC–V subtests suggests that relationships among retained and new subtests might result in associations and latent structure unanticipated by a priori conceptualizations (Strauss, Spreen, & Hunter, 2000). In consideration of all these issues the lack of EFA results is most disappointing given prior criticism of their absence in other Wechsler manuals (Canivez, 2010, 2014a).

In fact, prior independent investigations of intelligence test factor structures using EFA methods have produced serious and substantial challenges to the CFA–based latent structures of standardization data promoted in technical manuals. For example, two studies (Canivez, 2008; DiStefano & Dombrowski, 2006) found that only one or two factors described the structure of the SB–5 standardization sample in contrast to the five factors specified in the SB–5 technical manual (Roid, 2003b). Likewise, EFA analyses of the WISC–IV, WAIS–IV, RIAS, and Woodcock–Johnson III Tests of Cognitive Ability (McGrew & Woodcock, 2001) normative data arrived at conclusions discrepant from test publishers (Canivez & Watkins, 2010a, 2010b; Dombrowski, 2013; Dombrowski, 2014a, 2014b; Dombrowski, Watkins, & Brogan, 2009; Watkins, 2006). After reviewing earlier, similar evidence, Frazier and Youngstrom (2007) concluded that commercial ability tests are "substantially overfactored" (p. 178), identifying minor factors that "may not possess sufficient reliability to make decisions on the individual level" (p. 181).

Perhaps as a consequence of the profusion of factors created by overfactoring, technical manuals have advanced clinical interpretation schemes that focus on first-order factors. In contrast, independent analyses of data from clinical samples have typically favored interpretation based on the higher-order factor. For example, three EFA investigations of the WISC–IV and two EFA studies of the WAIS–IV found that most variance was associated with general intelligence (Bodin, Pardini, Burns, & Stevens, 2009; Canivez & Watkins, 2010a, 2010b; Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), suggesting that interpretation of the WISC–IV and WAIS–IV should focus on the global FSIQ score because it accounts for most of the common variance. Additionally, the FSIQ has been shown to be superior to factor index scores in predicting academic achievement with little incremental predictive validity offered by the factor index scores (Canivez, 2014b; Canivez, Watkins, James, Good, & James, 2014; Glutting, Watkins, Konold, & McDermott, 2006; Glutting et al., 1997; Nelson, Canivez, & Watkins, 2013). In fact, the limited unique variance captured by the first-order factors may be responsible for the poor incremental predictive validity of Wechsler factor scores.

Given the hypothesis that the WISC–V was overfactored (Frazier & Youngstrom, 2007) combined with copious empirical research evidence supporting the eminence of general intelligence and the enumerated problems with the CFA results reported by Wechsler (2014b), the present study utilized summary data from the WISC–V standardization sample subtest correlation matrix to examine its factor structure using EFA procedures. Primary research questions included (a) how many factors should be extracted and retained; (b) what are the subtest associations and relations with latent factors and is there evidence for the publisher's claim of five first-order factors; and (c) when extracting correlated theoretical factors and applying the Schmid and Leiman procedure (Schmid & Leiman, 1957), what proportion of variance is because of general intelligence versus group ability factors?

Method

Participants

Participants were members of the WISC–V standardization sample and included a total of 2,200 individuals ranging in age from 6–16 years. Detailed demographic characteristics are provided in the WISC–V Technical and Interpretive Manual (Wechsler, 2014b). The standardization sample was obtained using stratified proportional sampling across variables of age, sex, race/ethnicity, parental education level, and geographic region. Education level was a proxy for SES where accurate information about income is often difficult to obtain. Examination of tables in the Technical and Interpretive Manual (Wechsler, 2014b) revealed a close match to the U.S. census across stratification variables.

Instrument

The WISC–V is an individual test of general intelligence for children aged 6–16 years. The WISC–V, like the WISC–IV, overlaps in age with the WPPSI–IV at age 6 years and the WAIS–IV at age 16 years to allow clinicians the opportunity to select the most appropriate instrument depending on the referral question and child characteristics.

Organization and subtest administration order of the WISC–V reflect a new four level organization. The Full Scale IQ (FSIQ) is composed of seven primary subtests across the five factors (VC, VS, FR, WM, and PS), but if one of the FSIQ subtests is invalid or missing, that subtest may be substituted by a secondary subtest from within the same factor. Only one substitution is allowed. The Primary Index Scale level is composed of 10 WISC–V subtests (primary subtests) and are used to estimate the five WISC–V factor index scores (VCI, VSI, FRI, WMI, and PSI). No substitutions are allowed for the Primary Index Scales. The Ancillary Index level is composed of five scales that are not factorially derived: Quantitative Reasoning (QR), Auditory Working Memory (AWM), Nonverbal (NV), General Ability (GA), and Cognitive Proficiency (CP) and reflect various combinations of primary and secondary subtests. The Complementary Index level is composed of three scales: Naming Speed, Symbol Translation, and Storage and Retrieval derived from the newly created complementary subtests (Naming Speed Literacy, Naming Speed Quality, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation). Complementary subtests are not intelligence subtests and may not be substituted for primary or secondary subtests.

Procedure and Analyses

The WISC–V subtest correlation matrix for the total standardization sample (Table 5.1) was extracted from the Technical and Interpretive Manual (Wechsler, 2014b). Multiple criteria (Gorsuch, 1983) were examined to determine the number of factors to retain and included eigenvalue >1 (Kaiser, 1960), the scree test (Cattell, 1966), standard error of scree (SE_{Scree}; Zoski & Jurs, 1996), Horn's parallel analysis (HPA; Horn, 1965), minimum average partials (MAP; Velicer, 1976), Bayesian Information Criterion (BIC; Schwarz, 1978), and sample size adjusted BIC (SSBIC; Sclove, 1987). Simulation studies have found HPA and MAP to be the most accurate a priori empirical criteria with scree sometimes a useful adjunct (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). Random data and resulting eigenvalues for HPA were produced using the Monte Carlo PCA for Parallel Analysis computer program (Watkins, 2000) with 100 replications to provide stable eigenvalue estimates. A detailed analysis of HPA performance found that it tends to underfactor in the presence of a strong general factor (Crawford et al., 2010). BIC and SSBIC were not included in prior simulation studies but have proven useful in classification simulations (Morgan, 2015). BIC and SSBIC were estimated with the *psych* package within the R statistical system (R Development Core Team, 2015). The scree test is a subjective criterion so the SE_{Scree} as programmed by Watkins (2007) was used because it was reportedly the most accurate objective scree method (Nasser, Benson, & Wisenbaker, 2002).

Principal axis exploratory factor analyses (Fabrigar, Wegener, MacCallum, & Strahan, 1999) were used to analyze the WISC–V standardization sample correlation matrix using SPSS 21 for Macintosh OSX. Retained factors were subjected to promax rotation ($k = 4$; Gorsuch, 1983). Salient factor pattern coefficients were defined as those ≥.30 (Child, 2006). Factor solutions were examined for interpretability and theoretical plausibility (Fabrigar et al., 1999) with the empirical requirement that each factor should be marked by two or more salient loadings and no salient cross-loadings (Gorsuch, 1983).

Subtest scores on cognitive ability tests reflect combinations of both first-order *and* second-order factors and because of this Carroll (1993, 1995, 1998, 2003) argued that variance from the higher-order factor must be extracted first to residualize the lower-order factors, leaving them orthogonal to the higher-order factor. The Schmid and Leiman (1957) procedure has

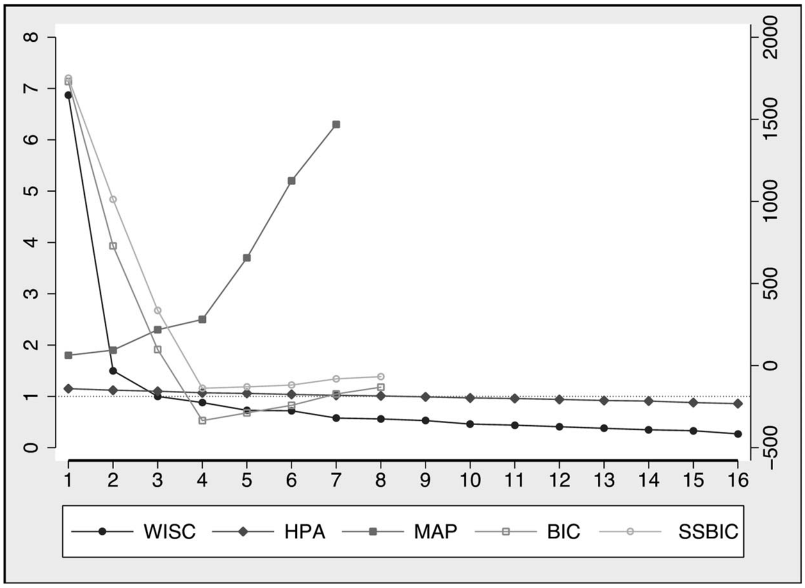

Figure 2. Graphic representation of various criteria for factor extraction for WISC-V standardization sample (N = 2,200). WISC and HPA are eigenvalues from standardization sample data and random data, respectively, and scaled on left *y*-axis; while MAP (multiplied by 100 for convenient display), BIC, and SSBIC values are scaled on the right *y*-axis. Smallest MAP, BIC, and SSBIC value indicates number of factors. WISC-V = Wechsler Intelligence Scale for Children--Fifth Edition; HPA = Horn's parallel analysis; MAP = minimun average partials; BIC = Bayesian Information Criteria; SSBIC = sample size adjusted BIC.

been recommended as the statistical method to accomplish this residualization (Carroll, 1993, 1995, 1997, 2003; Carretta & Ree, 2001; Gustafsson & Snow, 1997; McClain, 1996; Ree, Carretta, & Green, 2003; Thompson, 2004). It is a reparameterization of a higher-order model and an approximate bifactor solution (Reise, 2012). Additionally, numerous studies of its application with Wechsler scales (Canivez & Watkins, 2010a, 2010b; Golay & Lecerf, 2011; Watkins, 2006) and with other intelligence tests (Canivez, 2008, 2011; Canivez et al., 2009; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007) have been published. Accordingly, the Schmid and Leiman (1957) procedure as programmed in the MacOrtho program (Watkins, 2004) was subsequently applied to EFA solutions. For convenience, this method is labeled SL bifactor (Reise, 2012).

Using the factor pattern coefficients from the first-order obliquely rotated EFA solution and the second-order factor loading coefficients produced by EFA of the first-order factor intercorrelations, the Schmid and Leiman (1957) procedure applied in the MacOrtho program (Watkins, 2004) apportions common variance first to the higher-order factor and the residual common variance is then apportioned to the group factors. Schmid and Leiman noted that this "not only preserves the desired interpretation characteristics of the oblique solution, but also discloses the hierarchical structuring of the variables" (p. 53). It is this feature that led Carroll (1995) to insist on SL orthogonalization of higher-order models.

Omega-hierarchical and ω_s (Reise, 2012) were estimated as model-based reliability estimates of the latent factors (Gignac & Watkins, 2013). Chen, Hayes, Carver, Laurenceau, and Zhang (2012) noted that "for multidimensional constructs, the alpha coefficient is complexly determined, and McDonald's ω_h (McDonald, 1999) provides a better estimate for the composite score and thus should be used" (p. 228). Omega-hierarchical is the model-based reliability estimate for the hierarchical general intelligence factor independent of the variance of group factors. Omega-subscale is the model-based reliability estimate of a group factor with all other group *and* general factors removed (Reise, 2012). Omega estimates (ω_h and ω_s) may be obtained from EFA SL bifactor solutions and were produced using the *Omega* program (Watkins, 2013), which is based on the tutorial by Brunner, Nagy, and Wilhelm (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 would be preferred (Reise, 2012; Reise, Bonifay, & Haviland, 2013).

Results

Factor Extraction Criteria Comparisons

Figure 2 illustrates the scree plots from HPA for the WISC–V for the total standardization sample as well as MAP, BIC, and SSBIC results. MAP suggested one factor; eigenvalue >1, scree, and HPA suggested two or three factors; and BIC and SSBIC suggested four factors. In contrast, the publisher recommended five factors. Given that it is better to overextract than underextract (Wood, Tataryn, & Gorsuch, 1996), EFA began by extracting five factors to examine subtest associations based on the publisher's suggested structure and to allow examination of the performance

Table 1

Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) Exploratory Factor Analysis: Five Oblique Factor Solution for the Total Standardization Sample ($N = 2,200$)

WISC–V subtest	*g*	F1: Verbal Comprehension	F2: Working Memory	F3: Visual Spatial	F4: Processing Speed	F5: Inadequate	h^2
Similarities	.751	**.776** (.805)	.049 (.595)	.022 (.575)	−.005 (.343)	−.026 (.584)	.650
Vocabulary	.773	**.887** (.855)	−.032 (.584)	.042 (.602)	−.055 (.309)	−.021 (.606)	.735
Information	.754	**.790** (.816)	−.055 (.573)	−.007 (.581)	.002 (.338)	.095 (.624)	.669
Comprehension	.660	**.721** (.708)	.081 (.534)	−.026 (.480)	.052 (.341)	−.101 (.479)	.510
Block Design	.667	.033 (.555)	−.009 (.515)	**.598** (.732)	.127 (.445)	.079 (.582)	.554
Visual Puzzles	.686	.046 (.586)	.022 (.516)	**.857** (.824)	−.071 (.331)	−.063 (.579)	.684
Matrix Reasoning	.635	.068 (.546)	.168 (.557)	.267 (.595)	.035 (.361)	.216 (.592)	.430
Figure Weights	.648	.027 (.570)	−.014 (.584)	.173 (.614)	−.038 (.295)	**.619** (.739)	.560
Picture Concepts	.518	.266 (.489)	.101 (.431)	.208 (.465)	.028 (.284)	−.008 (.418)	.275
Arithmetic	.725	.219 (.466)	**.311** (.629)	.008 (.446)	.071 (.337)	.258 (.452)	.551
Digit Span	.703	−.039 (.565)	**.852** (.814)	.028 (.515)	−.042 (.392)	−.009 (.569)	.664
Picture Span	.572	.004 (.652)	**.593** (.682)	.095 (.563)	.000 (.423)	−.039 (.660)	.399
Letter–Number Sequencing	.690	.094 (.584)	**.821** (.792)	−.064 (.469)	−.043 (.373)	−.047 (.538)	.634
Coding	.419	−.031 (.276)	.075 (.388)	−.063 (.296)	**.758** (.745)	−.025 (.264)	.560
Symbol Search	.477	.023 (.337)	−.004 (.405)	.038 (.377)	**.769** (.777)	−.042 (.308)	.605
Cancellation	.219	.014 (.151)	−.133 (.152)	.035 (.189)	**.455** (.418)	.024 (.149)	.183
Eigenvalue		6.87	1.50	1.00	.88	.73	
% Variance		40.31	6.39	3.55	3.02	.89	
Factor correlations							
Verbal Comprehension		1.00					
Working Memory		.713	1.00				
Visual Spatial		.700	.634	1.00			
Processing Speed		.417	.518	.463	1.00		
F5: Inadequate		.724	.707	.726	.401	1.00	

Note. *g* = general structure coefficients based on first unrotated factor coefficients (*g*–loadings); h^2 = Communality. Factor pattern coefficients (structure coefficients) based on principal factors extraction with promax rotation ($k = 4$). Salient pattern coefficients presented in bold (pattern coefficient $\geq .30$).

of smaller factors. Models with four, three, and two factors were subsequently examined for adequacy.

Exploratory Factor Analyses

Five–factor model. Extracting five WISC–V factors with promax rotation produced a fifth factor with only one salient factor pattern coefficient (see Table 1). Block Design and Visual Puzzles saliently loaded on a common Visual Spatial factor but Matrix Reasoning had no salient pattern coefficients on any factor. Thus, Matrix Reasoning and Figure Weights did not share sufficient common variance to constitute a Fluid Reasoning dimension. Picture Concepts also failed to achieve a salient pattern coefficient on any factor. This pattern of psychometrically unsatisfactory results is emblematic of overextraction (Gorsuch, 1983; Wood et al., 1996).

Four–factor model. Table 2 presents the results of extracting four WISC–V factors with promax rotation. The *g*–loadings ranged from .220 (Cancellation) to .774 (Vocabulary) and all were within the fair to good range (except Coding, Symbol Search, and Cancellation) based on Kaufman's (1994) criteria ($\geq$.70 = good, .50–.69 = fair, $<$.50 = poor). Table 2 illustrates robust Verbal Comprehension (Similarities, Vocabulary, Information, and Comprehension), Working Memory (Arithmetic, Digit Span, Picture Span, and Letter–Number Sequencing), Perceptual Reasoning (Block Design, Visual Puzzles, Matrix Reasoning, and Figure Weights), and Processing Speed (Coding, Symbol Search, and Cancellation) factors with theoretically consistent subtest associations. None of the subtests loaded saliently on more than one factor, but Picture Concepts was not substantially associated with any of the group factors. The moderate to high factor correlations presented in Table 2 (.387 to .747) suggest the presence of a general intelligence factor (Gorsuch, 1983).

Two- and three–factor models. Table 3 presents results from the two and three WISC–V factor models with promax rotation. For the three-factor model, the VC and PR factors merged, leaving distinct WM and PS factors, and the Arithmetic subtest cross-loaded on two factors, increasing model complexity. There were no cross-loadings on the two-factor model but the WM factor merged with the VC and PR factors to create a 13–subtest factor, leaving only the separate PS factor. The two- and three-factor models clearly display a fusion of theoretically meaningful constructs that is symptomatic of underextraction, thereby rendering them unsatisfactory (Gorsuch, 1983; Wood et al., 1996).

Hierarchical EFA: SL Bifactor Model

Given these results, the four-factor EFA solution appeared to be the most reasonable and was accordingly transformed with the SL orthogonalization procedure (see Table 4). After transformation, all subtests were properly associated with their theoretically proposed factor except Picture Concepts, which had minor loadings on both VC and PR factors. The hierarchical *g*–factor accounted for 35.5% of the total variance and 67.1% of the common variance. The general factor also accounted for between 3.9% (Cancellation) and 50.0% (Vocabulary) of individual subtest variability.

At the group factor level, VC accounted for an additional 4.8%, WM for an additional 3.4%, PR for an additional 3.0%, and PS for an additional 6.2% of the total variance. The general and group factors combined to measure 53% of the common variance in WISC–V scores, leaving 47% unique variance (combination of

Table 2

Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) Exploratory Factor Analysis: Four Oblique Factor Solution for the Total Standardization Sample (N = 2,200)

WISC–V subtest	*g*	F1: Verbal Comprehension	F2: Working Memory	F3: Perceptual Reasoning	F4: Processing Speed	h^2
Similarities	.752	**.766** (.805)	.037 (.597)	.017 (.615)	.001 (.323)	.649
Vocabulary	.774	**.878** (.856)	−.046 (.586)	.039 (.641)	−.047 (.288)	.735
Information	.754	**.794** (.815)	−.033 (.577)	.062 (.630)	−.004 (.316)	.666
Comprehension	.660	**.703** (.707)	.057 (.534)	−.082 (.511)	.064 (.326)	.506
Block Design	.669	−.011 (.550)	−.051 (.511)	**.738** (.750)	.119 (.427)	.573
Visual Puzzles	.679	.024 (.582)	−.045 (.514)	**.815** (.781)	−.045 (.314)	.612
Matrix Reasoning	.636	.071 (.544)	.198 (.559)	**.436** (.634)	.016 (.340)	.431
Figure Weights	.638	.126 (.569)	.134 (.537)	**.503** (.657)	−.072 (.272)	.454
Picture Concepts	.519	.249 (.488)	.080 (.431)	.227 (.483)	.031 (.270)	.274
Arithmetic	.724	.248 (.464)	**.373** (.627)	.157 (.476)	.050 (.322)	.534
Digit Span	.704	−.049 (.564)	**.845** (.813)	.029 (.562)	−.035 (.373)	.663
Picture Span	.573	−.012 (.652)	**.572** (.684)	.085 (.623)	.008 (.400)	.396
Letter–Number Sequencing	.690	.085 (.584)	**.814** (.792)	−.096 (.517)	−.033 (.355)	.634
Coding	.420	−.025 (.273)	.085 (.384)	−.070 (.311)	**.747** (.747)	.562
Symbol Search	.477	.023 (.333)	−.007 (.401)	.030 (.387)	**.756** (.776)	.603
Cancellation	.220	.018 (.149)	−.124 (.150)	.064 (.194)	**.443** (.419)	.182
Eigenvalue		6.87	1.50	1.00	.88	
% Variance		40.23	6.38	3.51	2.85	
Factor correlations						
F1: Verbal Comprehension		1.00				
F2: Working Memory		.716	1.00			
F3: Perceptual Reasoning		.747	.693	1.00		
F4: Processing Speed		.387	.490	.456	1.00	

Note. *g* = general structure coefficients based on first unrotated factor coefficients (*g*–loadings); h^2 = Communality. Factor pattern coefficients (structure coefficients) based on principal factors extraction with promax rotation (k = 4). Salient pattern coefficients presented in bold (pattern coefficient $\geq$.30).

Table 3

Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) Exploratory Factor Analysis: Two and Three Oblique Factor Solutions for the Total Standardization Sample (N = 2,200)

	Two oblique factors				Three oblique factors				
WISC–V subtest	*g*	F1	F2: PS	h^2	*g*	F1	F2: WM	F3: PS	h^2
SI	.751	**.812** (.767)	−.094 (.297)	.595	.750	**.797** (.784)	.007 (.571)	−.040 (.325)	.616
VC	.768	**.868** (.792)	−.158 (.259)	.646	.772	**.919** (.828)	−.069 (.558)	−.090 (.292)	.696
IN	.751	**.821** (.769)	−.108 (.287)	.601	.754	**.863** (.799)	−.062 (.550)	−.040 (.320)	.643
CO	.658	**.681** (.724)	−.029 (.418)	.445	.657	**.652** (.685)	.025 (.677)	.012 (.415)	.457
BD	.659	**.571** (.634)	.158 (.362)	.438	.658	**.528** (.671)	.036 (.526)	.196 (.357)	.448
VP	.666	**.661** (.671)	.017 (.391)	.448	.664	**.608** (.676)	.052 (.509)	.055 (.321)	.455
MR	.637	**.598** (.646)	.074 (.292)	.406	.634	**.424** (.611)	.215 (.559)	.065 (.368)	.403
FW	.638	**.658** (.669)	−.025 (.334)	.418	.636	**.526** (.644)	.159 (.525)	−.016 (.455)	.414
PC	.522	**.511** (.561)	.026 (.357)	.274	.520	**.441** (.635)	.079 (.536)	.045 (.305)	.275
AR	.728	**.680** (.680)	.091 (.416)	.530	.727	**.399** (.519)	**.359** (.425)	.048 (.286)	.536
DS	.687	**.624** (.647)	.116 (.433)	.473	.707	.004 (.598)	**.840** (.818)	−.048 (.387)	.671
PS	.569	**.507** (.523)	.113 (.272)	.325	.575	.079 (.498)	**.567** (.628)	.008 (.336)	.398
LN	.675	**.628** (.667)	.089 (.298)	.456	.690	.068 (.596)	**.757** (.776)	−.060 (.361)	.607
CD	.423	−.017 (.402)	**.751** (.759)	.551	.420	−.082 (.302)	.071 (.388)	**.731** (.731)	.537
SS	.479	.048 (.343)	**.736** (.742)	.578	.479	.034 (.372)	−.017 (.406)	**.773** (.779)	.608
CA	.221	−.023 (.177)	**.416** (.405)	.164	.220	.057 (.174)	−.126 (.153)	**.460** (.421)	.185
Eigenvalue		6.87	1.50			6.87	1.50	1.00	
% Variance		39.72	6.21			39.99	6.29	3.40	
Factor correlations									
	F1	1.00			F1	1.00			
	F2: PS	.481	1.00		F2: WM	.773	1.00		
					F3: PS	.454	.516	1.00	

Note. WISC–V 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; *g* = general structure coefficients based on first unrotated factor coefficients (*g*–loadings); h^2 = Communality; PS = Processing Speed; WM = Working Memory. Factor pattern coefficients (structure coefficients) based on principal factors extraction with promax rotation ($k = 4$). Salient pattern coefficients presented in bold (pattern coefficient ≥.30).

specific and error variance). Notably, the Cancellation and Picture Concepts subtests accounted for very little common variance, being heavily influenced by unique variance.

Omega-hierarchical and ω_s coefficients were estimated based on the SL results in Table 4. Because Picture Concepts failed to load on any factor, it was not included in ω coefficient estimation. The ω_h coefficient for general intelligence (.833) was high and sufficient for scale interpretation; however, the ω_s coefficients for the four WISC–V group factors (VC, WM, PR, and PS) were considerably lower (.167–.505). Thus, the four WISC–V group factors, with the possible exception of PS, likely possess too little true score variance for clinical interpretation (Reise, 2012; Reise et al., 2013).

Discussion

Support for a five first-order and one higher-order (*g*) factor model for the 16 primary and secondary subtests of the WISC-V was claimed in the WISC–V Technical and Interpretive Manual (Wechsler, 2014b) based on CFA analyses. However, there were numerous problems with those CFAs and an EFA model with five factors was found to be psychometrically unsatisfactory, producing a singleton factor and two subtests with no salient loadings (see Table 1). In contrast, the preferred EFA model had four first-order factors very similar to the WISC–IV. Following transformation with the Schmid and Leiman (1957) procedure, the WISC–V *g*–factor accounted for 6 to 12 times more variance than any single group factor and twice the variance of all four group factors combined. Whether this structure is also observed across smaller age groupings is not yet known and should also be examined as it is possible that some age groups might evidence variable structures.

The preeminence of general intelligence found in this study is similar to other studies of Wechsler scales using both EFA and CFA methods (Bodin et al., 2009; Canivez, 2014b; Canivez & Watkins, 2010a, 2010b; Gignac & Watkins, 2013; Nelson et al., 2013; Watkins, 2006, 2010; Watkins & Beaujean, 2014; Watkins et al., 2013, 2006) and other intelligence tests (Canivez, 2008; Canivez et al., 2009; DiStefano & Dombrowski, 2006; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007). Likewise, these results are consistent with the broader professional literature on the importance of general intelligence (Deary, 2013; Jensen, 1998; Lubinski, 2000; Ree et al., 2003).

Gustafsson (1984) noted that, "individual differences in cognitive performance can be understood in terms of several sources of variance, some of which are broad and some of which are narrow" (p. 67). Gorsuch (1983) explained that, "in science, the concern is with generalizing as far as possible and as accurately as possible. Only when the broad and not so broad generalities do not apply to a given solution does one move to the narrowest, most specific level of generality" (p. 249). Given that most of the WISC–V variance was contributed by a broad and general factor, the WISC–V general factor is "of definite interest" (Gorsuch, 1983, p. 253) but its "lower order factors may be of little interest" (Wolff & Preising, 2005, p. 50). As predicted by Frazier and Youngstrom (2007), the low/limited

Table 4

Sources of Variance in the Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) for the Total Standardization Sample (N = 2,200) According to a SL Exploratory Bifactor Model (Orthogonalized Higher–Order Factor Model) With Four First–Order Factors

	General		F1: Verbal Comprehension		F2: Working Memory		F3: Perceptual Reasoning		F4: Processing Speed			
WISC–V Subtest	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2
Similarities	.690	.476	**.416**	**.173**	.020	.000	.009	.000	.001	.000	.650	.350
Vocabulary	.707	.500	**.476**	**.227**	−.024	.001	.020	.000	−.040	.002	.729	.271
Information	.690	.476	**.431**	**.186**	−.018	.000	.032	.001	−.003	.000	.663	.337
Comprehension	.602	.362	**.381**	**.145**	.030	.001	−.042	.002	.055	.003	.513	.487
Block Design	.642	.412	−.006	.000	−.027	.001	**.382**	**.146**	.101	.010	.569	.431
Visual Puzzles	.656	.430	.013	.000	−.024	.001	**.421**	**.177**	−.038	.001	.610	.390
Matrix Reasoning	.609	.371	.039	.002	.105	.011	**.225**	**.051**	.014	.000	.434	.566
Figure Weights	.612	.375	.068	.005	.071	.005	**.260**	**.068**	−.061	.004	.456	.544
Picture Concepts	.487	.237	*.135*	*.018*	.043	.002	.117	.014	.026	.001	.272	.728
Arithmetic	.685	.469	.135	.018	**.198**	**.039**	.081	.007	.043	.002	.535	.465
Digit Span	.681	.464	−.027	.001	**.449**	**.202**	.015	.000	−.030	.001	.667	.333
Picture Span	.551	.304	−.007	.000	**.304**	**.092**	.044	.002	.007	.000	.398	.602
Letter–Number Sequencing	.661	.437	.046	.002	**.433**	**.187**	−.050	.003	−.028	.001	.630	.370
Coding	.383	.147	−.014	.000	.045	.002	−.036	.001	**.636**	**.404**	.555	.445
Symbol Search	.435	.189	.012	.000	−.004	.000	.016	.000	**.644**	**.415**	.604	.396
Cancellation	.197	.039	.010	.000	−.066	.004	.033	.001	**.377**	**.142**	.186	.814
Total Variance		.355		.049		.034		.030		.062	.529	.471
Common Variance		.671		.092		.065		.056		.116		
	ω_h = .828		ω_s = .251		ω_s = .185		ω_s = .166		ω_s = .505			

Note. b = loading of subtest on factor; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ω_h = omega hierarchical; ω_s = omega subscale. Bold type indicates coefficients and variance estimates consistent with the theoretically proposed factor. Italic type indicates coefficients and variance estimates associated with an alternate factor (where cross-loading b was larger than for the theoretically assigned factor). Because of the failure of Picture Concepts to achieve a salient factor pattern coefficient on any factor it was not included in ω_h and ω_s estimation.

portions of variance uniquely captured by the four group factors along with the low ω_s coefficients indicated that too little true score variance is associated with the four specific group factors, with the possible exception of PS, to warrant confident clinical interpretation (Reise, 2012; Reise et al., 2013).

Although the intention of the publisher was to separate the former Perceptual Reasoning factor into separate Visual Spatial and Fluid Reasoning factors, it appears that this was not successful. EFA results, along with the redundant loading of FR on general intelligence in Wechsler's (2014b) CFA, suggest substantial problems for separate VS and FR factors given the available 16 WISC–V subtests. In addition, the Picture Concepts subtest did not substantially contribute to any group factor and the Arithmetic subtest was associated with WM, the theoretically appropriate factor, but its pattern coefficient was considerably lower than the other WM subtests. There have been numerous problems observed with the Arithmetic subtest (perhaps better thought of as a quantitative reasoning task and good general intelligence measure) and it is likely time to abandon it as a WM measure in the Wechsler scales (Canivez & Kush, 2013; Watkins & Ravert, 2013). Further, the Cancellation and Picture Concepts subtests accounted for very little common variance (communality = .18 and .27, respectively), calling into question their value as indicators of intelligence (Child, 2006).

In summary, EFA of the 16 primary and secondary subtests from the WISC–V standardization data did not support the five-factor structure, and therefore the interpretation, recommended by its publisher. Researchers have also presented results disparate with those provided in the technical manuals of other ability tests (cf., Canivez, 2008; Canivez & Watkins, 2010a, 2010b; DiStefano & Dombrowski, 2006; Dombrowski, 2013; 2014a, 2014b; Dombrowski, Watkins, & Brogan, 2009; Watkins, 2006). When this pattern of conflicting results is considered in conjunction with professional standards that identify "the ultimate responsibility for appropriate test use and interpretation lies predominantly with the test user" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 141), it is clear that reliance on publisher provided manuals will be inadequate to ensure valid assessment. It has been 76 years since Kelley (1939) warned against accepting "mental factors of no importance" and 50 years since Buros (1965) warned that, "counselors, personnel directors, psychologists, and school administrators seem to have an unshakable will to believe the exaggerated claims of test authors and publishers" (p. xxiv). It is our fond hope that today's professionals will consult the results of studies and reviews published by independent researchers and apply rigorous psychometric standards so that they "know what their tests can do and act accordingly (Weiner, 1989, p. 829).

References

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

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. http://dx.doi.org/10.1080/09297040802603661

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

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

Buros, O. K. (1965). Preface. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (p. xxiv). Highland Park, NJ: Gryphon Press.

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

Canivez, G. L. (2010). Review of the Wechsler Adult Intelligence Test. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (4th ed., 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. http://dx.doi.org/10.1037/a0025973

Canivez, G. L. (2014a). Review of the Wechsler Preschool and Primary Scale of Intelligence. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook* (4th ed., pp. 732–737). Lincoln, NE: Buros Institute of Mental Measurements.

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. http://dx.doi.org/10.1037/spq0000032

Canivez, G. L. (in press). 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*. 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. http://dx.doi.org/10.1037/a0018030

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

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. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0022046

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

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325–335. http://dx.doi.org/10.1111/1468-2389.00184

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

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

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.

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. http://dx.doi.org/10.1016/B978-008043793-4/50036-2

Carroll, J. B. (2012). The three–stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 883–890). New York, NY: Guilford Press.

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

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

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. http://dx.doi.org/10.1111/j.1467-6494.2011.00739.x

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

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

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

Deary, I. J. (2013). Intelligence. *Current Biology, 23,* R673–R676. http://dx.doi.org/10.1016/j.cub.2013.07.021

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

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

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

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

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. http://dx.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. http://dx.doi.org/10.1177/0734282909333179

Elliott, C. D. (2007). *Differential Ability Scales–2nd ed.: Introductory and technical handbook*. San Antonio, TX: Harcourt Assessment, Inc.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299. http://dx.doi.org/10.1037/1082-989X.4.3.272

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

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment, 12,* 320–329. http://dx.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. http://dx.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., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model–based reliability in the WAIS–IV. *Multivariate Behavioral Research, 48,* 639–662. http://dx.doi.org/10.1080/00273171.2013.804398

Glutting, J. J., Adams, W., & Sheslow, D. (2000). *Wide Range Intelligence Test: Manual*. Wilmington, DE: Wide Range, Inc.

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. *The Journal of Special Education, 40,* 103–114. http://dx.doi.org/10.1177/00224669060400020101

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. (1997). Incremental efficacy of WISC–III factor scores in predicting achievement: What do they tell us? *Psychological Assessment, 9,* 295–301. http://dx.doi.org/10.1037/1040-3590.9.3.295

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. http://dx.doi.org/10.1037/a0021230

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

Gustafsson, J. E. (1984). A unifying model for structure of intellectual abilities. *Intelligence, 8,* 179–203. http://dx.doi.org/10.1016/0160-2896(84)90008-4

Gustafsson, J.-E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107–135). Westport, CT: Greenwood Press.

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

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

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

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73–98). New York, NY: Guilford Press.

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

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

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

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

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Kelley, T. L. (1939). Mental factors of no importance. *Journal of Educational Psychology, 30,* 139–142. http://dx.doi.org/10.1037/h0056336

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

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. http://dx.doi.org/10.1016/j.obhdp.2010.02.003

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "sinking shafts at a few critical points". *Annual Review of Psychology, 51,* 405–444. http://dx.doi.org/10.1146/annurev.psych.51.1.405

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology, 4,* 229–240.

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

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside.

Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences, 42,* 875–881. http://dx.doi.org/10.1016/j.paid.2006.09.021

Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling, 22,* 76–86. http://dx.doi.org/10.1080/10705511.2014.935751

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

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

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

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

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau–equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32,* 329–353. http://dx.doi.org/10.1207/s15327906mbr3204_2

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ree, M. J., Carretta, T. R., & Green, M. T. (2003). The ubiquitous role of *g* in training. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 261–274). New York, NY: Pergamon Press. http://dx.doi.org/10.1016/B978-008043793-4/50051-9

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.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. http://dx.doi.org/10.1080/00223891.2012.725437

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

Reynolds, C. R., & Kamphaus, R. W. (2003a). *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources, Inc.

Roid, G. (2003a). *Stanford–Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside Publishing.

Roid, G. (2003b). *Stanford–Binet Intelligence Scales: Fifth Edition, technical manual*. Itasca, IL: Riverside Publishing.

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

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464. http://dx.doi.org/10.1214/aos/1176344136

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

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

Strauss, E., Spreen, O., & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment, 12,* 237–244.

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

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

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

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A view and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: A festschrift to Douglas Jackson at seventy* (pp. 41–71). Norwell, MA: Kluwer Academic. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

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

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

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

Watkins, M. W. (2007). SE_{scree} [Computer Software]. State College, PA: Ed.& Psych Associates.

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

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. http://dx.doi.org/10.1037/spq0000038

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

Watkins, M. W., & Ravert, C. M. (2013). Subtests, factors, and constructs: What is being measured by tests of intelligence? In J. Kush (Ed.), *Intelligence quotient: Testing, role of genetics and the environment and social outcomes* (pp. 55–68). New York, NY: Nova Science.

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. http://dx.doi.org/10.1177/0013164406288168

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins. http://dx.doi.org/10.1037/10020-000

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: NCS Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: NCS Pearson.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children* (5th ed.). San Antonio, TX: NCS Pearson.

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

Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment, 53,* 827–831. http://dx.doi.org/10.1207/s15327752jpa5304_18

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

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

Yuan, K.-H., & Chan, W. (2005). On nonequivalence of several procedures of structural equation modeling. *Psychometrika, 70,* 791–798. http://dx.doi.org/10.1007/s11336-001-0930-9

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. http://dx.doi.org/10.1007/s11336-003-0974-7

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

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

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

Received April 22, 2015
Revision received July 9, 2015
Accepted September 4, 2015 ■