INTERNATIONAL JOURNAL OF SCHOOL & EDUCATIONAL PSYCHOLOGY
2021, VOL. 9, NO. 1, 73–87
https://doi.org/10.1080/21683603.2019.1638854

ARTICLE

Construct validity of the WISC–IV Italian edition: A bifactor examination of the standardization sample: Chi niente sa, di niente dubita

Joseph C. Kush [a] and Gary L. Canivez [b]

[a]Department of Instruction and Leadership in Education, Duquesne University, Pittsburgh, Pennsylvania, USA; [b]Department of Psychology, Eastern Illinois University, Charleston, Illinois, USA

ABSTRACT

This study utilized confirmatory factor analyses to examine the latent factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition, Italian adaptation (WISC–IV Italian) standardization sample. One through five, oblique first-order factor models and higher-order as well as bifactor models were examined and compared using CFA. The bifactor model provided the best explanation of the scale's factor structure. Across all models, general intelligence accounted for the largest amount of explained common variance, with group factors accounting for non-significant and trivial amounts of explained common variance. Omega-hierarchical subscale coefficients indicated that unit-weighted composites that would be generated by group factors would contain miniscule unique variance and therefore be of little clinical utility. Results are consistent with numerous empirical studies examining other European adaptations of the instrument and are at odds with the model put forward by the publisher. Clinicians who use the WISC–IV Italian adaptation are warned against attempting to interpret factor index scores independent of general intelligence.

KEYWORDS

WISC–IV Italian; confirmatory factor analysis; bifactor analysis; structural validity

Over the past 70 years the Wechsler Intelligence Scales for Children have become the most commercially popular scales of intelligence for children (Alfonso, Oakland, LaRocca, & Spanakos, 2000; Groth-Marnat, 2009; Kush, 1996; Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Zhu & Weiss, 2005), transforming from a two-factor (WISC: Wechsler, 1949) to a four-factor tool for assessing the cognitive skills of children (WISC–IV; Wechsler, 2003a). This popularity extends beyond the United States, to Europe, where the Wechsler scales for children have been translated and adapted for use in the United Kingdom (WISC–IV[UK]; Wechsler, 2004a; WISC–V[UK]; Wechsler, 2016a), France (French WISC–IV; Wechsler, 2005a; French WISC–V, 2016b) and Italy (WISC–IV Italian; Orsini, Pezzuti, & Picone, 2012). International popularity notwithstanding, the Wechsler scales remain contentious and have been criticized for their theoretical incongruence for over the past 50 years for reasons including increased numbers of factors despite their lack of theoretical support, weak factorial invariance, inadequate long-term stability, and miniscule incremental validity (Beaujean & Benson, 2018).

The development of the WISC–IV attempted to address some of these criticisms by anchoring the construction of the instrument to the theories of intelligence offered by Carroll, Cattell, and Horn (Carroll, 1993, 2003, 2012; Cattell & Horn, 1978; Horn, 1988; Horn & Blankson, 2005) that were later merged and referred to as the Cattell-Horn-Carroll (CHC) model of cognitive abilities (McGrew & Woodcock, 2001). Specifically, the WISC–IV included subtests that were analogous to Carroll's (1993, 2003, 2012) Stratum I, narrow abilities, factor indexes that were analogous to Stratum II broad abilities, and a composite estimate (i.e., FSIQ) of general intelligence (Stratum III) consistent with Wechsler's (1939) notion of global capacity, and to Spearman's *g* (1927). The WISC–IV CHC-based models (Chen, Keith, Chen, & Chang, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Lecerf, et al., 2010b; Weiss, Keith, Zhu, & Chen, 2013b) included some of the basic Wechsler structure for subtests and associations with Verbal Comprehension (VC; CHC *Gc*), Working Memory (WM; CHC *Gsm*) sans Arithmetic, and Processing Speed (PS; CHC *Gs*); however, the WISC–IV Perceptual Reasoning (PR) dimension was split into two CHC factors with Block Design and Picture Completion intending to measure visual processing (*Gv*) and Matrix Reasoning and Picture Concepts purportedly measuring fluid reasoning (*Gf*). These CHC models are not without critics, including Canivez and Kush (2013), who pointed out multiple deficiencies and diminished utility of the proposed CHC models for the WAIS–IV and WISC–IV (Weiss, Keith, Zhu, & Chen, 2013a, 2003b).

CONTACT Joseph C. Kush kush@duq.edu Duquesne University, 327 Fisher Hall, 600 Forbes Avenue, Pittsburgh, PA 15282, USA.

In a departure from the framework bequeathed by Wechsler, Verbal and Performance IQs were eliminated and were replaced by four-factor index scores: Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed (PS) and Full-Scale IQ was retained as a composite general ability estimate. The factor structure of the WISC–IV was examined by the publisher using both exploratory (EFA) and confirmatory (CFA) factor analyses of the normative sample with both core (10 subtest) and supplemental (15 subtest) batteries (Wechsler, 2003b). These analyses produced the expected four, first-order factors. Independent analyses of the normative data (Sattler, 2008), produced comparable factor structure and measurement invariance across gender (Chen & Zhu, 2012) as well as age (Keith et al., 2006) and for both clinical and non-clinical populations (e.g., Chen & Zhu, 2012; Keith et al., 2006).

Curiously, the publisher failed to perform a higher-order factor analysis to evaluate the proposed multi-level structure of the WISC–IV (Wechsler, 2003b). Subsequently, three independent analyses of the WISC–IV standardization sample were conducted. The first two analyses (Keith, 2005; Keith et al., 2006) were fraught with numerous methodological shortcomings (e.g., limited number of models considered, abandonment of simple structure, factor loadings of 1.00) although a more rigorous third study (Watkins, 2006), demonstrated the dominance of the general intelligence factor when compared to the four first-order factors where the general factor accounted for almost three-fourths of the common variance, and the largest first-order group factor contributed just over 10% of the common variance.

In examinations of European translations and adaptations of the WISC–IV, Watkins, Canivez, James, Good, and James (2013) examined the latent factor structure of 10 core subtests of the United Kingdom WISC–IV version (WISC–IV^{UK}; Wechsler, 2004a) using confirmatory factor analytic (CFA) methods and concluded that a resulting bifactor model provided the best explanation of WISC–IV^{UK} factor structure with a referred sample. Subsequently, Canivez, Watkins, Good, James, and James (2017) completed CFAs with all 15 core *and* supplemental WISC–IV^{UK} subtests and determined that bifactor and higher-order representations of Wechsler and CHC structures explained these data equally well with another referred sample. However, in all models the general intelligence factor captured substantially more variance than the four (Wechsler) or five (CHC) group factors. Nothing is publicly known about the latent factor structure of the WISC–IV^{UK} standardization sample as the *Administration and Scoring Manual* (Wechsler, 2004b) was devoid of such analyses and only the US-based WISC–IV *Technical and Interpretive Manual* (Wechsler, 2003b) was included in the sale of the WISC–IV^{UK}. Further, the publisher denied requests for standardization sample raw data or correlation matrices that would have allowed for such assessment (Canivez et al., 2017; Watkins et al., 2013). Comparable findings were found with the Spanish translation of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC–IV Spanish; Wechsler, 2005b), developed for Spanish-speakers in the United States. McGill and Canivez (2016) utilized higher-order exploratory factor analytic techniques with the Schmid and Leiman (Schmid & Leiman, 1957) procedure not included in the WISC–IV Spanish *Technical Manual* and found again that the *g* factor accounted for large portions of total and common variance. Using CFA techniques, McGill and Canivez (2018) obtained similar results where bifactor models were judged best, the *g* factor contained large portions of explained variance, and the four (10 subtest) or five (14 subtest) group factors contained minimal amounts of explained variance.

The Wechsler Intelligence Scale for Children is the most widely used intelligence test for children in Italy (Cianci, Orsini, Hulbert, & Pezuti, 2013). Published in 2012, the WISC–IV, Italian adaptation (WISC–IV Italian; Orsini et al., 2012) mirrored the factor structure reported in the United States edition; however, to date, only one independent research study exists examining the factor structure of the WISC–IV Italian adaptation based on the standardization sample. While the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014a, 2014b) recently became available in the United States, the WISC–IV Italian adaptation is the current version in that country and remains widely used by Italian practitioners. Information regarding the release of the next revision of the WISC–IV Italian has not been made publicly available.

In an extension of previous European studies, Kush and Canivez (2018) examined the latent factor structure of the WISC–IV Italian adaptation standardization sample using Watkins (2018) best practices in exploratory factor analyses (EFA) not included in the WISC–IV Italian *Technical Manual*. None of the extraction criteria supported the retention of four factors, as suggested by the WISC–IV Italian *Technical Manual*. However, when the forced four-factor structure was subjected to second-order factor analysis and transformed with the Schmid and Leiman (1957) orthogonalization procedure, the hierarchical *g* factor accounted for large portions of total and common variance, while the four first-order group factors accounted for small portions of total and common variance, rendering interpretation at the factor index level of dubious value.

Collectively, studies of both American and European WISC–IV factor structure are consistent across both EFA or CFA approaches (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Nakano & Watkins, 2013; Styck & Watkins, 2016; Watkins, 2006, 2010; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), and with other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Canivez, Watkins, & Dombrowski, 2016, 2017; Canivez, Watkins, & McGill, 2019; Golay & Lecerf, 2011; Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Lecerf & Canivez, 2018; Lecerf, Rossier, Faves, Reverte, & Coleaux, 2010b; McGill & Canivez, 2016; Nelson, Canivez, & Watkins, 2013; Watkins & Beaujean, 2014) in demonstrating that the largest portions of variance is captured by the *g* factor and only relatively small allotments of variance were uniquely associated with group factors.

Bifactor models of human intelligence examine group factors derived from the residual correlations that remain after extracting the general factor (Holzinger & Swineford, 1937), and have a long history in the field of cognitive assessment (Gustafsson & Balke, 1993). While both bifactor models and hierarchical models include a general factor, there are important differences between the two approaches. In a hierarchical model, a higher-order factor accounts for common variance across lower-order oblique factors, and common subtest variance is attributed to the intercorrelations between factors (Brown, 2006). In contrast, with bifactor models, the general factor reflects common variance across subtest indicators while multiple group factors reflect additional shared variance that is not attributed to the general factor. Additionally, in bifactor models, the general and group factors are orthogonal and compete equally to explain subtest variance; they are not considered as either "higher" or "lower" relative to each other (Brunner, Nagy, & Wilhelm, 2012; Gignac, 2008; Schmid & Leiman, 1957). Finally, with bifactor models, the identification of a general factor that explains the subtest indicator intercorrelations is retained, and in addition, *nuisance* factors are identified that express the subtest covariation that is independent of the covariation due to the general factor.

When considerations are made across fit indices (e.g., AIC, BIC, CFI, TLI, RMSEA, SRMR) bifactor models of intelligence have demonstrated considerable advantages (Canivez, 2014; Gignac, 2006) and when compared with alternative CFA models, bifactor model are often preferred (Mansolf & Reise, 2017; Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013). Additionally, as demonstrated by Gignac (2016), fit indices which incorporate a penalty for model complexity do not favor the bifactor model and proportionality constraints must be taken into consideration when comparing the higher-order and bifactor models. Consistent with other CFA approaches, statistical fit should not be the sole criterion for choosing a bifactor model, rather, the selection process should be theory-driven.

The exact number of factors claimed to be measured by the WISC–IV remains nebulous with some studies suggesting the instrument measures five (Golay et al., 2013; Grégoire, 2006; Keith et al., 2006) while others, perhaps six abilities (Flanagan & Kaufman, 2004; Lecerf et al., 2010). CFA bifactor models fit WISC–IV data as well or better than higher-order models and variance estimates for the general intelligence factor has far exceeded variance estimates of the group factors (Canivez, 2014; Canivez et al., 2017, 2017; Gignac & Watkins, 2013; Golay et al., 2013; Kush & Canivez, 2018; Watkins, 2010; Watkins & Beaujean, 2014; Watkins et al., 2013). The present study examines and compares both Wechsler- and CHC-based measurement models, using both higher-order and bifactor models, to determine best fit to WISC–IV Italian adaptation standardization sample data using CFA.

Method

Participants

The Italian standardization sample for the WISC–IV Italian adaptation (WISC–IV Italian; Orsini et al., 2012; Wechsler, 2012) consists of 2,200 participants divided into 11 age groups, each of 12 months range, ranging from 6 to 16 years and 11 months old. The sample was reported to be representative of the Italian population across parental educational level.

Instrument

The WISC–IV, Italian adaptation (WISC–IV Italian; Orsini et al., 2012), is a test of general intelligence that consists 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 manual for the Italian adaptation of the WISC–IV (Orsini et al., 2012) provides the subtest correlation matrix and describes relationships with Full-Scale IQ, the four-factor indexes, and two additional indexes (GAI and CPI). In reviewing the WISC–IV Italian test manual, internal consistencies, test–retest stability, inter-rater agreement, and standard errors of measurement are comparable with those of the American version (Wechsler, 2003a, 2003b).

Procedure/analyses

EQS 6.3 (Bentler & Wu, 2016) was used to conduct CFA using maximum likelihood estimation. Covariance matrices were produced for CFA using the correlation matrix, means, and *SD*s from the total WISC–IV Italian standardization sample presented in the WISC–IV Italian *Technical and Interpretative Manual* (Table 5.1). Some first-order factors were underidentified because they were measured by only two subtests. In those CFAs, the two subtests were constrained to equality before estimating bifactor models to ensure accurate identification (Little, Lindenberger, & Nesselroade, 1999).

Consistent with previous WISC–IV structural analyses, four first-order models and two hierarchical models were specified and examined: (a) one factor; (b) two oblique verbal and nonverbal factors; (c) three oblique verbal, perceptual, and combined working memory/processing speed factors; (d) four oblique verbal, perceptual, working memory, and processing speed factors; (e) an indirect hierarchical (higher-order) model (as per Bodin et al., 2009), with four first-order factors; and (f) a direct hierarchical (bifactor) model (as per Watkins, 2010), with four first-order factors. Gignac (2008) has provided a detailed description of direct and indirect hierarchical models, but the direct hierarchical model is the bifactor model described by Holzinger and Swineford (1937).

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 indicated better fit for the CFI, whereas lower values indicated better fit for the SRMR and RMSEA. 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). Combinatorial heuristics of Hu and Bentler (1999) were applied and criteria for adequate model fit were CFI ≥ .90, SRMR ≤ .09, and RMSEA ≤ .08. Good model fit required CFI ≥ 0.95 with SRMR and RMSEA ≤ 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 (ΔCFI > .01, ΔRMSEA < .015, and ΔAIC > 10) than alternative models (Burnham & Anderson, 2004; Chen, 2007; Cheung & Rensvold, 2002). All models were examined for presence of local fit problems (e.g., negative, too high, or too low standardized path coefficients, coefficients exceeding limits [−1, 1], negative variance estimates) as models should never be retained "solely on global fit testing" (Kline, 2016, p. 461).

Model-based estimates of variance proportions that would be obtained in unit-weighted composite scores from associated indicators were estimated with coefficients omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}) (Reise, 2012; Rodriguez, Reise, & Haviland, 2016). The ω_H is the unique general intelligence factor variability estimate with variability of group factors removed. The ω_{HS} is the unique group factor variability estimate with the influence of all other group and general factors removed (Brunner et al., 2012; Reise, 2012). Omega estimates (ω_H and ω_{HS}) 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, Bonifay, & Haviland, 2013). The value of ω_H and ω_{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. The Hancock and Mueller (2001) construct reliability or construct replicability coefficient (*H*) supplemented omega coefficients and estimated the latent construct adequacy represented by the indicators, using a criterion value of .70 (Hancock & Mueller, 2001; Rodriguez et al., 2016). *H* coefficients were produced by the *Omega* program (Watkins, 2013).

Results

Model fit statistics presented in Table 1 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), with the one- and two-factor models producing RMSEAs ≥ .08. The

Table 1. CFA fit statistics for the WISC–IV Italian standardization sample (N = 2,200).

Model	χ^2	*df*	CFI	RMSEA	90% CI RMSEA	SRMR	AIC
One factor	1,751.35	90	.848	.092	[.088, .095]	.067	156,878.12
Two oblique factors	1,396.79	89	.880	.082	[.078, .086]	.060	156,525.56
Three oblique factors	1,126.86	87	.905	.074	[.070, .078]	.055	156,259.63
Four oblique factors (Wechsler)	525.29	84	.960	.049	[.045, .053]	.035	155,664.06
Five oblique factors (CHC)	503.25	80	.961	.049	[.045, .053]	.034	155,650.02
Wechsler higher-order	549.82	86	.958	.050	[.046, .053]	.036	155,684.59
CHC higher-order	573.98	85	.955	.051	[.047, .055]	.037	155,710.75
Wechsler bifactor	**416.36**	**75**	**.969**	**.045**	**[.041, .050]**	**.030**	**155,573.13**
CHC bifactor[1]	451.52	77	.966	.047	[.043, .051]	.031	155,608.29

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. [1]Two indicators of the second (*Gv*) and third (*Gf*) factors were constrained to equality to ensure model identification. Best model presented in bold text.

oblique four-factor (VC, PR, WM, PS) Wechsler-based model and oblique five-factor (Gc, Gv, Gf, Gsm, Gs) CHC-based model provided the best fits to these data, however, meaningful differences in fit statistics (CFI, RMSEA, SRMR) were not observed. The oblique five-factor CHC model produced a statistically significant better fit than the oblique four-factor Wechsler model, $\Delta ML\chi^2 = 22.04$, $\Delta df = 4$, $p < .0002$; and the oblique five-factor CHC model had a meaningfully lower AIC ($\Delta AIC = 14.04$). Although both the Wechsler- and CHC-based oblique models fit these data well, the latent factor correlations (Table 2) for both models (Wechsler correlations ranging .384–.803; CHC correlations ranging .384–.965) 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 *Gv* and *Gf* 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 produced a statistically significant better fit than the Wechsler higher-order model, $\Delta ML\chi^2 = 133.46$, $\Delta df = 11$, $p < .0001$. The CHC bifactor model produced a statistically significant better fit than the CHC higher-order model, $\Delta ML\chi^2 = 122.46$, $\Delta df = 8$, $p < .0001$. The Wechsler bifactor model produced a statistically significant better fit than the CHC bifactor model, $\Delta ML\chi 2 = 35.16$, $\Delta df = 2$, $p < .0001$; and the Wechsler bifactor model produced the lower AIC. There were no meaningful fit statistics differences ($\Delta CFI > .01$ and $\Delta RMSEA > .015$) between the Wechsler bifactor (Figure 1), CHC bifactor (Figure 1), Wechsler higher-order (Figure 2), and CHC higher-order (Figure 2) models so all are presented for comparison and illustration. However, ΔAIC values showed meaningful differences ($\Delta AIC > 10$) and improvements from the CHC higher-order to Wechsler higher-order to CHC bifactor to Wechsler bifactor (best).

Tables 3–6 present decomposed variance estimates based on the four different models including *g* and four or five group factors for comparison. Explained common variance (ECV) was dominated by the *g* factor in all four models ranging from .669 to .739. The ω_H coefficients for the *g* factor in all four models were high, ranging from .898 to .906, and exceeded the .75 criterion for confident interpretation (Reise, 2012; Reise et al., 2013). Explained common variance was considerably lower for both the Wechsler group factors (VC, PR, WM, PS; .037 to .137) and CHC group factors (*Gc, Gv, Gf, Gsm, Gs*; .011 to .136). The ω_{HS} coefficients for the Wechsler- and CHC-based group factors were also low, ranging from .053 to .517; almost all falling short of the suggested minimum .50 criterion (Reise, 2012; Reise et al., 2013). Consistent with the ECV estimates, ω_{HS} coefficients were also particularly quite low for CHC-based *Gv* and *Gf* group factors with $\omega_{HS} < .10$, meaning unit-weighted composite scores based on *Gv* and *Gf* subtest indicators would account for <10% unique true score variance. Tables 3–6 also present *H* coefficients that reflect

Table 2. Latent factor correlations for WISC–IV Italian standardization sample (N = 2,200) Wechsler and CHC-based oblique models.

Wechsler Model	VC	PR	WM	PS	
VC	–				
PR	.803	–			
WM	.776	.791	–		
PS	.384	.502	.478	–	
CHC Model	*Gc*	*Gv*	*Gf*	*Gsm*	*Gs*
Gc	–				
Gv	.808	–			
Gf	.782	.965	–		
Gsm	.776	.735	.820	–	
Gs	.384	.520	.473	.478	–

Note. VC = Verbal Comprehension, PR = Perceptual Reasoning, WM = Working Memory, PS = Processing Speed, CHC = Cattell-Horn-Carroll, *Gc* = Crystallized Intelligence/
Comprehension Knowledge, *Gv* = Visual-Spatial, *Gf* = Fluid Intelligence/ Fluid Reasoning, *Gsm* = Short-term Memory (Working Memory), *Gs* = Processing Speed.

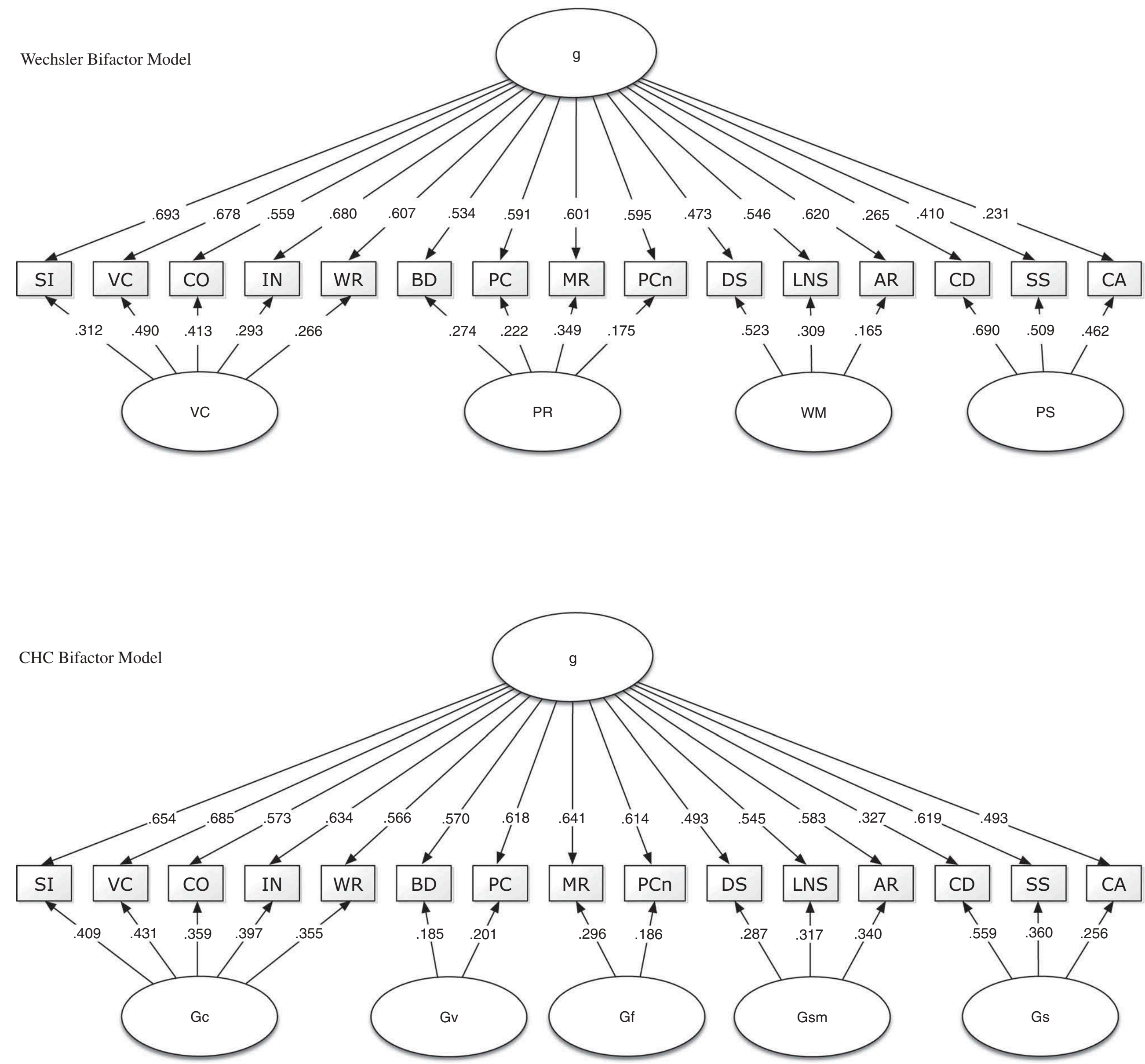

Figure 1. Bifactor measurement models (Wechsler bifactor model and CHC bifactor model), with standardized coefficients, for the WISC-IV Italian subtests with the standardization sample (N = 2,200). WISC–IV Italian subtests: SI = Similarities (Somiglianze), VC = Vocabulary (Vocabolario), CO = Comprehension (Comprensione), IN = Information (Informazione), WR = Word Reasoning (Ragionamento con le parole), BD = Block Design (Disegno con i cubi), PC = Picture Completion (Completamento di figure), MR = Matrix Reasoning (Ragionamento con le matrici), PCn = Picture Concepts (Concetti illustrate), DS = Digit Span (Memoria di cifre), LNS = Letter–Number Sequencing (Riordinamento di lettere e numeri), AR = Arithmetic (Ragionamento aritmetico), CD = Coding (Cifrario), SS = Symbol Search (Ricerca di simboli), CA = Cancellation (Cancellazione). All standardized path coefficients statistically significant ($p < .05$).

correlations between the latent factors and optimally weighted composite scores based on the assigned indicators (Rodriguez et al., 2016). The H coefficients for the general factor ranged from .879 to .888 and indicated the general factor would be well defined by the 15 WISC–V Italian subtests, but the H coefficients for the four (Wechsler) or five (CHC) group factors ranged from .072 to .605 and indicated that the group factors in either Wechsler or CHC configurations were inadequately defined by their subtest indicators.

Discussion

Two primary conclusions are borne out of the current results. First, the factor structure of the WISC–IV Italian adaptation is better characterized by a model

Figure 2. Higher-order measurement models (Wechsler higher-order model and CHC higher-order model), with standardized coefficients, for the 15 WISC-IV Italian subtests with the standardization sample (N = 2,200). WISC–IV Italian subtests: SI = Similarities (Somiglianze), VC = Vocabulary (Vocabolario), CO = Comprehension (Comprensione), IN = Information (Informazione), WR = Word Reasoning (Ragionamento con le parole), BD = Block Design (Disegno con i cubi), PC = Picture Completion (Completamento di figure), MR = Matrix Reasoning (Ragionamento con le matrici), PCn = Picture Concepts (Concetti illustrate), DS = Digit Span (Memoria di cifre), LNS = Letter–Number Sequencing (Riordinamento di lettere e numeri), AR = Arithmetic (Ragionamento aritmetico), CD = Coding (Cifrario), SS = Symbol Search (Ricerca di simboli), CA = Cancellation (Cancellazione). All standardized path coefficients statistically significant (p < .05).

that is substantially different than the one forwarded by the publisher. This conclusion is not unique to the present study and is supported by numerous independent research findings examining both the American and European versions of the test. Secondly, and also at odds with the claims of the test publisher, the underlying structure of the WISC–IV Italian is best explained primarily by general intelligence.

The current findings add to a growing body of evidence supporting a bifactor structure of the fourth edition of the Wechsler Intelligence Scale for Children (Canivez, 2014; Canivez et al., 2017; Watkins, 2010; Watkins et al., 2013) as well as with other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Gignac, 2005, 2006). By specifying a bifactor model, influences of *g* are *direct* to the subtests as are influences of the four primary factors (VC, PR, WM, and PS), rather than subtest influences of *g* being fully mediated by the four group factors prescribed by a higher-order model. The bifactor model allows *g* to be closer to the indicators (subtests) and *g* is conceptualized more as a breadth factor, rather than a superordinate factor (Gignac, 2008), allowing an examination of the distortion that may occur when unidimensional models are fit to multidimensional data. This seems more consistent with Spearman's (1904, 1927) conceptualization of general intelligence. One of the distinctive features of the bifactor model

Table 3. Decomposed CFA variance sources for the WISC–IV Italian standardization sample (N = 2,200): Wechsler bifactor model (Figure 1).

	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed				
WISC-IV Italian Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2	*ECV*
Similarities	.693	.480	.312	.097							.595	.405	.405
Vocabulary	.678	.460	.490	.240							.860	.140	.140
Comprehension	.559	.312	.413	.171							.482	.518	.518
Information	.680	.462	.293	.086							.725	.275	.275
Word Reasoning	.607	.368	.266	.071							.549	.451	.451
Block Design	.534	.285			.274	.075					.652	.348	.348
Picture Completion	.591	.349			.222	.049					.539	.461	.461
Matrix Reasoning	.601	.361			.349	.122					.640	.360	.360
Picture Concepts	.595	.354			.175	.031					.416	.584	.584
Digit Span	.473	.224					.523	.274			.975	.025	.025
Letter-Number Sequencing	.546	.298					.309	.095			.493	.507	.507
Arithmetic	.620	.384					.165	.027			.576	.424	.424
Coding	.265	.070							.690	.476	.456	.544	.544
Symbol Search	.410	.168							.509	.259	.606	.394	.394
Cancellation	.231	.053							.462	.213	.434	.566	.566
Total Variance		.309		.044		.018		.026		.063	.461	.539	
ECV		.669		.096		.040		.057		.137			
ω		.901		.857		.730		.684		.670			
ω_H/ω_{HS}		.803		.200		.118		.185		.517			
Relative ω		.892		.233		.162		.270		.771			
H		.881		.444		.233		.338		.605			
PUC		.790											

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

Table 4. Decomposed CFA variance sources for the WISC–IV Italian standardization sample (N = 2,200): CHC bifactor model (Figure 1).

	General		*Gc*		*Gv*		*Gf*		*Gsm*		*Gs*				
WISC-IV Italian Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2	ECV
Similarities	.678	.460	.346	.120									.579	.421	.793
Vocabulary	.661	.437	.509	.259									.696	.304	.628
Comprehension	.543	.295	.432	.187									.481	.519	.612
Information	.654	.428	.342	.117									.545	.455	.785
Word Reasoning	.592	.350	.298	.089									.439	.561	.798
Block Design	.568	.323			.201	.040							.363	.637	.889
Picture Completion	.616	.379			.201	.040							.420	.580	.904
Matrix Reasoning	.639	.408					.195	.038					.446	.554	.915
Picture Concepts	.613	.376					.195	.038					.414	.586	.908
Digit Span	.464	.215							.520	.270			.486	.514	.443
Letter-Number Sequencing	.540	.292							.326	.106			.398	.602	.733
Arithmetic	.609	.371							.188	.035			.406	.594	.913
Coding	.269	.072									.690	.476	.548	.452	.132
Symbol Search	.412	.170									.506	.256	.426	.574	.399
Cancellation	.240	.058									.457	.209	.266	.734	.216
Total Variance		.309		.051		.005		.005		.027		.063	.461	.539	
ECV		.670		.112		.012		.011		.060		.136			
ω		.901		.857		.562		.601		.682		.671			
ω_H/ω_{HS}		.805		.236		.058		.053		.199		.512			
Relative ω		.893		.275		.103		.088		.291		.763			
H		.880		.486		.078		.073		.345		.603			
PUC		.829													

Note. *Gc* = Crystallized Intelligence/Comprehension Knowledge, *Gv* = Visual-Spatial, *Gf* = Fluid Intelligence/Fluid Reasoning, *Gsm* = Short-term Memory (Working Memory), *Gs* = Processing Speed, *b* = standardized loading of subtest on factor, S^2 = variance, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω = Omega, ω_H= Omega-hierarchical (general factor), ω_{HS}= Omega-hierarchical subscale (group factors), *H* = construct reliability or replicability index, PUC = percentage of uncontaminated correlations.

is that both *g and* first-order group factors are simultaneous abstractions derived from the observed subtest indicators and therefore a more parsimonious and less complicated conceptual model (Canivez, 2016; Cucina & Byle, 2017; Gignac, 2008).

By placing the general factor at the same level as the group factors, the bifactor model is not really "hierarchical" as is the higher-order model that has dominated research on the structure of intelligence tests in the United States. The present results also support Carroll's theory due to the large contributions of *g* in WISC–IV Italian measurement and further supports previous commentary by Cucina and Howardson (2017) who also concluded from their analyses support for Carroll but not Horn-Cattell. These and other factors led Canivez and Youngstrom (2019) to call for an annulment of the arranged marriage of Cattell and Horn theory and Carroll's three stratum theory.

Consistent with the abundance of findings from the Wechsler scales for both American and international

Table 5. Decomposed CFA variance sources for the WISC–IV Italian standardization sample (N = 2,200): Wechsler higher-order model (Figure 2).

	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed				
WISC-IV Italian Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2	ECV
Similarities	.668	.446	.384	.147							.594	.406	.752
Vocabulary	.701	.491	.403	.162							.654	.346	.752
Comprehension	.586	.343	.338	.114							.458	.542	.750
Information	.650	.423	.373	.139							.562	.438	.752
Word Reasoning	.578	.334	.333	.111							.445	.555	.751
Block Design	.543	.295			.348	.652					.348	.652	.848
Picture Completion	.589	.347			.409	.591					.409	.591	.847
Matrix Reasoning	.615	.378			.446	.554					.446	.554	.848
Picture Concepts	.583	.340			.401	.599					.401	.599	.848
Digit Span	.503	.253					.269	.072			.325	.675	.778
Letter-Number Sequencing	.554	.307					.297	.088			.395	.605	.777
Arithmetic	.598	.358					.320	.102			.460	.540	.777
Coding	.326	.106							.560	.314	.420	.580	.253
Symbol Search	.360	.130							.619	.383	.513	.487	.253
Cancellation	.255	.065							.437	.191	.256	.744	.254
Total Variance		.308		.045		.016		.018		.059	.576	.424	
ECV		.691		.101		.037		.039		.133			
ω		.898		.855		.728		.659		.659			
ω_H/ω_{HS}		.804		.213		.111		.147		.492			
Relative ω		.895		.249		.152		.223		.747			
H		.879		.439		.206		.224		.568			
PUC		.790											

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

Table 6. Decomposed CFA variance sources for the WISC–IV Italian standardization sample (N = 2,200): CHC higher-order model (Figure 2).

	General		*Gc*		*Gv*		*Gf*		*Gsm*		*Gs*				
WISC-IV Italian Subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2	ECV
Similarities	.654	.428	.409	.167									.595	.405	.719
Vocabulary	.685	.469	.431	.186									.655	.345	.716
Comprehension	.573	.328	.359	.129									.457	.543	.718
Information	.634	.402	.397	.158									.560	.440	.718
Word Reasoning	.566	.320	.355	.126									.446	.554	.718
Block Design	.570	.325			.185	.034							.359	.641	.905
Picture Completion	.618	.382			.201	.040							.422	.578	.904
Matrix Reasoning	.641	.411					.296	.088					.498	.502	.824
Picture Concepts	.614	.377					.186	.035					.412	.588	.916
Digit Span	.493	.243							.287	.082			.325	.675	.747
Letter-Number Sequencing	.545	.297							.317	.100			.398	.602	.747
Arithmetic	.583	.340							.340	.116			.455	.545	.746
Coding	.327	.107									.559	.312	.419	.581	.255
Symbol Search	.619	.383									.360	.130	.513	.487	.747
Cancellation	.439	.193									.256	.066	.258	.742	.746
Total Variance		.334		.051		.005		.008		.020		.034	.452	.548	
ECV		.739		.113		.011		.018		.044		.075			
ω		.906		.855		.561		.624		.659		.646			
ω_H/ω_{HS}		.833		.241		.054		.080		.167		.270			
Relative ω		.919		.282		.095		.129		.253		.419			
H		.888		.476		.072		.117		.249		.402			
PUC		.829													

Note. Gc = Crystallized Intelligence/Comprehension Knowledge, *Gv* = Visual-Spatial, *Gf* = Fluid Intelligence/Fluid Reasoning, *Gsm* = Short-term Memory (Working Memory), *Gs* = Processing Speed, *b* = standardized loading of subtest on factor, S^2 = variance, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω = Omega, ω_H= Omega-hierarchical (general factor), ω_{HS}= Omega-hierarchical subscale (group factors), *H* = construct reliability or replicability index, PUC = percentage of uncontaminated correlations.

samples, the underlying structure of the WISC–IV Italian is best explained primarily by general intelligence (Bodin et al., 2009; Canivez, 2014; Canivez & Watkins, 2010a, 2010b; Gignac & Watkins, 2013; Kush & Canivez, 2018; Nelson et al., 2013; Watkins, 2006, 2010, 2013; Watkins & Beaujean, 2014; Watkins et al., 2006). Present results have also been replicated in independent assessments of the Canadian, Spanish, French, UK, and US versions of the WISC–V (Canivez et al., 2016; 2017; Canivez et al., 2019; Fennollar-Cortés & Watkins, 2018; Lecerf & Canivez, 2018; Watkins, Dombrowski, & Canivez, 2018). Current results also directly align with the recent Zaboski, Kranzler, and Gage (2018) meta-analysis of cognitive ability measures, which demonstrated that psychometric *g* had by far the strongest relationships with academic achievement, with *g* explaining more than 50% of the variance in achievement–more than all broad cognitive abilities combined, and with the meta-analytic findings of Warne and Burningham (2019) who found that *g* accounted for approximately half of the variance (45.9%) in the cognitive scores of individuals in non-Western and non-industrialized countries.

Decomposed variance estimates based on the bifactor models (Tables 3 and 4) and higher-order models (Tables 5 and 6) illustrated that the greatest portions of subtest variance were associated with the *g* factor and smaller portions of variance were associated with the four or five 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 the second-order *g* dimension (or bifactor/direct hierarchical *g*), which is estimated by the FSIQ score, and much smaller portions of total and common variance are apportioned to the first-order or group factors, estimated by the respective factor index scores. However, factor index scores in Wechsler scales or other first-order factor-based scores do not include only variance unique to that factor. Factor index scores or first-order factor-based scores conflate variance due to *g and* unique variance from the group factor; necessitating examination of ω_{HS} coefficients. This has been documented for the WISC–IV (Bodin et al., 2009; Watkins, 2006; Watkins et al., 2006), French WISC–IV (Golay et al., 2013), French WAIS–III (Golay & Lecerf, 2011), and the WAIS–IV (Canivez & Watkins, 2010a, 2010b). The implication of these findings is that the overall, omnibus FSIQ score should retain primary (if not exclusive) interpretive weight, as once the variance due to *g* is removed, the first-order factors do not provide sufficient information for interpretation.

Examination of the proportions of variance of the latent constructs indicated that the broad *g* factor had strong estimates allowing confident individual interpretation (ω = .901, ω_H = .803), but the ω_{HS} estimates for the four WISC–IV Italian adaptation group factors were low (.118-.517) and extremely limited for measuring unique constructs (Brunner et al., 2012; Reise, 2012). Therefore, the group factors (factor index scores) do not contain sufficient unique true score measurement variance for individual interpretation. For comparison purposes, standardized path coefficients from Watkins (2010) were used to calculate ω_H and present results were quite similar. The ω_H estimates for the four WISC–IV group factors from Watkins (2010) were also very low (.112-.388). Canivez (2014) also reported very low ω_H coefficients for the four WISC–IV group factors (.098-.330) in a sample of referred children demographically similar to Watkins (2010). This finding is consistent with studies from the US general population and referred samples (Bodin et al., 2009; Canivez, 2014; Styck & Watkins, 2016; Watkins, 2006, 2010; Watkins et al., 2006), and with Native American children (Nakano & Watkins, 2013). Present results are also consistent with several European WISC versions. Specifically, similar results were found for the WISC–IV^{UK} with Irish students (Canivez et al., 2017; Watkins et al., 2013), and with the WISC–IV Spanish (McGill & Canivez, 2016, 2018).

Psychometric *g* remains one of the most powerful constructs across the spectrum of psychological domains due to its presence in almost all scales of mental ability as well as its ability to predict a myriad of socially significant variables. It is critical however to remember the distinction between *g* and intelligence. The *g* factor is a hypothetical source of individual differences in performance on tasks on almost all cognitive measures regardless of their diversity and *g* is clearly related to the degree of complexity of the task's demands. The goal of this paper is not to promote any particular theory of intelligence but rather to point out that regardless of the theory of intelligence, the inclusion of the respective group factors (e.g., attention, inspection time, short-and long-term memory, analytical problem solving), will indeed improve predictive power but the influence will be relatively small when compared to *g*.

The evolution of the Wechsler scales in general, and the children's edition specifically, has followed a nonlinear trajectory. The selection of subtests is diverse and reflects a tremendous improvement over early scales (e.g., Galton's emphasis on sensory input – "Keenness of Sight and of Hearing; Colour Sense, Judgement of Eye; Breathing Power; Reaction Time; Strength of Pull and of Squeeze; Force of Blow; Span of Arms; Height, both standing and setting; and Weight" (Galton, 1908, p. 245). Yet in other ways, the composition of the scales have become bloated, attempting and claiming to do too much. The inclusion of many of the subtests ignores one of the critical characteristics of an intelligence test, put forward over 100 years ago, "Our purpose is to evaluate a child's level of intelligence. It should be understood that this means separating natural intelligence from instruction. It is his intelligence alone that we seek to measure, by disregarding as far as possible the degree of instruction which the child has enjoyed" (Binet, 1905, p. 196). Unfortunately, most modern revisions of the Wechsler scales are not supported by sound psychological theory. The identification of composite scores often appears to be done post-hoc; after the completion of factor analyses. Similarly, subtests appear to be treated as interchangeable and are added or deleted across revisions based on short-term influences. For example, between 1947 and 2002, Americans gained 24 points on Similarities, but only 2 points on Arithmetic and Information, while they all loaded heavily on *g* and on the Verbal factor (Flynn, 2009). However, the 10 core subtests of the WISC-IV used to measure Full-Scale IQ no longer include Information and Arithmetic.

Critics of commercially available intelligence tests (Gould, 1996; Murdoch, 2007) have argued that they include content that is influenced by nonintellectual factors, including reading ability, socioeconomic status, test-taking strategies, and cultural familiarity. The inclusion of these components improves the predictive power of the instruments because intelligence and other factors are being assessed (Watkins et al., 2006). Clearly, the knowledge of an individual's cognitive ability, as well as other factors related to academic achievement, offers an advantage over the sole knowledge of intellectual skills (Kush, Spring, & Barkand, 2012). However, as Jensen (1979) pointed out over 30 years ago, intelligence must be distinguished from learning, memory, and achievement. Intelligence is related to, but not the same as, academic achievement, and as Naglieri has cogently opined (Naglieri & Das, 1997; Naglieri & Rojahn, 2004), most current intelligence tests are contaminated with achievement content that confounds their interpretability.

We are not intelligence test apologists and believe that tests of intelligence represent one of the most important and well-researched diagnostic tools in a psychologist's battery. Global intelligence remains one of the single best predictors of academic and occupational success (Freberg, Vandiver, Watkins, & Canivez, 2008; Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012; Parker & Benedict, 2002; Roth et al., 2015; Sattler, 2008) and the general intelligence factor, as a construct, appears invariant and serves as an unbiased predictor across gender, disability, and ethnic groups (Kush & Watkins, 2007; Kush et al., 2001; Nakano & Watkins, 2013; Watkins & Kush, 2002). We believe problems occur only when the tests are asked to do more than they are psychometrically capable of.

Exaggerating the problem is that, many test publishers attempted to persuade test users that the strength of their convictions obviates the need for inquiry. We hope the profession has evolved sufficiently to recognize the fallacy in the circular logic of Boring's definition of intelligence (1923, p. 35) " … measurable intelligence is simply what the tests of intelligence test … " Despite repeated, feeble claims positing the value of clinical interpretations of factor index scores, results of the current study add to a considerable body of evidence that cautions against this practice (Bodin et al., 2009; Canivez, 2014; Canivez & Watkins, 2010a, 2010b; Canivez et al., 2016, 2017, 2019; Canivez et al., 2017, 2019; Fennollar-Cortés & Watkins, 2019; Gignac & Watkins, 2013; Lecerf & Canivez, 2018; McGill & Canivez, 2016, 2018; Nelson et al., 2013; Watkins, 2006, 2010; Watkins & Beaujean, 2014; Watkins et al., 2013; 2006).

Not unique to the Wechsler family of tests, publishers of commercial scales of intelligence continue to inflate the claims of what their instruments are capable of performing. Clearly, an instrument that boasts of the ability to assess multiple numbers of intelligence will have greater commercial value than an instrument that can only measure a single "type" of intelligence. Students of the scientific method recognize that paradigmatic shifts rarely develop from within the discipline. In this regard, the repeated, data-driven, findings from peer-reviewed publications warning against the practice of the over-interpretation of cognitive profiles (cf., McGill, Dombrowski, & Canivez, 2018) may not be enough for the test publishers to alter their practices. Unfortunately, these changes may not occur until external factors, such as litigation, for example, force publishers to reconsider claims that their scales have diagnostic utility that can be empirically substantiated. The WISC–IV Italian measures *g* quite well, but unique measurement of group factors is poor. Interpretations beyond Full-Scale IQ are a risky proposition that will likely be influenced by clinician's illusory correlation and confirmatory bias in decision-making.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Joseph C. Kush, Ph.D. is a Fulbright Scholar and Professor of Education at Duquesne University in Pittsburgh, PA. He serves on the editorial boards of several school psychology and assessment journals. His research interests include topics related to intellectual assessment and issues of test bias and test fairness for children from minority backgrounds. He is also strongly committed to issues of social justice. He is an accomplished guitarist and uses McIntosh computers exclusively.

Gary L. Canivez is Professor of Psychology at Eastern Illinois University principally involved in the Specialist in School Psychology program. Dr. Canivez is a Fellow of the

American Psychological Association Division of Quantitative and Qualitative Methods, a Charter Fellow of the Midwestern Psychological Association, a member of the Society for the Study of School Psychology, and past president of the Arizona Association of School Psychologists. He is an Associate Editor for *Archives of Scientific Psychology* and is an editorial board member for several school psychology and assessment journals. His research interests are in applied psychometrics in evaluating psychological and educational tests (including international applications); and empirically supported test interpretation.

ORCID

Joseph C. Kush http://orcid.org/0000-0001-9614-6351
Gary L. Canivez http://orcid.org/0000-0002-5347-6534

References

Alfonso, V. C., Oakland, T. D., LaRocca, R., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review*, *29*, 52–64. doi:10.1037.h0090497

Beaujean, A. A., & Benson, N. (2018 March). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*. Advance online publication. doi:10.1007/s40688-018-0182-1

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

Binet, A. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Annee Psychologique, 12, 191–244* (Translated in 1916 by E. S. Kite in The Development of Intelligence in Children). Vineland, NJ: Publications of the Training School at Vineland.

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

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic*, *35*, 35–36.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

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

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

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 multidimensionality 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., & 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., & 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., Good, R., James, K., & James, T. (2017). Construct validity of the WISC–IV^{UK} with a referred Irish sample: Wechsler and CHC model comparisons with 15 subtests. *British Journal of Educational Psychology*, *87*, 383–407. doi:10.1111/bjep.12155

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler intelligence scale for children–fifth UK edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, 89, 195–224. http://dx.doi.org/10.1111/bjep.12230

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York, NY: 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.

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., & 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, 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

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.

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

Cianci, L., Orsini, A., Hulbert, S., & Pezuti, L. (2013). The influence of parents' education in the Italian standardization sample of the WISC-III. *Learning and Individual Differences*, *28*, 47–53. doi:10.1016/j.lindif.2013.09.009

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

Cucina, J. M., & Howardson, G. N. (2017). Woodcock-Johnson–III, Kaufman Adolescent and Adult Intelligence Test (KAIT), Kaufman Assessment Battery for Children (KABC), and Differential Ability Scales (DAS) Support Carroll but Not Cattell-Horn. *Psychological Assessment*, *29*, 1001–1015. doi:10.1037/pas0000389

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

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC–IV assessment*. Hoboken, NJ: Wiley & Sons, Inc.

Flynn, A. S. (2009). *What is intelligence?* New York, NY: Cambridge University Press.

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC–III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology*, *15*, 131–139. doi:10.1080/09084280802084010

Galton, F. (1908). *Memories of my life*. London, UK: Methuen.

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.

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

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

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.

Gould, S. J. (1996). *The mismeasure of man* (*2nd* ed.). New York, NY: Norton.

Grégoire, J. (2006). *L'examen clinique de l'intelligence de l'enfant. Fondements et pratiques du WISC–IV [The clinical examination of the intelligence of the child. Foundations and practice of the WISC–IV]*. Sprimont, Belgium: Mardaga.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.

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

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

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

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York, NY: Plenum.

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 Press.

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

Jensen, A. R. (1979). The nature of intelligence and its relationship to learning. *Journal of Research and Development in Education*, *12*, 79–85.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? *Intelligence*, *40*, 123–138. doi:10.1016/j.intell.2012.01.009

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 Press.

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.

Kush, J. C. (1996). Factor structure of the WISC–III for students with learning disabilities. *Journal of Psychoeducational Assessment*, *14*, 32–40. doi:10.1177/073428299601400103

Kush, J. C., & Canivez, G. L. (2018). The higher order structure of the WISC–IV Italian adaptation using hierarchical exploratory factor analytic procedures. *International Journal of School & Educational Psychology*. Advance online publication. doi: 10.1080/21683603.2018.1485601.

Kush, J. C., Spring, M. B., & Barkand, J. (2012). Advances in the assessment of cognitive skills using computer based measurement. *Behavior Research Methods*, *44*, 125–134. doi:10.3758/s13428-011-0136-2

Kush, J. C., & Watkins, M. W. (2007). Structural validity of the WISC–III for a national sample of Native American students. *Canadian Journal of School Psychology*, *22*, 235–248. doi:10.1177/0829573507303056

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC–III for White and Black students from the WISC–III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review*, *30*, 70–88.

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

Lecerf, T., Reverte, I., Coleaux, L., Favez, N., & Rossier, J. (2010a). Indice d'aptitude général pour le WISC–IV: Normes francophones. (General ability index for the WISC–IV: French norms). *Pratiques Psychologiques*, *16*, 109–121. doi:10.1016/j.prps.2009.04.001

Lecerf, T., Rossier, J., Faves, N., Reverte, I., & Coleaux, L. (2010b). The four- vs. alternative six-factor structure of the French WISC–IV: Comparison using confirmatory factor analyses. *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

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

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

McGill, R. J., & Canivez, G. L. (2018). Confirmatory factor analyses of the WISC–IV Spanish core and supplemental Subtests: Validation evidence of the Wechsler and CHC models. *International Journal of School and Educational Psychology*, *6*, 239–251. doi:10.1080/21683603.2017.1327831

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, *71*, 108–121. doi:10.1016/j.jsp.2018.10.007

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

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

Murdoch, S. (2007). *IQ: A smart history of a failed idea*. Hoboken, NJ: Wiley.

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

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System: Interpretive handbook*. Itasca, IL: Riverside.

Naglieri, J. A., & Rojahn, J. (2004). Construct validity of the PASS theory and CAS: Correlations with achievement. *Journal of Educational Psychology*, *96*, 174–181. doi:10.1037/0022-0663.96.1.174

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., & 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

Orsini, A., Pezzuti, L., & Picone, L. (2012). *WISC–IV. Contributo alla taratura italiana*. Firenze, IT: Giunti OS.

Parker, D. R., & Benedict, K. B. (2002). Assessment and intervention: Promoting successful transitions for college students with ADHD. *Assessment for Effective Intervention*, *27*(3–24). doi:10.1177/073724770202700302

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of Intelligence 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

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

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. doi:10.1016/j.intell.2015.09.002

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Jerome M. Sattler.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. doi:10.1007/bf02289209

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

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

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

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

Warne, R. T., & Burningham, C. (2019). Spearman's *g* found in 31 non-Western nations: Strong evidence that *g* is a universal phenomenon. *Psychological Bulletin*, *145*, 237–272. doi:10.1037/bul0000184

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. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*, 219–246. doi:10.1177/0095798418771807

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

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

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

Watkins, M. W., & Kush, J. C. (2002). Confirmatory factor analysis of the WISC–III for students with learning disabilities. *Journal of Psychoeducational Assessment*, *20*, 4–19. doi:10.1177/073428290202000101

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. (1939). *Measurement of adult intelligence*. Baltimore, MD: Williams & Williams.

Wechsler, D. (1949). *Manual for the Wechsler intelligence scale for children*. 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. (2004a). *Wechsler intelligence scale for children–Fourth UK Edition*. London, UK: Harcourt Assessment.

Wechsler, D. (2004b). *Wechsler intelligence scale for children–Fourth UK Edition: Administration and scoring manual*. London: Harcourt Assessment.

Wechsler, D. (2005a). WISC–IV. In *Echelle d'intelligence de wechsler pour enfants-4e édition*. Paris, France: Pearson France-ECPA.

Wechsler, D. (2005b). *Wechsler intelligence scale for children–Fourth edition, Spanish*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2012). *WISC–IV Wechsler intelligence scale for children Quarta Edizione. Manuale di somministrazione e scoring*. Firenze, IT: Giunti OS.

Wechsler, D. (2014a). *Wechsler intelligence scale for children–Fifth Edition*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2014b). *Wechsler intelligence scale for children–Fifth Edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2016a). *Wechsler intelligence scale for children–Fifth UK Edition*. London, UK: Harcourt Assessment.

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

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

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of School Psychology*, *71*, 42–56. doi:10.1016/j.jsp.2018.10.001

Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 297–324). New York, NY: Guilford.

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 ω_H. *Applied Psychological Measurement*, *30*, 121–144. doi:10.1177/0146621605278814