

ARTICLE



## Confirmatory factor analyses of the WISC-IV Spanish core and supplemental subtests: Validation evidence of the Wechsler and CHC models

Ryan J. McGill <sup>a</sup> and Gary L. Canivez <sup>b</sup>

<sup>a</sup>School of Education, College of William & Mary, Williamsburg, Virginia, USA; <sup>b</sup>Department of Psychology, Eastern Illinois University, Charleston, Illinois, USA

### ABSTRACT

The present study examined the factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition, Spanish (WISC–IV Spanish, Wechsler, 2005a) with normative sample participants aged 6–16 years ( $N = 500$ ) using confirmatory factor analytic techniques not reported in the WISC–IV Spanish Manual (Wechsler, 2005b). For the 10 core subtest configuration, 1 through 4, first-order factor models, and higher-order versus bifactor models were compared using confirmatory factor analyses. The correlated four-factor Wechsler model provided good fit to these data, but the bifactor model showed statistically significant improvement over the higher-order model and correlated four-factor model. For the 14 core and supplemental subtest configuration, an alternative five-factor model based upon Cattell-Horn-Carroll (CHC; as per Weiss, et al., 2013b) configuration was also estimated. Results indicated that for the 14 subtest configuration, the alternative CHC model was preferred to the four-factor Wechsler model and the bifactor version of the CHC model also fit these data best. Across both configurations, variance apportionment and model-based reliability estimates illustrate well the dominance of the general intelligence factor when compared to the influence of the various combinations of group factors. Implications for clinical interpretation and the anticipated revision of the measurement instrument are discussed.

### KEYWORDS

bifactor model; clinical assessment; intelligence; model-based reliability; WISC–IV Spanish

Over the course of the last 50 years, the Wechsler intelligence scales have been among the most frequently utilized intelligence tests by school (educational) and clinical psychologists (Alfonso, Oakland, La Rocca, & Spanakos, 2000; Groth-Marnat, 2009; Sattler, 2008; Sotelo-Dynega & Dixon, 2014). Given their popularity, Wechsler scales have been adapted and translated for use in several countries with evidence provided to support measurement invariance across culture and between normative and referred samples (e.g., Chen, Keith, Weiss, Zhu, & Li, 2010; Chen & Zhu, 2012; Georgas, van de Vijver, Weiss, & Saklofske, 2003; San Miguel Montes, Allen, Puente, & Neblina, 2010).

The Wechsler Intelligence Scale for Children–Fourth Edition, Spanish (WISC–IV Spanish, Wechsler, 2005a) is a translation and adaptation of the WISC–IV (Wechsler, 2003) for use with Spanish-speaking children and adolescents ages 6–16 years. It includes 14 subtests (10 core and four supplemental), four first-order factor index scores (Verbal Comprehension [VC], Perceptual Reasoning [PR], Working Memory [WM], and Processing Speed [PS]), and a higher-order Full Scale score (FSIQ). The manual (Wechsler, 2005b) recommends that the first-order indexes should be the primary level of clinical

interpretation, especially when there is considerable variability in performance across those indicators.

According to Wechsler (2005b), the goal of the publisher was to develop a Spanish-language version of the WISC–IV that reflected the content, structure, and theoretical foundations of the parent instrument. As a result, users are encouraged to become familiar with the rationale for the content and structural revisions made to the WISC–IV from the WISC–III. More specifically, Harris, Muñoz, and Llorente (2008) suggested that the WISC–IV Spanish was developed to meet the needs of examinees learning English as a second language and in the process of acculturating to the United States. That is, attempts to develop a Spanish version of the WISC–III were never completed (Sanchez-Escobedo, Esquivel-Ancona, & Hollingworth, 2016). Although the WISC–IV was recently revised, the WISC–V Spanish is slated to be published in 2017; thus, the WISC–IV Spanish continues to be widely utilized by practitioners (Sotelo-Dynega & Dixon, 2014).

Although both exploratory and confirmatory factor analytic techniques were used to evaluate the core and supplemental subtests to evaluate the WISC–IV Spanish internal structure and reported in the manual,

considerable problems are noted. Several of these concerns involve the choice of the procedures employed for exploratory factor analysis (EFA).

The subtest intercorrelation matrix for the entire WISC-IV Spanish normative sample ( $N = 500$ ) was subjected to EFA using the principal axis extraction method (PAF) followed by oblique rotation (rotation unspecified) with a forced extraction based upon an *expected* four-factor solution. Initial results indicated that all subtest alignments were salient and consistent with the *predicted* model, thus demonstrating desired simple structure. These results were later replicated when applying the forced four-factor extraction procedure to two age groups (6–11 and 12–16; Table 5.3). Nevertheless, the use of a constrained analytical approach in which factor extraction is based upon subjective inference rather than more robust empirical criteria (e.g., parallel analysis, minimum average partials) is problematic, as it is essentially using EFA in a confirmatory context, a practice that has been critiqued extensively within the empirical literature (Costello & Osborne, 2005; Haig, 2005; Mulaik, 1987; Thompson, 2004).

More concerning was the failure to explicate the hierarchical structure of the WISC-IV Spanish despite the implied influence of Carroll's (1993) *three-stratum* model and the provision of scores (i.e., FSIQ) that imply a hierarchical measurement model. While oblique rotation was correctly employed under the assumption of correlated factors, it is not singularly sufficient for examining hierarchical structure and an additional step is required. To account for a potential hierarchical structure, Gorsuch (1983) and Thompson (2004) recommended that second-order factors be extracted and examined. Unfortunately, second-order factor analysis in the EFA procedures were not employed (Wechsler, 2005b). It has long been noted that interpretation of a second-order factor on the basis of first-order factors will typically overestimate the importance of lower-order factors at the expense of the higher-order factor (Carretta & Ree, 2001; Gignac, 2007, 2016; Watkins, 2006).

Confirmatory factor analytic (CFA) support for the WISC-IV Spanish structure was also reported in the manual (Wechsler, 2005b). Four models were examined: (a) one general factor; (b) two oblique (verbal and perceptual) factors; (c) three oblique (verbal, perceptual combined with working memory, and processing speed) factors; (d) four oblique (verbal, performance, processing speed, and working memory) factors. Tables 5.5 and 5.6 illustrate the increasingly better fit from one to four factors. Despite adequate

fit provided by the four-factor model, a correlated factors (oblique) model is considered insufficient because correlated factors, depending upon the nature of the data under consideration, may imply the presence of a hierarchical factor that may be worth additional investigation (Canivez, 2016; Gorsuch, 1983; Thompson, 2004). Also missing from the *Manual* were the hypothesized second-order subtest loadings, subtest specificity estimates, standardized CFA loading coefficients for any of the models that were estimated, and model-based reliability estimates including omega coefficients ( $\omega$ ; Canivez, 2016; Rodriguez, Reise, & Haviland, 2016). The body of literature on factor analysis methodology (e.g., Carroll, 1993, 1995; Gorsuch, 1983; Thompson, 2004) and model-based reliability (e.g., Reise, 2012; Reise, Bonifay, & Haviland, 2013) recommends the inclusion of this information because it assists test users in determining how the instrument should be interpreted.

Since its publication, independent psychometric investigations of WISC-IV Spanish have been scarce. Although San Miguel Montes et al. (2010) found that WISC-IV Spanish scores adequately discriminated between Puerto Rican children with various forms of brain dysfunction and normal controls, a hierarchical EFA (McGill & Canivez, 2016) using the aforementioned techniques suggested by Carroll (1993, 1995) on the WISC-IV Spanish normative sample correlations supported the subtest associations with the four factors reported in the manual, but found that the total and common variance accounted for by the four lower-order factors was dwarfed by that explained by the general factor of intelligence ( $g$ ). As a result, McGill and Canivez (2016) recommended that users largely forgo interpretation of factor indexes in clinical practice. It should be noted that these results were consistent with similar EFA research on the WISC-IV (Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), WAIS-IV (Canivez & Watkins, 2010a, 2010b), and the recently revised WISC-V (Canivez, Watkins, & Dombrowski, 2016; Dombrowski, Canivez, Watkins, & Beaujean, 2015).

However, the findings produced by McGill and Canivez (2016) are not without limitations. Although the authors noted that empirical criteria (e.g., parallel analysis, minimum average partials) did not support a four-factor solution for the WISC-IV Spanish normative data, they elected to proceed with a forced four-factor extraction, consistent with publisher theory, and only those results are reported. That is, other plausible

models for these data (i.e., two- and three-factor solutions) was not explored. Additionally, the authors limited their analyses to the 10 subtest core battery configuration and did not examine the alignment of the supplementary subtests with the four extracted factors.

An additional limitation is the fact that most hierarchical structure studies (e.g., McGill & Canivez, 2016) of the Wechsler scales have examined the relationship between higher-order  $g$  and the first-order factors with the effects of  $g$  on the subtests fully mediated through the first-order factors (i.e., indirect hierarchical model). According to Beaujean (2015), such models obfuscate the role and influence of  $g$  on the manifest variables (MV). As an alternative, a bifactor model (Holzinger & Swineford, 1937), sometimes referred to as a direct hierarchical (Canivez, 2014; Gignac, 2008) or nested-factors (Gustafsson & Balke, 1993) model, suggests that  $g$  and the group-specific factors have simultaneous direct effects on the MVs. Reise (2012) noted several advantages of the bifactor model, including (a) estimation of direct effects makes the influence of  $g$  easier to interpret, (b) both general and specific influences on subtests can be examined simultaneously, and (c) more precise estimation of model-based reliability coefficients that account for scale multidimensionality (e.g.,  $\omega$ ; Rodriguez et al., 2016). Although the bifactor model has been found to be a preferred solution when compared to the higher-order model for several versions of Wechsler scales (e.g., Canivez, 2014; Canivez, Watkins, & Dombrowski, 2017; Canivez, Watkins, Good, James, & James, 2017; Gignac & Watkins, 2013; Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Styck & Watkins, 2016; Watkins & Beaujean, 2014; Watkins, Canivez, James, Good, & James, 2013), additional replication of these results is needed, given the debate that this model has engendered within the psychometric community (e.g., Beaujean, 2015; Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013).

### Purpose of the present investigation

Consequently, the goal of the present study was to test the factor structure of the WISC-IV Spanish core and supplemental subtests using CFA. The present study adds to the current literature by exploring alternative measurement models for the WISC-IV Spanish core and supplemental subtests. More specifically, analyses (a) compared bifactor versus higher-order measurement models and (b) examined the tenability of a rival five-factor model based upon Cattell-Horn-Carroll configuration (CHC; Schneider & McGrew, 2012) for the total WISC-IV Spanish

battery that some researchers have suggested is preferred for Wechsler scales (e.g., Weiss, Keith, Zhu, & Chen, 2013a, 2013b). Although EFA and CFA are considered to be complementary procedures, Gorsuch (1983) noted that they provide answers to different empirical questions and that CFA is generally preferred when the goal of an investigation is to test theory. This is the first independent CFA of the WISC-IV Spanish, and thus, it is believed that the results obtained from the present investigation will be instructive for furthering our understanding of the structuring of WISC-IV Spanish variables, establishing evidence-based Wechsler interpretive procedures for clinical practice, and informing anticipated structural changes in the forthcoming revision of the instrument (i.e., Weiss et al., 2013a, 2013b).

## Method

### Participants

Participants were the 500 individuals aged 6–16 years included in the WISC-IV Spanish standardization sample. Detailed demographic characteristics are provided in the WISC-IV Spanish manual (Wechsler, 2005b). The standardization sample was obtained using stratified proportional sampling across variables of age, gender, parent educational level, and geographic region. Inspection of the normative information provided in the manual revealed a close match to U.S. Census 2000 estimates across the stratification variables, although, in an independent review, Braden and Iribarren (2007) noted that some geographic regions (i.e., Midwest) were undersampled. It should be noted that, although the manual indicates that the instrument was standardized using a larger sample of 851 participants, validation studies and norming commenced on a smaller subset of this sample. The reason for this discrepancy was not disclosed.

### Instrument

The WISC-IV Spanish (Wechsler, 2005a) is an individual test of general intelligence for ages 6–16 and is composed of 14 subtests ( $M_s = 10$ ,  $SD_s = 3$ ), 10 of which contribute to the measurement of four factor-based index scores: Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed (PS). Each of the four index scores is expressed as a standard score ( $M_s = 100$ ,  $SD_s = 15$ ). The FSIQ, thought to reflect psychometric  $g$ , is linear combination of the 10 core subtests (three VC, three PR, two WM, and two PS).

Internal consistency estimates ranged from .74 to .90 for the subtests and from .82 to .97 for index and FSIQ composite scores. Additional psychometric information is provided in the manual and independent test reviews are available (e.g., Braden & Iribarren, 2007; Clinton, 2007).

## Procedure and data analyses

### Confirmatory factor analysis

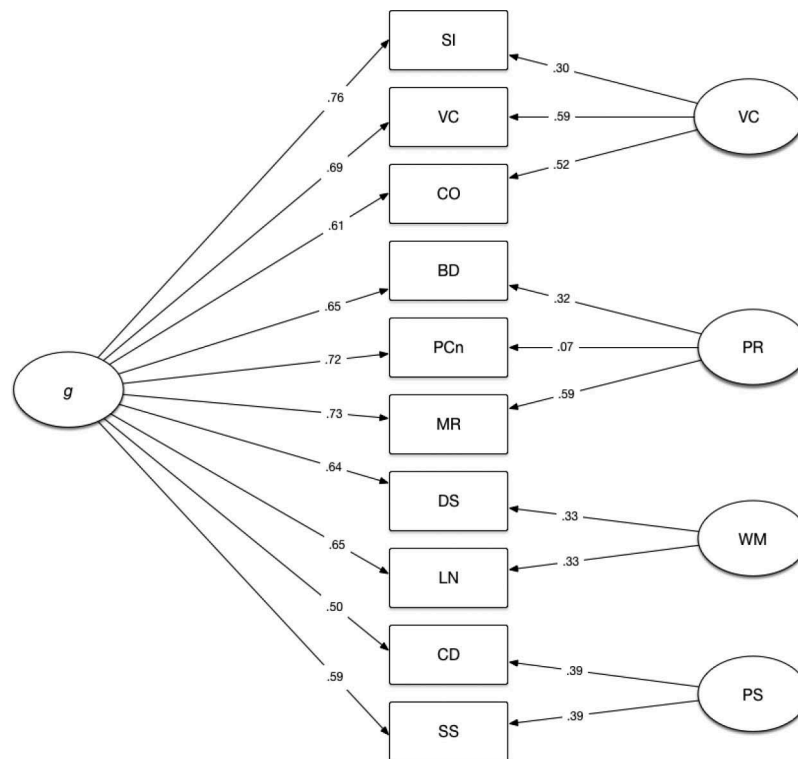
The WISC-IV Spanish subtest correlation matrix for the entire normative sample was extracted from the manual (p. 85, Table 5.1) and converted to a covariance matrix prior to conducting CFA. EQS, Version 6.2 (Bentler & Wu, 2012) was used to conduct CFA using maximum likelihood estimation. Separate analyses were conducted on the 10 core subtest and the 14 core and supplemental subtest configurations.

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 a combined processing speed or working memory) factors; (d) four oblique (verbal, performance, processing speed, and working memory) factors; (e) a higher-order model;

and (f) a bifactor model (as per Canivez, 2014), with four (Wechsler) group-specific factors (see Figure 1). For the 14 core and supplemental subtest configuration, an alternative five-factor CHC-based oblique model (Crystallized Ability [Gc], Fluid Reasoning [Gf], Visual Processing [Gv], Short-Term Memory [Gsm], and Processing Speed [Gs]) was also examined (as per Weiss et al., 2013b). Specification of a five-factor model with the core subtest configuration was not possible due to the fact that the Visual Processing factor could not be identified (i.e., only one aligned variable, Brown, 2016). In the bifactor model, factors with only two subtest indicators were constrained to equality to ensure specification as conducted by Canivez (2014) and Watkins and Beaujean (2014). Beaujean (2015) and Canivez (2016) provided a detailed description of the salient differences between bifactor and higher-order measurement models and assumptions regarding the appropriate structuring and influence of hypothesized cognitive dimensions, but the bifactor model is a variant of the so called *bifactor* model described by Holzinger and Swineford (1937).

### Model fit

To comport with best practice (e.g., Marsh, Hau, & Grayson, 2005; Mueller & Hancock, 2008), multiple indices



**Figure 1.** Wechsler bifactor (Direct hierarchical) measurement model, with standardized coefficients, for the WISC-IV Spanish Core Battery. SI = Similarities, VC = Vocabulary, CO = Comprehension, BD = Block Design, PCn = Picture Concepts, MR = Matrix Reasoning, DS = Digit Span, LN = Letter-Number Sequencing, CD = Coding, SS = Symbol Search.

were examined to evaluate the adequacy of model fit. Specifically, the (a) chi-square ( $\chi^2$ ), (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), (d) standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC) were used. Although there are no golden rules for evaluating model fit indices (Markland, 2007; West, Taylor, & Wu, 2012), the following guidelines were used for good model-fit criteria: (a) CFI  $\geq 0.95$ ; (b) SRMR and RMSEA  $\leq 0.06$  (Hu & Bentler, 1999). Higher CFI values and lower RMSEA values indicate better model fit, and these two indices were supplemented with  $\chi^2$  and AIC values. There are no specific criteria for information-based indices like the AIC, but smaller values may indicate better approximations of the true measurement model after accounting for model complexity (Vrieze, 2012). Meaningful differences between well-fitting models were evaluated based upon the following criteria: (a) change in chi-square ( $\Delta\chi^2$ ) for nested models, (b) exhibit good fit according to CFI, RMSEA, and SRMR indices; and (c) display the smallest AIC value (Burnham & Anderson, 2004).

### Model-based reliability

Finally, the bifactor model hypothesizes that each WISC-IV Spanish subtest is influenced simultaneously by two orthogonal constructs: a general ability factor ( $g$ ) and a first-order domain-specific group factor (e.g., verbal, perceptual, etc.). As a consequence, omega ( $\omega$ ) and omega-hierarchical ( $\omega_H$ ) for the general factor and omega-hierarchical subscale ( $\omega_{HS}$ ) for the group factors, were estimated as model-based reliability estimates of the factors

(Gignac & Watkins, 2013). Whereas  $\omega$  estimates the variance accounted for by both of the constructs in a given domain,  $\omega_H$  (and  $\omega_{HS}$ ) estimates the variance accounted for by a single target construct. Chen, Hayes, Carver, Laurenceau, and Zhang (2012) stressed that "for multi-dimensional constructs, the alpha coefficient is complexly determined, and McDonald's omega-hierarchical ( $\omega_h$ ; 1999) provides a better estimate for the composite score and thus should be used" (p. 228). The  $\omega_H$  coefficient is the model-based reliability estimate for the general intelligence factor with variability of group factors removed while the  $\omega_{HS}$  coefficient is the model-based reliability estimate of a group factor with all other group and the general factor removed (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012). Omega estimates ( $\omega_H$  and  $\omega_{HS}$ ) may be obtained from CFA bifactor solutions or decomposed variance estimates from higher-order models and were produced using the *Omega* program (Watkins, 2013), which is based on the tutorial by Brunner et al. (2012) and the work of Zinbarg, Revelle, Yovel, and Li (2005) and Zinbarg, Yovel, Revelle, and McDonald (2006). Although subjective, it has been suggested that omega coefficients should at a minimum exceed .50, but .75 is preferred (Reise, 2012; Reise et al., 2013).

## Results

### 10 core subtest configuration

Model fit statistics for the 10 core subtest configuration presented in Table 1 illustrate the increasingly better fit from one to four factors; however, fit statistics

**Table 1.** Confirmatory factor analysis fit statistics for WISC-IV Spanish normative sample ( $N = 500$ ).

Core Battery (10 Subtests)	$\chi^2$	df	$p$	CFI	SRMR	RMSEA	90% CI	AIC
							RMSEA	
First-order models								
One factor	312.96	35	< .01	.88	.06	.13	[.11, .14]	247.96
Two oblique factors (V, NV)	181.12	34	< .01	.94	.05	.09	[.08, .11]	113.12
Three oblique factors (V, P, PS/WM)	139.05	32	< .01	.95	.05	.08	[.07, .10]	75.05
Four oblique factors (Wechsler)	94.67*	29	< .01	.97	.04	.07	[.05, .08]	36.67
Hierarchical models								
Wechsler Higher-Order	107.93	31	< .01	.97	.04	.07	[.06, .09]	45.93
<b>Wechsler Bifactor<sup>a</sup></b>	<b>51.51**</b>	<b>27</b>	<b>&lt; .01</b>	<b>.99</b>	<b>.03</b>	<b>.04</b>	<b>[.02, .06]</b>	<b>2.50</b>
Core and Supplemental Battery (14 Subtests)								
First-order models								
One factor	528.21	77	< .01	.87	.06	.11	[.10, .12]	374.21
Two oblique factors (V, NV)	355.93	76	< .01	.92	.05	.09	[.08, .10]	203.93
Three oblique factors (V, P, PS/WM)	267.46	74	< .01	.94	.05	.07	[.06, .08]	119.46
Four oblique factors (Wechsler)	183.86	71	< .01	.97	.04	.06	[.05, .07]	41.86
Five oblique factors (CHC)	169.34*	67	< .01	.97	.04	.06	[.05, .07]	35.34
Hierarchical models								
CHC Higher-Order	207.99	72	< .01	.96	.04	.06	[.05, .07]	63.99
<b>CHC Bifactor<sup>a</sup></b>	<b>158.25**</b>	<b>65</b>	<b>&lt; .01</b>	<b>.97</b>	<b>.03</b>	<b>.05</b>	<b>[.04, .06]</b>	<b>28.25</b>

Note. V = Verbal, NV = Nonverbal, P = Perceptual, PS = Processing Speed, WM = Working Memory, CHC = Cattell-Horn-Carroll. In the Wechsler four-factor model, correlations ranged from .63 to .81. In the CHC five-factor model, correlations ranged from .62 to .96. CFI = comparative fit index, SRMR = standardized root mean square residual, RMSEA = root-mean-square error of approximation, AIC = Akaike information criterion. All values rounded to the nearest hundredth.

<sup>a</sup>Factors with less than three indicators were constrained to equality in order to ensure identification.

\*Statistically different ( $p < .01$ ) from previous four models.

\*\*Statistically different ( $p < .01$ ) from previous two models.

indicated that the one- two- and three-factor models were inadequate ( $CFI < .95$  or  $RMSEA > .06$ ). The oblique four-factor Wechsler model provided the best fit to the normative data among the first-order models. However, because the four WISC-IV Spanish factors were highly correlated (.63–.81), a higher-order dimension may be present (Gignac, 2016; Gorsuch, 1983; Thompson, 2004). As a result, alternative models with a general factor were examined. The bifactor model illustrated in Figure 1 provided a statistically better fit to these data than both the correlated factors model ( $\Delta\chi^2 = 43.16$ ,  $\Delta df = 2$ ,  $p < .01$ ) and the higher-order model ( $\Delta\chi^2 = 56.42$ ,  $\Delta df = 4$ ,  $p < .01$ ). Additionally, based on Hu and Bentler's (1998, 1999) dual criteria, the bifactor model produced good fit and meaningful differences (e.g., Cheung & Rensvold, 2002; Chen, 2007) in fit statistics from the two previous models. The bifactor model also produced the lowest AIC value of all the models that were tested, and, given the advantages (Brunner et al., 2012; Reise, 2012), it was selected as the best explanation for the WISC-IV Spanish 10 core subtest configuration.

Table 2 presents the decomposed subtest variance estimates of the WISC-IV Spanish 10 core subtest configuration based upon the bifactor model. As illustrated in Table 2, the general factor accounted for 72.2% of the common and 43.3% of the total variance. The VC factor accounted for 11.3% of the common variance and 6.7% of the total variance, the PR factor accounted for 7.6% of the common variance and 4.6% of the total variance, the PS factor accounted for 5.2% of the common variance and 3.1% of the total variance, and the WM factor accounted for 3.7% of the common variance and 2.2% of the total variance. Thus, the  $g$  factor accounted for substantially greater portions of WISC-IV Spanish common and total variance relative

to the four group factors. The  $\omega$ ,  $\omega_H$ , and  $\omega_{HS}$  coefficients were estimated based on the standardized loadings in Table 2. The  $\omega_H$  coefficient for general intelligence (.84) was high and sufficient for confident scale interpretation; however, the  $\omega_{HS}$  coefficients for the four group factors were considerably lower (.15–.27). Thus, unit-weighted scores based on the four group factors likely possess too little true score variance for confident clinical interpretation (Reise, 2012; Reise et al., 2013).

#### 14 core and supplemental subtest configuration

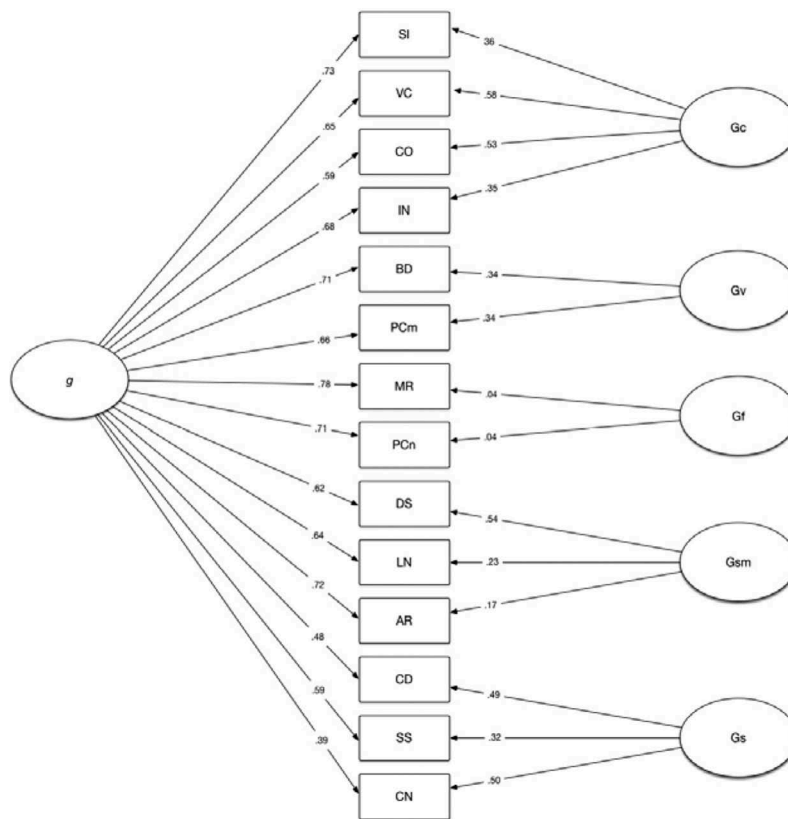
Model fit statistics for the 14 core and supplemental subtest configuration presented in Table 1 illustrate the increasingly better fit from one to five factors. Although both the four-factor (Wechsler) and five-factor (CHC) models provided good fit to these data, the alternative CHC-based model provided a statistically significant improvement in fit based upon the chi-square difference test ( $\Delta\chi^2 = 14.52$ ,  $\Delta df = 4$ ,  $p < .01$ ). Additionally, the five-factor CHC model provided the best fit to these data and yielded the lowest AIC value of all of the first-order models. However, factor correlations (.62–.96) suggested that a higher-order dimension may be present (Gignac, 2016; Gorsuch, 1983; Thompson, 2004). As a result, alternative models that included a general factor were examined. While fit statistics for the CHC higher-order model suggested it was an adequate solution for these data, specification of that model resulted in an isomorphic standardized coefficient (.99) between the first-order Fluid Reasoning ( $Gf$ ) factor and the second-order  $g$  factor. According to Brown (2016), loadings that approach unity suggest the presence of a Heywood case, rendering the model an untenable explanation for the data. Conversely, the CHC bifactor

**Table 2.** Sources of variance in the WISC-IV Spanish core battery according to the Wechsler bifactor model.

Subtest	General		VC		PR		PS		WM		$h^2$	$u^2$
	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$		
Vocabulary	.69	.47	.56	.31							.78	.22
Comprehension	.61	.37	.52	.27							.64	.36
Similarities	.76	.58	.30	.09							.67	.33
Matrix Reasoning	.72	.52			.59	.35					.88	.12
Picture Concepts	.72	.51			.07*	.01					.52	.48
Block Design	.65	.43			.32	.10					.53	.47
Coding	.50	.25					.39	.15			.40	.60
Symbol Search	.59	.35					.39	.15			.51	.49
Digit Span	.64	.41							.33	.11	.52	.48
Letter-Number Sequencing	.65	.43							.33	.11	.54	.46
% Total Variance	43.3		6.7		4.6		3.1		2.2		59.9	40.1
%Common Variance	72.2		11.3		7.6		5.2		3.7		100.0	
$\omega$	.92		.87		.83		.62		.69			
$\omega_H/\omega_{HS}$	.84		.27		.15		.21		.15			

Note. WISC-IV Spanish Wechsler Factors: VC = Verbal Comprehension, PR = Perceptual Reasoning, PS = Processing Speed, WM = Working Memory.  $b$  = standardized loading of subtest on factor,  $S^2$  = variance explained,  $h^2$  = communality,  $u^2$  = uniqueness.  $\omega$  = Omega,  $\omega_H$  = Omega-hierarchical (general factor),  $\omega_{HS}$  = Omega-hierarchical subscale (group factors). All values rounded to the nearest hundredth.

\* $p < .05$ .



**Figure 2.** CHC bifactor (Direct hierarchical) measurement model, with standardized coefficients, for the WISC-IV Spanish Core and Supplemental Battery. SI = Similarities, VC = Vocabulary, CO = Comprehension, IN = Information, BD = Block Design, PCm = Picture Completion, MR = Matrix Reasoning, PCn = Picture Concepts; DS = Digit Span, LN = Letter-Number Sequencing, AR = Arithmetic, CD = Coding, SS = Symbol Search, CN = Cancellation.

model illustrated in Figure 2 provided a statistically better fit to these data than both the correlated factors model ( $\Delta\chi^2 = 11.09$ ,  $\Delta df = 2$ ,  $p < .01$ ) and the higher-order model ( $\Delta\chi^2 = 49.74$ ,  $\Delta df = 7$ ,  $p < .01$ ).

Additionally, based on Hu and Bentler's (1998, 1999) dual criteria, the bifactor model produced good fit and meaningful differences (e.g., Chen, 2007; Cheung & Rensvold, 2002) in fit statistics from the higher-order

**Table 3.** Sources of variance in the WISC-IV Spanish core and supplemental battery according to the CHC bifactor model.

Subtest	General		Gc		Gf		Gv		Gs		Gsm		$h^2$	$u^2$
	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$	$b$	$S^2$		
Vocabulary	.66	.43	.58	.34									.77	.23
Comprehension	.59	.35	.54	.29									.64	.36
Similarities	.74	.55	.36	.13									.68	.32
Information	.69	.47	.35	.12									.60	.40
Matrix Reasoning	.78	.62			.04*	.00							.62	.38
Picture Concepts	.71	.51			.04*	.00							.51	.49
Block Design	.71	.51					.35	.12					.63	.37
Picture Completion	.66	.44					.35	.12					.56	.44
Coding	.59	.35							.54	.27			.64	.36
Symbol Search	.59	.35							.33	.11			.46	.54
Cancellation	.40	.16									.50	.25	.41	.59
Digit Span	.63	.39									.54	.29	.68	.32
Letter-Number Sequencing	.64	.42									.24	.06	.47	.53
Arithmetic	.73	.53							.17	.03			.56	.44
% Total Variance		43.3		6.3		0.0		1.7		4.6		2.7	58.6	41.4
%Common Variance		73.8		10.7		0.0		2.9		7.9		4.6	100.0	
$\omega$		.94		.89		.72		.74		.74		.79		
$\omega_H/\omega_{HS}$		.87		.28		.00		.15		.32		.15		

Note. WISC-IV Spanish CHC Factors: Gc = Crystallized Ability, Gf = Fluid Reasoning, Gv = Visual Processing, Gs = Processing Speed; Gsm = Short-Term Memory.  $b$  = standardized loading of subtest on factor,  $S^2$  = variance explained,  $h^2$  = communality,  $u^2$  = uniqueness.  $\omega$  = Omega,  $\omega_H$  = Omega-hierarchical (general factor),  $\omega_{HS}$  = Omega-hierarchical subscale (group factors). All values rounded to the nearest hundredth.

\* $p < .05$

model. The CHC bifactor model also produced the lowest AIC value of all the models that were tested, and, given the advantages (Brunner et al., 2012; Reise, 2012), it was selected as the best explanation for the WISC-IV Spanish 14 core and supplemental subtest configuration.

Table 3 presents the decomposed subtest variance estimates of the WISC-IV Spanish 14 core and supplemental subtest configuration based upon the CHC bifactor model. As illustrated in Table 3, the general factor accounted for 73.8% of the common and 43.3% of the total variance. The Crystallized Ability factor ( $G_c$ ) accounted for 10.7% of the common variance and 6.3% of the total variance, the  $G_f$  factor failed to account for any meaningful variance, the Visual Processing ( $G_v$ ) factor accounted for 2.9.2% of the common variance and 3.1% of the total variance, the Processing Speed ( $G_s$ ) factor accounted for 7.9% of the common variance and 4.6% of the total variance, and the Short-Term Memory ( $G_{sm}$ ) factor accounted for 4.6% of the common variance and 2.7% of the total variance. Thus, the  $g$  factor accounted for substantially greater portions of WISC-IV Spanish common and total variance relative to the five CHC group factors. The  $\omega$ ,  $\omega_H$ , and  $\omega_{HS}$  coefficients were estimated based on the standardized loadings in Table 3. The  $\omega_H$  coefficient for general intelligence (.87) was high and sufficient for confident scale interpretation; however, the  $\omega_{HS}$  coefficients for the five CHC group factors were considerably lower (.00–.32). Thus, unit-weighted scores produced from the five CHC group factors likely possess too little true score variance for confident clinical interpretation (Reise, 2012; Reise et al., 2013).

## Discussion

The WISC-IV Spanish (Wechsler, 2005a) is an individually administered test of intelligence that was developed for a number of applications for Spanish-speaking examinees. As such, it is frequently utilized by bilingual school and clinical psychologists, and previous versions have served as a reference for back translations for development of cross-cultural measures (Harris et al., 2008). Inexplicably, since its publication over a decade ago, its structure has yet to be independently evaluated using CFA and its 14 core and supplemental subtest configuration has evaded empirical scrutiny entirely. The results from the present independent CFA challenge the WISC-IV Spanish structure promoted in the manual. Although the CFA results furnished by Wechsler (2005b) supported a four-factor oblique model for both the 10 core and 14 core and supplemental subtest configurations, the potential influence of

the general intelligence dimension ( $g$ ) was only modeled for one of the first-order models (one-factor) that were examined. The present results confirm that all 14 core and supplemental subtests contribute to the measurement of a broad general intelligence dimension *and* a group factor (i.e., lower-order Wechsler or CHC-related abilities).

During test validation, the selection of a measurement model is crucial, as it provides the statistical rationale for the scores that are provided to users of that instrument (Brunner et al., 2012). Thus, it is difficult to understand why authors of tests such as the WISC-IV Spanish continue to rely on a correlated factors model to validate internal structure when a more complex multilevel structure is consistent with the variables thought to be sampled by that measure and in the scores (i.e., FSIQ) that are later developed and presented for interpretation by users. First-order measurement models focus only on broad abilities (that conflate general and group factor variance); as a consequence, they mask the potential influence of higher-order constructs such as  $g$  (Gignac, 2007, 2016). Additionally, Gignac (2016) argued that correlated factors models have unmodeled complexity and may potentially inflate and distort the importance of lower-order cognitive constructs.

Consistent with other investigations of Wechsler scales (e.g., Canivez, 2014; Canivez, Watkins, & Dombrowski, 2017; Canivez, Watkins, Good et al., 2017; Gignac & Watkins, 2013; Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Nelson, Canivez, & Watkins, 2013; Watkins & Beaujean, 2014; Watkins et al., 2013), the present results support the bifactor (direct hierarchical) model in both 10 and 14 WISC-IV Spanish configurations. Although several CFA investigations (e.g., Bodin, Pardini, Burns, & Stevens, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Weiss et al., 2013b) have suggested that a higher-order solution is preferred for the WISC-IV, it is important to note that these studies failed to model rival bifactor solutions. In the higher-order model, the general factor's influence is fully mediated by the first-order factors and the general factor is produced from the correlations between the group factors and thus conceptualized as a superordinate construct (Gignac, 2008). As a result, this model produces a constraint that restricts the general and specific variance within a group factor to be proportional (Gignac, 2016; Yung, McLeod, & Thissen, 1999). As a result, it has been argued that the bifactor model, in which the effects of the general and specific factors on measured variables are estimated directly, is preferred as the effects of the variables are easier to

interpret (Canivez, 2016; Gignac, 2007, 2016) and general intelligence is conceptualized as a breadth construct (Gignac, 2008).

Decomposed variance estimates based on the bifactor models (see Figures 1 and 2) presented in Tables 2 and 3 illustrate well that the greatest portions of variance were associated with the *g* factor and that the resulting four (core) and five (core and supplemental) group factors accounted for much smaller proportions of variance. Examination of model-based reliability estimates for WISC-IV Spanish constructs indicated that, although the general factor had strong estimates that were sufficient for individual interpretation for both the core ( $\omega_H = .84$ ) and supplemental batteries ( $\omega_H = .87$ ), the omega estimates for the four Wechsler group factors in the core battery (.15–.27) and the five CHC group factors identified in the supplemental battery (.00–.32) were all quite low. The dominance of the general factor supports primary interpretation of the FSIQ for the WISC-IV Spanish rather than the factor index scores regardless of whether the supplemental subtests are utilized to augment the 10 subtest core battery configuration.

Additionally, the present results suggest that the 14 subtest core and supplemental configuration may be better represented by a five-factor CHC-based measurement model, as opposed to the four-factor Wechsler model that represents the 10 subtest core configuration. However, as previously mentioned, the relatively small portions of variance accounted for by those group factors raise questions about the potential clinical utility of these and other related CHC indices (see Kranzler, Floyd, Benson, Zabolski, & Thibodaux, 2016). Nevertheless, this finding is consistent with the results from a reexamination of the WISC-IV normative data by Weiss and colleagues (2013b) that was commissioned as part of a special edition of the *Journal of Psychoeducational Assessment* to debate whether a four- or five-factor structure best fit Wechsler scales. However, Canivez and Kush (2013) challenged the final higher-order validation model produced by Weiss et al. (2013b) due to the fact that it included subtest cross-loading, the specification of an intermediary factor (quantitative reasoning) in order to disentangle *Gf* from *Gv*, and similar isomorphism between *Gf* and the second-order *g* factor. Additionally, it was noted that Weiss et al. failed to evaluate rival bifactor models. Nevertheless, these results appear to have been instrumental in the decision to move from a four-factor Wechsler structure to a five-factor CHC-based structure in the latest WISC revision. As the veracity of the five-factor CHC based WISC-V structure has been substantially challenged in several recent independent

factor analytic studies (e.g., Canivez et al., 2016, 2017; Dombrowski et al., 2015), this important debate appears to be far from settled. As it is anticipated that similar structural changes will accompany the forthcoming revision of the WISC-IV Spanish, the present results will be useful for evaluating the sufficiency of these potential modifications.

### Study limitations

While the present study is the first independent analysis of the WISC-IV Spanish internal structure using CFA techniques since its publication, it is not without limitations that must be considered when evaluating these results. Most notably, the results were derived from a single intercorrelation matrix representing the entire WISC-IV Spanish normative sample spanning an expansive age range. Although this is the only correlation matrix provided in the manual, additional examination of the structural validity of the WISC-IV Spanish at different points of the age span would be beneficial, including examination of invariance.

Additionally, although the present results challenge the interpretive recommendation for primary interpretation of the first-order factor scores, it is important to note that structural validity examinations are important, but not singularly sufficient for informing the construct validity of scores from intelligence tests such as the WISC-IV Spanish. While additional examinations of the incremental utility of the factor scores in capturing unique variance beyond the FSIQ have yet to be conducted with the WISC-IV Spanish, it is hard to imagine these specific group factors providing useful incremental information when predicting performance in academic achievement or relations with other external criteria, given the results produced from related analyses of other Wechsler scales (e.g., Canivez, 2013; Canivez, Watkins, James, James, & Good, 2014; Glutting, Watkins, Konold, & McDermott, 2006; Glutting, Youngstrom, Ward, Ward, & Hale, 1997), primarily because of the small portions of unique true-score variance such group factors contain.

Also, it is important to note that the WISC-IV Spanish was influenced by aspects of Carroll's (1993) three-stratum theory, which predated modern-day incarnations of CHC configuration (e.g., Schneider & McGrew, 2012). Although the five-factor CHC-based structure was consistent with other related WISC-IV CFA studies (e.g., Weiss et al., 2013b), additional research examining the relationships between WISC-IV Spanish constructs and external measures would be useful for determining the degree to which they cohere with other CHC dimensions (i.e., Reynolds, Keith, Flanagan, & Alfonso, 2013).

While the present results suggest that a bifactor model is preferred for the WISC-IV Spanish subtest configurations, some researchers have questioned whether the bifactor model is a tenable structure for human cognitive abilities (e.g., Murray & Johnson, 2013; Reynolds & Keith, 2013). Specifically, Murray and Johnson (2013) suggest caution in interpreting differences between higher-order and bifactor models, as model fit statistics may be biased in favor of the latter. However, Gignac (2016) suggests that this “bias” is due to the fact that the higher-order model imposes a proportionality constraint. However, the bifactor model “did not generally produce a better fit when the true underlying structure was not a bifactor one” (Morgan et al., 2015, p. 15) based on Monte Carlo simulations. Moreover, Murray and Johnson concluded that, when attempting to estimate or account for domain-specific abilities (something the WISC-IV Spanish is claimed to do), the “bifactor model factor scores should be preferred” (Murray & Johnson, 2013, p. 420). This seems critical in evaluation of the construct validity of the WISC-IV Spanish because of publisher claims of what factor index scores measure and primary emphasis on factor index score interpretation. While adjudication of this issue is beyond the scope of the present discussion, this important methodological debate is far from settled.

Finally, as previously mentioned, the WISC-IV Spanish is currently being revised with an anticipated publication date later this year. As it is anticipated that many clinicians will continue to utilize the previous version over the next one to two years, the present results will continue to have important implications for clinical interpretation of the measurement instrument. Furthermore, factor-analytic investigations of intelligence tests have additional implications beyond clinical assessment for the understanding of the nature of human cognitive abilities. In some ways, each investigation is inherently a replication of accepted theoretical structures of intelligence (e.g., higher-order model, Wechsler model, CHC). Furthermore, the so-called “reproducibility crisis” (Pashler & Wagenmakers, 2012) in scientific psychology illustrates well that it is sometimes beneficial to reevaluate the evidence base for widely accepted theories (or recommended application of those theories) in light of new developments by researchers. Thus, the results of these investigations continue to have merit beyond the publication date of a particular measure (McGill, 2017).

### Conclusion and implications for practice

The present study provides clinicians with important information substantially qualifying interpretive recommendations of the WISC-IV Spanish (Wechsler, 2005b). As “the ultimate responsibility

for appropriate test use and interpretation lies predominantly with the test user” (AERA, APA, & NCME, 2014, p. 141), clinicians using the WISC-IV Spanish in clinical evaluations must seriously consider the present information to make informed decisions about which WISC-IV Spanish scores have satisfactory psychometric utility. Although the present study located group-specific factors more successfully than a previous EFA (McGill & Canivez, 2016), variance apportionment and model-based reliability estimates suggested that these dimensions lack enough reliable target construct variance for individual decision making. Conversely, the estimates associated the general intelligence factor support primary clinical interpretation of the FSIQ. As a consequence, it is recommended that users of the WISC-IV Spanish focus most, if not all, of their interpretive weight at that level of measurement.

### Acknowledgment

A preliminary version of this research was presented at the 10th conference of the International Test Commission, Vancouver, British Columbia, Canada.

### About the authors

Ryan J. McGill, PhD, BCBA-D, NCSP, is assistant professor and director of the school psychology program at the College of William & Mary. His scholarly interests include psychological assessment and measurement, judgment and decision-making in applied psychology, and assessment and identification of specific learning disability.

Gary L. Canivez, PhD, is professor of psychology at Eastern Illinois University. He is a fellow of Division 5 (Quantitative and Qualitative Methods) of the American Psychological Association and an elected member of the Society for the Study of School Psychology. His scholarly interests include applied psychological measurement, individual intellectual assessment, and child and adolescent behavioral assessment.

### ORCID

Ryan J. McGill  <http://orcid.org/0000-0002-5138-0694>  
 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. Retrieved from <http://www.nasponline.org>
- American Educational Research Association, American Psychological Association, & National Council on

- Measurement on Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence*, 3, 121–136. doi:10.3390/jintelligence3040121
- Bentler, P. M., & Wu, E. J. C. (2012). *EQS for Windows*. Encino, CA: Multivariate Software.
- Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*, 15, 417–424. doi:10.1080/09297040802603661
- Braden, J. P., & Iribarren, J. A. (2007). Test review: Wechsler, D. (2005). Wechsler Intelligence Scale for Children-Fourth Edition Spanish. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment*, 25, 292–299. doi:10.1177/0734282907302955
- Brown, T. A. (2016). *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. 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. (2013). Incremental validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment*, 25, 484–495. doi:10.1037/a0032092
- Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*, 29, 38–51. doi:10.1037/spq0000032
- Canivez, G. L. (2016). Bifactor modeling in construct validation of multifaceted tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247–271). Gottingen, Germany: Hogrefe.
- Canivez, G. L., & 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 Wechsler Intelligence Scale for Children-Fourth UK Edition with a referred Irish sample: Wechsler and Cattell-Horn-Carroll model comparisons with 15 subtests. *British Journal of Educational Psychology*. Advance online publication. doi:10.1111/bjep.12155
- Canivez, G. L., Watkins, M. W., James, T., James, K., & Good, R. (2014). Incremental validity of WISC-IV<sup>UK</sup> factor index scores with a referred Irish sample: Predicting performance on the WIAT-II<sup>UK</sup>. *British Journal of Educational Psychology*, 84, 667–684. doi:10.1111/bjep.12056
- Carretta, T. R., & Ree, M. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment*, 9, 325–335. doi:10.1111/1468-2389.00184
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, 30, 429–452. doi:10.1207/s15327906mbr3003\_6
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. doi: 10.1080/10705510701301834
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80, 219–251. doi:10.1111/j.1467-6494.2011.00739.x
- Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, 49, 677–682. doi:10.1016/j.paid.2010.06.004
- Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, 52, 161–166. doi:10.1016/j.paid.2011.10.006
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. doi: 10.1207/S15328007SEM0902\_5
- Clinton, A. (2007). Test review: Wechsler, D. (2005). Wechsler Intelligence Scale for Children-Fourth Edition Spanish. San Antonio, TX: Psychological Corporation. *Journal of Psychoeducational Assessment*, 25, 285–292. doi:10.1177/0734282907300448
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9. Retrieved from <http://pareonline.net/getvn.asp?v=10&n=7>
- Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children-Fifth Edition with the 16 primary and secondary subtests. *Intelligence*, 53, 194–201. doi:10.1016/j.intell.2015.10.009
- Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In

- J. Georgas, L. G. Weiss, & F. J. R. van de Vijver (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 278–312). San Diego, CA: Academic Press.
- 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. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences*, 42, 37–48. doi:10.1016/j.paid.2006.06.019
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bi-factor 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
- 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 WIAT-II. *Journal of Special Education*, 40, 103–114. doi:10.1177/00224669060400020101
- Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of the WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment*, 9, 295–301. doi:10.1037/1040-3590.9.3.295
- 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. *Psychological Assessment*, 25, 496–508. doi:10.1037/a0030676
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: John Wiley.
- Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434. doi:10.1207/s15327906mbr2804\_2
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, 40, 303–329. doi:10.1207/s15327906mbr4003\_2
- Harris, J. G., Muñoz, M. R., & Llorente, A. M. (2008). Considerations in using the WISC-IV with Hispanic children. In A. Prifiteria, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 409–442). San Diego, CA: Academic Press.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41–54. doi:10.1007/BF02287965
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. *Psychological Methods*, 3, 424–453. doi:10.1037/1082-989X.3.4.424
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi:10.1080/10705519909540118
- Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review*, 35, 108–127. Retrieved from <http://www.nasponline.org>
- Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School and Educational Psychology*, 4, 124–136. doi:10.1080/21683603.2016.1155515
- Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, 42, 851–858. doi:10.1016/j.paid.2006.09.023
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Erlbaum.
- McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McGill, R. J. (2017). Re(examining) relations between CHC broad and narrow cognitive abilities and reading achievement. *Journal of Educational and Developmental Psychology*, 7, 265–282. doi:10.5539/jedp.v7n1p265
- 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
- Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in factor 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, 2–30. doi:10.3390/jintelligence3010002
- Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In J. W. Osborne (Ed.), *Best practices in quantitative methods*. Thousand Oaks, CA: Sage.
- Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, 22, 267–305. doi:10.1207/s15327906mbr2203\_3
- 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*, 37, 487–494. doi:10.1016/j.intell.2013.06.004
- 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
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530. doi:10.1177/1745691612465253
- 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
- Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwane (Eds.), *The Oxford handbook of child psychological assessment* (pp. 48–83). New York: Oxford University Press.
- Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology*, 51, 535–555. doi:10.1016/j.jsp.2013.02.003
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98, 223–237. doi:10.1080/00223891.2015.1089249
- San Miguel Montes, L. E., Allen, D. N., Puente, A. E., & Nebolina, C. (2010). Validity of the WISC–IV Spanish for a clinically referred sample of Hispanic children. *Psychological Assessment*, 22, 465–469. doi:10.1037/a0018895
- Sanchez-Escobedo, P., Esquivel-Ancona, F., & Hollingworth, L. (2016). Intellectual assessment of children and youth in Mexico: Past, present, and future. *International Journal of School and Educational Psychology*, 4, 247–253. doi:10.1080/21683603.2016.1163745
- Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). La Mesa, CA: Sattler.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford.
- Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools*, 51, 1031–1045. doi:10.1002/pits.21800
- 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.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17, 228–243. doi:10.1037/a0027127
- 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. (2013). *Omega* [Computer software]. Phoenix, AZ: Ed & Psych Associates.
- Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition. *School Psychology Quarterly*, 29, 52–63. doi:10.1037/spq0000038
- Watkins, M. W., Canivez, G. L., James, T., Good, R., & James, K. (2013). Construct validity of the WISC–IV<sup>UK</sup> with a large referred Irish sample. *International Journal of School and Educational Psychology*, 1, 102–111. doi:10.1080/21683603.2013.794439
- Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition among referred students. *Educational and Psychological Measurement*, 66, 975–983. doi:10.1177/0013164406288168
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2005a). *Wechsler Intelligence Scale for Children–Fourth Edition, Spanish*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2005b). *Wechsler Intelligence Scale for Children–Fourth Edition, Spanish: Manual*. San Antonio, TX: Psychological Corporation.
- 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, 93–113. 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
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. New York, NY: Guilford Press.
- Yung, Y. F., McLeod, L. D., & Thissen, D. (1999). On the relationship between the higher-order model and the hierarchical factor model. *Psychometrika*, 64, 113–128. doi:10.1007/BF02294531
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$ : 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. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for  $\omega$ H. *Applied Psychological Measurement*, 30, 121–144. doi:10.1177/0146621605278814