INTERNATIONAL JOURNAL OF SCHOOL & EDUCATIONAL PSYCHOLOGY
2019, VOL. 7, NO. S1, 2–14
https://doi.org/10.1080/21683603.2018.1474820

ARTICLE

Structural validity of the Spanish Wechsler Intelligence Scale for Children–Fourth Edition in a large sample of Spanish children with attention-deficit hyperactivity disorder

Javier Fenollar-Cortés[a], Carlos López-Pinar[b], and Marley W. Watkins[c]

[a]Department of Psychology, Universidad Loyola Andalucía, Sevilla, Spain; [b]Department of Psychobiology, University of Valencia, Valencia, Spain; [c]Department of Educational Psychology, School of Education, Baylor University, Waco, Texas, USA

ABSTRACT

The factorial structure of the WISC-IV for 859 Spanish children diagnosed with ADHD was examined. A bifactor model with the four factors first identified by Wechsler (2003a) was the best fit to the data. The Coding and Symbol Search subtests were particularly poor measures of *g* but relatively strong measures of the Processing Speed factor. In contrast, the Block Design (BD) and Picture Concepts (PC) subtests were relatively strong measures of *g* but weak measures of the Perceptual Reasoning factor. In fact, 80% of the BD variance and 97.8% of the PC variance and was due to the general factor. Additionally, the Wechsler bifactor model was invariant across ADHD-Combined and ADHD-Inattentive groups, permitting a direct comparison of WISC-IV across children diagnosed with these subtypes of ADHD. Only the FSIQ exhibited robust estimates of reliability (ω = .85 and ω_h = .70). In contrast, the group factor scores were unreliable measures of their proposed underlying factors (ω_{hs} coefficients ranging from .14 to .50). It is unlikely that WISC-IV index score profiles can validly contribute to ADHD assessments. Consequently, clinicians must produce psychometric evidence to justify the interpretation of Wechsler score profiles for children with ADHD.

KEYWORDS

ADHD; WISC-IV; CFA; bifactor; intelligence; validity; invariance

Some researchers have suggested that Wechsler Intelligence Scales for Children (WISC) score patterns could be valuable indicators for attention-deficit hyperactivity disorder (ADHD) assessment (Flanagan & Kaufman, 2009; Mayes & Calhoun, 2004; Schwean & Saklofske, 2005). Specifically, it has been hypothesized that children with ADHD are characterized by lower Processing Speed Index (PSI) or Working Memory Index (WMI) scores than children who do not have ADHD (Mayes & Calhoun, 2006; Thaler, Bello, & Etcoff, 2013). In fact, groups of children with ADHD have been found to exhibit a cognitive profile marked by lower PSI and WMI scores (Fenollar-Cortés, Navarro-Soria, González-Gómez, & García-Sevilla, 2015; Mayes & Calhoun, 2006; Scheirs & Timmers, 2009; Snow & Sapp, 2000; Yang et al., 2013). Consequently, clinicians have been advised that a unique WISC profile with low WMI or PSI scores is indicative of ADHD (Smitha, Dennis, Varghese, & Vinayan, 2014; Walg, Hapfelmeier, El-Wahsch, & Prior, 2017). Nevertheless, the diagnostic accuracy of WISC cognitive patterns for individual children with ADHD seems to be low (Devena & Watkins, 2012).

Most recent studies of WISC score patterns have applied the WISC-IV (Wechsler, 2003b) and have implicitly assumed that the factor structure of the WISC-IV is identical across clinical and normative groups. If the WISC-IV is not measuring the same constructs with equal fidelity across clinical and normative groups, then scores may differ due to influences such as time, gender, age, or assessment method, rather than clinical condition. That is, there would be “no clear basis for drawing inferences from the scores” (Kline, 2016, p. 396) and reliance on score patterns would be invalid. Thus, the structural validity of the WISC-IV among children with ADHD must be established before any analysis of score patterns can be undertaken (Wicherts, 2016).

Factor structure of the WISC-IV

Standardization sample

According to Wechsler (2003a), the WISC-IV has 10 core subtests that are indicators of four correlated factors: Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed

CONTACT Javier Fenollar-Cortés jfenollar@uloyola.es Department of Psychology, Universidad Loyola Andalucía, Sevilla 41014, Spain.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/usep.

(PS). Five supplemental subtests can be administered if needed for additional information or to substitute for one of the core subtests. Wechsler reported that this four-factor oblique structure was supported by exploratory (EFA) and confirmatory (CFA) factor analyses with both the 10 core subtest and the 15 subtest complete battery. A similar structure was proposed for the WISC-IV in other countries (e.g., Wechsler, 2004a, 2004b, 2005a, 2005b, 2005c, 2007, 2008, 2012).

Although the scoring structure of the WISC-IV includes a Full Scale IQ score based on its 10 core subtests, Wechsler (2003a) did not test a hierarchical factor structure with a general intelligence factor corresponding to the FSIQ. As noted by Gorsuch (1983), correlated first-order factors imply a hierarchical structure that should be "examined so that the investigator may gain the fullest possible understanding of the data" (p. 255). More recently, Gignac and Kretzschmar (2017) demonstrated that correlated-factor models do not necessarily ensure that distinctive group-level factors exist and recommended that researchers "extend their analyses beyond a correlated-factor model, whether from a higher-order modeling strategy or a bifactor modeling strategy" (p. 145).

There are two major types of hierarchical structures: higher-order and bifactor. The higher-order model assumes that the subtests are explained in part by first-order factors and that these factors are in turn explained by the higher-order general factor. In this model, general intelligence (*g*) is seen as a superordinate factor having a direct effect on several group factors but an indirect effect on the measured variables. In contrast, the bifactor model conceptualizes *g* as a breadth factor having direct effects on the measured variables (Reise, 2012). See Gignac (2008) for a detailed description of these models.

The omission of a hierarchical factor in analyses of the WISC-IV was soon rectified by independent researchers. Watkins (2006) applied EFA to the WISC-IV standardization sample by transforming an oblique higher-order structure into an orthogonal structure using the Schmid and Leiman (1957) transformation as recommended by Carroll (1993) and found a strong general factor of intelligence as well as the four group factors defined by Wechsler (2003a). That structure was also supported by Keith (2005), who applied CFA to the WISC-IV standardization sample and tested both higher-order and bifactor models.

Subsequently, researchers posited that a five-factor higher-order model aligned with the Cattell-Horn-Carroll (CHC; Schneider & McGrew, 2012) theoretical model of intelligence might better describe the standardization sample data (Keith, Fine, Taub, Reynolds, & Kranzler, 2006). In the CHC model, the Wechsler PR factor was split into fluid intelligence and visual-spatial reasoning factors. Empirically, the general intelligence factor explained about two thirds of the common variance in both the Wechsler four-factor and the CHC five-factor models and both models were good fits to the WISC-IV standardization sample data. Some researchers argued for interpretation of the CHC model (Weiss, Keith, Zhu, & Chen, 2013), whereas others argued for interpretation of the Wechsler model (Canivez & Kush, 2013). Similarly, both higher-order and bifactor models were good fits to the WISC-IV standardization sample data and both models revealed the dominant influence of the general factor. Some researchers argued for interpretation of the higher-order model (Reynolds & Keith, 2013) whereas others argued for interpretation of the bifactor model (Canivez, 2016). Notably, analyses of the WISC-IV standardization sample usually included all 15 subtests. This allowed each factor to be defined by two or more indicators, which was sufficient for statistical identification (McDonald & Ho, 2002).

Mixed clinical samples

IQ tests like the WISC-IV are seldom administered to nonclinical samples. Rather, they are almost always administered to children suspected of having a disability whose IQ scores may exhibit a factorial structure different from children without disabilities (Strauss, Sherman, & Spreen, 2006). Therefore, empirical investigations of the factorial structure of the WISC-IV among groups of children with disabilities are necessary (Chen & Zhu, 2012).

Multiple studies have been conducted with mixed clinical groups (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Chen & Zhu, 2012; Devena, Gay, & Watkins, 2013; Nakano & Watkins, 2013; San Miguel Montes, Allen, Puente, & Neblina, 2010; Watkins, 2010; Watkins, Canivez, James, James, & Good, 2013; Watkins, Wilson, Kotz, Carbone, & Babula, 2006; Weiss et al., 2013) as well as children with learning disabilities (Giofrè & Cornoldi, 2015; Styck & Watkins, 2016), low IQ (Gomez, Vance, & Watson, 2017), and high IQ (Molinero, Mata, Calero, García-Martín, & Araque-Cuenca, 2015; Rowe, Dandridge, Pawlush, Thompson, & Ferrier, 2014). As with the standardization sample, Wechsler four-factor and CHC five-factor models in both higher-order and bifactor structures tended to exhibit roughly equivalent fit across the clinical samples. In consequence, some researchers argued for interpretation of the CHC model (Weiss et al., 2013) whereas others argued for interpretation of the Wechsler model (Bodin et al., 2009). Likewise, some researchers argued for interpretation of the higher-order model (Chen & Zhu, 2012) whereas others argued for interpretation of the bifactor model (Canivez, 2014).

Regardless of the structure and number of factors, *g* was the predominant source of common variance among all the clinical samples. Typically, these studies only included analyses of the 10 core WISC-IV subtests. With only 10 subtests to measure five factors, statistical identification becomes tenuous because the factor loadings may not form independent clusters (McDonald & Ho, 2002). That is, there may be too few pure indicators of some factors. CHC five-factor models also abandoned the parsimony of simple structure (Thurstone, 1947) by allowing cross-loadings of subtests. Simple structure honors "the purpose of science [which] is to uncover the relatively simple deep structure principles or causes that underlie the apparent complexity observed at the surface structure level" (Le, Schmidt, Harter, & Lauver, 2010, p. 112).

ADHD samples

Although a few children with ADHD were included in some of the mixed clinical samples, there has been little research focused exclusively on these children. One study supported a four-factor oblique structure for the Chinese version of the WISC-IV among 334 Taiwanese children and adolescents diagnosed with ADHD but did not test higher-order or bifactor models, nor a five-factor CHC model (Yang et al., 2013). However, the first-order factor intercorrelations ranged from .36 to .79 (*Md* = .64), indicating the presence of a general factor (Gorsuch, 1983).

Another study applied CFA to the WISC-IV scores of 233 students diagnosed with ADHD by school multidisciplinary teams in the United States (Styck & Watkins, 2017) and found that the Wechsler four-factor model in both higher-order and bifactor structures were good fits to the data. However, the higher-order four-factor model was slightly superior in fit. Unfortunately, CHC five-factor models were not tested and the applicability of school-based diagnoses limits comparability with other clinical samples.

A third study included 314 children with ADHD diagnosed by a single clinical neuropsychologist (Thaler et al., 2015) and found that higher-order versions of the Wechsler four-factor model and the CHC five-factor model were both good fits to the data. However, one of the factors in the CHC model loaded at .97 on the general factor. This indicates that those factors were empirically redundant (Le et al., 2010), which constitutes a major threat to discriminant validity (Brown, 2015). Unfortunately, a bifactor structure was not tested and this study relied on participants from one clinical practice in one U.S. city, so its generalizability may be limited.

The final study that focused on the factor structure of the WISC-IV included 812 children with ADHD recruited from an Australian outpatient psychiatric clinic (Gomez, Vance, & Watson, 2016). Four-factor and five-factor models in both higher-order and bifactor structures exhibited good fit to the data, but the bifactor structure with four general factors was slightly superior in fit. However, all participants were from a single clinic in Australia, so the results may not generalize to other countries.

As with the mixed clinical samples, Wechsler four-factor and CHC five-factor models in both higher-order and bifactor structures tended to exhibit relatively equivalent model-to-data fits across the ADHD samples; but researchers disagreed on the preferred WISC-IV structure. These studies were consistent in finding that the general intelligence factor accounted for about twice the common variance as all the group factors combined. For the CHC five-factor models, the independent clusters assumption was violated and the parsimony of simple structure was abandoned because there were too few subtests to uniquely identify all factors. Each study was marked by distinct limitations but all shared a common weakness: Participants included a heterogeneous mixture of children diagnosed with inattentive (ADHD-I) and combined (ADHD-C) subtypes of ADHD.

To date, investigations of the factor structure of the WISC-IV among children with ADHD have been limited to versions of the WISC-IV normed in the United States, China, and Australia. Given the sparcity of this research, further investigations are needed. For instance, the global Spanish-speaking community is composed of 572 million people (Instituto Cervantes, 2017) and there have been no published studies of the factor structure of the Spanish version of the WISC-IV among children with ADHD. It is also important to examine the possible influence of ADHD subtypes because no prior study has considered that the factor structure of the WISC-IV might differ across subtypes (Gomez et al., 2016; Thaler et al., 2015; Wicherts, 2016). This lacuna is puzzling because developmental, treatment, and genetic differences among ADHD subtypes have been documented (Greven, Asherson, Rijsdijk, & Plomin, 2011; Greven, Harlaar, Dale, & Plomin, 2011; Hodgson, Hutchinson, & Denson, 2014). Accordingly, the current study had two primary aims: (a) identify the factorial structure of the core Spanish WISC-IV (WISC-IV^{spa}; Wechsler, 2005c) among children with ADHD, and (b) determine if that structure is invariant across ADHD subtypes.

Method

Participants

Participants were 859 children (79.3% male) newly diagnosed with ADHD, aged 6 to 16 years (M = 9.6; SD = 2.81). All participants were of Spanish origin but information on their race or ethnicity and socioeconomic status was not available. Of these, 494 children (74.5% male) met criteria for ADHD predominantly inattentive (ADHD-I) and 365 children (88.8% male) met criteria for ADHD-combined (ADHD-C). Complete WISC-IV data were available for all 859 cases. Given doubts about the validity of the ADHD predominantly hyperactive–impulsive subtype (Willcutt et al., 2012), cases with this diagnosis were not included in this study.

The ADHD subtype groups were statistically different in age (t = 4.77, df = 788, p < .001). Specifically, there were more very young children in the ADHD-C group and more teenagers in the ADHD-I group. The ADHD subtype groups also differed significantly in gender (χ^2 = 16.2, df = 1, p < .001) with the ADHD-C group having a larger proportion of males. Comorbidity with secondary diagnoses were not statistically significant (χ^2 = 10.39, df = 9, p = .32).

Instrument

The WISC-IV was normed in Spain in 2005 (WISC-IV^{spa}; Wechsler, 2005c). As with versions for other countries, its core subtest battery includes 10 subtests (M = 10, SD = 3) that contribute to the computation of a FSIQ score as well as four-factor index scores (M = 100, SD = 15). Specifically, the Similarities (SI), Vocabulary (VO), and Comprehension (CO) subtests load on the Verbal Comprehension factor (VC); the Block Design (BD), Picture Concepts (PC), and Matrix Reasoning (MR) subtests account for the Perceptual Reasoning factor (PR); the Digit Span (DS) and Letter–Number Sequencing (LN) subtests make up the Working Memory factor (WM); and the Coding (CD) and Symbol Search (SS) subtests compose the Processing Speed factor (PS). The modifications made from the English version were marginal. For example, monetary references were replaced by object references, names of localities, replacement of polysemic words, and so forth.

The standardization sample included 1,590 children aged 6–16 years stratified by age, gender, and region of the country. The sample was representative of the Spanish population. Wechsler (2005d) provided information on the reliability and validity of WISC-IV^{spa} scores. For example, the average split-half reliability of the Full Scale Intelligence Quotient (FSIQ) for the standardization sample was .95 whereas the average reliability of factor index scores ranged from .86 (PSI) to .92 (VCI). The reliability of subtest scores ranged from .72 (SS) to .92 (Cancellation). Concurrent validity was supported by a comparison of WISC-IV^{spa} scores to other cognitive and achievement tests. Factorial validity evidence was presented via a series of CFA with the final structural model consisting of four correlated first-order factors (i.e., VCI, PRI, WMI, and PSI). Neither higher-order nor bifactor models were examined.

Procedure

Participants were selected from cases referred for neuropsychological evaluation to a university-affiliated mental health unit in Valencia, Spain, that specialized in the assessment and treatment of ADHD. The diagnostic algorithm was based on parent and teacher rating scales (Conners and ADHD-RS-IV scales), sustained attention tasks (CPT), and a clinical interview (following ADHD DSM criteria). The WISC-IV^{spa} was routinely included in those clinical assessments. No supplementary subtests were administered. ADHD diagnoses were established by master's-level trained psychologists, based on the information from parents and teachers (through observation scales and interviews) and a neuropsychological test battery. The 10 core subtests of the WISC-IV^{spa} were administered according to standardized methods by the same master's-level trained psychologists. Exclusion criteria were (a) FSIQ below 70, (b) history of severe mental health problems, (c) comorbid intellectual disability, and (d) pharmacological treatment at the time of assessment.

Analyses

All CFA were conducted with Mplus 8.0 for Macintosh (Muthén & Muthén, 2017) from covariance matrices using the maximum likelihood (ML) estimator. Latent variables were scaled by setting a reference indicator or by setting the variance of latent variables (Brown, 2015). Parameter estimates were constrained to equality in bifactor models when there with only two indicators per factor.

Consistent with previous WISC-IV factorial analyses, only higher-order and bifactor models were examined because theory (Carroll, 1993) and the WISC-IV score structure (i.e., implied by a FSIQ score) signify a hierarchical structure. Four-factor models followed the scoring structure of the WISC-IV whereas in five-factor CHC models, the structure was hypothesized to consist of (a) crystallized intelligence

(Gc) indicated by the VO, SI, and CO subtests, (b) fluid reasoning (Gf) loaded by the MR and PC subtests, (c) visual processing (Gv) marked by the BD and SS subtests, (d) short-term memory (Gsm) identified by the DS and LN subtests, and (e) processing speed (Gs) specified by the CD and SS subtests. Consistent with prior research (Gomez et al., 2016; Thaler et al., 2015), the SS subtest was cross-loaded on both Gv and Gs factors.

As recommended by Kline (2016), global model fit was evaluated with multiple fit indices: model chi-square (χ^2), Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), and Akaike's Information Criterion (AIC; Akaike, 1987). Based on the combinational rules of Hu and Bentler (1999), good fit required CFI ≥ .95 *and* RMSEA ≤ .06. For AIC, lower values identify models more likely to generalize (Akaike, 1987). Meaningful differences between well-fitting models was also evaluated using ΔCFI > .01, ΔRMSEA > .015 (Chen, 2007; Cheung & Rensvold, 2002), and ΔAIC ≥ 10 (Anderson, 2008). Given that global fit indices are averages that can mask areas of local misfit and potentially invalidate a model (McDonald & Ho, 2002), parameter estimates were also examined to ensure that they made statistical and substantive sense (Brown, 2015).

After establishing a baseline model with these methods, measurement invariance was tested between ADHD subtypes to explore the possible differences in the WISC-IV^{spa} structure across subtypes. A hierarchical series of models that imposed increasingly more stringent constraints were tested (Brown, 2015). Configural invariance is the least stringent model: It only imposes the constraint that variables load on the same factors in each group. Metric invariance adds the constraint that factor loadings are equal across groups. Scalar invariance adds constraints on the unique factor variances. Finally, strict invariance adds the constraint that the observed residuals are equal across groups.

Although Wechsler (2005c) provided internal consistency estimates of reliability, it is now widely recognized that split-half and alpha coefficients are based on statistical assumptions that are unrealistic and likely to be violated in practice (Cho & Kim, 2015), resulting in biased estimates of reliability (Green & Hershberger, 2000). Model-based coefficient omega provides an alternative reliability estimate that makes fewer and more realistic assumptions (Reise, 2012). It replaces the classical test theory hypothesis of true and error variance with the factor analytic conceptualization of common and unique variance. Coefficient omega (ω) is an estimate of the proportion of variance in the unit-weighted FSIQ score attributable to both general and group factors, whereas coefficient omega hierarchical (ω_h) is an estimate of the proportion of variance in the unit-weighted FSIQ score attributable to the general factor alone. Omega subscale (ω_s) is an estimate of the proportion of variance in a unit-weighted group factor score (i.e., VCI, PRI, etc.) explained by both the general factor and the targeted group factor. Omega hierarchical subscale (ω_{hs}) is an estimate of the proportion of variance in the unit-weighted group factor score (i.e., VCI, PRI, etc.) explained solely by that group factor independent of the general factor. Thus, ω and ω_s reflect the variance due to multiple factors, whereas ω_h and ω_{hs} account for the variance contributed by a single factor. A comparison of ω, ω_h, ω_s, and ω_{hs} coefficients can assist in the interpretation of general and group factor scores (Reise, 2012; Reise, Bonifay, & Haviland, 2013; Rodriguez, Reise, & Haviland, 2016; Watkins, 2017).

There is no generally accepted guideline for the sufficiency of ω and ω_s coefficients for making clinical decisions about individuals. However, they should probably be judged like coefficient alpha values because both index the variance of multiple common factors. There is no universally accepted guideline for coefficient alpha values sufficient for high-stakes decisions about individuals, but values ≥ .90 are commonly recommended and "scores with values below .90 should not be interpreted" (Kranzler & Floyd, 2013, p. 71). Likewise, there are no guidelines for clinical interpretation of ω_h or ω_{hs} coefficients, but it has been recommended that they should exceed .50, although .75 would be preferred (Reise, 2012).

Results

WISC-IV^{spa} subtest and index scores for ADHD-C, ADHD-I, and total groups are presented in Table 1. The mean scores for WMI, PSI, and FSIQ were lower than the U.S. and Spanish averages ($d \approx -.57$, $-.56$, and $-.33$, respectively). These results are consistent with previous studies with clinical samples (Fenollar-Cortés et al., 2015; Gomez et al., 2016; Mayes & Calhoun, 2006). Nevertheless, subtest scores were univariate normal and exhibited nonsignificant multivariate kurtosis ($\chi^2[1] = 0.43$, $p = .051$).

Confirmatory factor analyses (Cfas)

Global model fit statistics are presented in Table 2. The CHC higher-order model was not a good fit to the data. Additionally, its Gf and *g* factors were empirically redundant with a loading of .98 and the Gf residual variance term was nonsignificant, making five group factors implausible (Gignac & Kretzschmar, 2017). The

Table 1. WISC-IV scores for ADHD-C, ADHD-I, and total groups.

	ADHD-C		ADHD-I			Total			
Score	Mean	*SD*	Mean	*SD*	*d*	Mean	*SD*	Skew	Kurtosis
BD	9.79	2.72	9.51	2.85	0.10	9.63	2.79	0.10	−0.18
Si	10.43	2.77	10.41	2.68	0.01	10.42	2.72	0.36	0.18
DS	8.82	2.58	8.55	2.60	0.11	8.66	2.59	0.30	0.15
PC	11.43	2.75	10.91	2.58	0.20*	11.13	2.67	−0.08	−0.25
CD	8.07	3.11	7.68	2.67	0.13	7.85	2.87	0.18	−0.06
VO	10.21	2.62	10.02	2.49	0.08	10.10	2.55	0.27	0.05
LN	8.90	2.51	8.86	2.57	0.16	8.88	2.54	−0.06	0.06
MR	10.06	2.76	9.78	2.77	0.10	9.90	2.77	0.21	−0.25
CO	9.81	2.79	9.87	2.87	−0.02	9.84	2.83	0.21	−0.13
SS	9.06	2.74	8.67	2.55	0.15*	8.84	2.64	0.28	0.17
VCI	101.76	12.85	101.21	12.78	0.04	101.45	12.81	0.28	0.33
PRI	101.85	13.17	99.74	13.18	0.16*	100.64	13.21	0.24	−0.16
WMI	92.37	12.82	91.73	13.24	0.05	92.00	13.06	0.01	0.33
PSI	93.31	13.67	91.27	12.43	0.16*	92.14	13.00	0.00	−0.24
FSIQ	96.47	12.74	94.66	11.71	0.15*	95.43	12.18	0.30	−0.19

Note. 365 children in ADHD–Combined sample, 494 children in ADHD–Inattentive sample, and 859 in total sample. BD is Block Design, SI is Similarities, DS is Digit Span, PC is Picture Concepts, CD is Coding, VO is Vocabulary, LN is Letter-Number Sequencing, MR is Matrix Reasoning, CO is Comprehension, SS is Symbol Search, VCI is Verbal Comprehension Index, PRI is Perceptual Reasoning Index, WMI is Working Memory Index, PSI is Processing Speed Index, FSIQ is Full Scale IQ, and *d* is Cohen's standardized effect size.

*Statistically significant at $p < .05$. However, no single comparison was statistically significant at $p < .0035$ (to control experimentwise error rate at .05 [i.e., .0035 x 15 = .0525]).

Table 2. Fit statistics for the Spanish WISC-V 10 core subtests for the total ADHD sample (N = 859).

Model	χ^2	df	CFI	RMSEA	90% CI RMSEA	AIC	ΔCFI	ΔRMSEA	ΔAIC
Wechsler Models									
Higher-Order	65.55	31	.983	.036	.024-.048	39,453	.006	.005	9
Bifactor	49.35	27	**.989**	**.031**	.017-.045	**39,444**	0	0	0
CHC Models									
Higher-Order	199.46	33	.918	.077	.067-.087	39,583	.071	.046	139
Bifactor[a]	60.62	27	.983	.038	.025-.051	39,456	.006	.007	12

Note. CFI is comparative fit index, RMSEA is root mean square error of approximation, CI is confidence interval, and AIC is Akaike information criterion. Wechsler models contain four first-order factors whereas CHC models contain five first-order factors. Loadings on factors with only two indicators were constrained to equality to ensure identification. Indictors with best fit in bold.

[a]Inadmissible results due to nonpositive definite matrix. Negative variance estimate for Matrix Reasoning subtest and for Picture Concepts subtest if Matrix Reasoning loading subsequently constrained to zero.

CHC bifactor model produced a statistically improper solution. That is, it had negative variance estimates for subtests loading on the Gf factor. When that model was modified by removing the SS subtest from the Gs factor and fixing the BD loading on the Gv factor by specifying its reliability and variance as per Hayduk and Littvay (2012), the model converged, but the MR and PC subtests did not load on the Gf factor. Thus, the Gf factor was superfluous, as in the higher-order model. Additionally, the five-factor models are unlikely to generalize, given their ΔAIC values (Anderson, 2008; Preacher, Zhang, Kim, & Mels, 2013). Thus, they did not appear to be optimal solutions for the WISC-IV[spa] among this sample of children with ADHD.

Both Wechsler four-factor models exhibited good global fit to the data. The bifactor Wechsler model was a statistically better fit than its higher-order variant ($\Delta\chi^2[4] = 16.2$, $p = .003$) but their fit was not meaningfully different (ΔCFI = .006, ΔRMSEA = .005). However, the higher-order model is somewhat less likely than the bifactor model to generalize (ΔAIC = 9). Both higher-order and bifactor models are illustrated in Figure 1. Carroll (1993) recommended that higher-order models be transformed via the Schmid and Leiman (1957) technique to allow direct interpretation of subtests akin to a bifactor model. When that transformation was effected, the results were not substantially different from these bifactor model results. This approximate equivalence can be verified by tracing the paths in the higher-order model in Figure 1 (e.g., *g* to PR of .89 x PR to MR of .70 = .62 whereas the path from *g* to PR in the bifactor model was .60).

Bifactor fit might be discounted because it has been found to take advantage of small unmodeled cross-loadings (Murray & Johnson, 2013). However, it is currently impossible to distinguish between higher-order and bifactor structures when the "true" generating model is unknown (Mansolf & Reise, 2017). To its credit, the bifactor model provides interpretational clarity because of the direct relationships between subtests and factors and thereby offers conceptual parsimony: Group and general factors can be interpreted in terms of their constituent subtests, whereas the higher-order model involves interpretation of an abstraction of an abstraction (Gignac, 2008). Given its slightly better global fit and greater conceptual parsimony,

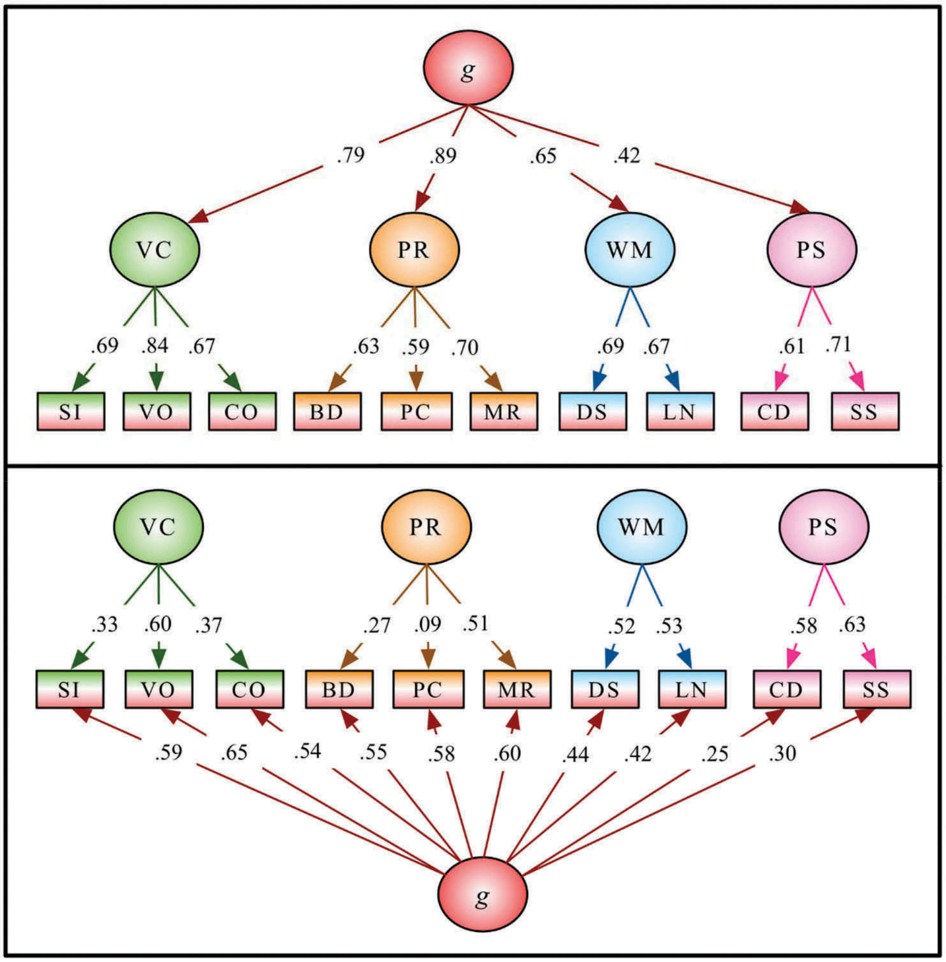

Figure 1. Higher-order (top) and bifactor (bottom) models for the WISC-IV for 859 Spanish children with ADHD.

Note. g = general intelligence; VC = Verbal Comprehension Factor; PR = Perceptual Reasoning Factor; WM = Working Memory Factor; PS = Processing Speed Factor; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PC = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter-Number Sequencing; CD = Coding; and SS = Symbol Search.

the bifactor model was retained as the best fit to this data even though it was marked by low loadings on the PR factor.

Table 3 displays the sources of variance in the Wechsler bifactor model. The general factor accounted for more common and total variance than the group factors

Table 3. Sources of variance in the Wechsler Intelligence Scale for Children–Fourth Edition ADHD sample (N = 859) in Spain.

	General		VC		PR		WM		PS				
Subtest	b	b^2	b	b^2	b	b^2	b	b^2	b	b^2	h^2	u^2	ECV
Similarities	.59	.346	.33	.110							.455	.545	.759
Vocabulary	.65	.421	.60	.362							.784	.216	.538
Comprehension	.54	.288	.37	.135							.423	.577	.682
Block Design	.55	.298			.27	.075					.373	.627	.800
Picture Concepts	.58	.340			.09	.008					.348	.652	.978
Matrix Reasoning	.60	.365			.51	.264					.629	.371	.580
Digit Span	.44	.194					.52	.274			.468	.532	.416
Letter-Number Sequencing	.42	.180					.53	.285			.465	.535	.387
Coding	.25	.063							.58	.332	.395	.605	.160
Symbol Search	.30	.087							.63	.392	.479	.521	.182
Total Variance	.258		.061			.035		.056		.072	.482	.518	
Common Variance	.536		.126			.072		.116		.150			
ω/ω_s	.849		.783			.695		.636		.607			
ω_h/ω_{hs}	.703		.274			.141		.381		.504			

Note. b = standardized loading of subtest on factor, b^2 = variance explained in the subtest, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, VC = Verbal Comprehension, PR = Perceptual Reasoning, WM = Working Memory, PS = Processing Speed, ECV is explained common variance, ω = omega total, ω_s = omega subscale ω_h = omega hierarchical, ω_{hs} = omega hierarchical subscale.

combined (53.6% vs. 46.4% and 25.8% vs. 22.4%, respectively). Altogether, 48.2% of the WISC-IV variance was explained by the general and group factors whereas 51.8% was due to error and specific factors. The CD and SS subtests were particularly poor measures of *g* but relatively strong measures of the PS factor. In contrast, the BD and PC subtests were relatively strong measures of *g* and relatively weak measures of the PR factor. The explained common variance of a subtest (ECV; Rodriguez et al., 2016) expresses the proportion of variance in its score due to the general factor. Thus, 80% of the BD variance and 97.8% of the PC variance were due to the general factor.

Invariance

Based on the Wechsler bifactor model, measurement invariance testing was conducted across ADHD subtypes. None of the models were statistically or practically significantly different (see Table 4), suggesting that WISC-IV^{spa} scores can be directly compared across ADHD-C and ADHD-I groups. Of note, similar nonsignificant invariance results were obtained when oblique four-factor models were compared. Also of note, the WISC-IV^{spa} scores obtained for ADHD-C and ADHD-I groups (Table 1) were almost identical given that .20 was the largest standardized score difference.

Reliability

None of the WISC-IV^{spa} composite scores met the .90 standard for omega coefficients but the FSIQ score was close at .85. Only the FSIQ and PSI scores met the omega hierarchical minimum standard of .50. Given these low ω_{hs} coefficients, "to interpret subscale scores as representing the precise measurement of some latent variable that is unique or different from the general factor, clearly, is misguided" (Rodriguez et al., 2016, p. 225).

Omega hierarchical coefficients also index one facet of factorial validity. Specifically, the square root of omega hierarchical is the correlation between the factor and the factor score. Thus, the FSIQ and PSI scores seems to be a fair representation of their factors with correlations of .84 and .71, respectively. However, the group factor index scores of VCI, PRI, and WMI with factor correlations of .52, .38, and .62, respectively, were poor representations of their underlying factors.

Table 4. Invariance tests for WISC-IV baseline bifactor model for ADHD-C (N = 365) and ADHD-I (N = 494) groups.

Model	χ^2	df	CFI	ΔCFI	RMSEA	ΔRMSEA	$\Delta\chi^2$	df	p
Baseline									
ADHD-C	42.92	27	.983	–	.040	–	–	–	–
ADHD-I	34.30	27	.993	–	.023	–	–	–	–
Configural	77.22	54	.988	–	.032	–	–	–	–
Metric	94.38	72	.989	.001	.027	.005	17.16	18	.512
Scalar	111.64	82	.985	.004	.029	.002	9.77	5	.082
Strict	121.54	92	.984	.001	.030	.001	10.00	10	.441

Note: Invariance tested with immediately prior model (e.g., metric versus configural p = .512). ΔCFI for all comparisons < .01 and ΔRMSEA < .01 (Cheung & Rensvold, 2002).

Discussion

The factorial validity of the WISC-IV^{spa} in 859 Spanish children diagnosed with ADHD was examined. A bifactor model with the four factors first identified by Wechsler (2003a) was the best fit to the data. In this model, the general factor accounted for more common and total variance than the group factors combined (53.6% vs. 46.4% and 25.8% vs. 22.4%, respectively). The CD and SS subtests were particularly poor measures of *g* but relatively strong measures of the PS factor. In contrast, the BD and PC subtests were relatively strong measures of *g* but weak measures of the PR factor. Additionally, the Wechsler bifactor model was invariant across ADHD-C and ADHD-I groups, permitting a direct comparison of WISC-IV^{spa} across children diagnosed with these subtypes of ADHD. Only the FSIQ exhibited robust estimates of reliability (ω = .85 and ω_h = .70). In contrast, the group factor scores were unreliable measures of their proposed underlying factors (ω_{hs} = .27 for VCI, .14 for PRI, .38 for WMI, and .50 for PSI scores).

Consistent with our results, previous studies have also supported the Wechsler bifactor model in ADHD samples in Australia and the United States (Gomez et al., 2016; Styck & Watkins, 2017). These results are also in accord with previous research investigating the factorial validity of the WISC-IV with other clinical samples in several countries (Canivez, 2014; Devena et al., 2013; Gomez et al., 2017; Molinero et al., 2015; Styck & Watkins, 2016; Watkins, 2010; Watkins et al., 2013) and its U.S. normative sample (Canivez & Kush, 2013; Watkins, 2006) as well as with the U.S. normative samples of other cognitive batteries (Cucina & Howardson, 2017). Furthermore, a Wechsler bifactor structure has also been preferred for the recently published WISC-V (Canivez, Watkins, & Dombrowski, 2016, 2017; Wechsler, 2014).

Some researchers have suggested that interpretation of WISC index scores (score profiles, differences between primary or secondary index scores, etc.) could provide clinically meaningful information for assessment and diagnosis of ADHD (Fenollar-Cortés et al., 2015; Loh, Piek, & Barrett, 2011; Mayes & Calhoun, 2006; Parke, Thaler, Etcoff, & Allen, 2015;

Smitha et al., 2014; Thaler et al., 2013; Theiling & Petermann, 2014; Walg et al., 2017). This approach assumes that factor index scores are sufficiently reliable for such comparisons and that each WISC factor index score can be interpreted as a unidimensional measure of its purported factor (Glass, Ryan, Charter, & Bartels, 2009). For example, the Processing Speed latent construct is consistently and solely responsible for PSI scores. To the contrary, the WISC-IV is multidimensional with multiple latent variables contributing to its factor index scores. For instance, the PSI score is composed of variance contributed by the general factor, the PS factor, random error, and systematic subtest variance not shared with other subtests. The current study demonstrated that the majority of the PSI score variance was due to random error and nonshared systematic variance (see the uniqueness values in Table 3 for Coding and Symbol Search subtests), making it an unreliable measure of the processing speed construct.

Further, it is a mistake to interpret multidimensional scales as if they are unidimensional (Reise, 2012). To avoid this mistake, omega coefficients can be contrasted with omega hierarchical coefficients to gauge the dimensionality and interpretability of each composite score (Gignac & Kretzschmar, 2017; Reise, 2012; Rodriguez et al., 2016). For instance, Reise (2012) concluded that with low ω_{hs} coefficients, "the interpretation of the subscales as precise indicators of unique constructs is extremely limited—very little reliable variance exists beyond that due to the general factor" (p. 691). Rodriguez et al. (2016) noted that low ω_{hs} coefficients indicate that "unit-weighted subscale scores are highly troublesome because their variance mostly reflects a general trait and not the specific trait intended to be assessed" (p. 233). Similarly, DeMars (2013) judged that "when loadings on the [group] factors are low, only the general factor score carries a reliable interpretation" (p. 375). The low ω_{hs} coefficients found in the current study make it highly unlikely that WISC-IV index score profiles can validly contribute to ADHD assessments. Notably, Gomez et al. (2016) arrived at the same conclusion after studying the factor structure of the WISC-IV in their sample of children with ADHD.

As with all research, this study has limitations. First, the categorical diagnostic approach for ADHD has been questioned due to doubts about the reliability and validity of ADHD subtypes (Willcutt et al., 2012). A dimensional diagnostic approach might result in different ADHD groups. Second, the study did not include the five WISC-IV supplemental subtests, making it difficult to adequately evaluate CHC models. Nevertheless, the supplemental subtests of the WISC-IV are rarely used in clinical practice, so our results may be useful for typical clinical applications. Third, some sociodemographic variables such as parent education level, region of country, family income, race, and so forth, were not considered. These variables could be included in future studies. Fourth, there was no objective method to verify the accuracy of our ADHD diagnoses. However, concern over this limitation is reduced because our results were generally similar to other studies with ADHD samples (e.g., Gomez et al., 2016). Fifth, there was no objective way to verify that our assessors administered and scored the WISC-IV accurately. First noticed by Terman (1918), subsequent research in applied settings has shown nontrivial amounts of score variation due to differences among examining psychologists rather than among children (McDermott, Watkins, & Rhoad, 2014). This source of error should be mitigated in future studies. Sixth, further research is needed to explore the measurement invariance of WISC scales between ADHD and non-ADHD samples and between male and female samples. Finally, we also consider it important to replicate the current study in other Spanish speaking countries given that invariance was not obtained when the French WISC-IV was administered to French and Swiss samples (Reverte, Golay, Favez, Rossier, & Lecerf, 2015).

Conclusion

Our results are consistent with previous studies that applied the WISC-IV with clinical samples in highlighting the predominant influence of the general intelligence factor on WISC-IV scores (Bodin et al., 2009; Canivez, 2014; Devena et al., 2013; Gomez et al., 2016, 2017; Nakano & Watkins, 2013; Styck & Watkins, 2016, 2017; Thaler et al., 2015; Watkins, 2010; Watkins et al., 2013, 2006). It appears that the WISC-IV is structured hierarchically, with a broad general intelligence factor impacting all subtests, and four narrower group factors independently impacting separate groups of subtests for children diagnosed with ADHD-C and ADHD-I. However, those four group factors account for small amounts of the common and total variance and are, therefore, relatively unreliable.

The desire to interpret WISC-IV group factor scores as reliable and univariate measures of their underlying factors for identification of cognitive strengths and weaknesses in clinical practice is understandable but not supported psychometrically. Profiles of WISC-IV group factor scores do not provide distinct and reliable information beyond the FSIQ score (DeMars, 2013). Similar issues have been reported for subscales in educational assessment (Sinharay, Puhan, & Haberman, 2011) and seems to be an intractable problem when subscores are highly intercorrelated (Bulut, Davison, & Rodriguez, 2017; Feinberg & Jurich, 2017). Accordingly, clinicians applying the WISC-IV to children with ADHD should adhere to professional standards and justify the application of score profiles with empirical evidence on their psychometric properties (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014; International Test Commission, 2014).

Acknowledgments

Our appreciation to Almudena Hernández-Ruiz, Bonastre Gabasa-Gracia, Susana Lázaro-Mañes, and Ana Lozano-Monterde; all were members of Apunt Psicòlegs Mental Health Unit. We offer our special thanks to Juan Carlos Moyano-Gálvez.

About the authors

Javier Fenollar-Cortés, PhD, is a professor and clinical psychologist from the Department of Psychology at Universidad Loyola Andalucía (Spain). His teaching and research focus on ADHD and neurocognitive performance and cognitive profiles, in relation to improving ADHD conceptualization and diagnosis through clinical scales and neuropsychological tasks.

Carlos López-Pinar is a clinical psychologist and PhD student from the Department of Psychobiology at the University of Valencia (Spain). His research interests involve individual differences in child psychopathology and professional issues in clinical psychology.

Marley W. Watkins, PhD, is professor and chairman of the Department of Educational Psychology at Baylor University. His research interests include professional issues, the psychometrics of assessment and diagnosis, individual differences, and computer applications.

ORCID

Marley W. Watkins http://orcid.org/0000-0001-6352-7174

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332. doi:10.1007/BF02294359

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

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. New York, NY: Springer.

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

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

Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research*, *52*, 86–104. doi:10.1080/00273171.2016.1253452

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

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

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

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

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

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

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

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, *18*, 207–230. doi:10.1177/1094428114555994

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

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, *13*, 354–378. doi:10.1080/15305058.2013.799067

Devena, S. E., Gay, C. E., & Watkins, M. W. (2013). Confirmatory factor analysis of the WISC-IV in a hospital referral sample. *Journal of Psychoeducational Assessment*, *31*, 591–599. doi:10.1177/0734282913483981

Devena, S. E., & Watkins, M. W. (2012). Diagnostic utility of WISC-IV General Abilities Index and Cognitive Proficiency Index difference scores among children with ADHD. *Journal of Applied School Psychology*, *28*, 133–154. doi:10.1080/15377903.2012.669743

Feinberg, R. A., & Jurich, D. P. (2017). Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*, *36*, 5–13. doi:10.1111/emip.12142

Fenollar-Cortés, J., Navarro-Soria, I., González-Gómez, C., & García-Sevilla, J. (2015). Cognitive profile for children with ADHD by using WISC-IV: Subtype differences? *Journal of Psychodidactics*, *20*, 157–176. doi:10.1387/RevPsicodidact.12531

Flanagan, D. P., & Kaufman, A. S. (2009). *Essentials of WISC-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.

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., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factors models: Limitations and suggestions. *Intelligence*, *62*, 138–147. doi:10.1016/j.intell.2017.04.001

Giofrè, D., & Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence*, *52*, 36–43. doi:10.1016/j.intell.2015.07.002

Glass, L. A., Ryan, J. J., Charter, R. A., & Bartels, J. M. (2009). Discrepancy score reliabilities in the WISC-IV standardization sample. *Journal of Psychoeducational Assessment*, *27*, 138–144. doi:10.1177/0734282908325158

Gomez, R., Vance, A., & Watson, S. (2017). Bifactor model of WISC-IV: Applicability and measurement invariance in low and normal IQ groups. *Psychological Assessment*, *29*, 902–912. doi:10.1037/pas0000369

Gomez, R., Vance, A., & Watson, S. D. (2016). Structure of the Wechsler Intelligence Scale for Children–Fourth Edition in a group of children with ADHD. *Frontiers in Psychology*, *7*(737), 1–11. doi:10.3389/fpsyg.2016.00737

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

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7*, 251–270. doi:10.1207/S15328007SEM0702_6

Greven, C. U., Asherson, P., Rijsdijk, F. V., & Plomin, R. (2011). A longitudinal twin study on the association between inattentive and hyperactive-impulsive ADHD symptoms. *Journal of Abnormal Child Psychology*, *39*, 623–632. doi:10.1007/s10802-011-9513-7

Greven, C. U., Harlaar, N., Dale, P. S., & Plomin, R. (2011). Genetic overlap between ADHD symptoms and reading is largely driven by inattentiveness rather than hyperactivity-impulsivity. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*, 6–14.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*(159), 1–17. doi:10.1186/1471-2288-12-159

Hodgson, K., Hutchinson, A. D., & Denson, L. (2014). Nonpharmacological treatments for ADHD: A meta-analytic review. *Journal of Attention Disorders*, *18*, 275–282. doi:10.1177/1087054712444732

Hu, L., & Bentler, P. (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

Instituto Cervantes. (2017). *El español en el mundo*. Retrieved from https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2017.pdf

International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. *International Journal of Testing*, *14*, 195–217. doi:10.1080/15305058.2014.918040

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

Keith, T. Z, Fine, J. G, Taub, G. E, Reynolds, M. R, & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the wechsler intelligence scale for children—fourth edition: What does it measure. *School Psychology Review*, *35*(1), 108-127.

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

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. New York, NY: Guilford.

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

Loh, P. R., Piek, J. P., & Barrett, N. C. (2011). Comorbid ADHD and DCD: Examining cognitive functions using the WISC-IV. *Research in Developmental Disabilities*, *32*, 1260–1269. doi:10.1016/j.ridd.2011.02.008

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

Mayes, S. D., & Calhoun, S. L. (2004). Similarities and differences in Wechsler Intelligence Scale for Children–Third Edition (WISC-III) profiles: Support for subtest analysis in clinical referrals. *The Clinical Neuropsychologist*, *18*, 559–572. doi:10.1080/13854040490888530

Mayes, S. D., & Calhoun, S. L. (2006). WISC-IV and WISC-III profiles in children with ADHD. *Journal of Attention Disorders*, *9*, 486–493. doi:10.1177/1087054705283616

McDermott, P. A., Watkins, M. W., & Rhoad, A. M. (2014). Whose IQ is it?–Assessor bias variance in high-stakes psychological assessment. *Psychological Assessment*, *26*, 207–214. doi:10.1037/a0034832

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64–82. doi:10.1037//1082-989X.7.1.64

Molinero, C., Mata, S., Calero, D., García-Martín, B., & Araque-Cuenca, A. (2015). Usefulness of WISC-IV in determining intellectual giftedness. *Spanish Journal of Psychology*, *18*, 1–10. doi:10.1017/sjp.2015.63

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

Muthén, B. O., & Muthén, L. K. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

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

Parke, E. M., Thaler, N. S., Etcoff, L. M., & Allen, D. N. (2015). Intellectual profiles in children With ADHD and comorbid learning and motor disorders. *Journal of Attention Disorders*. doi:10.1177/1087054715576343

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*, 28–56. doi:10.1080/00273171.2012.710386

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

Reverte, I., Golay, P., Favez, N., Rossier, J., & Lecerf, T. (2015). Testing for multigroup invariance of the WISC-IV structure across France and Switzerland: Standard and CHC models. *Learning and Individual Differences*, *40*, 127–133. doi:10.1016/j.lindif.2015.03.015

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

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

Rowe, E. W., Dandridge, J., Pawlush, A., Thompson, D. F., & Ferrier, D. E. (2014). Exploratory and confirmatory factor analyses of the WISC-IV with gifted students. *School Psychology Quarterly*, *29*, 536–552. doi:10.1037/spq0000009

San Miguel Montes, L. E., Allen, D. N., Puente, A. E., & Neblina, C. (2010). Validity of the WISC–IV Spanish for a clinically referred sample of Hispanic children. *Psychological Assessment*, *22*, 465. doi:10.1037/a0018895

Scheirs, J. G. M., & Timmers, E. A. (2009). Differentiating among children with PDD-NOS, ADHD, and those with a combined diagnosis on the basis of WISC-III profiles. *Journal of Autism and Developmental Disorders*, *39*, 549–556. doi:10.1007/s10803-008-0657-4

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

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

Schwean, V. L., & Saklofske, D. H. (2005). Assessment of attention deficit hyperactivity disorder with the WISC-IV. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 235–280). San Diego, CA: Elservier Academic Press.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*, 29–40. doi:10.1111/emip.2011.30.issue-3

Smitha, V. S., Dennis, D. I. M., Varghese, P. K., & Vinayan, K. P. (2014). Cognitive profile of children with attention deficit-hyperactivity disorder. *Amrita Journal of Medicine*, *10*, 28–32.

Snow, J. B., & Sapp, G. L. (2000). WISC-III subtest patterns of ADHD and normal samples. *Psychological Reports*, *87*, 759–765.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York, NY: Oxford University Press.

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

Styck, K. M., & Watkins, M. W. (2017). Structural validity of the WISC-IV for students with ADHD. *Journal of Attention Disorders*, *21*, 921–928. doi:10.1177/1087054714553052

Terman, L. M. (1918). Errors in scoring Binet tests. *Psychological Clinic*, *12*, 33–39.

Thaler, N. S., Barchard, K. A., Parke, E., Jones, W. P., Etcoff, L. M., & Allen, D. N. (2015). Factor structure of the Wechsler Intelligence Scale for Children: Fourth Edition in children with ADHD. *Journal of Attention Disorders*, *19*, 1013–1021. doi:10.1177/1087054712459952

Thaler, N. S., Bello, D. T., & Etcoff, L. M. (2013). WISC-IV profiles are associated with differences in symptomatology and outcome in children with ADHD. *Journal of Attention Disorders*, *17*, 291–301. doi:10.1177/1087054711428806

Theiling, J., & Petermann, F. (2014). Neuropsychological profiles on the WAIS-IV of ADHD adults. *Journal of Attention Disorders*. doi:10.1177/1087054713518241

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

Walg, M., Hapfelmeier, G., El-Wahsch, D., & Prior, H. (2017). The faster internal clock in ADHD is related to lower processing speed: WISC-IV profile analysis and time estimation tasks facilitate the distinction between real ADHD and pseudo-ADHD. *European Child & Adolescent Psychiatry*, *26*, 1177–1186. doi:10.1007/s00787-017-0971-5

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

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC–IV^{UK} with a large referred Irish sample. *International Journal of School & Educational Psychology*, *1*, 102–111. 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*, 976–983. doi:10.1177/0013164406288168

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

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2004a). *Wechsler Intelligence Scale for Children–Fourth edition (WISC-IV^{CDN})*. Torono, Canada: Harcourt Assessment.

Wechsler, D. (2004b). *Wechsler Intelligence Scale for Children–Fourth UK Edition*. London, UK: Harcourt Assessment.

Wechsler, D. (2005a). *Wechsler Intelligence Scale for Children–Fourth Edition*. Paris, France: Editions du Centre de Psychologie Appliquée.

Wechsler, D. (2005b). *Wechsler Intelligence Scale for Children–Fourth edition: Australian Standardised Edition (WISC-IV Australian)*. Sydney, Australia: PsychCorp.

Wechsler, D. (2005c). *Escala de inteligencia de Wechsler para niños–IV (WISC-IV)*. Madrid, España: Pearson Educación.

Wechsler, D. (2005d). *WISC-IV. Manual Técnico y de Interpretación*. Madrid, España: Pearson Educación.

Wechsler, D. (2007). *Escala Wechsler de lnteligencia para Niños–IV*. Mexico City, Mexico: Manua l Moderno.

Wechsler, D. (2008). *Wechsler Intelligence Scale for Children–Fourth Edition, Mainland China version*. Zhuhai, Guangdong, China: King-May Psychological Assessment Technology Development Limited Company.

Wechsler, D. (2012). *WISC-IV. Wechsler Intelligence Scale for Children Quarta Edizione*. Firenze, Italy: Giunti-OS.

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

Wechsler, D. (2015). *Escala de inteligencia de Wechsler para niños–V*. Madrid, España: Pearson Educación.

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

Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *30*, 1006–1016. doi:10.1080/13854046.2016.1205136

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., . . . Lahey, B. B. (2012). Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of Abnormal Psychology*, *121*, 991–1010. doi:10.1037/a0027347

Yang, P., Cheng, C.-P., Chang, C.-L., Liu, T.-L., Hsu, H.-Y., & Yen, C.-F. (2013). Wechsler Intelligence Scale for Children 4th edition–Chinese version index scores in Taiwanese children with attention-deficit/hyperactivity disorder. *Psychiatry and Clinical Neurosciences*, *67*, 83–91. doi:10.1111/pcn.12014