HAMMILL INSTITUTE ON DISABILITIES

Journal of Learning Disabilities
2016, Vol. 49(2) 216–224

DOI: 10.1177/0022219414539565
journaloflearningdisabilities.sagepub.com
SAGE

Structural Validity of the WISC-IV for Students With Learning Disabilities

Kara M. Styck, PhD[1] and Marley W. Watkins, PhD[2]

Abstract

The structural validity of the *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV) was evaluated using confirmatory factor analysis for a clinical sample of 1,537 students diagnosed with specific learning disabilities (SLD) by school psychologists in two large southwestern school districts. Results indicated that a bifactor model consisting of four first-order domain specific factors and a general intelligence breadth factor fit the data best. Consequently, the structural validity of the WISC-IV for students with SLD was supported by the results of the present study. The general intelligence factor contributed the most information, accounting for 48% of the common variance. Given this structure, it was recommended that score interpretation should emphasize the Full-Scale IQ score because of the marginal contributions of the first-order domain-specific factors and their low precision of measurement independent of the general factor.

Keywords

intelligence, factor analysis, learning disabilities

Standardized individual IQ test scores are frequently consulted for the identification of specific learning disabilities (SLD) largely due to the routine use of an ability–achievement discrepancy identification procedure (Reschly & Hosp, 2004; Zirkel, 2013). Therefore, the validity of high-stakes decisions regarding the presence of SLD is directly influenced by the validity of IQ test score interpretations. One aspect of validity evidence that is particularly important is the degree to which the scoring structure of the test matches the theoretical structure of the underlying construct the test purports to measure, referred to as structural validity (Messick, 1995).

Structural Validity of the WISC-IV for the Standardization Sample

The Wechsler intelligence scales are among the most popularly used IQ tests for assessment of children and adolescents (Stinnett, Havey, & Oehler-Stinnett, 1994) and the structural validity of the *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV; Wechsler, 2003a) has been investigated extensively by its developers and by independent researchers (Chen & Zhu, 2008; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Watkins, 2006; Wechsler, 2003b). Both exploratory (EFA) and confirmatory (CFA) factor analyses were conducted to assess the fidelity of the WISC-IV scoring structure with the standardization sample for the 10 subtests that make up the core battery as well as for all 15 core and supplementary subtests (Wechsler, 2003b). EFA results suggested an oblique four-factor model fit the core battery best, which was also supported by results of the CFA after comparing several competing models with one-, two-, three-, four-, and five-correlated factors. The oblique four-factor model was also found to generally measure the same construct across gender (Chen & Zhu, 2008) and age (Keith et al., 2006), extending empirical support for the structural validity of the WISC-IV.

However, the oblique four-factor model did not include a higher-order *g* factor and this has been criticized because the Wechsler theoretical structure of intelligence posits a multilevel structure (Wechsler, 2003b). Two groups of independent researchers recognized this issue (Keith et al., 2006; Watkins, 2006). Keith et al. (2006) and Weiss, Keith, Zhu, and Chen (2013) used the WISC-IV standardization sample data from all 15 subtests and compared the fit of a second-order factor model with four first-order factors matching those specified in Wechsler and a second-order *g* factor to the fit of five first-order factors as specified by the Cattell–Horn–Carroll (CHC; McGrew, 1997; 2005) theoretical model of intelligence. Both models fit the data equally well with negligible differences in the reported fit

[1]University of Texas at San Antonio, USA
[2]Baylor University, Waco, TX, USA

Corresponding Author:
Kara M. Styck, Department of Educational Psychology, University of Texas at San Antonio, 501 W. Cesar E. Chavez Blvd., San Antonio, TX 78207-4415, USA.
Email: Kara.Styck@utsa.edu

statistics (Chen, 2007), but Keith et al. favored the CHC model over the adapted Wechsler model on theoretical grounds. Nevertheless, the higher-order Wechsler model was more parsimonious because the CHC model contained a perfect loading of *Gf* onto *g*, suggesting redundancy of one factor, as well as numerous cross-loadings (i.e., Symbol Search, Picture Completion, and Matrix Reasoning subtests all cross-loaded on multiple first-order factors) that violated simple structure (Thurstone, 1940) and would make interpretation challenging.

Although the higher-order models tested by Keith et al. (2006) were more congruent with theory than the original first-order model proposed by Wechsler (2003a), higher-order factor models do not permit direct assessment of the relationships between the first-order factors (i.e., the four WISC-IV factors) and external criterion (i.e., achievement test scores) because a portion of the variance of the first-order factors is explained by the second-order *g* factor (Brunner, 2008; Chen, West, & Sousa, 2006; Schmiedek & Li, 2004). This results in an inability to identify the variance in the WISC-IV subtest scores that is explained by the domain-specific factors over and above the *g* factor without application of the Schmid-Leiman (1957) orthogonalization procedure. Given the predominance of IQ testing for identifying SLD by assessing the degree to which students' exhibit predicted academic achievement gains (Reschly & Hosp, 2004), this is a serious limitation.

Watkins (2006) followed the advice of Carroll (1993) and applied the Schmid-Leiman (1957) orthogonalization procedure, which transforms the first-order factors so that they are orthogonal to each other and to the general factor, to the results of an EFA of the10 core subtest scores of the standardization sample. This method allowed a decomposition of the variance of each WISC-IV subtest into three components: general intelligence, domain specific intelligence (e.g., verbal comprehension), and uniqueness (variance specific to each subtest alone and error). When viewed from this perspective, the general factor accounted for 71% of the common variance in the WISC-IV and for most of its predictive validity.

Structural Validity of the WISC-IV for Clinical Samples

IQ tests are seldom administered to nonclinical samples in applied practice. Rather, they are almost exclusively administered to students who are suspected of having a disability. It is well documented that IQ test scores obtained from clinical samples tend to have different distributional characteristics than IQ test scores obtained from nonclinical samples (Canivez & Watkins, 1998; Chen & Zhu, 2012; Devena, Gay, & Watkins, 2013; Nakano & Watkins, 2013; Watkins, Wilson, Kotz, Carbone, & Babula, 2006). Therefore, it is prudent to investigate the structural validity of the WISC-IV with clinical samples because it is those individuals who will most likely be administered the test.

Only one EFA has been conducted on the WISC-IV with a clinical sample (Watkins et al., 2006). It used the scores from the core battery with a clinical sample of 432 students referred for special education evaluations. Results supported the oblique four-factor model proposed by Wechsler (2003b) and highlighted the contributions of *g* through the application of the Schmid-Leiman (1957) orthogonalization procedure. The *g* factor accounted for 76% of the common variance and 47% of the total variance explained by the model, leaving the combined four domain-specific factors only contributing an additional 15% of the total variance.

The structural validity of the WISC-IV for clinical samples has been investigated in seven CFA studies. Four of the seven CFA studies (Canivez, 2014; Devena et al., 2013; Nakano & Watkins, 2013; Watkins, 2010) tested first-order oblique models, a higher-order model (see Figure 1), and a bifactor model (see Figure 2; Holzinger & Swineford, 1937). In bifactor models, the general factor directly affects all measured variables and is orthogonal to the domain specific factors, each of which affects a subset of the measured variables. This structure led Gignac (2008) to label this a direct hierarchical structure and to suggest that *g* is a breadth factor. In the higher-order model, *g* directly affects the first-order factors, which in turn, directly affect the measured variables. General intelligence in this model is a superordinate construct that indirectly affects measured variables. That is, *g* is fully mediated by the first-order factors in the higher-order model.

Participants in Nakano and Watkins (2013), Watkins (2010), and Canivez (2014) included 176, 355, and 345 students referred for special education evaluations, respectively, and participants in Devena et al. (2013) consisted of 297 children and adolescents administered comprehensive psychological evaluations at a pediatric hospital outpatient facility. Watkins, Devena et al., and Canivez all identified the bifactor model as fitting the data best, but Nakano and Watkins favored the second-order Wechsler model on the basis of empirical criteria—though differences in model fit statistics were small. In spite of these interpretive differences, the *g* factor accounted for 69% to 76% of the common variance in all analyses and an overwhelming 33% to 50% of the total variance explained by the model. The combined four domain-specific factors only contributed an additional 1.4% to 17% of the total variance, thereby accruing additional support for the predominance of *g* in explaining IQ test performance on the WISC-IV. Empirical research investigating the structural validity of the French and British versions of the WISC-IV with referred and nonclinical samples have noted similar conclusions regarding the superiority of the bifactor model for explaining the relationships between subtest scores (Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Watkins, Canivez, James, James, & Good,

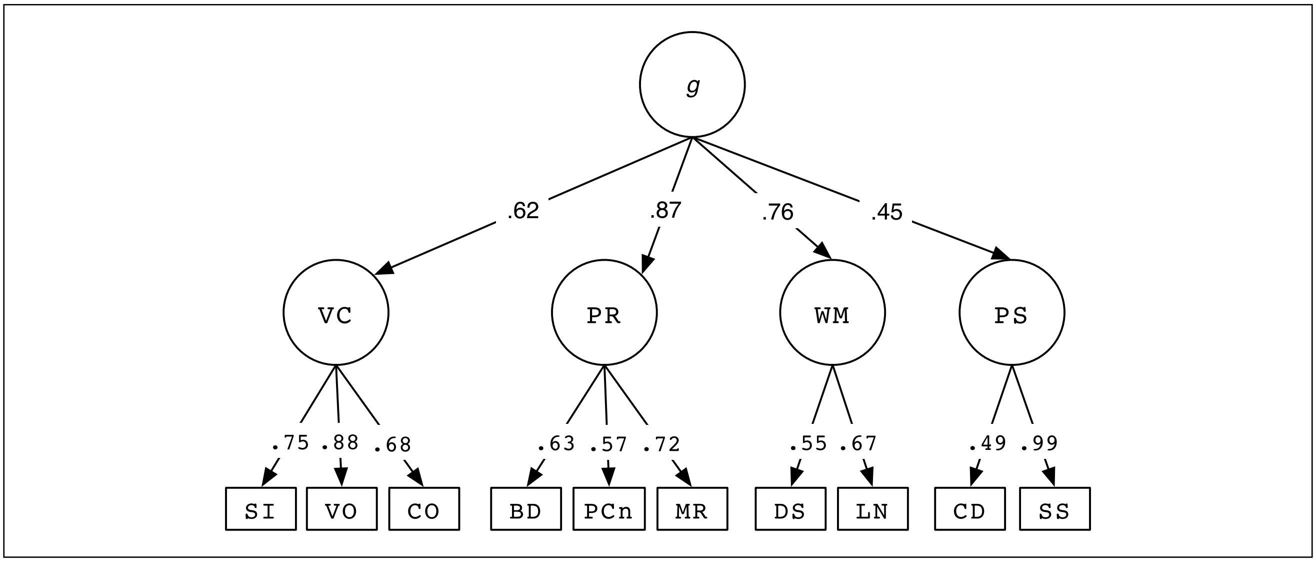

Figure 1. Higher-order structure of the WISC-IV for 1,537 students with learning disabilities.
Note. BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; *g* = General Intelligence; LN = Letter-Number Sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

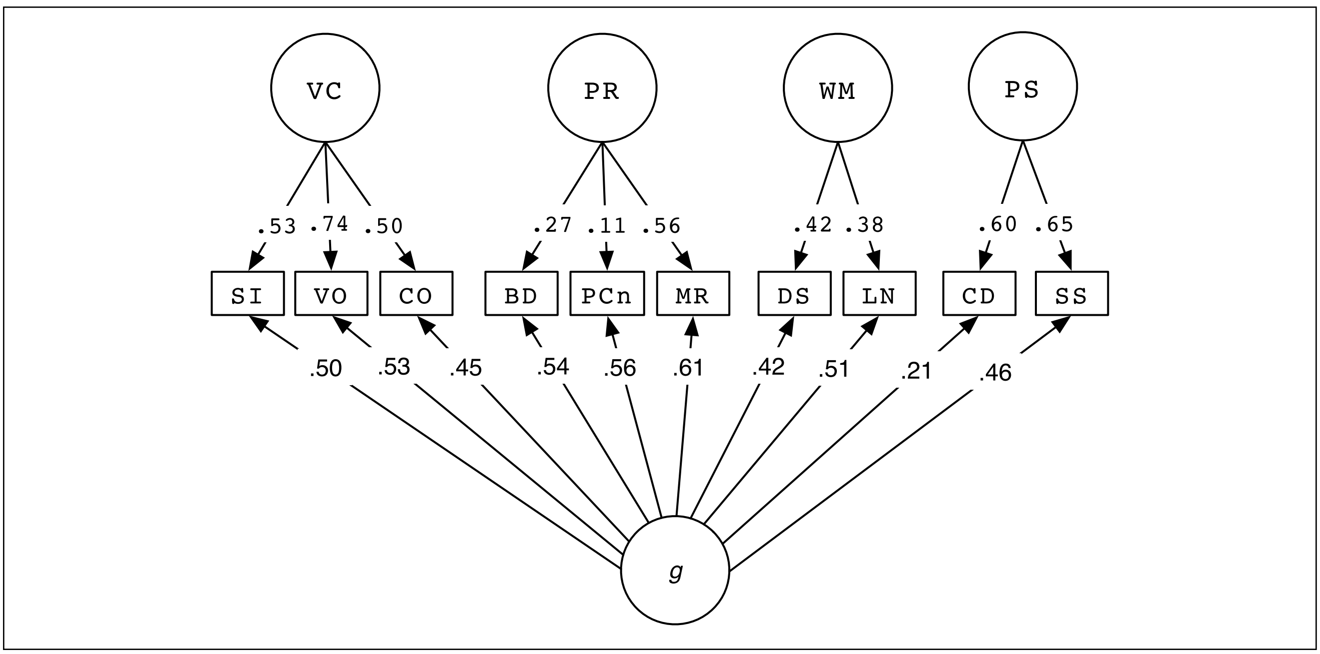

Figure 2. Bifactor structure of the WISC-IV for 1,537 students with learning disabilities.
Note. BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; *g* = General Intelligence; LN = Letter-Number Sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

2013) and negligible contributions of domain specific factors over and above *g* when this information is reported (Watkins et al., 2013).

The remaining three CFA studies on the WISC-IV with clinical samples evaluated the fit of a series of second-order models, but did not evaluate the fit of a bifactor model (Bodin, Pardini, Burns, & Stevens, 2009; Chen & Zhu, 2012; Weiss et al., 2013). Bodin et al. (2009) determined the adapted Wechsler (2003b) model with four first-order factors and a superordinate *g* factor best fit a clinical sample of

344 children and adolescents who were administered the core battery of the WISC-IV as part of a neuropsychological evaluation at a pediatric hospital. The Schmid-Leiman (1957) orthogonalization procedure was also used to assess the individual contributions of each factor. Bodin et al. reported that 77% of the common variance was attributed to *g* along with 48% of the total variance, whereas the combined four domain-specific factors contributed only an additional 15% of the total variance explained by the model.

The second-order Wechsler (2003b) model supported by Bodin et al. and Nakano and Watkins (2013) has also been reported to measure the same construct across clinical and nonclinical groups of the WISC-IV standardization sample (Chen & Zhu, 2012) as well as a second-order CHC model with both four- and five-first-order factors (Weiss et al., 2013). However, the CHC models included in Weiss et al. (2013) required use of all 15 subtests and contained numerous cross-loadings analogous to Keith et al. (2006). Canivez and Kush (2013) noted these and other concerns about the Weiss et al. study in a commentary published within the same special issue of the *Journal of Psychoeducational Assessment*. For example, the literature review neglected to cite empirical evaluations of rival models, several intermediate latent factors were required to support the general CHC structure (i.e., a Quantitative Reasoning factor was created to explain the relationship between the Figure Weights and Arithmetic subtests and an Inductive Reasoning factor was created to explain the relationship between the Picture Concepts and Matrix Reasoning subtests), and the variance attributed to the CHC factors above and beyond *g* was not reported.

Although the structural validity of the WISC-IV has been investigated with clinical samples, the generalizability of the results do not extend to specific homogenous clinical subpopulations because each clinical sample consisted of a heterogeneous group of children and adolescents diagnosed with myriad disabilities (Bodin et al., 2009; Canivez, 2014; Chen & Zhu, 2012; Devena et al., 2013; Nakano & Watkins, 2013; Watkins, 2006, 2010; Watkins et al., 2006; Weiss et al., 2013). It is possible that the theoretical structure of the WISC-IV is different for various homogenous clinical populations and it is important that this possibility be tested. Unfortunately, no study has evaluated the structural validity of the WISC-IV for a single clinical subsample. Given the prevalent use of IQ test scores for determining the presence of SLD (Reschly & Hosp, 2004; Zirkel, 2013) and the popularity of the WISC-IV (Stinnett et al., 1994), the purpose of the present investigation is to evaluate the structural validity of the WISC-IV for children and adolescents diagnosed with SLD.

Method

Participants

Participants were 1,537 (62.5% male) students with learning disabilities enrolled in two large Southwestern school districts. Educational records indicated that participants' ethnic/racial backgrounds was 72.0% Caucasian, 16.6% Hispanic, 5.3% African American, 2.1% Native American, 1.5% Asian/Pacific, and 2.5% Other/Missing. Students were 6 to 16 years of age (M = 10.4, SD = 2.3 years) and enrolled in grades kindergarten through eleven (Md = 4).

Most participants were given multiple learning disability diagnoses. For example, 70.7% of the participants had a learning disability in reading, 48.1% had a learning disability in math, and 57.0% had a learning disability in written expression. Of the students with a reading disability, 76.2% had an additional learning disability diagnosis in math, written expression, oral expression, or listening comprehension. Likewise, a large proportion of the students with a math disability (73.9%) had an additional learning disability diagnosis in another academic area. In addition, 238 students with learning disabilities had a secondary diagnosis (e.g., speech-language impairment, other health impairment, emotional disability, etc.).

Instruments

The WISC-IV contains 10 core and 5 supplemental subtests, each with a population mean of 10 and standard deviation of 3. The core subtests are used to form four factor indexes, where the Verbal Comprehension Index (VC) is based on the Similarities, Vocabulary, and Comprehension subtests; the Perceptual Reasoning Index (PR) is based on Block Design, Matrix Reasoning, and Picture Concepts subtests; the Working Memory Index (WM) on the Digit Span and Letter-Number Sequencing subtests; and the Processing Speed Index (PS) on the Coding and Symbol Search subtests. The FSIQ is also formed from the 10 core subtests. The factor indexes and FSIQ each have a population mean of 100 and standard deviation of 15. The supplemental subtests were not included in this study because their infrequent application precluded requisite statistical power for multivariate analyses.

Examiners used a variety of academic achievement measures, but the majority of scores were from a version of the *Wechsler Individual Achievement Test* (47.3%) and a version of the *Woodcock–Johnson Tests of Achievement* (49.8%). Both measures are well-developed scales with nationally representative normative samples and strong psychometric characteristics (Thorndike & Thorndike-Christ, 2010).

Procedures

Participants were enrolled in two large southwestern public school districts. The first district had an enrollment of 32,500 students and included 31 elementary, 8 middle, and 6 high schools. Ethnic composition for the 2009–2010 academic year was 67.2% Caucasian, 23.8% Hispanic, 4.0% African American, 3.9% Asian, and 1.1% Native American. The second district served 26,000 students in 2009–2010, with 16 elementary schools, 3 kindergarten through 8th grade schools, 6 middle schools, 5 high schools, and 1 alternative school. Caucasian students composed 83.1% of enrollments, Hispanic 10.5%, Asian 2.9%, African American 1.7%, and other ethnic minorities 1.8%.

After obtaining Internal Review Board and school district approval, eight trained school psychology doctoral students examined approximately 7,500 student special education files and retrieved assessment data from all special education files spanning the years 2003 to 2010 where psychologists had administered the WISC-IV. Following this procedure, data were obtained on 2,783 students with a primary diagnoses of learning disabilities (57.6%), emotional disability (11.6%), attention-deficit/hyperactivity disorder (8.0%), intellectual disability (2.6%), other disabilities (12.1%), and no diagnosis (8.0%). Of the 1,603 students with learning disabilities, 66 did not have scores from all 10 core WISC-IV subtests.

The 1,537 students with complete WISC-IV scores were determined to be learning disabled by local multidisciplinary teams following district and state guidelines that permitted the use of ability–achievement discrepancies. In general, diagnostic criteria resulted in expected reading–math score patterns for students with learning disabilities in reading and math. For example, students with a learning disability in reading exhibited lower reading scores than students with a learning disability in math (M = 79.3 vs. 83.0, respectively) and students with a learning disability in math exhibited lower math scores than students with a learning disability in reading (M = 80.0 vs. 88.4, respectively). This reading–math pattern was magnified when students with an exclusive learning disability in reading (M reading = 82.8, M math = 95.5) were compared to students with an exclusive learning disability in math (M reading = 93.2, M math = 81.3).

Analyses

Mplus 7 for the Macintosh (Muthén & Muthén, 2012) was used to conduct CFA using maximum likelihood estimation with Satorra-Bentler (1994) scaling to correct for minor departures of multivariate normality. Consistent with previous WISC-IV structural analyses, four first-order models and two hierarchical models were specified and examined (a) one factor; (b) two oblique verbal and nonverbal factors; (c) three oblique verbal, perceptual, and combined working memory/processing speed factors; (d) four oblique verbal, perceptual, working memory, and processing speed factors as per Wechsler (2003b); (e) a higher-order model (as per Bodin et al., 2009) with four first-order factors; and (f) a bifactor (sometimes called direct hierarchical or nested) model (as per Watkins, 2010) with four domain specific factors. See Gignac (2008) for a detailed description of higher-order and bifactor models and Figures 1 and 2 for illustrations of each model.

Given the complexity of the statistical models and the extreme degree of comorbidity among participants, it was not possible to divide the learning disability sample into discrete subtypes (i.e., reading, math, written expression) and retain sufficient power for subsequent CFAs (Kline, 2011). Comorbidity has been found among other large samples of students with learning disabilities (Benson & Taub, 2013). Consequently, analyses were conducted with the total sample of students with learning disabilities.

Although there are no universally accepted cutoff values for model fit indices, multiple indices that represented a variety of fit criteria were examined (Kline, 2011), specifically the (a) χ^2, (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC). The standards for good model fit were (a) CFI $\geq$ .95, (b) RMSEA $\leq$ .06, and (c) SRMR $\leq$ .06 (Hu & Bentler, 1999). There are no specific criteria for information-based fit indices like the AIC, but smaller values indicate better approximations of the true model (Vrieze, 2012). For a model to be deemed superior, it had to (a) exhibit good fit according to CFI, RMSEA, and SRMR standards and (b) display the smallest AIC value.

Finally, factor reliabilities were estimated with coefficient omega (ω) and omega hierarchical (ω_h) as per Watkins (2013). The traditional coefficient alpha reliability estimate has long been known to be biased (Sijtsma, 2009; Yang & Green, 2011) and coefficient omega has been recommended as its replacement (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012). Omega estimates reliability based on the total systematic variance in each factor including variance from the general factor and the domain specific factor, whereas omega hierarchical estimates the reliability of each factor with variance from the general factor removed. Thus, ω_h controls for that part of reliability due to the general factor and is useful for judging the utility of factor index scores (Reise, 2012). There are no absolute standards for evaluating the magnitude of ω or ω_h, but it has been tentatively suggested that values near .75 might be preferred, and values greater than .50 might be a minimum (Reise, Bonifay, & Haviland, 2013).

Results

As reflected in Table 1, WISC-IV subtest and index scores were lower than the national average (d = −0.6). These descriptive statistics were expected because lower cognitive and achievement scores are frequently observed in clinical samples (Benson & Taub, 2013; Watkins, 2010). Univariate skewness and kurtosis values were under 1.0, but multivariate normality was not supported (χ^2 = 326.9, df = 20, p < .001; Doornik & Hansen, 2008).

The model fit statistics presented in Table 2 illustrate the increasingly better fit obtained from 1 to 4 first-order factors. Based on the standards for CFI, RMSEA, and SRMR

Table 1. Descriptive Statistics for 1,537 Students With Learning Disabilities Tested on the Wechsler Intelligence Scale for Children–Fourth Edition.

Score	*M*	*SD*	Skewness	Kurtosis
Block Design	9.19	2.69	0.20	0.16
Similarities	8.82	2.67	0.33	0.12
Digit Span	7.88	2.44	0.19	0.17
Picture Concepts	9.86	2.86	−0.11	0.14
Coding	8.36	3.03	0.44	0.31
Vocabulary	8.45	2.48	0.20	0.21
Letter-Number Sequencing	8.30	2.70	−0.53	0.03
Matrix Reasoning	9.33	2.69	0.13	0.18
Comprehension	9.04	2.51	−0.12	0.63
Symbol Search	8.84	2.81	−0.30	0.23
Verbal Comprehension Index	92.76	12.13	0.16	0.77
Perceptual Reasoning Index	96.80	13.08	−0.08	0.19
Working Memory Index	88.64	11.92	−0.19	0.18
Perceptual Speed Index	92.19	14.17	0.19	−0.16
Full-Scale IQ	91.20	11.51	−0.06	−0.06
Reading	82.84	11.40	−0.30	0.75
Math	87.62	12.66	−0.16	0.61

Table 2. CFA Fit Statistics for the Wechsler Intelligence Scale for Children–Fourth Edition Among 1,537 Students With Learning Disabilities.

Model	χ^2	*df*	CFI	RMSEA	90% RMSEA	SRMR	AIC
One factor	1169.59	35	.704	.145	.138–.152	.091	71,228
Verbal and nonverbal	651.52	34	.839	.109	.101–.116	.075	70,704
Three factors	421.83	32	.898	.089	.082–.097	.056	70,471
Wechsler four factors	131.37	29	.973	.048	.040–.056	.029	70,181
Higher-order	150.45	31	.969	.050	.042–.058	.033	70,196
Bifactor[a]	126.15	27	.974	.049	.040–.058	.030	70,177

Note. AIC = Akaike information criterion; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
[a]Two indicators of third and fourth factors were constrained to equality to allow model identification.

indices, the correlated four-factor, higher-order, and bifactor models were all good fits to the data. However, the four factors were highly correlated (*Md* = .52) in the oblique first-order model, suggesting the presence of a general factor (Gorsuch, 1983), making that model an inadequate explanation of the data. When comparing the higher-order and bifactor models, the bifactor model exhibited superior fit according to all the fit indices, including the AIC. Thus, it was selected as the best explanation of the WISC-IV factor structure among students with learning disabilities.

Table 3 presents decomposed WISC-IV subtest variance estimates based on the bifactor model. The general factor accounted for 47.9% of the common variance and 23.7% of the total variance, the VC factor accounted for 21.8% of the common variance and 10.8% of the total variance, the PR factor accounted for 8.1% of the common variance and 4.0% of total variance, the WM factor accounted for 6.4% of the common variance and 3.2% of the total variance, and the PS factor accounted for 15.7% of the common variance and 7.8% of the total variance (see Table 3). Thus, the *g* factor accounted for substantially greater portions of WISC-IV common and total variance relative to the domain specific factors. In addition, communality was lower than uniqueness for six of the ten WISC-IV subtests.

The omega and omega hierarchical coefficients presented in Table 3 are estimates of the reliability of the WISC-IV factors. In the case of the four domain specific factors, ω_h coefficients estimated the scale reliabilities with the effects of the general factor removed and ranged from .166 (PR) to .525 (PS). In contrast, the *g* factor exhibited a ω_h coefficient of .667. These values suggest that if both total and factor scores are formed, interpretation of the factor scores "as precise indicators of unique constructs is extremely limited—very little reliable variance exists beyond that due to the general factor" (Reise, 2012, p. 691).

Table 3. Sources of Variance in the Wechsler Intelligence Scale for Children–Fourth Edition Among 1,537 Children With Learning Disabilities.

	General		VC		PR		WM		PS			
Subtest	*b*	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var	h^2	u^2
SI	.504	.254	.534	.285							.539	.461
VO	.528	.279	.735	.540							.819	.181
CO	.446	.199	.504	.254							.453	.547
BD	.536	.287			.273	.075					.362	.638
PCn	.558	.311			.114	.013					.324	.676
MR	.606	.367			.562	.316					.683	.317
DS	.416	.173					.418	.175			.348	.652
LN	.508	.258					.378	.143			.401	.599
CD	.212	.045							.601	.361	.406	.594
SS	.445	.198							.647	.419	.617	.383
% total variance		23.7		10.8		4.0		3.2		7.8	49.5	50.5
% common variance		47.9		21.8		8.1		6.4		15.7		
ω		.851		.818		.699		.543		.671		
ωh		.667		.482		.166		.231		.525		

Note. b = standardized loading of subtest on factor; BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; h^2 = communality; LN = Letter-Number Sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search; u^2 = uniqueness; Var = percentage variance explained in the subtest; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

Discussion

Results of the present study generally support the structural validity of the WISC-IV for a clinical sample of students identified as having SLD. Moreover, results indicate that the bifactor model was a superior fit to the data, which consisted of four domain specific factors matching Wechsler's (2003b) structure and a general intelligence breadth factor. This substantiates results of previous research on the WISC-IV with heterogeneous clinical samples (Canivez, 2014; Devena et al., 2013; Watkins, 2010). Nakano and Watkins (2013) is the only published study in which a second-order model exhibited a better fit than the bifactor model. However, Nakano and Watkins reported that both models with a higher-order *g* factor fit the data well and differences in the model fit statistics were minimal.

The bifactor model has numerous advantages over other models of intelligence. Principally, it elucidates the individual contributions of all first-order factors and it enables researchers to study the external criterion validity of domain specific subabilities (Brunner, 2008; Chen et al., 2006; Schmiedek & Li, 2004). Bifactor models are also more parsimonious than second-order factor models (Gustafsson, 2001) and resemble Carroll's (1993) conceptualization of *g* as a breadth factor. Applying a bifactor model to the WISC-IV data in this study, approximately 48% of the common variance was attributed to *g*, along with 24% of the total variance. However, the general intelligence factor explained more than twice the amount of the total variance of any single domain specific factor.

Some popular contemporary theories of intelligence unwittingly emphasize interpretation of domain specific factors over *g* (Flanagan, Ortiz, & Alfonso, 2013; McGrew, 1997, 2005), but the results of the present study do not support those claims. Specifically, 6 of the 10 subtests contained more specific and error variance than general or group factor variance. In addition, the ω_h values of .166 to .525 for the four first-order factors indicate that those factors did not retain much precision of measurement independent of the general factor. This implies large confidence intervals around the factor scores and makes interpretation of a person's specific ability very uncertain (Brunner et al., 2012).

The present study represents the first CFA with a homogeneous clinical sample, but it is not without limitations. Participants were recruited from a single geographic location, which may limit generalizability. In addition, approximately 15% of participants were concurrently diagnosed with a second educational disability. SLD are highly comorbid with other conditions, such as attention-deficit/hyperactivity disorder (DuPaul, Gormley, & Laracy, 2013; Nelson & Canivez, 2012), developmental coordination disorders (Kaplan, Wilson, Dewey, & Crawford, 1998), and epilepsy (Fastenau, Shen, Dunn, & Austin, 2008). Therefore, participants are representative of the population of students diagnosed with SLD. Furthermore, the present investigation included a single sample of students with SLD. This precluded an empirical evaluation of measurement invariance with a nonclinical sample. The WISC-IV has demonstrated measurement invariance with a heterogeneous group of students with disabilities (Chen & Zhu, 2012), and future research should investigate the degree to which measurement invariance is sustained across a nonclinical sample of students and a homogeneous sample of students with SLD. Notwithstanding results of the present investigation verify those of numerous others that detect the dominating influence of a general intelligence factor (Bodin et al., 2009; Canivez, 2014; Devena et al., 2013; Watkins, 2006, 2010; Watkins et al., 2006). As a result, interpretation of the WISC-IV in applied settings should emphasize the Full-Scale IQ score and interpretation of domain specific factor scores should be undertaken only after considering their precision of measurement and with the understanding that "each specific factor score estimate provides little information beyond that provided by the general factor estimate" (DeMars, 2013, p. 374).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Benson, N., & Taub, G. E. (2013). Invariance of Woodcock-Johnson III scores for students with learning disorders and students without learning disorders. *School Psychology Quarterly*, *28*, 256–272. doi:10.1037/spq0000028

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

Brunner, M. (2008). No g in education? *Learning and Individual Differences*, *18*, 152–165. doi:10.1016/j.lindif.2007.08.005

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

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/ssq0000032

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

Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment*, *10*, 285–291. doi:1040-3590/98

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

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

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

Chen, H., & Zhu, J. (2008). Factor invariance between genders of the Wechsler Intelligence Scale for Children–Fourth Edition. *Personality and Individual Differences*, *45*, 260–266. doi:10.1016/j.paid.2008.04.008

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

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*. Advance online publication. doi:10.1177/0734282913483981

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939. doi:10.1111/j.1468-0084.2008.00537.x

DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications of DSM-5 for assessment and treatment. *Journal of Learning Disabilities*, *46*, 43–51. doi:10.1177/0022219412464351

Fastenau, P. S., Shen, J., Dunn, D. W., & Austin, J. K. (2008). Academic underachievement among children with epilepsy. *Journal of Learning Disabilities*, *41*, 195–207. doi:10.1177/0022219408317548

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). How to interpret test data. In A. S. Kaufman & N. L. Kaufman (Series Ed.), *Essentials of cross-battery assessment third edition* (3rd ed., pp. 121–226). Hoboken, NJ: John Wiley.

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

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: Lawrence Erlbaum.

Gustafsson, J. E. (2001). On the hierarchical structure of ability and personality. In J. M. Collis & S. Messick (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 25–42). Mahwah, NJ: Lawrence Erlbaum.

Holzinger, K. J., & Swineford, F. (1937). The bifactor method. *Psychometrika*, *2*, 41–54.

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

Kaplan, B. J., Wilson, B. N., Dewey, D. M., & Crawford, S. G. (1998). DCD may not be a discrete disorder. *Human Movement Science*, *17*, 471–490.

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.

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

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

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

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749. doi:0003-066X/95

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th 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

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

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

Reschly, D. J., & Hosp, J. L. (2004). State SLD identification policies and practices. *Learning Disability Quarterly*, *27*, 197–213.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structural analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399–419). Thousand Oaks, CA: Sage.

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

Schmiedek, F., & Li, S. C. (2004). Toward an alternative representation for disentangling age-associated differences in general and specific cognitive abilities. *Psychology and Aging*, *19*, 40–56.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120. doi:10.1007/S11336-008-9101-0

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

Thorndike, R. M., & Thorndike-Christ, T. M. (2010). *Measurement and evaluation in psychology and education* (8th ed.). New York, NY: Pearson.

Thurstone, L. L. (1940). Current issues in factor analysis. *Psychological Bulletin*, *37*, 189–236.

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:1040-3593/06

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., 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 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*, 976–983. doi:10.1177/0013164406288168

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

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

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

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*, 377–392. doi:10.1177/0734282911406668

Zirkel, P. A. (2013). The trend in SLD enrollments and the role of RTI. *Journal of Learning Disabilities*, *46*, 473–479. doi:10.1177/0022219413495297