Journal of Attention Disorders
2017, Vol. 21(11) 921–928

DOI: 10.1177/1087054714553052
journals.sagepub.com/home/jad
SAGE

Structural Validity of the WISC-IV for Students With ADHD

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

Abstract

Objective: To evaluate the structural validity of the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV). **Method:** Confirmatory factor analyses were applied to a sample of 233 students diagnosed with ADHD by school multidisciplinary evaluation teams to evaluate the relative fit of the following competing models: (a) one factor, (b) two oblique verbal and nonverbal factors, (c) three oblique verbal, perceptual, and combined working memory/processing speed factors, (d) four oblique verbal, perceptual, working memory, and processing speed factors, (e) a higher-order model with four first-order factors, and (f) a bifactor model with four domain-specific factors. **Results:** A higher-order four-factor model fit the data best, which was composed of a general intelligence factor and four domain-specific factors that matched the four factors specified in the WISC-IV technical and interpretive manual. Moreover, the general intelligence factor explained more than two times the total variance contributed by all four domain-specific factors combined. **Conclusions:** Results substantiate previous research on the WISC-IV, indicating that the general intelligence factor contributes the most reliable information. Consequently, it is recommended that interpretation of the WISC-IV remain at the Full-Scale IQ score level. *(J. of Att. Dis. 2017; 21(11) 921-928)*

Keywords

intelligence, factor analysis, ADHD, WISC-IV, validity

IQ test scores have been noted as clinically useful in the diagnosis and treatment planning for students with ADHD because they can be used to make differential diagnoses (i.e., ADHD compared with intellectual impairment), may contribute explanatory information regarding presenting symptoms (i.e., low working memory or processing speed scores), and can ostensibly aid intervention planning by identifying patterns of cognitive strengths and weaknesses (Schwean & McCrimmon, 2008; Schwean & Saklofske, 1998). However, the validity of these test score interpretations is contingent upon empirical evidence in support of the match between the scoring structure of the test and the underlying theoretical structure of the construct the test purportedly measures (Loevinger, 1957). This is referred to as structural validity, which is appraised using exploratory and confirmatory factor analytic (EFA and CFA) statistical techniques (Messick, 1995).

The Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) is the most recently published version of the Wechsler intelligence scales, which are among the most commonly administered standardized individual intelligence tests for children and adolescents (Braden, 2013). According to the WISC-IV technical and interpretive manual, the structure of intelligence measured by the WISC-IV is composed of four domain-specific factors measuring working memory, processing speed, perceptual reasoning, and verbal comprehension as well as a general intelligence factor (Wechsler, 2003b). This general structure has been supported in the standardization sample (Wechsler, 2003b), with clinical samples (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Devena, Gay, & Watkins, 2013; Nakano & Watkins, 2013; Styck & Watkins, 2014; Watkins, 2010), between samples with and without clinical disorders (Chen & Zhu, 2012; Weiss, Keith, Zhu, & Chen, 2013), across age (Keith, Fine, Taub, Reynolds, & Kranzler, 2006) and across sex (Chen & Zhu, 2008). Moreover, the structural validity of the WISC-IV has also been supported in international versions of the test (Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Watkins, Canivez, James, James, & Good, 2013).

However, only two studies to date have been published in peer-reviewed journals that evaluated the structural

[1]University of Texas at San Antonio, TX, 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, San Antonio, TX 78207, USA.
Email: Kara.Styck@utsa.edu

validity of the WISC-IV with a homogeneous sample of students with ADHD (Thaler et al., 2012; P. Yang et al., 2013)—only one of which was conducted on the U.S. version of the test (Thaler et al., 2012). Thaler et al. (2012) compared the four-factor model reported in the WISC-IV technical and interpretive manual (Wechsler, 2003b) with the Cattell–Horn–Carroll (CHC) theoretical structure of intelligence (McGrew, 1997, 2005) using the 10 core WISC-IV subtests for a sample of 314 children and adolescents aged 6 to 16.6 years old with a primary diagnosis of ADHD provided by a pediatric neuropsychologist at a private practice clinic following *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev; *DSM-IV-TR*; American Psychiatric Association [APA], 2000) diagnostic criteria. Participants consisted predominantly of males (70%) and carried comorbid diagnoses of reading disorder (28%), mathematics disorder (18%), disorder of written expression (25%), oppositional defiant disorder (6%), and developmental coordination disorder (19%), all of which are consistent with epidemiologic studies of ADHD with community and clinical samples (Bauermeister et al., 2007).

The CHC model evaluated in Thaler et al. (2012) consisted of five domain-specific factors, including *Gc* (Vocabulary, Similarities, and Comprehension subtests), *Gf* (Picture Concepts and Matrix Reasoning subtests), *Gv* (Block Design and Symbol Search subtests), *Gs* (Symbol Search and Coding subtests), and *Gsm* (Digit Span and Letter–Number Sequencing subtests) as well as a general intelligence factor, *g*. Results indicated that the Wechsler (2003b) model fits the data well, $\chi^2(30) = 54.17$, $p < .01$; comparative fit index (CFI) = .97; root mean square error of approximation (RMSEA) = 0.05, but the CHC model was determined to fit the data better despite marginal differences in fit statistics, $\chi^2(24) = 32.26$, $p = .12$, CFI = 0.99, RMSEA = 0.04. Nonetheless, CHC models contain numerous interpretive challenges. Thaler et al. reported that *Gf* loaded 0.97 onto *g*, which is consistent with other studies investigating the CHC factor structure with the Wechsler scales (Keith et al., 2006; Weiss et al., 2013) and suggests that these two factors may not be independent. Second, Symbol Search was permitted to cross-load onto *Gv* and *Gs*. This violates simple structure and complicates the interpretation of these two factors as measuring different latent constructs (Thurstone, 1940). Moreover, results of the Schmid and Leiman (1957) orthogonalization procedure applied to the CHC model revealed that *g* accounted for 62.5% of the total variance and the five CHC domain-specific factors contributed negligible information (i.e., *Gc* = 11.2%, *Gf* = 0.9%, *Gv* = 2.9%, *Gs* = 17.1%, and *Gsm* = 5.4%).

P. Yang et al. (2013) investigated the structural validity of the Chinese version of the WISC-IV (Wechsler, 2007) with a sample of 334 children and adolescents aged 7 to 11 years recruited from the outpatient service of a psychiatric hospital in Taiwan. ADHD was diagnosed according to *DSM-IV-TR* criteria and the majority of participants were male, similar to Thaler et al. (2012) and commonly reported in samples of children and adolescents with ADHD (Bauermeister et al., 2007). Results indicated that the four-factor model reported in the WISC-IV technical and interpretive manual fit the data well, $\chi^2(30) = 50.47$, $p = .008$, CFI = 0.99, RMSEA = 0.05, accruing additional support for the structural validity of the WISC-IV with children and adolescents diagnosed with ADHD.

However, previous research on the U.S. version of the WISC-IV is limited to one study conducted with a sample of children and adolescents diagnosed with ADHD from a single pediatric neuropsychologist in a specific geographic location (Thaler et al., 2012). The diagnosis of ADHD in school-based samples is markedly different because it involves adherence to the Individuals With Disabilities Education Improvement Act (2004) and eligibility decisions made by school multidisciplinary evaluation teams. Therefore, it is important to empirically evaluate the structural validity of the WISC-IV for school-based samples because different diagnostic criteria are used to identify ADHD in school settings, and these differences may produce samples with distinctive characteristics. Furthermore, Thaler et al. only compared the four-factor model delineated in the WISC-IV technical manual with a CHC model, whereas Wechsler's (2003b) study contained comparisons between models with one-, two-, three-, four-, and five-correlated factors. Likewise, P. Yang et al. (2013) only evaluated the fit of the four-factor model cited in Wechsler (2003b) as providing a superior fit to the standardization sample data and did not compare the fit of the four-factor solution to any competing models.

The relative fit of the alternative models presented in Wechsler (2003b) to a sample of school children and adolescents diagnosed with ADHD remains unknown. Consequently, the purpose of the present investigation is to evaluate the structural validity of the WISC-IV in a school-based sample to expand research on the structure of the WISC-IV to this target population.

Method

Participants

Participants were 233 (70% male) students with ADHD enrolled in two large Southwestern school districts. School multidisciplinary evaluation teams determined that participants met federal and state criteria for students with Other Health Impairment due to ADHD diagnoses. Diagnostic decisions were guided by recommendations made from special education eligibility evaluation results and verified by a physician in accordance with Arizona Administrative Code (R7-2-401). Educational records indicated that participants' ethnic/racial background was 80.3% Caucasian, 9.9%

Hispanic, 5.2% African American, 0.40% Native American, 0.40% Asian/Pacific, and 3.9% Other/Missing. Students were 6 to 16 years of age (M = 10.5, SD = 2.6 years) and enrolled in kindergarten through Grade 10 (Md = 4). Participants received secondary diagnoses of learning disabilities (7%), severe language impairments (4%), emotional disabilities (3%), and other (3.5%).

Instruments

The WISC-IV contains 10 core and 5 supplemental subtests, each with a population mean of 10 and a standard deviation of 3. The core subtests are used to create four-factor index scores: (a) the Verbal Comprehension Index (VCI) from the Similarities, Vocabulary, and Comprehension subtests; (b) the Perceptual Reasoning Index (PRI) from the Block Design, Matrix Reasoning, and Picture Concepts subtests; (c) the Working Memory Index (WMI) from the Digit Span and Letter–Number Sequencing subtests; and (d) the Processing Speed Index (PSI) from the Coding and Symbol Search subtests. The Full-Scale IQ (FSIQ) is computed from all 10 core subtests. The factor indices and FSIQ each have a population mean of 100 and a standard deviation of 15. The supplemental subtests were not included in this study because they were rarely administered to this sample of students (N = 3-19 per supplemental subtest).

Examiners used a variety of academic achievement measures, but the majority of scores were from a version of the Wechsler Individual Achievement Test (37%) and a version of the Woodcock–Johnson Tests of Achievement (59%). 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. The second district served 26,000 students in 2009 to 2010, with 16 elementary schools, 3 kindergarten through eighth-grade schools, 6 middle schools, 5 high schools, and 1 alternative school. School district demographics were collected from information provided by the National Center for Education Statistics (2012). The first district was comprised of approximately 84% non-Hispanic or Latino students, with 6% of their students identified as English Language Learners. The second district was comprised of approximately 88% non-Hispanic or Latino students, with 4% of their students identified as English Language learners.

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 233 students with a primary diagnosis of ADHD who had been administered all 10 core WISC-IV subtests.

Analyses

Mplus 7 for the Macintosh (Muthén & Muthén, 2012) was used to conduct CFA using maximum likelihood estimation with Satorra and Bentler (1994) scaling to correct for departures from 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 bi-factor (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 bi-factor models.

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, (a) χ^2, (b) CFI, (c) RMSEA, (d) standardized root mean square residual (SRMR), and (e) Akaike's Information Criterion (AIC) were examined. The standard for good model fit was as follows: (a) CFI $\geq$ 0.95, (b) RMSEA $\leq$ 0.06, and (c) SRMR $\leq$ 0.06 (Hu & Bentler, 1999). There are no specific criteria for information-based fit indices such as 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; Y. 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

Table 1. Descriptive Statistics for WISC-IV Scores of 233 Students With ADHD.

	M	*SD*	Skewness	Kurtosis
Block Design	9.41	3.06	0.37	−0.61
Similarities	9.52	2.88	0.19	−0.39
Digit Span	8.70	2.70	0.55	1.05
Picture Concepts	10.24	2.92	−0.12	−0.01
Coding	8.25	2.96	0.25	0.37
Vocabulary	9.40	2.53	0.34	0.03
Letter–Number Sequencing	8.80	2.60	−0.21	0.42
Matrix Reasoning	9.70	2.93	0.47	0.86
Comprehension	9.32	2.69	−0.10	1.29
Symbol Search	9.14	2.99	−0.07	0.69
Verbal Comprehension Index	96.35	12.56	0.40	0.50
Perceptual Reasoning Index	98.78	14.19	0.30	−0.16
Working Memory Index	92.42	12.47	0.16	1.13
Perceptual Speed Index	92.90	14.53	0.19	0.24
Full-Scale IQ	94.47	13.51	0.54	0.38
Reading	95.02	12.16	0.27	−0.21
Math	93.52	12.80	0.35	1.10

Note. WISC-IV = Wechsler Intelligence Scale for Children–Fourth Edition.

Table 2. CFA Fit Statistics for the WISC-IV Among 233 Students With ADHD.

Model	χ^2	*df*	CFI	RMSEA	90% RMSEA	SRMR	AIC
One factor	133.25	35	0.852	0.110	0.090-0.130	0.064	10888.2
Verbal and nonverbal	98.40	34	0.903	0.090	0.070-0.111	0.061	10851.8
Three factors	71.03	32	0.941	0.072	0.050-0.095	0.047	10825.8
Wechsler four factors	43.66	29	0.978	0.047	0.011-0.074	0.038	10802.4
Higher-order	44.89	31	0.979	0.044	0.000-0.070	0.038	10799.9
Bi-factor[a]	36.35	27	0.986	0.039	0.000-0.068	0.034	10798.9

Note. CFA = confirmatory factor analytic; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike's information criterion; WISC-IV = Wechsler Intelligence Scale for Children–Fourth Edition.
[a]Two indicators of third and fourth factors were constrained to equality to allow model identification and the residual variance of the Block Design subtest was constrained greater than 0.

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 generally lower than the national average ($d \approx -0.33$) with consistent reading and math scores. 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 unremarkable (Curran, West, & Finch, 1996), but multivariate normality was rejected (χ^2 = 92.81, $df = 20$, $p < .001$; 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 indices, the correlated four-factor, higher-order, and bi-factor models were all good fits to the data. The four factors were highly correlated (*Md* = .68) in the oblique first-order model, suggesting the presence of a general factor (Gorsuch, 1983), making that model an inadequate explanation of the data. In addition, the bi-factor model estimated a residual variance for the Block Design subtest that was less than 0. When the offending residual was constrained to be above 0, several individual parameters were not statistically significant. Given these difficulties, the bi-factor model was removed from consideration (Kline, 2011). Thus, the higher-order model was selected as the best explanation of the WISC-IV factor structure among students with ADHD (see Figure 1).

Table 3 presents decomposed WISC-IV subtest variance estimates based on the higher-order model illustrated in

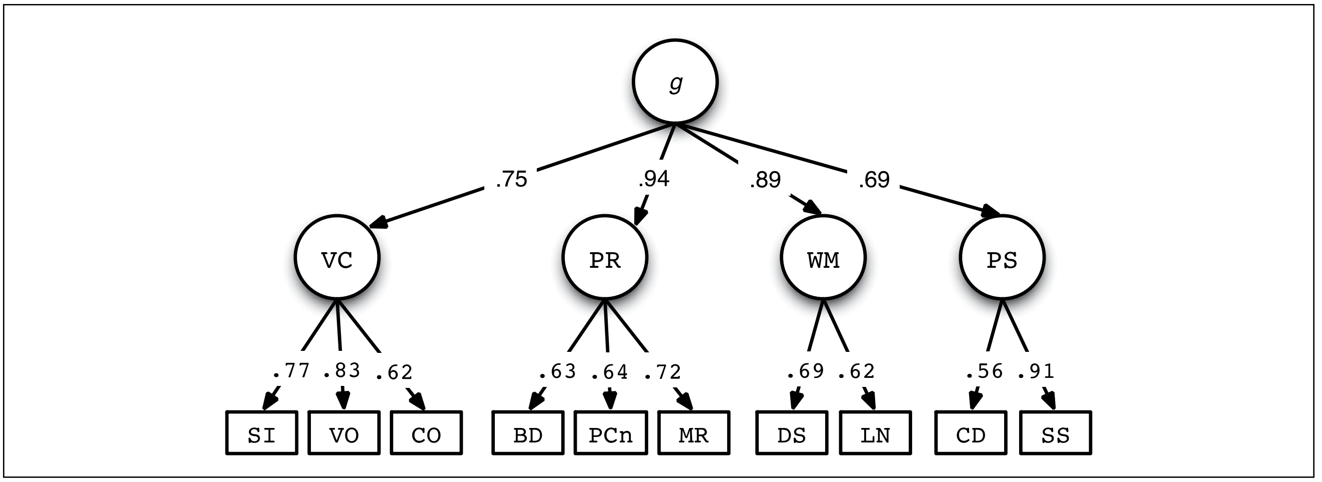

Figure 1. Higher-order structure of the WISC-IV for 233 students with ADHD.
Note. g = general intelligence; VC = Verbal Comprehension Factor; PR = Perceptual Reasoning Factor; WM = Working Memory Factor; PS = Processing Speed Factor; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search.

Table 3. Sources of Variance in the WISC-IV Among 233 Children With ADHD.

	General		VC		PR		WM		PS			
Subtest	*b*	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var	h^2	u^2
SI	.579	33.5	.511	26.1							0.596	0.404
VO	.623	38.8	.549	30.1							0.689	0.311
CO	.461	21.3	.406	16.5							0.378	0.622
BD	.586	34.3			.224	5.0					0.393	0.607
PCn	.597	35.6			.228	5.2					0.408	0.592
MR	.673	45.3			.255	6.5					0.518	0.482
DS	.614	37.7					.311	9.7			0.474	0.526
LN	.554	30.7					.283	8.0			0.387	0.613
CD	.385	14.8							.400	16.0	0.308	0.692
SS	.629	39.6							.653	42.6	0.822	0.178
% Total variance		33.16		7.27		1.67		1.77		5.86	49.7	50.3
% Common variance		66.67		14.62		3.36		3.55		11.79	100	
ω	.879		.786		.701		.601		.711			
ω_h	.781		.344		.089		.124		.369			

Note. WISC-IV = Wechsler Intelligence Scale for Children–Fourth Edition; VC = Verbal Comprehension Factor; PR = Perceptual Reasoning Factor; WM = Working Memory Factor; PS = Processing Speed Factor; h^2 = Communality; u^2 = Uniqueness; *b* = standardized loading of subtest on factor; Var = percent variance explained in the subtest; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search.

Figure 1. The general factor accounted for 66.7% of the common variance and 33.2% of the total variance, the VC factor accounted for 14.6% of the common variance and 7.3% of the total variance, the PR factor accounted for 3.4% of the common variance and 1.7% of the total variance, the WM factor accounted for 3.6% of the common variance and 1.8% of the total variance, and the PS factor accounted for 11.8% of the common variance and 5.9% of the total variance (see Table 3). Thus, the *g* factor accounted for about twice as much common and total variance as the domain-specific factors (66.7% vs. 33.3% and 33.2% vs. 16.6%, respectively). The general factor explained more variance in all subtests than their respective first-order factors with the exception of Coding and Symbol Search, where the PS factor and *g* factor were equally influential. In addition, communality was lower than uniqueness for 6 of the 10 WISC-IV subtests.

The omega and omega hierarchical coefficients presented in Table 3 are model-based 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 .09 (PR) to .37 (PS). These values suggest that "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" (Reise, 2012, p. 691). In contrast, the *g* factor exhibited a ω_h coefficient of .78.

Discussion

The present investigation empirically evaluated the structural validity of the WISC-IV with a sample of students diagnosed with ADHD by school multidisciplinary evaluation teams. Results indicated that a higher-order four-factor model with four domain-specific factors matching those delineated in Wechsler (2003b) offered a good fit to the data compared with multiple competing models. Thus, these results support the structural validity of the WISC-IV for a sample of students diagnosed with ADHD.

Particularly noteworthy is the amount of variance explained by each domain-specific factor over and above the general intelligence factor. The general intelligence factor contributed 33.2% of the total subtest variance, whereas the combined domain-specific factors only contributed an additional 16.5%. This indicates that the domain-specific factors accounted for less than half of the variance attributed to the general intelligence factor. The second largest portion of the total variance was contributed by the verbal comprehension factor, which amounted to a mere 7.3%. Moreover, omega hierarchical coefficients indicated that the general intelligence factor offered the highest estimate of reliability (ω_h = 0.78) with estimates for the remaining four domain-specific factors falling well below conventional notions of acceptability (Reise et al., 2013).

Some researchers claim that interpretation of WISC-IV index scores can yield clinically meaningful information that can be used to make diagnostic decisions and aid intervention planning (Flanagan, Ortiz, & Alfonso, 2013; Schwean & McCrimmon, 2008). However, results of the present study are consistent with previous research investigating the structural validity of the WISC-IV for clinical samples that strongly indicates interpretation should remain at the FSIQ score level (Bodin et al., 2009; Canivez, 2014; Devena et al., 2013; Golay et al., 2013; Nakano & Watkins, 2013; Styck & Watkins, 2014; Thaler et al., 2012; Watkins, 2010; Watkins et al., 2013). The general intelligence factor contributes the most reliable information for the WISC-IV, which is approximated by the FSIQ score. A basic tenet of measurement holds that "without some degree of consistency, the issue of accuracy may not be germane" (Viswanathan, 2005, p. 6). Given that empirical research on the WISC-IV has repeatedly demonstrated that the domain-specific factors contribute inconsequential information that is imprecise (Canivez, 2014; Devena et al., 2013; Styck & Watkins, 2014; Watkins et al., 2013), what information can the WISC-IV index scores provide that is accurate and of clinical importance?

The present study included a homogeneous sample of students diagnosed with ADHD. However, school multidisciplinary evaluation teams were responsible for making eligibility decisions and these decisions may not adhere to current diagnostic criteria outlined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; APA, 2013). Therefore, results may not generalize to samples outside of school-based settings. Furthermore, epidemiologic studies suggest that most children are diagnosed with ADHD-Predominantly Inattentive Type, with the prevalence of children diagnosed with ADHD-Predominantly Hyperactive-Impulsive Type or ADHD-Combined Type approximately equal (Froehlich et al., 2007). It is possible that the underlying factor structure of the WISC-IV may differ as a function of the particular subtype of ADHD diagnosis present (Greven, Rijsdijk, & Plomin, 2011; McKee, 2012).

An additional limitation of the present study includes the absence of information regarding the medication status of participants. It is possible that the factor structure of the WISC-IV may be different for children and adolescents with ADHD who are taking medication compared with those who are not taking medication. Future research on the structural validity of the WISC-IV should empirically evaluate the impact of these and other variables. Limitations notwithstanding, results of the present study corroborate previous research on the structure of the WISC-IV with clinical samples (Bodin et al., 2009; Canivez, 2014; Devena et al., 2013; Nakano & Watkins, 2013; Styck & Watkins, 2014; Watkins, 2010) and add support for the structural validity of the test with students diagnosed with ADHD (Thaler et al., 2012; P. Yang et al., 2013).

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

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Arizona Administrative Code 7 A.A.C. § 2-4

Bauermeister, J. J., Shrout, P. E., Ramírez, R., Bravo, M., Alegría, M., Martínez-Taboas, A., . . .Canino, G. (2007). ADHD correlates, comorbidity, and impairment in community and treated samples of children and adolescents. *Journal of Abnormal Child Psychology*, *35*, 883-898. doi:10.1007/s10802-007-9141-4

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

Braden, J. P. (2013). Psychological assessment in school settings. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: Assessment psychology* (2nd ed., Vol. 10, pp. 291-314). Hoboken, NJ: John Wiley.

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

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

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

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

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

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

Froehlich, T. E., Lanphear, B. P., Epstein, J. N., Barbaresi, W. J., Katusic, S. K., & Kahn, R. S. (2007). Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of US children. *Archives of Pediatrics & Adolescent Medicine*, *161*, 857-864.

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.

Greven, C. U., Rijsdijk, F. V., & Plomin, R. (2011). A twin study of ADHD symptoms in early adolescence: Hyperactivity-impulsivity and inattentiveness show substantial genetic overlap but also genetic specificity. *Journal of Abnormal Child Psychology*, *39*, 265-275. doi:10.1007/s10802-010-9451-9

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

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400.

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.

Loevinger, J. (1957). Objective tests as instruments of psychological theory [Monograph]. *Psychological Reports*, *3*(Pt. 9), 635-694.

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.

McKee, T. E. (2012). Examining the dimensionality of ADHD symptomatology in young adults using factor analysis and outcome prediction. *Journal of Attention Disorders*, *16*, 427-437. doi:10.1177/1087054710396039

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: Author.

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

National Center for Education Statistics. (2012, February 3). *Elementary/secondary information system* [database]. Available from http://nces.ed.gov

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

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.

Schwean, V. L., & McCrimmon, A. (2008). Attention-deficit/hyperactivity disorder: Using the WISC-IV to inform intervention planning. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (pp. 193-215). Burlington, MA: Elsevier.

Schwean, V. L., & Saklofske, D. H. (1998). WISC-III assessment of children with attention deficit/hyperactivity disorder. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 92-118). San Diego, CA: Academic Press.

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

Styck, K. M., & Watkins, M. W. (2014). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities*. Advance online publication. doi:10.1177/0022219414539565

Thaler, N. S., Barchard, K. A., Parke, E., Jones, W. P., Etcoff, L. M., & Allen, D. N. (2012). Factor structure of the Wechsler Intelligence Scale for Children: Fourth Edition in children with ADHD. *Journal of Attention Disorders*. Advance online publication. doi:10.1177/1087054712459952

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.

Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: SAGE.

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. (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-IVUK with a large referred Irish sample. *International Journal of School & Educational Psychology*, *1*, 102-111. doi:10.1080/21683603.2013.794439

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

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

Wechsler, D. (2007). *Wechsler Intelligence Scale for Children–Fourth Edition—Chinese version*. Taipei: The Chinese Behavioral Science 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*, 114-131. doi:10.1177/0734282913478032

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

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

Author Biographies

Kara M. Styck is an assistant professor in the School Psychology program at the University of Texas at San Antonio. Her research interests involve individual differences in child psychopathology, psychometrics, and professional issues in school psychology.

Marley W. Watkins 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.