

Structural and Incremental Validity of the Wechsler Adult Intelligence Scale–Fourth Edition With a Clinical Sample

Jason M. Nelson

University of Georgia Regents' Center for Learning Disorders

Gary L. Canivez

Eastern Illinois University

Marley W. Watkins

Baylor University

Structural and incremental validity of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV; Wechsler, 2008a) was examined with a sample of 300 individuals referred for evaluation at a university-based clinic. Confirmatory factor analysis indicated that the WAIS–IV structure was best represented by 4 first-order factors as well as a general intelligence factor in a direct hierarchical model. The general intelligence factor accounted for the most common and total variance among the subtests. Incremental validity analyses indicated that the Full Scale IQ (FSIQ) generally accounted for medium to large portions of academic achievement variance. For all measures of academic achievement, the first-order factors combined accounted for significant achievement variance beyond that accounted for by the FSIQ, but individual factor index scores contributed trivial amounts of achievement variance. Implications for interpreting WAIS–IV results are discussed.

Keywords: intelligence, Wechsler Adult Intelligence Scale–Fourth Edition, WAIS–IV, factor analysis

Slightly over a decade after the introduction of its predecessor, the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV; Wechsler, 2008a) was published and provided an update to what has historically been the most frequently used test for measuring adolescent and adult intelligence (Stinnett, Havey, & Oehler-Stinnett, 1994; C. E. Watkins, Campbell, Nieberding, & Hallmark, 1995). Relative to the Wechsler Adult Intelligence Scale–Third Edition (WAIS–III; Wechsler, 1997), 53% of the WAIS–IV items are new (Sattler & Ryan, 2009). Among many changes to the core battery, perhaps the most substantive include replacement of the Verbal IQ and Performance IQ with the Verbal Comprehension Index and Perceptual Reasoning Index, deletion of the Picture Arrangement and Object Assembly subtests, addition of the Visual Puzzles subtest, and required administration of 10 subtests to obtain the Full Scale IQ (FSIQ) rather than 13 (for extensive discussions of additional changes, see Sattler & Ryan, 2009; Wechsler, 2008b).

Tests with substantive changes such as those made to the WAIS–IV may result in the measurement of different constructs compared to previous test versions (Strauss, Spreen, & Hunter, 2000). Therefore, investigation of the internal structure of the WAIS–IV is necessary for determining the constructs it measures

because it cannot be assumed they are the same constructs measured by the WAIS–III. To that end, several empirical investigations of the internal structure of the WAIS–IV were conducted both prior to and after its publication. Prior to the publication of the WAIS–IV, confirmatory factor analyses (CFAs) were conducted to assess the structural validity of its scores. Six models were examined (see Wechsler, 2008b, for a discussion of these models) and included both first-order and second-order factor models. The first-order four-factor model, with factors labeled Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed (PS), was found to fit the standardization data well and to have a superior fit to models with fewer first-order factors. A second-order model with general intelligence as the second-order factor and the above-mentioned four factors as first-order factors was also examined and favored over the first-order four-factor model despite resulting in a slightly inferior fit. Referring to this decrease in fit, the test authors stated, “This is expected because the fit of a second-order model can never exceed the fit of the corresponding first-order model” (Wechsler, 2008b, p. 66).

Following publication of the WAIS–IV, several independent researchers conducted structural validity analyses with the standardization data. Using the 10 core and five supplemental WAIS–IV subtests, Benson, Hulac, and Kranzler (2010) conducted CFA to compare the model favored by the test authors to various models they argued were more aligned with the Cattell–Horn–Carroll (CHC) structure of intelligence. Results indicated that a CHC-inspired structure was a better fit to the data than the model favored by the test authors. More specifically, they found a five-factor model with factors labeled Crystallized Intelligence, Visual Processing, Fluid Reasoning, Short-Term Memory, and Processing Speed to have a significantly superior fit than the four-factor

This article was published Online First April 1, 2013.

Jason M. Nelson, Regents' Center for Learning Disorders, University of Georgia; Gary L. Canivez, Department of Psychology, Eastern Illinois University; Marley W. Watkins, Department of Educational Psychology, Baylor University.

Correspondence concerning this article should be addressed to Jason M. Nelson, Regents' Center for Learning Disorders, University of Georgia, 337 Milledge Hall, Athens, GA 30602. E-mail: jmnelson@uga.edu

model proposed by the test authors. Also using the full standardization data, Ward, Bergman, and Hebert (2012) found that the four-factor model with three orthogonal minor factors they labeled Spatial Visualization, Quantitative Reasoning, and Digit-Letter Memory Span was the best fitting and most theoretically defensible model.

Benson et al. (2010) also conducted factorial invariance analyses to determine whether the factor structure of the WAIS-IV remained consistent across age groups. They found differences in the magnitude of factor loadings across various age cohorts of the standardization sample. Other measurement invariance studies with the WAIS-IV have been conducted in which factorial invariance across the United States and Canadian standardization samples using data from the core subtests alone (Bowden, Saklofske, & Weiss, 2011b) and data from the core and supplemental subtests together (Bowden, Saklofske, & Weiss, 2011a) was investigated. The baseline model estimation phases of both studies indicated the four-factor model reported in the WAIS-IV *Technical and Interpretive Manual* (Wechsler, 2008b) was better fitting than models with fewer factors, although it should be noted that these models were slightly modified from those in the technical manual due to the inclusion of joint loadings and correlated residuals. Further examination indicated invariance of the four-factor model across the samples using the core and supplemental subtests.

A final set of studies of the factor structure of the WAIS-IV involved exploratory factor analyses of data from the total standardization sample (Canivez & Watkins, 2010b) and data from the adolescent subsample of the standardization sample (Canivez & Watkins, 2010a), including transformation via the Schmid and Leiman (1957) procedure. Incorporation of various factor extraction criteria generally resulted in the recommended extraction of fewer than four factors, with most resulting in the recommended extraction of one or two factors. In both studies, hierarchical analyses with the Schmid and Leiman (1957) procedure indicated that the second-order *g* factor accounted for large portions of common and total variance, whereas the first-order factors accounted for small portions of variance. Thus, Canivez and Watkins (2010a, 2010b) recommended that clinical interpretation of WAIS-IV scores be primarily focused on the second-order level of general intelligence rather than the first-order factors as advocated in the WAIS-IV *Technical and Interpretive Manual* (Wechsler, 2008b).

To date, all structural validity investigations of the WAIS-IV have used the standardization data of the WAIS-IV, and no studies have been conducted in which the WAIS-IV factor structure has been examined with data from a clinical sample. As acknowledged by Ward et al. (2012), examination of the WAIS-IV factor structure using data from a clinical sample could result in significantly different findings than investigation using the standardization data. The importance of further examination of the WAIS-IV using clinical samples has been emphasized in the extant literature (Bowden et al., 2011a; Holdnack, Zhou, Larrabee, Mills, & Salt-house, 2011). Additionally, replication of factor models in independent samples is an important step in thoroughly investigating structural validity (Raykov & Marcoulides, 2000). It is particularly important to investigate the structural validity in an independent sample using the 10 core subtests because clinicians typically do not administer all 15 subtests (Canivez & Watkins, 2010b), and

there are no composite score norms provided by the publisher based on the supplemental subtests.

Determining the best fitting model among several rival theoretically based models leads to a better understanding of structural validity and, in turn, to the most appropriate approach to test interpretation (Kline, 2005). One rival model for the WAIS-IV that has yet to be empirically investigated is the direct hierarchical (also called nested factors or bifactor) model. The authors of the WAIS-IV proposed that an indirect hierarchical model with a general intelligence (*g*) factor at the structure's apex and four first-order factors below was the best fitting and most theoretically defensible model. Gignac (2005, 2006, 2008) criticized this model on the grounds that it is unrealistic to assume that *g*'s influence on subtest performance is fully mediated by the first-order factors and therefore has no direct influence. In contrast, no such assumption is made within the direct hierarchical model, where both the *g* factor and the first-order factors are modeled orthogonally and therefore both directly influence subtest performance (Gignac, 2005). A direct hierarchical model has been found to be better fitting than an indirect hierarchical model in several factor-analytic studies of the Wechsler scales, including the WAIS-Revised (Gignac, 2005), WAIS-III (Gignac, 2006), Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; M. W. Watkins, 2010), and French version of the WISC-IV (Golay, Reverte, Rossier, Favez, & Lecerf, 2012). Comparing direct and indirect hierarchical models of the WAIS-IV would likely provide guidance regarding the most appropriate conceptualization of the *g* factor.

Research on the incremental validity of WAIS-IV scores would also likely aid in interpretive decision making (Sechrest, 1963). Structural validity studies alone are insufficient for adequately determining the importance of the FSIQ relative to the more narrowly constructed factor index scores (Canivez, Konold, Collins, & Wilson, 2009). *Incremental validity* has been defined as the "extent to which a measure adds to the prediction of a criterion beyond what can be predicted with other data" (Hunsley, 2003, p. 443). In the case of the WAIS-IV, it would be informative to know the incremental validity of the four factor index scores in predicting external criteria (e.g., academic achievement) beyond the FSIQ. No analyses of incremental validity were reported in the WAIS-IV *Technical and Interpretive Manual*, but zero-order correlations were reported between the WAIS-IV IQ and index scores and scores on the Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler, 2001) based on a sample of ninety-three 16- to 19-year-olds tested during the WAIS-IV standardization. Correlations ranged from the .50s to the .70s for the FSIQ and factor index scores. Similar zero-order correlations between the WAIS-IV and the Wechsler Individual Achievement Test-Third Edition (WIAT-III; NCS Pearson, 2009) from a study of fifty-nine 16- to 19-year-olds administered these measures during the WIAT-III standardization were also reported in the WIAT-III *Technical and Interpretive Manual*. These two data sets were obtained and reanalyzed by Canivez (in press) to assess the incremental validity of the WAIS-IV factor index scores. The WAIS-IV FSIQ accounted for statistically significant and generally large portions of WIAT-II and WIAT-III subtest and composite score variance, whereas the WAIS-IV factor index scores combined to predict statistically significant but trivial to medium increments in achievement scores. None of the individual index scores uniquely predicted significant variance beyond that pre-

dicted by the four index scores combined. The conclusion was that primary interpretation of WAIS-IV scores should be of the FSIQ as suggested by Canivez and Watkins (2010a, 2010b). The WAIS-IV factor index scores demonstrated greater incremental validity than the WISC-IV factor index scores (Glutting, Watkins, Konold, & McDermott, 2006) with these two small nonclinical samples, but replication is needed, and the incremental validity of WAIS-IV factor index scores among referred samples is unknown. If primary interpretation of the four factor index scores promoted by the publisher is to be followed, then the four factor index scores must demonstrate meaningful incremental validity beyond the FSIQ.

Purpose of Study and Research Questions

The purpose of the current study was to examine the structural and incremental validity of the WAIS-IV with a clinical sample. Three research questions were investigated: (a) What is the best fitting structural model among the four first-order models (i.e., one-factor, two-factor, three-factor, and four-factor models) reported in the WAIS-IV *Technical and Interpretive Manual*? (b) What is the best fitting hierarchical model—the direct or indirect hierarchical model? and (c) What is the incremental predictive validity of the four WAIS-IV factor index scores in predicting basic academic skills, academic fluency, and higher level academic skills beyond the FSIQ?

Method

Participants

Participants ($N = 300$) were individuals who sought comprehensive psychoeducational evaluations at a university-based clinic staffed by psychologists specializing in the assessment of learning disorders. All individuals who were administered the WAIS-IV as part of their evaluation between 2009 and 2012 were included in the study. Most participants were undergraduates attending four-year colleges or universities (75%; $n = 225$); the remainder of participants were high school students transitioning to college (13%; $n = 39$), students attending two-year technical schools (1.67%; $n = 5$), graduate students (6.33%; $n = 19$), and a small number of students who could not be classified (4%; $n = 12$). Participants ranged in age from 16 to 61 years ($M = 22.36$, $SD = 7.56$). Most participants in the sample were White (86%; $n = 258$); the remaining participants were African American (8.33%; $n = 25$), Hispanic (2.33%; $n = 7$), Asian American (1.66%; $n = 5$), American Indian (.3%; $n = 1$), multiracial (.3%; $n = 1$), and other (1%; $n = 3$). Diagnoses were wide-ranging, but the most common diagnoses were learning disabilities (LD) and attention-deficit/hyperactivity disorder (ADHD). Diagnostic status included 32% LD ($n = 96$), 34.67% ADHD ($n = 104$), 17.67% comorbid LD/ADHD ($n = 53$), and 15.67% other ($n = 47$).

Instruments

WAIS-IV. The WAIS-IV is an individually administered intelligence test for individuals between the ages of 16 and 90 years. It consists of 10 core and five supplemental subtests. Only the core subtests were administered for the current study. Four index scales

(VC, PR, WM, and PS) are derived from the 10 core subtests. The specific subtests contributing to each index include Similarities, Vocabulary, and Information on the VC Index; Block Design, Matrix Reasoning, and Visual Puzzles on the PR Index; Digit Span and Arithmetic on the WM Index; and Symbol Search and Coding on the PS Index. The indexes are also combined to derive the FSIQ. Average internal consistency estimates ranged from .78 to .94 for subtests, from .90 to .96 for factor index scores, and .98 for the FSIQ. Various validity estimates are also presented in the *Technical and Interpretive Manual* (Wechsler, 2008b).

Instruments for examining incremental validity. Basic academic skills, academic fluency, and higher level academic skills served as external criteria for the incremental validity analyses. The Woodcock-Johnson III Tests of Achievement Normative Update Form A (WJ-III; Woodcock, McGrew, & Mather, 2007) and the Nelson-Denny Reading Test Form H (NDRT; Brown, Fishco, & Hanna, 1993a) were used to quantify these academic skills.

Basic academic skills. The WJ-III Academic Skills Cluster served as an overall indicator of basic academic skills. It consists of the Letter-Word Identification, Calculation, and Spelling subtests. Letter-Word Identification was designed to measure skill at reading words in isolation. Examinees solve paper-and-pencil math computation problems ranging from simple addition to calculus on the Calculation subtest. On the Spelling subtest, examinees are asked to spell progressively more difficult words. The median Academic Skills Cluster score reliability coefficient was .96. Median split-half reliability estimates were .94, .86, and .90 for Letter-Word Identification, Calculation, and Spelling scores, respectively. Additional evidence for validity of scores from these measures and the measures presented below is available in the WJ-III technical manual (McGrew, Schrank, & Woodcock, 2007) and NDRT technical report (Brown, Fishco, & Hanna, 1993b).

Academic fluency. The WJ-III Reading Fluency, Math Fluency, and Writing Fluency subtests along with the associated composite score, the Academic Fluency Cluster, were used to measure academic fluency. On the Reading Fluency subtest, examinees are asked to read simple sentences as quickly as they can and decide if each statement is true. Math Fluency was designed to measure how quickly examinees can solve simple arithmetic problems. Examinees write simple, short sentences as quickly as they can on the Writing Fluency subtest. The median Academic Fluency Cluster score reliability coefficient was .96. Median split-half reliability estimates for scores from the Reading Fluency, Math Fluency, and Writing Fluency subtests were .95, .98, and .83, respectively.

Higher level academic skills. Higher level academic skills investigated in the current study included math reasoning and reading comprehension. The WJ-III Applied Problems subtest was used to measure math reasoning. On this subtest, examinees are required to apply mathematics to real-life situations. The median split-half reliability for Applied Problems scores was .93. The NDRT Comprehension scale was used to measure reading comprehension. Examinees are presented with eight passages and 38 multiple-choice questions and asked to answer as many questions as they can within a 20-min period. Kuder-Richardson 20 coefficients ranged from .85 to .89 for NDRT scores.

Procedure

Data were archival and extracted from a database used to store de-identified assessment information for all individuals evaluated at the clinic. Instruments were individually administered by licensed psychologists or unlicensed psychologists, master’s-level clinicians, or doctoral students in clinical psychology under the supervision of a licensed psychologist.

Data Analyses

Mplus 7.0 for Macintosh was used to conduct CFA with maximum likelihood estimation. Consistent with previous WAIS-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, nonverbal, and combined working memory/processing speed factors; (d) four oblique verbal, nonverbal, working memory, and processing speed factors; (e) an indirect hierarchical model (as per Bodin, Pardini, Burns, & Stevens, 2009) with four correlated first-order factors; and (f) a direct hierarchical model (as per M. W. Watkins, 2010) with four first-order factors. See Gignac (2008) for a detailed description of direct and indirect hierarchical models.

To judge model fit, Hu and Bentler (1998, 1999) recommended a dual criterion to guard against both Type I and Type II errors with comparative fit index (CFI) values greater than or equal to .95 and root-mean-square error of approximation (RMSEA) values less than or equal to .06. Higher CFI values and lower RMSEA values indicate better fit. Chi-square and Akaike information criterion (AIC) values were also considered. Nonstatistically significant chi-square values tend to indicate good model fit, and smaller AIC values indicate better fit after accounting for model complexity. Meaningful differences between well-fitting models were evaluated with $\Delta CFI > .01$ (Cheung & Rensvold, 2002), $\Delta RMSEA > -.015$ (Chen, 2007), and smallest AIC as standards. Latent factor reliabilities were estimated with coefficient omega (ω) and omega hierarchical (ω_h) as programmed by M. W. Watkins (2013). Omega estimated the reliability of the latent factor combining general and specific factor variance, whereas ω_h (what Reise, 2012, termed omega subscale) estimated the reliability of each latent factor with variance from other factors removed (Brunner, Nagy, & Wilhelm, 2012).

As suggested by Lubinski (2000), hierarchical multiple regression analyses were conducted to assess proportions of observed achievement subtest or composite score variance predicted by the observed WAIS-IV Full Scale IQ and factor index scores. Multiple regression analyses are appropriate for investigations that are predictive rather than explanatory in focus (Glutting et al., 2006; Pedhazur, 1997). The purpose of the incremental validity aspect of the current study was to investigate the applied usage of the WAIS-IV in predicting academic achievement. Therefore, the observed scores derived from the instruments (i.e., the scores practitioners have access to) were used rather than latent construct scores that can be derived from structural equation modeling. Although the latter approach is useful for explanatory research and testing theory, its results do not have direct practical applications (Glutting et al., 2006; Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004) and therefore was deemed inconsistent with the purpose of the current study.

The WAIS-IV FSIQ was singularly entered into the first block and the four WAIS-IV factor index scores were jointly entered into the second block in SPSS Version 20 for Mac linear regression analysis. The WJ-III subtest, WJ-III cluster, and NDRT scores served as dependent variables. The change in predicted achievement test score variance provided by the four WAIS-IV factor index scores in the second block provided an estimate of the incremental prediction beyond the WAIS-IV FSIQ from the first block. Unique incremental contribution of factor index score prediction of achievement test variance was estimated by the squared part correlations from the second block. Cohen’s (1988) criteria for effect sizes (small effect $R^2 = .03$ [3%], medium effect $R^2 = .10$ [10%], large effect $R^2 = .30$ [30%]) were used to evaluate effect size estimates.

Order of variable entry in multiple regression analysis influences variance attributed to predictors, as variables entered first capture greater criterion variance than variables entered later. Some have suggested entry of factor index scores into Block 1 and the FSIQ into Block 2 and reporting the incremental validity of the FSIQ above and beyond the first-order factors (Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007). This was not done in the present study because, as Glutting et al. (2006, p. 106) noted, such a procedure would “repeal [the] scientific law” of parsimony and would result in testing “a nonsensical hypothesis” (Schneider, 2008, p. 52).

Results

Descriptive statistics for the WAIS-IV and achievement measures are presented in Tables 1 and 2 and illustrate univariate normality. Mardia’s (1970) standardized multivariate kurtosis estimate for the WAIS-IV data was 2.77 and well under the criterion of 15.01 for multivariate normality (Byrne, 2006). With the high communalities and normality of these data, a sample size of 300 was deemed adequate for subsequent CFA analyses (Mundfrom & Shaw, 2005; Raykov & Marcoulides, 2000). WAIS-IV means for this sample were approximately 1 standard deviation lower than

Table 1
Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV)
Descriptive Statistics for 300 Referred College Students

Score	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Similarities	10.63	2.74	0.22	0.11
Vocabulary	10.79	2.90	0.00	-0.26
Information	10.23	2.89	0.31	-0.32
Block Design	9.41	3.15	0.32	-0.48
Matrix Reasoning	10.52	2.48	-0.29	0.49
Visual Puzzles	9.91	2.72	0.27	-0.48
Digit Span	9.00	2.66	0.53	0.29
Arithmetic	8.88	2.69	0.27	-0.66
Symbol Search	9.10	2.66	0.39	0.00
Coding	8.73	2.57	0.48	0.45
Verbal Comprehension Index	102.92	13.62	0.44	0.48
Perceptual Reasoning Index	99.41	13.54	0.04	-0.22
Working Memory Index	93.91	13.34	0.37	-0.11
Processing Speed Index	94.02	12.51	0.40	0.03
Full Scale IQ	97.74	12.00	0.09	-0.22

Note. Mardia’s (1970) normalized multivariate kurtosis estimate for WAIS-IV subtests was 2.77.

Table 2
Descriptive Statistics for Achievement Measures

Score	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Letter–Word Identification	94.46	10.13	−0.40	0.36
Calculation	96.56	13.71	−0.03	−0.54
Spelling	97.81	13.05	−0.25	−0.50
Academic Skills Cluster	95.38	12.22	−0.21	−0.17
Reading Fluency	92.96	12.66	0.30	1.13
Math Fluency	87.97	15.66	0.06	0.18
Writing Fluency	95.78	13.68	0.26	0.75
Academic Fluency Cluster	91.47	12.63	0.02	0.05
Applied Problems	95.11	10.53	0.44	0.20
Nelson–Denny Reading Test	193.14	25.55	0.13	−0.85

those typically obtained by college students (Beaujean, Firmin, Michonski, Berry, & Johnson, 2010; Lassiter, Bell, Hutchinson, & Matthews, 2001), and less variability was observed among the current sample than the standardization sample. Generally, lower scores on subtests, factor indices, and the FSIQ are observed in referred samples (Canivez & Watkins, 1998; M. W. Watkins, 2010). Not surprisingly, Table 1 illustrates that scores were lower for working memory and processing speed; previous research has indicated that individuals with ADHD and LD, on average, score lower on these factor indices than those without these disorders (Frazier, Demaree, & Youngstrom, 2004; Swanson & Hsieh, 2009; Trainin & Swanson, 2005).

Structural Validity

Model fit statistics presented in Table 3 illustrate the increasingly better fit from one to four factors; however, fit statistics indicated that the one-, two-, and three-factor models were clearly inadequate. The correlated four-factor (see Figure 1), indirect hierarchical (see Figure 2), and direct hierarchical (see Figure 3) models were all statistically better fits to these data than the one-, two-, and three-factor models. Of the three better models, the correlated four-factor model was a better fit to these data than the indirect hierarchical model ($\Delta\chi^2 = 21.48, p < .001$), as was the direct hierarchical model ($\Delta\chi^2 = 24.09, p < .001$). There were no statistically significant or meaningful differences in fit statistics between the correlated four-factor model and the direct hierarchical model ($\Delta\chi^2 = 2.61, p > .10, \Delta CFI = .001, \Delta RMSEA = .003$).

Table 3
Confirmatory Factor Analysis Fit Statistics for Wechsler Adult Intelligence Scale–Fourth Edition for 300 Referred College Students

Model	χ^2	<i>df</i>	CFI	RMSEA	90% CI RMSEA	AIC	ΔCFI	$\Delta RMSEA$
First-order models								
One factor	391.13	35	.683	.184	[.168, .201]	13808.6		
Two factors (V and NV)	301.13	34	.760	.162	[.146, .179]	13716.5	.077	.022
Three factors (V, NV, WM/PS)	198.38	32	.852	.132	[.114, .150]	13609.1	.092	.030
Four factors	78.45	29	.956	.075	[.056, .096]	13488.6	.104	.057
Hierarchical models								
Indirect hierarchical	99.93	31	.939	.086	[.068, .105]	13506.9	.017	.011
Direct hierarchical ^a	75.84	27	.957	.078	[.078, .099]	13488.9	.018	.008

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; V = Verbal; NV = Nonverbal; WM/PS = Working Memory/Processing Speed.

^a Two indicators of third and fourth factors were constrained to be equal to ensure identification.

Because the four WAIS–IV latent factors were highly correlated and imply a higher order structure (Gorsuch, 1988), such explication is necessary and the reason the direct hierarchical model was judged to provide the best explanation of the WAIS–IV factor structure.

Table 4 illustrates the portions of variance based on the direct hierarchical model (see Figure 3). The general factor accounted for 32.9% of the total and 52.5% of the common variance. The VC factor accounted for 10.9% of the total variance and 17.4% of common variance, the PR factor accounted for 2.8% of the total variance and 4.5% of common variance, the WM factor accounted for 5.7% of the total variance and 9.2% of the common variance, and the PS factor accounted for 10.2% of the total variance and 16.4% of the common variance. Thus, the higher order *g* factor accounted for substantially greater portions of WAIS–IV common and total variance relative to the factor index scores. Table 4 presents ω_h coefficients that estimated the latent construct reliability with the effects of other constructs removed. In the case of the four WAIS–IV factor indices, ω_h coefficients ranged from .107 (PR) to .638 (PS).

Incremental Validity

Table 5 presents results from hierarchical multiple regression analyses for WJ–III and NDRT scores with the WAIS–IV FSIQ in the first block and the four factor index scores (Verbal Comprehension Index, Perceptual Reasoning Index [PRI], Working Memory Index [WMI], Processing Speed Index [PSI]) in the second block. R^2 values are reported in percent.

Basic academic skills. The WAIS–IV FSIQ accounted for statistically significant ($p < .0001$) portions of the WJ–III Letter–Word Identification (23.0%), Calculation (33.5%), and Spelling (14.4%) scores that represented medium to large effect sizes. Also presented in Table 5 are the increases in achievement variance provided by the combined effects of WAIS–IV factor index scores beyond the achievement variance due to the FSIQ. Statistically significant portions of WJ–III Letter–Word Identification (19.9%, $p < .0001$), Calculation (4.4%, $p < .001$), and Spelling (17.6%, $p < .0001$) score variance was incrementally accounted for by the WAIS–IV factor index scores that represented small to medium effect sizes. The unique contributions of WAIS–IV factor index scores in predicting the WJ–III Academic Skills subtests ranged

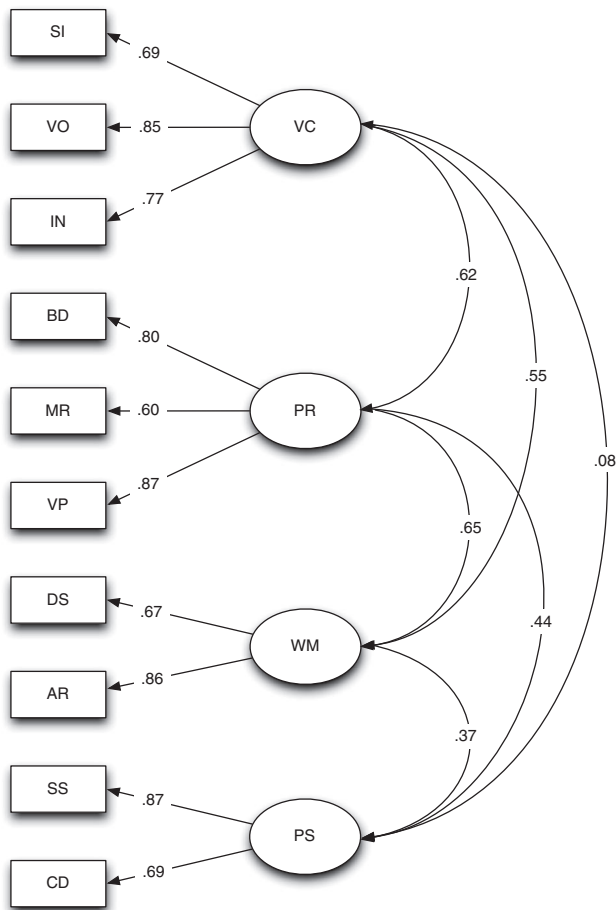


Figure 1. Correlated four-factor first-order measurement model, with standardized coefficients, for the Wechsler Adult Intelligence Scale–Fourth Edition for 300 referred college students. SI = Similarities; VO = Vocabulary; IN = Information; BD = Block Design; MR = Matrix Reasoning; VP = Visual Puzzles; DS = Digit Span; AR = Arithmetic; SS = Symbol Search; CD = Coding; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor.

from 0.0% (WMI & PSI in Letter–Word Identification) to 1.4% (WMI in Calculation).

The WAIS-IV FSIQ accounted for statistically significant ($p < .0001$) portions of the WJ-III Academic Skills Cluster (34.3%) that represented a large effect size. A statistically significant ($p < .0001$) portion of the WJ-III Academic Skills Cluster (16.2%) score variance was incrementally accounted for by the combined WAIS-IV factor index scores and represented a medium effect size. The unique contributions of WAIS-IV factor index scores in predicting the WJ-III Academic Skills Cluster ranged from 0.0% (PRI) to 0.9% (WMI).

Academic fluency. The WAIS-IV FSIQ accounted for statistically significant ($p < .0001$) portions of the WJ-III Reading Fluency (19.3%), Math Fluency (13.9%), and Writing Fluency (21.6%) scores that represented medium effect sizes. Table 5 also illustrates the increases in achievement variance provided by the combined effects of WAIS-IV factor index scores beyond the

achievement variance due to the FSIQ. Statistically significant portions of WJ-III Reading Fluency (17.8%, $p < .0001$), Math Fluency (29.9%, $p < .0001$), and Writing Fluency (9.1%, $p < .0001$) score variance was incrementally accounted for by the WAIS-IV factor index scores that represented small to medium effect sizes. The unique contributions of WAIS-IV factor index scores in predicting the WJ-III Academic Skills subtests ranged from 0.0% (PRI in Writing Fluency) to 2.2% (WMI in Math Fluency).

The WAIS-IV FSIQ accounted for statistically significant ($p < .0001$) portions of the Academic Fluency Cluster (28.0%) score that represented a medium effect size. A statistically significant ($p < .0001$) portion of the WJ-III Academic Fluency Cluster (22.8%) score variance was incrementally accounted for by the combined WAIS-IV factor index scores and represented a medium effect size. The unique contributions of WAIS-IV factor index scores in predicting the WJ-III Academic Fluency Cluster ranged from 0.1% (PRI) to 1.7% (PSI).

Higher level academic skills. The WAIS-IV FSIQ accounted for statistically significant ($p < .0001$) portions of the WJ-III Applied Problems (49.1%) and NDRT (29.2%) scores that represented large and medium effect sizes, respectively. Table 5 shows the increases in achievement variance provided by the combined effects of WAIS-IV factor index scores after achievement variance due to the FSIQ was accounted for with statistically significant ($p < .0001$) portions of achievement variance incrementally accounted for with the WJ-III Applied Problems (6.5%) and the NDRT (10.1%) scores, which represented small to medium effect sizes, respectively. The unique contributions of WAIS-IV factor index scores in predicting the WJ-III Applied Problems and NDRT scores ranged from 0.0% (PRI in the NDRT) to 0.8% (WMI in Applied Problems).

Discussion

The current study is the first to examine the factor structure of the WAIS-IV with data obtained from a sample that was not part of the standardization of the instrument. Of the four first-order factor models, the correlated four-factor model consisting of VC, PR, WM, and PS exhibited superior fit. Additionally, it was the only lower order model that produced adequate fit statistics. This finding replicates results reported in the WAIS-IV *Technical and Interpretive Manual* and a reexamination of the WAIS-IV standardization data on the 10 core subtests reported by Bowden et al. (2011b). Comparison to other empirical studies on the WAIS-IV factor structure examined via CFA could not be made because those studies (Benson et al., 2010; Bowden et al., 2011a; Ward et al., 2012) included the five supplemental subtests in their analyses. In the two other extant studies on the factor structure of the WAIS-IV (Canivez & Watkins, 2010a, 2010b), exploratory factor analysis was used. In those two studies, various factor extraction criteria suggested the extraction of no more than two factors and in most cases only one, although it was noted that the subtests were correctly associated with the four theoretically proposed first-order factors. In the current study, both the one-factor and two-factor models were poor fits to the data.

Results of the analysis comparing the direct and indirect hierarchical models indicated that the direct hierarchical model was the better fitting model. These results are consistent with those of

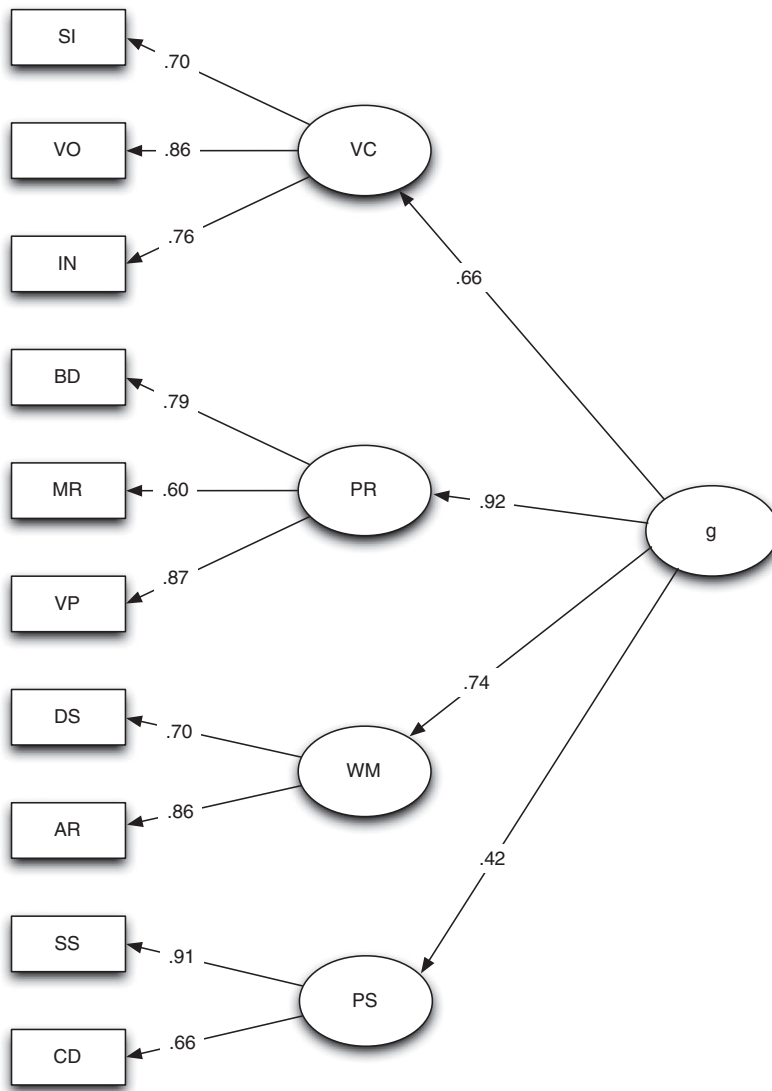


Figure 2. Indirect hierarchical measurement model, with standardized coefficients, for the Wechsler Adult Intelligence Scale–Fourth Edition for 300 referred college students. SI = Similarities; VO = Vocabulary; IN = Information; BD = Block Design; MR = Matrix Reasoning; VP = Visual Puzzles; DS = Digit Span; AR = Arithmetic; SS = Symbol Search; CD = Coding; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; g = general intelligence.

factor structure investigations of two previous versions of the WAIS (Gignac, 2005, 2006), a study of the WISC–IV (M. W. Watkins, 2010), and a study of the French WISC–IV (Golay et al., 2012). The present findings are at odds with those reported in the WAIS–IV *Technical and Interpretive Manual*, in which an indirect hierarchical model was presented as the favored model, although a direct hierarchical was apparently not examined or reported. Within the direct hierarchical model, the *g* factor was modeled as having a direct influence on subtest performance rather than an indirect influence fully mediated via the VC, PR, WM, and PS factors, as in the indirect hierarchical model. The direct hierarchical model has been described as a more defensible model than the indirect hierarchical model because the full mediation of the nar-

row group factors of the indirect model “should probably be considered unreasonable in the area of psychology, where test developers are not likely adept enough to develop subtests in such a way that they will share variance with the general factor to the extent that they will load onto a group-level factor” (Gignac, 2006, p. 85). Additionally, general intelligence is modeled as a *breadth* factor in a direct hierarchical model rather than as a *superordinate* factor in an indirect hierarchical model. Gignac (2008) argued that breadth rather than superordination has historically been a more fundamental aspect of general intelligence from a theoretical perspective.

Results of the analyses of the sources of variance according to a direct hierarchical model were similar to those of Canivez and Wat-

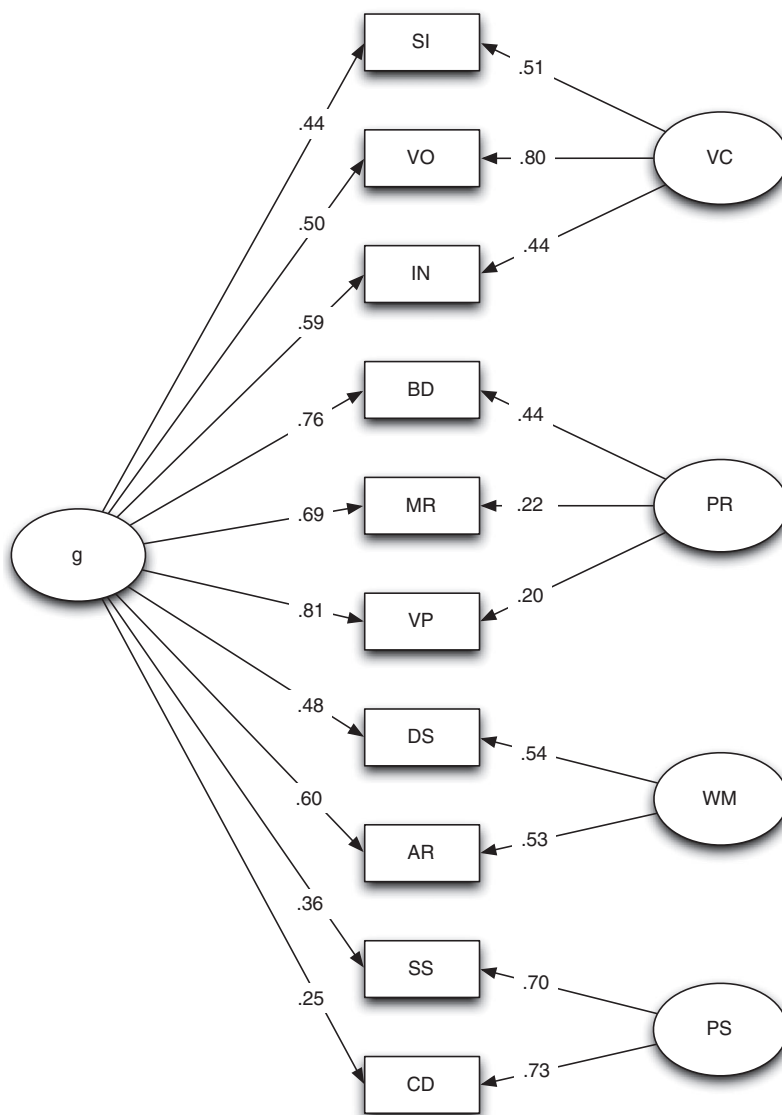


Figure 3. Direct hierarchical measurement model, with standardized coefficients, for the Wechsler Adult Intelligence Scale–Fourth Edition for 300 referred college students. SI = Similarities; VO = Vocabulary; IN = Information; BD = Block Design; MR = Matrix Reasoning; VP = Visual Puzzles; DS = Digit Span; AR = Arithmetic; SS = Symbol Search; CD = Coding; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; g = general intelligence.

kins (2010a, 2010b) and Niileksela, Reynolds, and Kaufman (2012) in that the *g* factor accounted for far more total and common variance than any of the four individual first-order factors. The *g* factor accounted for more total and common variance than that accounted for by the four first-order factors combined. These results are also consistent with those obtained from investigations of other intelligence tests, including studies of the Reynolds Intellectual Assessment Scales (Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007), Stanford-Binet Intelligence Scales–Fifth Edition (Canivez, 2008; DiStefano & Dombrowski, 2006), Cognitive Assessment System (Canivez, 2011), WISC–IV (Bodin et al., 2009; M. W. Watkins, 2006, 2010; M. W. Watkins, Wilson, Kotz, Carbone, & Babula, 2006), and French

WAIS–III (Golay & Lecerf, 2011). The four first-order factors in the present study accounted for more total (29.6%) and common (47.5%) variance than observed by Canivez and Watkins (2010a, 2012b) with the WAIS–IV standardization sample (20.4%–21.1% for total variance and 33%–35% for common variance). Additionally, the four first-order factors in the present study accounted for more total and common variance than that found in the WISC–IV (M. W. Watkins, 2010). In that study, total and common variance explained by the *g* factor was 3 times as large as that explained by the four first-order factors, which explained only 15.9% and 24.8% of the total and common variance, respectively.

The present analyses estimated the reliability of WAIS–IV latent factors with ω_h , with ω_h coefficients of .470 for VC, .107 for

Table 4
Sources of Variance in the Wechsler Adult Intelligence Scale–Fourth Edition for the Referred College Student Sample (N = 300)
According to the Confirmatory Factor Analysis Direct Hierarchical Model

Subtest	General		VC		PR		WM		PS		h^2	u^2
	b	Var	b	Var	b	Var	b	Var	b	Var		
Similarities	.435	.189	.509	.259							.448	.552
Vocabulary	.499	.249	.798	.637							.886	.114
Information	.587	.345	.441	.194							.539	.461
Block Design	.756	.572			.441	.194					.766	.234
Matrix Reasoning	.694	.482			.215	.046					.528	.472
Visual Puzzles	.814	.663			.200	.040					.703	.297
Digit Span	.478	.228					.539	.291			.519	.481
Arithmetic	.604	.365					.533	.284			.649	.351
Symbol Search	.362	.131							.702	.493	.624	.376
Coding	.254	.065							.729	.531	.596	.404
Total variance (%)	32.9		10.9		2.8		5.7		10.2		62.6	37.4
Common variance (%)	52.5		17.4		4.5		9.2		16.4		100	
ω	.908		.827		.854		.736		.757			
ω_h	.737		.470		.107		.365		.638			

Note. VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; b = standardized loading of subtest on factor; Var = variance explained in the subtest; h^2 = communality; u^2 = uniqueness; ω_h = omega hierarchical.

PR, .365 for WM, and .638 for PS. Lower coefficients were observed for the WAIS–IV standardization sample ages 70–90 years by calculating ω_h from Niileksela et al. (2012, Table 2), where ω_h for $G_c = .292$, $G_f = .000$, $G_v = .149$, $G_s = .350$, and $G_{sm} = .251$. Such low ω_h coefficients suggest that interpretation of the factor indices “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).

The prediction of academic achievement has been described as the most important application of intelligence tests (Brown, Reynolds, & Whitaker, 1999), and therefore the incremental predictive validity aspect of the current study using various dimensions of academic achievement has implications regarding the practical utility of the WAIS–IV. Results indicated that the FSIQ accounted for statistically significant portions of all WJ–III cluster and subtest scores investigated as well as NDRT scores. The FSIQ accounted for a large portion of WJ–III Academic Skills Cluster scores, a medium portion of WJ–III Academic Fluency Cluster scores, and medium to large portions of higher level academic achievement represented by the WJ–III Applied Problems (i.e., math reasoning) subtest and the NDRT (i.e., reading comprehension). Of the specific academic test scores, the untimed math subtests of the WJ–III were best predicted by the FSIQ, and the spelling and timed math subtests of the WJ–III were least accounted for by the FSIQ. Incremental prediction of the four index scores combined was statistically significant for all academic achievement composite and subtest scores. The effect size for math fluency approached the large range for the incremental prediction and was in the medium range for all academic achievement scores other than untimed mathematics and timed writing, of which the four factor index scores only accounted for small portions of variance. The unique incremental contributions to prediction of achievement by the four factor index scores were trivial.

The current study’s results are the first reported on the incremental validity of WAIS–IV factor index scores beyond the FSIQ in predicting academic achievement with an independent clinical

sample. Incremental validity analyses were not reported in the WAIS–IV *Technical and Interpretive Manual* or in such manuals of previous adult versions of the Wechsler scales. Several similar studies have been conducted using the child version of the Wechsler scales as well as other intelligence tests. In general, results from these studies indicated that the overall Full Scale score of the intelligence tests accounted for large portions of achievement variance, whereas the more specific index scores accounted for trivial to small amounts of variance beyond the overall score (Glutting et al., 2006; Glutting, Youngstrom, Ward, Ward, & Hale, 1997; Nelson & Canivez, 2012; M. W. Watkins, Glutting, & Lei, 2007; Youngstrom, Kogos, & Glutting, 1999). The results of the current study indicate greater incremental predictive validity of the WAIS–IV index scores than the index scores of other intelligence tests examined in extant empirical studies. Canivez’s (in press) study of incremental validity of the WAIS–IV using the standardization data provides the most direct comparison for the current results. Specific comparisons with the results of Canivez show that in the present study WAIS–IV factor index scores accounted for greater portions of WJ–III Letter–Word Identification than WIAT–II or WIAT–III Word Reading, greater portions of WJ–III Spelling than WIAT–II or WIAT–III Spelling, and greater portions of WJ–III Reading Fluency than WIAT–III Oral Reading Fluency. Other similar achievement subtests did not appear to differ.

The reasons for this finding are unclear. One possible explanation has to do with the characteristics of the current sample. This sample consisted mainly of college students who were diagnosed with LD and/or ADHD. These individuals often have discrepancies between their measured intelligence and academic achievement, in favor of the former (Gregg, 2009). That is, they are an anomalous group because their IQ scores predict higher achievement scores than they actually obtain. Although speculative, it might be that the more specific abilities represented by the WAIS–IV index scores have a greater influence on academic achievement beyond general intelligence for this specific group of individuals with disabilities than for the general population.

Table 5
Incremental Contribution of Observed Wechsler Adult Intelligence Scale–Fourth Edition Factor Index Scores in Predicting Woodcock–Johnson III Achievement and Nelson–Denny Reading Test Scores Beyond the Full Scale IQ (FSIQ)

Predictor	Letter–Word Identification (<i>n</i> = 289)		Calculation (<i>n</i> = 289)	
	Variance (%)	Increment ^a (%)	Variance (%)	Increment ^a (%)
FSIQ	23.0	23.0**	33.5	33.5**
Index scores (<i>df</i> = 4) ^b	42.8	19.9**	37.9	4.4*
VCI		0.1		0.7
PRI		0.2		0.6
WMI		0.0		1.4
PSI		0.0		0.6
	Spelling (<i>n</i> = 288)		Academic Skills Cluster (<i>n</i> = 287)	
FSIQ	14.4	14.4**	34.3	34.3**
Index scores (<i>df</i> = 4) ^b	30.8	17.6**	50.4	16.2**
VCI		1.1		0.7
PRI		0.1		0.0
WMI		1.3		0.9
PSI		0.6		0.2
	Reading Fluency (<i>n</i> = 299)		Math Fluency (<i>n</i> = 300)	
FSIQ	19.3	19.3**	13.9	13.9**
Index scores (<i>df</i> = 4) ^b	37.1	17.8**	43.8	29.9**
VCI		0.8		0.6
PRI		0.1		0.2
WMI		0.6		2.2
PSI		1.3		1.8
	Writing Fluency (<i>n</i> = 300)		Academic Fluency Cluster (<i>n</i> = 299)	
FSIQ	21.6	21.6**	28.0	28.0**
Index scores (<i>df</i> = 4) ^b	30.6	9.1**	50.8	22.8**
VCI		0.1		0.8
PRI		0.0		0.1
WMI		0.1		1.1
PSI		0.3		1.7
	Applied Problems (<i>n</i> = 275)		Nelson–Denny (<i>n</i> = 297)	
FSIQ	49.1	49.1**	29.2	29.2**
Index scores (<i>df</i> = 4) ^b	55.7	6.5**	39.3	10.1**
VCI		0.3		0.3
PRI		0.3		0.0
WMI		0.8		0.1
PSI		0.1		0.3

Note. The Academic Skills Cluster is composed of the Letter–Word Identification, Calculation, and Spelling subtests, and the Academic Fluency Cluster is composed of the Reading Fluency, Math Fluency, and Writing Fluency subtests. Variance percentages are $R^2 \times 100$. VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

^a Unless otherwise indicated, all unique contributions are squared part correlations equivalent to changes in R^2 if this variable was entered last in block entry regression procedure. ^b Partialing out FSIQ.

* $p < .001$. ** $p < .0001$.

It is important to note, however, that such differences in incremental validity based on referred versus nonreferred group status have not been found with similar child samples (Glutting et al., 1997; M. W. Watkins et al., 2007). Additionally, the current results were suggestive only that the four factor index scores combined predicted unique achievement variance beyond the FSIQ. Therefore, statements cannot be made that specific abilities (e.g., working memory or processing speed) were more influential in predicting achievement for the current sample than for previous samples. The unique predictions by individual factor index scores evidenced by squared part correlations were not statistically significant and were trivial in their effects.

Although it may be tempting to make comparison of the current results with those from studies of CHC theory using structural equation modeling (see McGrew & Wendling, 2010,

for a review), such comparisons would be misguided because latent constructs rather than observed scores were examined in the vast majority of these CHC studies. Unlike the observed scores used in the current study, latent constructs derived from SEM are error free. Score distributions differ and individuals are ranked differently based on whether latent constructs or observed scores are used (Oh et al., 2004). Beyond these quantitative differences, the purpose of the current study is fundamentally disparate from investigations of CHC construct–achievement relations. Whereas the current study was focused on examining the direct practical utility of WAIS–IV scores in predicting academic achievement, the purpose of the majority of CHC studies is to examine theoretical relationships. Both types of studies are important; however, their different purposes and methods limit comparability.

Limitations

The results of the current study should be interpreted considering the following limitations. First, several characteristics of the sample limit generalizability of these results. Ethnicity was not representative of the U.S. population at large. Furthermore, although the age range was broad, most individuals were of typical college age, and therefore age groups near the high end of the range were not well represented. A final characteristic that limits the generalizability of these findings is that the sample consisted primarily of adolescents and young adults with various learning or attention disorders who planned on attending, were attending, or had attended college. Second, a nonreferred comparison group was not included in the current study. Inclusion of such a group would have permitted more rigorous examination of the factor structure of the WAIS-IV and the incremental validity of the index scores. More specifically, inclusion of both referred and nonreferred groups would have allowed for not only examination of underlying factor structure but also investigation of whether metric relationships between subtest scores and associated latent constructs are equivalent across groups (Bowden et al., 2008).

Implications for Practice

Several implications for practice can be drawn from the current study's results. First, it is recommended that the factor structure of the WAIS-IV be interpreted as consisting of a general intelligence factor and the four index factors purported by the test developers when assessing similar individuals. Second, although these results support the four index factors, they do not support the statement made in the WAIS-IV *Technical and Interpretive Manual* that focus on these scores "is recommended as the primary level of clinical interpretation" (Wechsler, 2008b, p. 127). Rather, the current results suggest that the greatest interpretive weight should be afforded to the FSIQ representing the general intelligence construct because it captured the greatest amount of variance within the factor structure, had the highest latent factor reliability estimate, and was consistently a stronger predictor of academic achievement than the four factor indices. Third, these results provide greater support for interpretation of the four index scores at a secondary level of clinical interpretation than have results of previous studies (e.g., Canivez & Watkins, 2010a, 2010b). The VC, PR, WM, and PS factors combined accounted for more common and total factor variance than observed in previous studies. The current study's results indicated that the four index factors combined predicted meaningful academic achievement beyond the FSIQ, further supporting clinical attention to the index scores at a secondary level of interpretation. However, those making interpretations at the factor index score level must be mindful that although the index scores predicted meaningful academic achievement variance when examined together, none of the individual index scores uniquely predicted meaningful variance beyond that predicted by the four index scores together. Additionally, the latent factor reliability estimates of the specific factors once other factor variance was removed were low. Finally, our results suggest that *g* as measured by the WAIS-IV is better conceptualized as a breadth factor rather than a superordinate factor. Although this result does not appear to have direct implications for interpreting the FSIQ, it calls attention to a different conceptualization of *g* than that of the test publisher.

References

- Beaujean, A. A., Firmin, M. W., Michonski, J. D., Berry, T., & Johnson, C. (2010). A multitrait-multimethod examination of the Reynolds Intellectual Assessment Scales in a college sample. *Assessment, 17*, 347–360. doi:10.1177/1073191109356865
- Benson, N., Hulac, D. M., & Krantzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV): What does the WAIS–IV measure? *Psychological Assessment, 22*, 121–130. doi:10.1037/a0017767
- 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
- Bowden, S. C., Gregg, N., Bandalos, D., Davis, M., Coleman, C., Holdnack, J. A., & Weiss, L. G. (2008). Latent mean and covariance differences with measurement equivalence in college students with developmental difficulties versus the Wechsler Adult Intelligence Scale–III/Wechsler Memory Scale–III normative sample. *Educational and Psychological Measurement, 68*, 621–642. doi:10.1177/0013164407310126
- Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011a). Augmenting the core battery with supplemental subtests: Wechsler Adult Intelligence Scale–IV measurement invariance across the United States and Canada. *Assessment, 18*, 133–140. doi:10.1177/1073191110381717
- Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011b). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale–IV in the United States and Canada. *Educational and Psychological Measurement, 71*, 186–199. doi:10.1177/0013164410387382
- Brown, J. I., Fishco, V. V., & Hanna, G. (1993a). *Nelson–Denny Reading Test*. Chicago, IL: Riverside.
- Brown, J. I., Fishco, V. V., & Hanna, G. (1993b). *Nelson–Denny Reading Test Technical Report Forms G & H*. Chicago, IL: Riverside.
- Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since Bias in Mental Testing. *School Psychology Quarterly, 14*, 208–238. doi:10.1037/h0089007
- 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
- Byrne, B. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales–Fifth Edition for children and adolescents. *School Psychology Quarterly, 23*, 533–541. doi:10.1037/a0012884
- Canivez, G. L. (2011). Hierarchical factor structure of the Cognitive Assessment System: Variance partitions from the Schmid–Leiman (1957) procedure. *School Psychology Quarterly, 26*, 305–317. doi:10.1037/a0025973
- Canivez, G. L. (in press). Incremental criterion validity of WAIS–IV factor index scores: Relationships with WIAT–II and WIAT–III subtest and composite scores. *Psychological Assessment*.
- Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler Abbreviated Scale of Intelligence and Wide Range Intelligence Test: Convergent and structural validity. *School Psychology Quarterly, 24*, 252–265. doi:10.1037/a0018030
- Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment, 10*, 285–291.
- Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV) adolescent subsample. *School Psychology Quarterly, 25*, 223–235. doi:10.1037/a0022046
- Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition

- (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, 22, 827–836. doi:10.1037/a0020429
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet–Fifth Edition. *Journal of Psychoeducational Assessment*, 24, 123–136. doi:10.1177/0734282905285244
- Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, 27, 494–507. doi:10.1177/0734282909333179
- Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, 18, 543–555. doi:10.1037/0894-4105.18.3.543
- Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment*, 12, 320–329. doi:10.1177/1073191105278118
- Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, 27, 73–86. doi:10.1027/1614-0001.27.2.73
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science Quarterly*, 50, 21–43.
- Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education*, 40, 103–114. doi:10.1177/00224669060400020101
- Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment*, 9, 295–301. doi:10.1037/1040-3590.9.3.295
- Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment*, 23, 143–152. doi:10.1037/a0021230
- Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2012). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment*. Advance online publication. doi:10.1037/a0030676
- Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 231–258). New York, NY: Plenum Press.
- Gregg, N. (2009). *Adolescents and adults with learning disabilities and ADHD: Assessment and accommodation*. New York, NY: Guilford Press.
- Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified? A response to special issue authors. *Applied Neuropsychology*, 14, 37–51. doi:10.1080/09084280701280445
- Holdnack, J. A., Zhou, X., Larrabee, G. J., Millis, S. R., & Salthouse, T. A. (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, 18, 178–191. doi:10.1177/1073191110393106
- Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, 15, 443–445. doi:10.1037/1040-3590.15.4.443
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Lassiter, K. S., Bell, N. L., Hutchinson, M. B., & Matthews, T. D. (2001). College student performance on the General Ability Measure for Adults and the Wechsler Intelligence Scale for Adults—Third Edition. *Psychology in the Schools*, 38, 1–10. doi:10.1002/1520-6807(200101)38:1<1::AID-PITS1>3.0.CO;2-M
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: “Sinking shafts at a few critical points”. *Annual Review of Psychology*, 51, 405–444. doi:10.1146/annurev.psych.51.1.405
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530. doi:10.1093/biomet/57.3.519
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III Normative Update technical manual*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive–achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 47, 651–675.
- Mundfrom, D. J., & Shaw, D. G. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5, 159–168. doi:10.1207/s15327574ijt0502_4
- NCS Pearson. (2009). *Wechsler Individual Achievement Test—Third Edition*. San Antonio, TX: Author.
- 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
- Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology*, 45, 439–456. doi:10.1016/j.jsp.2007.03.003
- Niileksela, C. R., Reynolds, M. R., & Kaufman, A. S. (2012). An alternative Cattell–Horn–Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70–90. *Psychological Assessment*. Advance online publication. doi:10.1037/a0031175
- Oh, H.-J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., & McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: Implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *Journal of Special Education*, 38, 159–173. doi:10.1177/00224669040380030301
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Harcourt Brace.
- Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. doi:10.1080/00273171.2012.715555
- Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. San Diego, CA: Sattler.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53–61. doi:10.1007/BF02289209
- Schneider, W. J. (2008). Playing statistical Ouija Board with commonality analysis: Good questions, wrong assumptions. *Applied Neuropsychology*, 15, 44–53. doi:10.1080/09084280801917566
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153–158. doi:10.1177/001316446302300113

- 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
- Strauss, E., Spreen, O., & Hunter, M. (2000). Implications for test revisions for research. *Psychological Assessment, 12*, 237–244. doi:10.1037/1040-3590.12.3.237
- Swanson, H. L., & Hsieh, C. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research, 79*, 1362–1390. doi:10.3102/0034654309350931
- Trainin, G., & Swanson, H. L. (2005). Cognition, metacognition, and achievement of college students with learning disabilities. *Learning Disability Quarterly, 28*, 261–272. doi:10.2307/4126965
- Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*, 328–340. doi:10.1037/a0025614
- Watkins, C. E., Jr., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54–60. doi:10.1037/0735-7028.26.1.54
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment, 18*, 123–125. doi:10.1037/1040-3590.18.1.123
- Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children-Fourth Edition among a national sample of referred students. *Psychological Assessment, 22*, 782–787. doi:10.1037/a0020043
- Watkins, M. W. (2013). Omega [Computer software]. Phoenix, AZ: Ed & Psych Associates.
- Watkins, M. W., Glutting, J. J., & Lei, P. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14*, 13–20. doi:10.1080/09084280701280353
- Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred students. *Educational and Psychological Measurement, 66*, 975–983. doi:10.1177/0013164406288168
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2001). *Wechsler Individual Achievement Test-Second Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale-Fourth Edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale-Fourth Edition technical and interpretive manual*. San Antonio, TX: Pearson.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson III Tests of Achievement Normative Update*. Itasca, IL: Riverside
- Youngstrom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly, 14*, 26–39. doi:10.1037/h0088996

Received October 1, 2012

Revision received January 29, 2013

Accepted January 30, 2013 ■