Psychological Assessment
2013, Vol. 25, No. 2, 484–495

1040-3590/13/$12.00 DOI: 10.1037/a0032092

Incremental Criterion Validity of WAIS–IV Factor Index Scores: Relationships With WIAT–II and WIAT–III Subtest and Composite Scores

Gary L. Canivez
Eastern Illinois University

The present study examined the incremental validity of Wechsler Adult Intelligence Scale–4th Edition (WAIS–IV; Wechsler, 2008a) factor index scores in predicting academic achievement on the Wechsler Individual Achievement Test–2nd Edition (WIAT–II; Psychological Corporation, 2002a) and on the Wechsler Individual Achievement Test–3rd Edition (WIAT–III; Wechsler, 2009a) beyond that predicted by the WAIS–IV Full Scale IQ (FSIQ). As with previous intelligence test incremental validity studies, the WAIS–IV FSIQ accounted for statistically significant and generally large portions of WIAT–II and WIAT–III subtest and composite score variance. WAIS–IV factor index scores combined to provide statistically significant increments in variance accounted for in most WIAT–II and WIAT–III subtest and composite scores over and above the FSIQ score; however, the effect sizes ranged from trivial to medium as observed in investigations with other intelligence tests (i.e., Glutting, Watkins, Konold, & McDermott, 2006; Youngstrom, Kogos, & Glutting, 1999). Individually, the WAIS–IV factor index scores provided trivial to small unique contributions to predicting WIAT–II and WIAT–III scores. This finding indicated that the FSIQ should retain primacy and greatest interpretive weight in WAIS–IV interpretation, as previously indicated by WAIS–IV subtest variance partitions form hierarchical exploratory factor analyses (Canivez & Watkins, 2010a, 2012b).

Keywords: WAIS–IV, construct validity, incremental validity, hierarchical multiple regression, achievement prediction

The Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV; Wechsler, 2008a), one of the most frequently used intelligence tests for adults and older adolescents, includes 15 subtests (10 core and five supplemental), four first-order factor index scores (Verbal Comprehension [VCI], Perceptual Reasoning [PRI], Working Memory [WMI], and Processing Speed [PSI]), and the higher order Full Scale IQ score (FSIQ). The Object Assembly and Picture Arrangement subtests were deleted (thereby reducing subtests with manipulative objects), and Visual Puzzles, Figure Weights, and Cancellation subtests were created and added. Item coverage and range were increased, and the WAIS–IV theoretical foundation was updated. Like other contemporary intelligence tests such as the Wechsler Intelligence Scale for Children–Fourth Edition (WISC–IV; Wechsler, 2003), the Stanford–Binet Intelligence Scales–Fifth Edition (SB–5; Roid, 2003), Kaufman Assessment Battery for Children–Second Edition (KABC–II; Kaufman & Kaufman, 2004), Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003), and Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000); the WAIS–IV content and structure were modified in an attempt to reflect current conceptualizations of intellectual measurement articulated by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Cattell, 1966).

Carroll's (1993, 2003) three-stratum theory of cognitive abilities proposes 50–60 narrow abilities (Stratum I) at the bottom (subtests), 8–10 broad ability factors (Stratum II) in the middle (first-order factors), and the general (*g*) ability factor (Stratum III) at the top (second-order factor), which is estimated by the overall FSIQ or global IQ. The American Psychological Association task force study of intelligence (Neisser et al., 1996) noted that the hierarchical nature of intelligence measurement was the most widely accepted view and this still appears to be true. G*f*–G*c* theory (Cattell, 1943; Horn, 1988, 1991; Horn & Cattell, 1966; Horn & Noll, 1997) is similar and describes aspects of reasoning abilities that allow the individual to solve novel problems (*fluid intelligence* [G*f*]) and abilities acquired through the individual's exposure to aspects of his or her culture such as language and educational experiences (*crystallized intelligence* [G*c*]). Extension of G*f*-G*c* theory by Horn (1991) and Horn and Noll (1997) is similar to that

This article was published Online First May 6, 2013.

This research was partially supported by a 2011 Summer Research Grant from the Council on Faculty Research, Eastern Illinois University. Preliminary results and versions were presented at the 2012 annual conventions of the National Association of School Psychologists and the American Psychological Association. The author thanks NCS Pearson for graciously providing WAIS–IV–WIAT–II and WAIS–IV–WIAT–III data sets reported in the respective WAIS–IV and WIAT–III technical manuals for analyses.

Correspondence concerning this article should be addressed to Gary L. Canivez, Department of Psychology, Eastern Illinois University, 600 Lincoln Ave., Charleston, IL 61920-3099. E-mail: glcanivez@eiu.edu

of Carroll with some eight, nine, or more broad dimensions but *does not* include higher order *g,* as Horn argued there was insufficient construct validity evidence for singular *g*. Cattell–Horn–Carroll (CHC) theory is an approach in which the work of Cattell and Horn was merged with that of Carroll (Evans, Floyd, McGrew, & Leforgee, 2002; Flanagan, 2000; McGrew, 2005), but this seems an odd combination, given that Carroll provided evidence for higher order *g* while Horn argued that a singular *g* did not exist and was a statistical artifact.

While the WAIS–IV implies CHC constructs, the four-factor WAIS–IV structure does not comport exactly with CHC theory. The WAIS–IV structure appears to have CHC consistency with the VCI subtests measuring crystallized ability (G*c*), WMI subtests measuring short-term memory (G*sm*), and PSI subtests measuring processing speed (G*s*). The PRI subtests are divided into two CHC factors where Block Design, Visual Puzzles, and Picture Completion measure visual processing (G*v*) and Matrix Reasoning and Figure Weights measure fluid reasoning (G*f*). Because there are no quantitative reasoning subtests to combine with the WAIS–IV Arithmetic subtest, it often has variance split between the G*sm* and some other dimension. Benson, Hulac, and Kranzler (2010) provided an independent examination of the WAIS–IV structure and tested various alternative CHC models to the hierarchical model presented in the WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b) using all 15 subtests. Weiss, Keith, Zhu, and Chen (in press) also examined the WAIS–IV structure from a CHC perspective and noted a good fit to the standardization data and invariance for normative and clinical samples. While Benson et al. argued that the CHC alternative was superior and provided formulae for calculating CHC composites, Weiss et al. showed both the Wechsler four-factor model or the CHC five-factor model fit well and provided alternate interpretations (for a commentary, see Canivez & Kush, in press). Absent from both Benson et al.'s and Weiss et al.'s examinations was a bifactor (Holzinger & Swineford, 1937), nested factor (Gustafsson & Balke, 1993), or direct hierarchical (Gignac, 2008) model for comparison, where the higher order *g* factor has *direct* paths and influence on subtests rather than paths indirectly influencing subtests through the first-order factors, implying full mediation of *g* through the first-order factors. Also, there was no presentation of apportioned subtest variance to the higher order *g* factor, first-order CHC factors, and that which was unique. Little was said about interpretation of the higher order *g* factor. Finally, the standardized path coefficient of .99 between the second-order *g* factor and G*f* indicates, like other studies of CHC explanations of Wechsler scales, that G*f* is indistinguishable from *g* and thus problematic. This could, however, be an artifact from structural equation modeling (SEM) methods (Golay, Reverte, Rossier, Favez, & Lecerf, 2012). A related study by Ward, Bergman, and Hebert (2012) found that the four-factor model with three orthogonal minor factors (viz., Spatial Visualization, Quantitative Reasoning, and Digit–Letter Memory Span) was the best fitting model and most theoretically defensible, but their proposed CHC solution was incomplete due to a lack of inclusion of higher order *g*.

Interpretation of the four WAIS–IV factor index scores "is recommended as the primary level of clinical interpretation, especially in cases with considerable variability across the index and or subtest scores" (Wechsler, 2008b, p. 127). This recommendation was challenged by two empirical studies of the WAIS–IV standardization sample data (Canivez & Watkins, 2010a, 2010b). In both studies, researchers examined the apportioned WAIS–IV subtest variance to the higher order *g* factor and to the four lower order factors with the Schmid and Leiman (1957) orthogonalization procedure used as Carroll (1993, 1995, 1997, 2003) and others (Carretta & Ree, 2001; Gustafsson & Snow, 1997; McClain, 1996; Ree, Carretta, & Green, 2003; Thompson, 2004) recommended. Whether examining 10-, 15-, or 12-subtest configurations with the total standardization sample or the 16- to 19-year-old adolescent subsample, the second-order *g* factor accounted for the largest portions of total and common variance while the four first-order factors accounted for considerably smaller portions of total and common variance, and it was concluded that primary interpretation should reside with the FSIQ. However, compared with the WISC–IV, the four WAIS–IV first-order factors accounted for somewhat larger portions of common and total variance. Whether that additional first-order factor variance is meaningful cannot be determined from examination of the internal structure.

While interesting and important, examinations of the internal structure of tests using exploratory factor analysis or confirmatory factor analysis are insufficient as they cannot fully answer questions of validity or diagnostic utility or efficiency (Canivez, Konold, Collins, & Wilson, 2009; Carroll, 1997; Kline, 1994; Lubinski & Dawis, 1992). The factor structure of any measure must be further examined with relations to external criteria to see how well they perform. For intelligence tests, one important external criterion is academic achievement.

Predicting academic achievement is a primary use of intelligence tests, and some have suggested that predicting student achievement is the most important application of intelligence tests (Brown, Reynolds, & Whitaker, 1999; Weiss & Prifitera, 1995). Since the creation of the first Binet–Simon Scale of Intelligence (Binet & Simon, 1904), prediction of school performance (academic achievement) has been a primary use of intelligence tests. Research has consistently shown that intelligence tests account for meaningful levels of academic achievement variance (Brody, 2002; Carroll, 1993; Gottfredson, 1997, 2008; Jensen, 1998; Lubinski, 2000; Lubinski & Humphreys, 1997; Naglieri & Bornstein, 2003), and average IQ–achievement correlations are near .55 across age groups (Brody, 2002; Neisser et al., 1996). Approximately 85%–90% of predictable criterion variable variance is accounted for by a single general IQ score (Thorndike, 1986), and among the best normed intelligence and achievement tests, it is quite common to observe concurrent FSIQ–composite achievement correlations near .70 (Elliott, 2007; Glutting et al., 2000; Kaufman & Kaufman, 1993, 2004; Naglieri & Das, 1997b; NCS Pearson, 2009; Reynolds & Kamphaus, 2003; Roid, 2003; Wechsler, 2003, 2008b; Wechsler & Naglieri, 2006; Woodcock, McGrew, & Mather, 2001).

Perhaps due to the relative newness of the WAIS–IV, published peer-reviewed studies of the relationships between WAIS–IV scores and measures of academic achievement have not yet appeared. However, the WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b) presented results from a study where a sample of 93 adolescents (age range 16–19 years) were administered the WAIS–IV and Wechsler Individual Achievement Test–Second Edition (WIAT–II; Psychological Corporation, 2001) during the WAIS–IV standardization. Zero-order Pearson correlations between the WAIS–IV and WIAT–II scores were strong, ranging from the .50s to .70s for the FSIQ and factor index scores. The *Wechsler Individual Achievement Test–Third Edition Technical Manual* (WIAT–III; NCS Pearson, 2009b) presented results from a study where 59 adolescents (age range 16–19-years) were administered the WAIS–IV and WIAT–III during the WIAT–III standardization. As with the WAIS–IV and WIAT–II, zero-order Pearson correlations for the WAIS–IV and WIAT–III were generally strong and ranged from the .50s to the .70s for the FSIQ and factor index (VCI and PRI) scores. WAIS–IV WMI and PSI relationships with WIAT–III scores, however, were not presented.

Interpretation of WAIS–IV scores, as with the scores of other intelligence tests, generally begins at the highest level (FSIQ) and moves to lower levels of the test (factor indexes, index comparisons, subtests, subtest comparisons; Kamphaus, Winsor, Rowe, & Kim, 2005). While the WAIS–IV *Technical and Interpretive Manual* emphasized the factor index scores as the primary level of interpretation, these scores and comparisons must *each* be supported by sufficient reliability and validity evidence (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 1999).

The extent to which variance within WAIS–IV subtests is primarily second-order (FSIQ), first-order (factor index scores), or specific to the subtest is not readily apparent, and this information was not included in the WAIS–IV *Technical and Interpretive Manual*. Because subtests include variance attributed to the higher order *g* factor, first-order factors, specificity, and error, Carroll (1995) insisted on inspection of variance partitions within intelligence tests to understand how subtest variance is apportioned to the second-order and first-order dimensions. Subtest performance on cognitive ability measures reflects combinations of both first-order *and* second-order factors, which is why Carroll argued that the Schmid and Leiman (1957) procedure be used to first extract variance from the higher order factor to residualize the lower order factors, leaving them orthogonal to the higher order factor, to facilitate interpretation. Numerous studies of intelligence tests have consistently shown that the greatest portions of total and common variance are apportioned to the second-order dimension, which is represented by the Full Scale score, and much smaller portions of total and common variance is apportioned to the first-order dimensions, represented by the respective factor scores. This has been documented for the WISC–IV (Bodin, Pardini, Burns, & Stevens, 2009; Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), French WISC–IV (Golay et al., 2012), SB–5 (Canivez, 2008), WASI and WRIT (Canivez et al., 2009), RIAS (Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007), Cognitive Assessment System (CAS; Canivez, 2011; Naglieri & Das, 1997a), French WAIS–III (Golay & Lecerf, 2011), and WAIS–IV (Canivez & Watkins, 2010a, 2010b). The implication of these findings is that the overall, omnibus, Full Scale score should retain primary interpretive weight rather than the first-order factor-based scores. The viability of the first-order factor scores following consideration of the Full Scale score must be further evaluated with methods that compare the various intelligence test scores (viz., FSIQ and factor index scores) with external criteria (i.e., predictive validity, incremental validity, and diagnostic utility or efficiency).

Considering *external* validity investigations, such as predictive validity and incremental validity of lower order scores *beyond* that of higher order scores (Haynes & Lench, 2003; Hunsley, 2003; Hunsley & Meyer, 2003), is important when one is interpreting intelligence tests across multiple levels and scores. Incremental validity relates to 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). With this approach, the relative importance of WAIS–IV factor index scores versus the global FSIQ may be assessed. Hunsley and Meyer (2003) suggested that incremental validity is simple and straightforward; however, application to intelligence tests is complicated by their hierarchical nature and the fact that clinicians may be simultaneously interpreting scores at three different levels (i.e., Full Scale scores, factor scores, and subtest scores). Interpreting scores at all levels of the test ignores the fact illustrated previously that some reliable subtest variance is apportioned to the higher order *g* factor, some to the first-order factors, and some remains unique to the subtest (viz., specificity and error). Thus, interpreting all scores results in interpretive redundancy because for individual scores, such variance cannot be disaggregated.

For assessment of the incremental validity in intelligence tests, hierarchical multiple regression analysis is a well-established statistical procedure (e.g., Canivez, 2012; Freberg, Vandiver, Watkins, & Canivez, 2008; Glutting et al., 2006; Glutting, Youngstrom, Ward, Ward, & Hale, 1997; Kahana, Youngstrom, & Glutting, 2002; Nelson & Canivez, 2012; Ryan, Kreiner, & Burton, 2002; Watkins, Glutting, & Lei, 2007; Youngstrom et al., 1999) and “is probably the most common analytic strategy for quantifying the incremental contributions of specific methods, items, or measures to existing assessments” (McFall, 2005, p. 320). Entering the FSIQ in the first block and entering all first-order factor scores in the second block when predicting academic achievement test performance provides an examination of the additional achievement variance accounted for by the first-order factor scores after accounting for that predicted by the FSIQ.

Incremental validity of first-order factor scores over and above second-order Full Scale scores in predicting academic achievement in a variety of contexts has indicated that most of the reliable achievement variance is accounted for by Full Scale intelligence test scores and little additional prediction is provided by first-order factor scores (Canivez, 2012; Freberg et al., 2008; Glutting et al., 1997, 2006; Kahana et al., 2002; Nelson & Canivez, 2012; Ryan et al., 2002; Watkins et al., 2007; Youngstrom et al., 1999). It may be that these limited portions of achievement test score variance accounted for by first-order factor scores is related to the generally smaller portions of subtest variance apportioned to the first-order factor scores identified through hierarchical exploratory factor analyses.

Presently, there are no examinations of the incremental validity of the WAIS–IV factor index scores, and such examination is necessary to help determine interpretive weight and decision making of the factor index scores beyond the FSIQ (Sechrest, 1963). Because somewhat greater portions of subtest variance were observed in the first-order WAIS–IV factors (Canivez & Watkins, 2010a, 2010b), it is possible that greater incremental validity may be observed in the WAIS–IV factor index scores than was observed in the WISC–IV (Glutting et al., 2006). For proper assessment of the incremental predictive validity of WAIS–IV factor index scores, hierarchical multiple regression analyses must be conducted to first account for achievement test score variance (WIAT–II and WIAT–III) predicted by the WAIS–IV FSIQ and then to account for additional proportions of achievement variance predicted by the WAIS–IV factor index scores. If *primary* interpretation of the four factor index scores is to be followed, then the four factor index scores must demonstrate meaningful incremental validity beyond the FSIQ.

Method

Participants

Table 1 presents demographic characteristics of both the WAIS–IV and WIAT–II ($N = 93$) and the WAIS–IV and WIAT–III ($N = 59$) samples. The WAIS–IV and WIAT–II sample (Wechsler, 2008b) included 16- to 19-year-old high school students while the WAIS–IV and WIAT–III sample (NCS Pearson, 2009b) included 16- to 17-year-old students. Both samples matched closely the demographic characteristics of the respective population estimates from the U.S. census across major variables of race/ethnicity, parental education level/socioeconomic status, and geographic region.

Instruments

Wechsler Adult Intelligence Scale–Fourth Edition. The Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV; Wechsler, 2008a) is an individual test of general intelligence for ages 16–90 years, originating with the 1939 Wechsler–Bellevue Intelligence Scale (Wechsler, 1939b). Consistent with Wechsler's definition of intelligence (i.e., "global capacity"; Wechsler, 1939a, p. 229) and all versions of his tests, the WAIS–IV measures general intelligence through the administration of numerous subtests, each of which is an indicator and estimate of intelligence. The WAIS–IV has 10 core subtests that are used to produce the FSIQ. The Verbal Comprehension Index (VCI) and Perceptual Reasoning Index (PRI) are each composed of three subtests while the Working Memory Index (WMI) and Processing Speed Index (PSI) are each composed of two subtests. Supplemental subtests are provided as substitutes for core subtests when necessary (one each for the VC, WM, and PS scales and two for the PR scale); however, three of the supplemental subtests (Figure Weights, Letter–Number Sequencing, and Cancellation) are not available for 70- to 90-year-olds. Average internal consistency estimates ranged from .78 to .94 for subtests and from .90 to .96 for factor index scores; the average internal consistency estimate was .98 for the FSIQ. Various validity estimates are also presented in the *Technical and Interpretive Manual* (Wechsler, 2008b), and reviews are available (Canivez, 2010; Schraw, 2010).

Wechsler Individual Achievement Test–Second Edition. The Wechsler Individual Achievement Test–Second Edition (WIAT–II; Psychological Corporation, 2002a) is an individually administered achievement test used for diagnostic purposes and measures a variety of skills in reading, mathematics, writing, listening, and speaking. It includes nine subtests and five composite scores and is a major revision of the original version. The WIAT–II was standardized on students from preschool through Grade 16 and on a representative sample of the U.S. population stratified on major variables such as age, grade, sex, race/ethnicity, parental education level, and geographic location. Extensive information regarding psychometric features (reliability, validity, norms) is presented in the *Examiner's Manual* (Psychological Corporation, 2002b). Subtest and composite scores are reported as commonly scaled standard scores ($M = 100$, $SD = 15$).

Wechsler Individual Achievement Test–Third Edition. The Wechsler Individual Achievement Test–Third Edition (WIAT–III; NCS Pearson, 2009a) is a diagnostic and individually administered achievement test in a variety of areas related to reading, mathematics, writing, listening, and speaking. It includes

Table 1
Demographic Characteristics of the WAIS–IV and WIAT–II Sample and WAIS–IV and WIAT–III Sample

Sample/demographic variable	n	%	U.S. %
WAIS–IV/WIAT–II sample ($N = 93$)			
Sex			
Male	48	51.6	50.0
Female	45	48.4	50.0
Race/ethnicity			
White	59	63.4	63.0
Black/African American	12	12.9	14.4
Hispanic	15	16.1	16.1
Asian American	3	3.2	3.7
Other	4	4.3	3.0
Parental education level			
< 12 years	11	11.8	12.5
12 years	26	28.0	28.3
13–15 years	30	32.3	33.0
≥ 16 years	26	28.0	25.8
Geographic region			
Midwest	18	19.4	23.9
Northeast	10	10.8	18.5
South	23	24.7	34.6
West	42	45.2	22.0
WAIS–IV/WIAT–III sample ($N = 59$)			
Sex			
Male	31	52.5	50.0
Female	28	47.5	50.0
Race/ethnicity			
White	47	79.7	60.6
Black/African American	4	6.8	14.3
Hispanic	6	10.2	18.2
Asian American	2	3.4	3.9
Other	0	0.0	3.0
Parental education level			
< 12 years	5	8.5	11.5
12 years	18	30.5	25.4
13–15 years	22	37.3	33.6
≥ 16 years	14	23.7	29.5
Geographic region			
Midwest	20	33.9	22.6
Northeast	15	25.4	18.2
South	11	18.6	35.6
West	13	22.0	23.6

Note. WAIS–IV/WIAT–II U.S. population estimates based on data for 16- to 19-year-olds from WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b), which were obtained from the 2005 U.S. Census. WAIS–IV/WIAT–III U.S. population estimates based on total U.S. population estimates for 4- to 19-year-olds as reported in the WIAT-III *Technical Manual* (NCS Pearson, 2009b), which were obtained from the 2005 U.S. Census. WAIS–IV = Wechsler Adult Intelligence Scale–Fourth Edition; WIAT–II = Wechsler Individual Achievement Test–Second Edition; WIAT–III = Wechsler Individual Achievement Test–Third Edition.

17 subtest and eight composite scores and is a major revision of the WIAT–II. The WIAT–III was standardized on 2,775 students (preschool through Grade 12) who were demographically representative of the U.S. population stratified on variables of grade, age, sex, race/ethnicity, parent education level, and geography. The WIAT–III *Technical Manual* (NCS Pearson, 2009b) presents detailed information regarding reliability (internal consistency, short-term stability, interrater agreement), validity, and norms. Subtest and composite scores are reported as commonly scaled standard scores (M = 100, SD = 15).

Procedure

According to the WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b), the 93 participants in the WAIS–IV and WIAT–II data set were 16- to 19-year-old high school students, and most (n not reported) were administered the WAIS–IV first. The mean test interval between WAIS–IV and WIAT–II was 11 days, with a range of 0–60 days. According to the WIAT–III *Technical Manual* (NCS Pearson, 2009b), the 59 participants who were administered the WAIS–IV and WIAT–III were 16- to 17-year-old students, and all were first administered the WIAT–III, with the WAIS–IV administered from 1 to 53 days later.

Data Analyses

Hierarchical multiple regression analyses were conducted to assess proportions of WIAT–II and WIAT–III achievement subtest and composite scores variance accounted for by the observed WAIS–IV FSIQ and factor index scores. 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 via SPSS Version 19 for Mac linear regression analysis. WIAT–II analyses included subtest scores (Math Reasoning, Numerical Operations, Pseudoword Decoding, Spelling, Word Reading, Reading Comprehension, Written Expression, Listening Comprehension, and Oral Expression) and composite scores (Mathematics Composite, Oral Language Composite, Reading Composite, Written Language Composite, and the Total Achievement Composite) as dependent variables. For the WIAT–III analyses, subtest scores (Word Reading, Pseudoword Decoding, Reading Comprehension, Oral Reading Fluency, Oral Reading Rate, Oral Reading Accuracy, Numerical Operations, Math Problem Solving, Math-Addition, Math-Subtraction, Math-Multiplication, Essay Composition: Grammar & Mechanics, Spelling, Sentence Composition, Essay Composition, Listening Comprehension, and Oral Expression) and composite scores (Total Reading Composite, Basic Reading Composite, Reading Comprehension and Fluency Composite, Mathematics Composite, Math Fluency Composite, Written Expression Composite, Oral Language Composite, and Total Achievement Composite) served as dependent variables. The change in WIAT–II and WIAT–III achievement 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. As noted by Glutting et al. (2006), multiple regression analyses are appropriate due to the *predictive* focus of the study (Pedhazur, 1997). 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.

Results

WAIS–IV/WIAT–II Sample (N = 93)

WIAT–II subtests. Table 2 presents results from hierarchical multiple regression analyses for WIAT–II subtest scores. The WAIS–IV FSIQ accounted for statistically significant (p < .001) portions of each of the WIAT–II subtest scores ranging from 18% (Pseudoword Decoding) to 63.5% (Math Reasoning). These portions of WIAT–II achievement variance predicted from the FSIQ represented large effect sizes.

Also illustrated in Table 2 are R^2 increases (reported in percentages) provided by the combined and unique effects of WAIS–IV factor index scores in predicting each of the nine WIAT–II subtests after achievement variance due to the FSIQ was accounted for. Statistically significant (p < .05) portions of WIAT–II subtest variance was incrementally accounted for by the combined WAIS–IV factor index scores for all subtests except Spelling and Written Expression and ranged from 1.7% (Written Expression) to 12.3% (Listening Comprehension). These increased variance portions represented trivial (R^2 < .03) to medium effect sizes. The *unique* contributions of WAIS–IV factor index scores in predicting each of the nine WIAT–II subtests (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (0.2% to 1.5%), PRI (0.3% to 1.5%), WMI (0.1% to 2.2%), and PSI (0.2% to 1.7%).

WIAT–II composites. Table 3 presents results from hierarchical multiple regression analyses for the five WIAT–II composite scores. The WAIS–IV FSIQ accounted for statistically significant (p < .001) portions of each of the WIAT–II composite scores ranging from 42.7% (Written Language Composite) to 76.6% (Total Achievement Composite). These portions of WIAT–II achievement variance predicted from the FSIQ represented large effect sizes.

Table 3 also presents the R^2 increases (reported in percentages) provided by the combined and unique effects of WAIS–IV factor index scores in predicting each of the five WIAT–II composite scores after the achievement variance due to the FSIQ was accounted for. Statistically significant (p < .05) amounts of WIAT–II composite score variance was incrementally accounted for by the combined WAIS–IV factor index scores for all composites except the Written Language Composite and ranged from 1.4% (Written Language Composite) to 9.9% (Reading Composite). These increased variance amounts represented trivial (R^2 < .03) to small effect sizes. The *unique* contributions of WAIS–IV factor index scores in predicting each of the five WIAT–II composites (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (0%–1.4%), PRI (0%–1.1%), WMI (0%–2.0%), and PSI (0%–1.1%).

WAIS–IV/WIAT–III Sample (N = 59)

WIAT–III subtests. Table 4 presents results from hierarchical multiple regression analyses for the 17 WIAT–III subtest scores. The WAIS–IV FSIQ accounted for statistically significant (p < .05) portions of each of the WIAT–III subtest scores ranging from 8.3% (Oral Reading Accuracy) to 59.8% (Oral Expression). These portions of WIAT–III achievement variance predicted from

Table 2
Percentages of Incremental Contribution of Observed WAIS–IV Factor Index Scores in Predicting WIAT–II Achievement Subtest Scores

Predictor	Math Reasoning		Numerical Operations		Pseudoword Decoding		Spelling		Word Reading		Reading Comprehension		Written Expression		Listening Comprehension		Oral Expression	
	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]
FSIQ	63.5	63.5***	60.2	60.2***	18.0	18.0***	35.0	35.0***	42.3	42.3***	59.4	59.4***	36.3	36.3***	50.3	50.3***	41.0	41.0***
Index scores[b]	67.9	4.4*	65.1	4.9*	30.2	12.2**	38.9	3.9	50.8	8.5**	69.1	9.7***	38.0	1.7	62.6	12.3***	43.3	2.3
VCI		0.7		1.2		0.6		0.2		0.7		1.3		1.5		1.1		1.1
PRI		0.9		1.3		0.8		0.5		1.5		0.4		1.3		0.3		0.8
WMI		1.6		2.2		0.1		0.1		0.5		0.6		1.3		0.2		0.7
PSI		0.9		1.1		1.3		0.4		1.7		0.4		1.3		0.2		0.8

Note. WAIS–IV = Wechsler Adult Intelligence Scale–Fourth Edition; WIAT–II = Wechsler Individual Achievement Test–Second Edition; FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index. Variance percentages are R^2*100.
[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] Degrees of freedom = 4; partialing out FSIQ.
* $p < .05$. ** $p < .01$. *** $p < .001$.

the FSIQ represented medium to large effect sizes for all subtests except Oral Reading Accuracy, which had a small effect size.

Also presented in Table 2 are R^2 increases (reported in percentages) provided by the combined and unique effects of WAIS–IV factor index scores in predicting each of the 17 WIAT–III subtests after achievement variance due to the FSIQ was accounted for. Statistically significant ($p < .05$) portions of WIAT–III subtest variance incrementally accounted for by the combined WAIS–IV factor index scores were observed for only the Spelling and Listening Comprehension subtests. However, increases in WIAT–III subtest prediction ranged from 2.4% (Essay Composition: Grammar and Mechanics) to 13.7% (Oral Reading Accuracy), and the small sample size likely reduced power and affected statistical significance. These increased variance portions represented trivial ($R^2 < .03$) to medium effect sizes. The *unique* contributions of WAIS–IV factor index scores in predicting each of the 17 WIAT–III subtests (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (0%–3.5%), PRI (0%–4.0%), WMI (0%–5.6%), and PSI (0%–3.9%).

WIAT–III composites. Table 5 presents results from hierarchical multiple regression analyses for the eight WIAT–III Composite scores. The WAIS–IV FSIQ accounted for statistically significant ($p < .001$) portions of each of the WIAT–III composite scores ranging from 25.9% (Basic Reading Composite) to 62.8% (Total Achievement Composite). These portions of WIAT–III achievement variance predicted from the FSIQ represented large effect sizes for all but the Basic Reading Composite, which had a medium effect size.

Table 5 also presents R^2 increases (reported in percentages) provided by the combined and unique effects of WAIS–IV factor index scores in predicting each of the eight WIAT–III composite scores after the achievement variance due to the FSIQ was accounted for. Statistically significant ($p < .05$) portions of WIAT–III subtest variance incrementally accounted for by the combined WAIS–IV factor index scores were observed for only the Written Expression Composite, Oral Language Composite, and Total Achievement Composite scores. However, increases in WIAT–III composite score prediction ranged from 5.1% (Mathematics Composite) to 11.7% (Written Expression Composite), and the small sample size likely reduced power and affected statistical significance. These increased variance portions represented small to medium effect sizes. The *unique* contributions of WAIS–IV factor index scores in predicting each of the eight WIAT–III subtests (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (0%–2.1%), PRI (0%–3.6%), WMI (0%–1.5%), and PSI (0%–2.8%).

Discussion

The present study assessed the incremental validity of WAIS–IV factor index scores in predicting academic achievement beyond that provided by the FSIQ. Hierarchical multiple regression analyses were used to determine the extent to which WAIS–IV factor index scores provided meaningful improvements in prediction of WIAT–II and WIAT–III scores *beyond* the FSIQ. For the WIAT–II, the WAIS–IV FSIQ provided statistically significant prediction with large effect sizes for all WIAT–II subtests (except Pseudoword Decoding) and all WIAT–II composite

Table 3
Percentages of Incremental Contribution of Observed WAIS–IV Factor Index Scores in Predicting WIAT–II Achievement Composite Scores

	Mathematics Composite		Oral Language Composite		Reading Composite		Written Language Composite		Total Achievement Composite	
Predictor	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]
FSIQ	71.1	71.1***	61.7	61.7***	57.8	57.8***	42.7	42.7***	76.6	76.6***
Index scores[b]	76.0	4.9**	69.5	7.7**	67.7	9.9***	44.1	1.4	79.2	2.6*
VCI		1.0		1.4		0.0		0.6		0.1
PRI		1.1		0.7		0.4		0.9		0.0
WMI		2.0		0.5		0.0		0.5		0.1
PSI		1.1		0.6		0.5		0.8		0.0

Note. Variance percentages are $R^{2*}100$. WAIS–IV = Wechsler Adult Intelligence Scale–Fourth Edition; WIAT–II = Wechsler Individual Achievement Test–Second Edition; FSIQ = Full Scale IQ; 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] Degrees of freedom = 4; partialing out FSIQ.
* $p < .05$. ** $p < .01$. *** $p < .001$.

scores. This finding is consistent with observations from many different intelligence tests and measures of academic achievement. Statistically significant improvements in prediction of WIAT–II subtests by the combined WAIS–IV factor index scores were observed for all subtests (except Spelling and Written Expression) and all composite scores (except Written Language). Effect sizes ranged from trivial ($R^2 < .03$) to medium.

These results also are similar to previous findings (Glutting et al., 1997, 2006; Youngstrom et al., 1999). However, direct comparison with the incremental validity results from Glutting et al. (2006) showed that in the present study, the combined WAIS–IV factor index scores provided increased prediction of the WIAT–II Reading Composite (9.9%) and WIAT–II Mathematics Composite (4.9%) beyond the predictions by the WAIS–IV FSIQ that were much larger than that produced by the combined contributions WISC–IV factor index scores (1.8% for Reading Composite and 0.3% for Mathematics Composite) observed by Glutting et al. Additional direct comparisons are not possible as the Glutting et al. study concentrated solely on the WIAT–II Reading and Mathematics Composite scores. The present results are more similar to larger portions of incremental prediction by the combined first-order factors (5%–16%) reported by Glutting et al. (1997) using the WISC–III (Wechsler, 1991) to predict WIAT scores in the linking sample. The increased amounts of WIAT–II achievement predicted by WAIS–IV factor index scores are likely related to the increased portions of first-order factor variance reported by Canivez and Watkins (2010a, 2010b).

WAIS–IV factor index score incremental validity results for the WIAT–III were similar to those for the WIAT–II in the present study as proportions of WIAT–III achievement accounted for by the WAIS–IV FSIQ were large and higher than that accounted for by the combined factor index scores. Generally, WIAT–III subtest and composite score variance predicted by the WAIS–IV FSIQ and factor index scores combined was quite a bit less than that observed with the WIAT–II for comparable subtests. This may be sample specific, and the sample for the WIAT–III was quite small. Thus, firm conclusions should be avoided, and replication with additional and larger samples is needed.

Direct comparisons of WAIS–IV factor index score incremental validity in predicting WIAT–III performance are not possible as there appeared to be no published studies of incremental validity for the WAIS–IV or WIAT–III for comparison. Comparing results from the present study with other studies in the literature is difficult because of differences in the achievement tests used as the criterion for prediction.

The FSIQ accounts for more achievement variance as it benefits from the aggregate across numerous subtests that produces greater true score variance and less error variance (Cronbach, 1951; Gottfredson, 2008; Gustafsson & Undheim, 1996; Lubinski & Dawis, 1992). Multicollinearity of the FSIQ and factor index scores in multiple regression analyses was observed in the present study as in other investigations (i.e., Glutting et al., 2006) due to the linear combination of subtests to produce factor index scores and the FSIQ. However, this redundancy is precisely the problem practitioners must confront in their predictive (and explanatory) interpretations of both second-order factor (g) and first-order factors in clinical assessments. As noted in the introduction, clinicians interpreting the FSIQ and then interpreting factor index scores are counting variance twice because subtest variance is in part second-order g, first-order factor, and unique. Such variance cannot be disaggregated for individuals, and thus interpretation of observed FSIQ and factor index scores would result in redundancy. With respect to the use of hierarchical multiple regression analysis to determine improvements in R^2 provided by WAIS–IV factor index scores, such multicollinearity does not invalidate the method and directly answers the question of improved prediction of an external criterion (Dana & Dawes, 2007; Schneider, 2008).

Hale, Fiorello, Kavanagh, Holdnack, and Aloe (2007) pointed out an issue in multiple regression analysis that relates to the order of variable entry as variables entered first capture greater criterion variance than variables entered later. Their suggestion and illustration were entry of first-order 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 that showed a similar result of little incremental validity of the FSIQ in the second block. This is due to the multicollinearity issue noted previously. While possible to do, this should not be done and *was not* done in the present study because as Glutting et al. (2006) duly noted, such a procedure would require rejecting the law of parsimony and to "repeal scientific law" (p. 106). Schneider (2008) also

Table 4
Percentages of Incremental Contribution of Observed WAIS–IV Factor Index Scores in Predicting WIAT–III Achievement Subtest Scores

	Word Reading		Pseudoword Decoding		Reading Comprehension		Oral Reading Fluency		Oral Reading Rate		Oral Reading Accuracy	
Predictor	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr
FSIQ	27.0	27.0^{***}	19.1	19.1^{***}	36.3	36.3^{*}	24.3	24.3^{***}	23.1	23.1^{***}	8.3	8.3^{*}
Index scores[a]	38.3	11.3	28.0	8.9	41.5	5.1	34.8	10.5	30.7	7.7	22.0	13.7
VCI		0.1		3.5		0.5		0.5		0.9		2.4
PRI		1.3		3.8		1.2		0.0		0.2		4.0
WMI		0.0		1.4		1.4		1.2		1.5		0.6
PSI		0.3		3.9		0.7		0.5		1.0		1.7

	Numerical Operations		Math Problem Solving		Math–Addition		Math–Subtraction		Math–Multiplication		Essay Composition: Grammar & Mechanics	
	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr
FSIQ	51.3	51.3^{***}	44.6	44.6^{***}	36.6	36.6^{***}	35.0	35.0^{***}	24.2	24.2^{***}	19.9	19.9^{***}
Index scores[a]	54.1	2.8	52.1	7.5	42.7	6.1	42.5	7.4	31.6	7.4	22.3	2.4
VCI		0.0		0.5		0.0		0.0		0.2		0.3
PRI		0.0		0.2		0.0		0.1		0.3		0.1
WMI		0.3		0.0		0.5		0.6		0.0		0.4
PSI		0.0		0.6		0.1		0.0		0.0		0.4

	Spelling		Sentence Composition		Essay Composition		Listening Comprehension		Oral Expression	
	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr	Variance	Incr
FSIQ	32.8	32.8^{***}	22.2	22.2^{***}	16.2	16.2^{**}	39.5	39.5^{***}	59.8	59.8^{***}
Index scores[a]	45.6	12.9^{*}	33.2	11.1	21.6	5.3	52.5	13.0^{*}	64.8	5.1
VCI		1.3		4.0		1.0		0.8		0.4
PRI		2.8		2.0		0.2		0.1		0.0
WMI		0.2		5.6		0.9		0.6		0.3
PSI		1.6		3.5		0.9		0.0		0.3

Note. Variance percentages are $R^{2*}100$. WAIS-IV = Wechsler Adult Intelligence Scale–Fourth Edition; WIAT-III = Wechsler Individual Achievement Test–Third Edition; FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index. For increments (Incr): 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.
[a] Degrees of freedom = 4; partialing out FSIQ.
* $p < .05$. ** $p < .01$. *** $p < .001$.

rejected this and referred to the reversing of variable entry by Hale et al. (2007) as testing "a nonsensical hypothesis: Does the weighting used to compute the FSIQ predict more variance than the near optimal weighting chosen by regression? Unsurprisingly, the answer is no." (p. 52).

The importance of incremental validity of multilevel intelligence tests is based on an important scientific principle articulated by William of Ockham (alternate spelling: Occam): the law of parsimony states "what can be explained by fewer principles is needlessly explained by more" Jones, 1952, p. 620). Less complex explanations, rather than more complex explanations, for phenomena are preferred, and in the case of intelligence test interpretation, the Full Scale score, an estimate of *g*, is the more parsimonious index than the lower level factor or broad ability scores (and subtest scores). For the factor index scores to be relevant in prediction of external criteria, they must demonstrate *meaningful* predictive validity *beyond* that provided by the Full Scale score. Application of incremental validity of WAIS–IV factor index scores with samples of disabled youths might provide an interesting contrast to results presented here but to date has not been done.

Due to the predictive nature of the present study and assessment of *observed* variables (FSIQ and factor index scores), hierarchical multiple regression analyses were used. It might be argued that an alternative means for analysis would be SEM of latent first- and second-order dimensions in *explaining* achievement variance. While this is an interesting *theoretical* question (and one that might also be used to examine rival CHC-based first-order factors suggested by Benson et al., 2010, and Weiss et al., in press), there are a number of problems regarding this approach as it applies to practitioner use of observed scores the tests provide. Oh, Glutting, Watkins, Youngstrom, and McDermott (2004) have noted that latent construct scores used in SEM analyses are not equivalent to the observed standard scores practitioners use and are not provided by tests, and their distributions differ. Use of SEM and explanatory analyses in the present study would also be complicated (ill advised) given the small sample sizes (Lei & Wu, 2007), particularly those for the WAIS–IV and WIAT–III. The interest in the present study, however, was the incremental prediction provided by *observed* factor index scores beyond the FSIQ—available scores practitioners actually use. It would be interesting to examine

Table 5
Percentages of Incremental Contribution of Observed WAIS–IV Factor Index Scores in Predicting WIAT–III Achievement Composite Scores

Predictor	Total Reading Composite		Basic Reading Composite		Reading Comprehension & Fluency Composite		Mathematics Composite		Math Fluency Composite		Written Expression Composite		Oral Language Composite		Total Achievement Composite	
	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]	Variance	Increment[a]
FSIQ	40.4	40.4***	25.9	25.9***	39.7	39.7***	52.7	52.7***	36.0	36.0***	38.8	38.8***	60.5	60.5***	62.8	62.8***
Index scores[b]	48.5	8.1	34.5	8.6	46.0	6.3	57.8	5.1	43.1	7.1	50.5	11.7*	69.5	9.0**	69.0	6.2*
VCI		1.6		2.1		0.0		0.1		0.0		0.6		0.7		0.0
PRI		3.2		3.6		0.6		0.0		0.1		0.0		0.0		0.5
WMI		0.9		1.0		0.0		0.1		0.3		1.5		0.5		0.0
PSI		1.8		2.8		0.1		0.2		0.0		0.4		0.1		0.2

Note. Variance percentages are R^2*100. WAIS–IV = Wechsler Adult Intelligence Scale–Fourth Edition; WIAT–III = Wechsler Individual Achievement Test–Third Edition; FSIQ = Full Scale IQ; 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] Degrees of freedom = 4; partialing out FSIQ.
* $p < .05$. ** $p < .01$. *** $p < .001$.

incremental validity of first-order CHC factors suggested by Benson et al. (2010) but not presently possible because the WAIS–IV FSIQ is based only on the 10 core subtests, and a 15-subtest FSIQ is not available.

Clinical utility of WAIS–IV factor index scores for individual use is important and also of interest. Incremental validity as examined in the present study pertains to *groups* and may set limits for clinical utility but does not assess clinical utility. Given the somewhat larger effect sizes for WAIS–IV factor index scores improvement in predicting achievement compared with for the WISC–IV, one might expect better utility. However, clinical utility is a question that must be examined with alternate methods. Clinical utility is concerned with the ability of a test to correctly predict an *individual's* membership in a clinical group or a response to differential treatment (Meehl, 1959; Mullins-Sweatt & Widiger, 2009; Wiggins, 1988). Such diagnostic utility or efficiency of WAIS–IV scores is as yet unknown. A number of distinct group differences comparisons are reported in the WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b), but examination of diagnostic utility/efficiency of WAIS–IV factor index scores for use with these groups has not yet been reported and should be examined. Weiner (2003) noted, however, that even when research on group differences has shown statistically significant differences, it has been rare that such results translated to predictive power for individual use. Given the small portions of *unique* factor index score prediction of achievement, it is hard to imagine there would be incremental utility to these scores.

Limitations

Results of the present study must be considered in relation to significant limitations. The most important and significant limitation in the present study is the sample size of these two data sets. While nationally representative based on respective population estimates from the U.S. census as part of WAIS–IV and WIAT–III standardizations, the data sets are small, particularly the WAIS–IV and WIAT–III data sets. This resulted in limited power for statistical significance testing (for WIAT–III scores); however, focus on effect sizes (increases in achievement variance accounted for beyond the FSIQ) is more relevant for consideration of incremental validity and importance of WAIS–IV factor index scores for interpretation. These appear to be the only data available comparing the WAIS–IV with the WIAT–II and WIAT–III, so replication is certainly necessary to better inform clinicians on the relative contributions of WAIS–IV scores for judging the importance of factor index scores for interpretation. Also, the present data sets are not composed of individuals with disabilities, and it could be that incremental contribution of WAIS–IV factor index scores may be higher in samples of individuals with particular disabilities. Broad generalization to such groups is not recommended, and the question of incremental validity of WAIS–IV factor index scores for those with disabilities should also be examined.

Conclusion

At this time, based on the present incremental validity analyses and previous results from hierarchical exploratory factor analyses (Canivez & Watkins, 2010a, 2010b), the primary level of

WAIS–IV interpretation and greatest interpretive weight should be the FSIQ rather than the factor index scores as directed in the WAIS–IV *Technical and Interpretive Manual* (Wechsler, 2008b). Also, when interpretations of WAIS–IV factor index scores are made, they should be done cautiously and in light of peer-reviewed empirical studies in addition to analyses reported in the *Technical and Interpretive Manual* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Benson, N., Hulac, D. M., & Kranzler, 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.

Binet, A., & Simon, T. (1904). Methodes nouvelle pour le diagnostic du niveau intellectual des anormaux [New methods for the diagnosis of the abnormal intellectual level]. *L'Annee Psychologique, 11,* 191–244. doi:10.3406/psy.1904.3675

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

Brody, N. (2002). *g* and the one–many problem: Is one enough? In G. R. Bock, J. A. Goode, & K. Webb, *The nature of intelligence: Novartis Foundation Symposium 233* (pp. 122–135). New York, NY: Wiley.

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

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford–Binet Intelligence Scales for children and adolescents. *School Psychology Quarterly, 23,* 533–541. doi:10.1037/a0012884

Canivez, G. L. (2010). Review of the Wechsler Adult Intelligence Test—Fourth Edition. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 684–688). Lincoln, NE: Buros Institute of Mental Measurements.

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. (2012). Incremental validity of the cognitive assessment system: Predicting achievement with CAS Full Scale and PASS Scale scores. Manuscript submitted for publication.

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., & Kush, J. C. (in press). WISC–IV and WAIS–IV structural validity: Alternate methods, alternate results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). *Journal of Psychoeducational Assessment*.

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

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325–335. doi:10.1111/1468-2389.00184

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511571312

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429–452. doi:10.1207/s15327906mbr3003_6

Carroll, J. B. (1997). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to the bell curve* (pp. 125–156). New York, NY: Springer–Verlag. doi:10.1007/978-1-4612-0669-9_6

Carroll, J. B. (2003). The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). New York, NY: Pergamon Press. doi:10.1016/B978-008043793-4/50036-2

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193. doi:10.1037/h0059973

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139–164. doi:10.1111/j.1745-3984.1978.tb00065.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. doi:10.1007/BF02310555

Dana, J., & Dawes, R. M. (2007). Comment on Fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology, 14,* 21–25. doi:10.1080/09084280701280379

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

Elliott, C. D. (2007). *Differential Ability Scales: Introductory and technical handbook* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31,* 246–262.

Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly, 15,* 295–329. doi:10.1037/h0088789

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC–III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology, 15,* 131–139. doi:10.1080/09084280802084010

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., Adams, W., & Sheslow, D. (2000). *Wide Range Intelligence Test: Manual*. Wilmington, DE: Wide Range.

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. *The 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 (BSEM). *Psychological Assessment*. Advance online publication. doi:10.1037/a0030676

Gottfredson, L. S. (1997). Intelligence and social policy. *Intelligence, 24,* 288–320. doi:10.1016/S0160-2896(97)90010-6

Gottfredson, L. S. (2008). Of what value is intelligence? In A. Prifitera, D. Saklofske, & L. G. Weiss (Eds.), *WISC–IV clinical assessment and intervention* (2nd ed., pp. 545–564). Amsterdam, the Netherlands: Elsevier.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434. doi:10.1207/s15327906mbr2804_2

Gustafsson, J.-E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107–135). Westport, CT: Greenwood Press.

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York, NY: Macmillan.

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

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456–466. doi:10.1037/1040-3590.15.4.456

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41–54. doi:10.1007/BF02287965

Horn, J. L. (1988). Thinking about human abilities. In R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York, NY: Plenum Press. doi:10.1007/978-1-4613-0893-5_19

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock–Johnson technical manual* (Rev. ed., pp. 197–232). Itasca, IL: Riverside.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57,* 253–270. doi:10.1037/h0023816

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York, NY: Guilford Press.

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

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446–455. doi:10.1037/1040-3590.15.4.446

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jones, W. T. (1952). *A history of Western philosophy*. New York, NY: Harcourt Brace.

Kahana, S. Y., Youngstrom, E. A., & Glutting, J. J. (2002). Factor and subtest discrepancies on the Differential Abilities Scale: Examining prevalence and validity in predicting academic achievement. *Assessment, 9,* 82–93. doi:10.1177/1073191102009001010

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 23–38). New York, NY: Guilford Press.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kline, P. (1994). *An easy guide to factor analysis*. London, England: Routledge.

Lei, P.-W., & Wu, Q. (2007, Fall). Introduction to structural equation modeling: Issues and practical considerations. *ITEMS: Instructional Topics in Educational Measurement,* 33–43.

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

Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills, and proficiencies. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 1–59). Palo Alto, CA: Consulting Psychologists Press.

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24,* 159–201. doi:10.1016/S0160-2896(97)90016-7

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology, 4,* 229–240.

McFall, R. M. (2005). Theory and utility: Key themes in evidence-based assessment. Comment on the special section. *Psychological Assessment, 17,* 312–323. doi:10.1037/1040-3590.17.3.312

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

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology/Revue canadienne de psychologie, 13,* 102–128. doi:10.1037/h0083769

Mullins-Sweatt, S. N., & Widiger, T. A. (2009). Clinical utility and DSM–V. *Psychological Assessment, 21,* 302–312. doi:10.1037/a0016607

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment, 21,* 244–260. doi:10.1177/073428290302100302

Naglieri, J. A., & Das, J. P. (1997a). *Cognitive Assessment System*. Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (1997b). *Cognitive Assessment System: Interpretive handbook*. Itasca, IL: Riverside.

NCS Pearson. (2009a). *Wechsler Individual Achievement Test–Third Edition*. San Antonio, TX: Author.

NCS Pearson. (2009b). *Wechsler Individual Achievement Test–Third Edition technical manual*. San Antonio, TX: Author.

Neisser, U., Boodoo, G., Bouchard, Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101. doi:10.1037/0003-066X.51.2.77

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

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.

Psychological Corporation. (2002a). *Wechsler Individual Achievement Test–Second Edition*. San Antonio, TX: Author.

Psychological Corporation. (2002b). *Wechsler Individual Achievement Test–Second Edition examiner's manual*. San Antonio, TX: Author.

Ree, M. J., Carretta, T. R., & Green, M. T. (2003). The ubiquitous role of g in training. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 262–274). New York, NY: Pergamon Press. doi:10.1016/B978-008043793-4/50051-9

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales: Fifth Edition: Technical manual*. Itasca, IL: Riverside.

Ryan, J. J., Kreiner, D. S., & Burton, D. B. (2002). Does high scatter affect the predictive validity of WAIS–III IQs? *Applied Neuropsychology, 9,* 173–178. doi:10.1207/S15324826AN0903_5

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

Schneider, J. (2008). Playing statistical ouija board with commonality analysis (and other errors). *Applied Neuropsychology, 15,* 44–53. doi:10.1080/09084280801917566

Schraw, G. (2010). Test review of the Wechsler Adult Intelligence Scale–Fourth Edition. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 688–690). Lincoln, NE: Buros Institute of Mental Measurements.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23,* 153–158. doi:10.1177/001316446302300113

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association. doi:10.1037/10694-000

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29,* 332–339. doi:10.1016/0001-8791(86)90012-6

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, M. W. (2006). Orthogonal higher order structure of the Weschsler Intelligence Scale for Children—Fourth edition. *Psychological Assessment, 18,* 123–125. doi:10.1037/1040-3590.18.1.123

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (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. (1939a). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins. doi:10.1037/10020-000

Wechsler, D. (1939b). *Wechsler Bellevue Intelligence Scale*. New York, NY: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition technical and interpretive manual*. 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.

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Psychological Corporation.

Weiner, I. B. (2003). Prediction and postdiction in clinical decision making. *Clinical Psychology: Science and Practice, 10,* 335–338. doi:10.1093/clipsy.bpg030

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (in press). WAIS–IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*.

Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC–III FSIQ across ethnic and gender groups. *Journal of School Psychology, 33,* 297–304. doi:10.1016/0022-4405(95)00016-F

Wiggins, J. S. (1988). *Personality and prediction: Principles of personality assessment*. Malabar, FL: Krieger.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities*. 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 March 19, 2012
Revision received December 17, 2012
Accepted December 17, 2012 ■