British Journal of Educational Psychology (2014), 84, 667–684

The British Psychological Society

www.wileyonlinelibrary.com

Incremental validity of WISC–IV^{UK} factor index scores with a referred Irish sample: Predicting performance on the WIAT–II^{UK}

Gary L. Canivez[1]*, Marley W. Watkins[2], Trevor James[3], Rebecca Good[3] and Kate James[3]

[1]Department of Psychology, Eastern Illinois University, Charleston, Illinois, USA
[2]Department of Educational Psychology, Baylor University, Waco, Texas, USA
[3]Éirim: The National Assessment Agency, Ltd., Dublin, Ireland

Background. Subtest and factor scores have typically provided little incremental predictive validity beyond the omnibus IQ score.

Aims. This study examined the incremental validity of Wechsler Intelligence Scale for Children – Fourth UK Edition (WISC–IV^{UK}; Wechsler, 2004a, *Wechsler Intelligence Scale for Children – Fourth UK Edition*, Harcourt Assessment, London, UK) and factor index scores in predicting academic achievement on the Wechsler Individual Achievement Test – Second UK Edition (WIAT–II^{UK}; Wechsler, 2005a, *Wechsler Individual Achievement Test-Second UK Edition*, Pearson, London, UK), beyond that predicted by the WISC–IV^{UK} FSIQ.

Sample. The sample included 1,014 Irish children (ages 6–0 to 16–9) who were referred for evaluation of learning difficulties.

Method. Hierarchical multiple regression analyses were used with the WISC–IV^{UK} FSIQ (Block 1) and factor index scores (Block 2) as predictors and WIAT–II^{UK} subtest and composite scores as dependent variables.

Results. The WISC–IV^{UK} FSIQ accounted for statistically significant and generally large portions of WIAT–II^{UK} subtest and composite score variance. WISC–IV^{UK} factor index scores combined to provide statistically significant increments in prediction of most WIAT–II^{UK} subtest and composite scores over and above the FSIQ; however, the effect sizes were mostly small as previously observed (i.e., Canivez, 2013a, *Psychol. Assess.*, *25*, 484; Glutting *et al.*, 2006, *J. Spec. Educ.*, *40*, 103; Nelson *et al.*, 2013, *Psychol. Assess.*, *25*, 618). Individually, the WISC–IV^{UK} factor index scores provided small *unique* contributions to predicting WIAT–II^{UK} scores.

Conclusion. This, in combination with studies of apportioned variance from bifactor confirmatory factor analysis (Watkins *et al.*, 2013, *Int. J. Sch. Educ. Psychol.*, *1*, 102), indicated that the WISC–IV^{UK} FSIQ should retain the greatest weight in WISC–IV^{UK} interpretation.

Wechsler intelligence scale popularity has resulted in various adaptations, translations, and norming for use in different countries and varying cultures. When the Wechsler Intelligence Scale for Children – Fourth Edition (WISC–IV; Wechsler, 2003a,b) was under

*Correspondence should be addressed to Gary L. Canivez, Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920–3099, USA (email: glcanivez@eiu.edu).

DOI:10.1111/bjep.12056

development, a parallel version was developed for the United Kingdom (UK). Due to similarities in language, only minor changes to items or language and spelling, and some reordering of items based on item difficulty were reportedly required (Wechsler, 2004b). After the WISC–IV was anglicized and adapted for the United Kingdom in 2002, it was then standardized on a representative UK sample stratified by geographic region, sex, race/ethnicity, and parent education level (Wechsler, 2004b). Like the WISC–IV, the Wechsler Intelligence Scale for Children – Fourth UK Edition (WISC–IV^{UK}; Wechsler, 2004a) is a test of general intelligence composed of 15 subtests, 10 of which are mandatory and contribute to measurement of four factor-based index scores: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI).

While the WISC–IV^{UK} *Administration and Scoring Manual* (Wechsler, 2004b) is based on the UK version and standardization sample; in the roughly 10 years since standardization data were collected, Pearson UK has failed to produce a technical manual or technical report disclosing important reliability and validity information based on the UK standardization sample. Thus, WISC–IV^{UK} psychometric properties from the standardization sample are still unknown, nor have the psychometric properties of the WISC–IV^{UK} been adequately reported for clinical samples (e.g., Ford *et al.*, 2008; Gordon, Duff, Davidson, & Whitaker, 2010).

To date, only one study has examined the WISC–IV^{UK} latent structure and was based on a large sample (n = 794) of referred children in Ireland (Watkins, Canivez, James, Good, & James, 2013) where the WISC–IV^{UK} is commonly used. Results indicated the latent factor structure of the 10 WISC–IV^{UK} core subtests was similar to other investigations, supporting the robust general intelligence dimension and four specific dimensions (factor indexes), and the bifactor (direct hierarchical) model was arguably the best representation of WISC–IV^{UK} structure. Although Murray and Johnson (2013) suggested that there may be inherent bias in favour of the bifactor model due to unmodelled complexity created by cross-loadings that are fixed to zero in CFA, Golay, Reverte, Rossier, Favez, and Lecerf (2013) found superiority of the bifactor model even when including subtest cross-loading estimates using Bayesian structural equation modelling. The Watkins *et al.* examination of latent factor reliabilities found the general intelligence dimension had a strong reliability estimate (ω_h = .802), while the four specific factors had poor reliabilities (ω_h) ranging from .143 (*Perceptual Reasoning*) to .376 (*Processing Speed*). Thus, it was recommended that primary WISC–IV^{UK} interpretation should be of the FSIQ, not the four-factor index scores.

These structural results were similar to those found in other studies of Wechsler scales (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Canivez & Watkins, 2010a,b; Gignac, 2005, 2006; Gignac & Watkins, 2013; Golay & Lecerf, 2011; Golay *et al.*, 2013; Nelson, Canivez, & Watkins, 2013; Niileksela, Reynolds, & Kaufman, 2013; Watkins, 2006; Watkins & Beaujean, 2014; Watkins, Wilson, Kotz, Carbone, & Babula, 2006) and other intelligence tests (Canivez, 2008, 2011; Canivez, Konold, Collins, & Wilson, 2009; Dombrowski, 2013, 2014a,b; Dombrowski & Watkins, 2013; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007).

Latent factor structure examinations of intelligence tests using EFA or CFA are interesting and important but insufficient because they cannot fully answer questions of validity and do not address diagnostic utility or efficiency (Canivez *et al.*, 2009; Carroll, 1997; Kline, 1994; Lubinski & Dawis, 1992). The factor structure of any measure must be further evaluated against relations with external criteria. For intelligence tests, one important external criterion is academic achievement.

Some have suggested that predicting student achievement is the most important application of intelligence tests (Brown, Reynolds, & Whitaker, 1999; Weiss & Prifitera, 1995). Prediction of school performance or academic achievement has been a primary use of intelligence tests since the creation of the first Binet–Simon Scale of Intelligence (Binet & Simon, 1904). Research consistently shows 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) with average IQ–achievement correlations near .55 across age groups (Brody, 2002; Neisser *et al.*, 1996). The general IQ score generally accounts for approximately 85–90% of criterion variable variance (Thorndike, 1986), and among the best normed intelligence and achievement tests, it is common to observe concurrent Full Scale IQ–composite achievement correlations near .70 (Elliott, 2007; Glutting, Adams, & Sheslow, 2000; Kaufman & Kaufman, 1993, 2004; Naglieri & Das, 1997; Reynolds & Kamphaus, 2003; Roid, 2003; Wechsler, 2003a,b, 2008a,b; Wechsler & Naglieri, 2006; Woodcock, McGrew, & Mather, 2001). Longitudinal investigations of general intelligence estimates by Deary, Strand, Smith, and Fernandes (2007) and Watkins, Lei, and Canivez (2007) also demonstrated the strong prediction of future academic achievement. Watkins *et al.* also demonstrated support for the temporal and causal influence of intelligence on academic achievement.

Table H.6 from the WIAT–II^{UK} *UK Scoring and Normative Supplement* (Wechsler, 2005c) presents zero-order Pearson correlations between WISC–IV^{UK} FSIQ, factor index scores, and subtest scores as well as WIAT–II^{UK} subtest and composite scores from the linking sample. WISC–IV^{UK} FSIQ and VCI correlations with WIAT–II^{UK} composite scores were generally strong, ranging from the .50s to .70s, and were slightly lower for WIAT–II^{UK} subtests. The lowest correlations were between the WISC–IV^{UK} PSI and WIAT–II^{UK} subtests and composites. Unfortunately, zero-order Pearson correlations for WISC–IV^{UK} factor index scores with WIAT–II^{UK} subtests and composites obfuscate portions of WIAT–II^{UK} achievement that is related to general intelligence (estimated by the FSIQ) and that which is related to the WISC–IV^{UK} group factors (estimated by factor index scores).

Consideration of *external* validity such as predictive or criterion-related validity and, more importantly, *incremental validity* of lower-order scores *beyond* that of higher-order scores (Haynes & Lench, 2003; Hunsley, 2003; Hunsley & Meyer, 2003) is critical when intelligence tests are interpreted across multiple levels and scores as is recommended for the WISC–IV^{UK}. Incremental validity is 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). Using this approach, the relative importance of WISC–IV^{UK} factor index scores versus the global FSIQ may be determined. Hunsley and Meyer (2003) suggested that incremental validity is simple and straightforward; however, application to intelligence tests is a bit more complicated because of the hierarchical nature of the various scores and the fact that clinicians often simultaneously interpret scores at as many as three different levels (i.e., Full Scale score, factor index scores, and subtest scores). Interpreting scores at all levels ignores the fact 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) as noted by Carroll (1995); so interpreting all WISC–IV^{UK}-obtained scores results in interpretive redundancy because such variance cannot be disaggregated for individual-obtained scores. Further, each score or comparison used in test interpretation must be supported by sufficient reliability and validity evidence (AERA, APA, & NCME, 1999).

Hierarchical multiple regression analysis (H-MRA) is a well-established statistical procedure for assessing incremental validity (e.g., Canivez, 2013a; Freberg, Vandiver, Watkins, & Canivez, 2008; Glutting, Youngstrom, Ward, Ward, & Hale, 1997; Glutting, Watkins, Konold, & McDermott, 2006; Kahana, Youngstrom, & Glutting, 2002; Nelson & Canivez, 2012; Nelson *et al.*, 2013; Ryan, Kreiner, & Burton, 2002; Watkins *et al.*, 2007; Youngstrom, Kogos, & Glutting, 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). The FSIQ is entered into the first block, and all factor index scores are entered into the second block when predicting academic achievement test performance to provide an examination of the additional achievement variance accounted for by the factor index scores *after* accounting for that predicted by the FSIQ.

Research regarding the incremental validity of intelligence test first-order factor scores over and above the higher-order Full Scale score when predicting academic achievement has indicated that most of the reliable achievement variance can be attributed to the omnibus score and little additional achievement variance is predicted by factor scores (Canivez, 2013a; Freberg *et al.*, 2008; Glutting *et al.*, 1997, 2006; Kahana *et al.*, 2002; Nelson & Canivez, 2012; Nelson *et al.*, 2013; Ryan *et al.*, 2002; Watkins *et al.*, 2007; Youngstrom *et al.*, 1999). It is very likely the limited portions of achievement test score variance accounted for by factor scores are related to the generally smaller portions of subtest variance uniquely apportioned to the first-order factors identified through exploratory and confirmatory factor analyses (i.e., Canivez, 2014; Watkins, 2006; Watkins 2010).

To date, there appears to be no examination of the incremental validity of the WISC–IV^{UK} factor index scores in prediction of achievement beyond that of the FSIQ, and such examination is necessary to help further determine interpretive value and use of the WISC–IV^{UK} factor index scores beyond the FSIQ (Sechrest, 1963). Given that an examination of the structure of the WISC–IV^{UK} found little reliable specific factor index score variance (Watkins *et al.*, 2013), it was expected that there would be limited incremental prediction of achievement beyond that provided by the FSIQ. To properly assess the incremental predictive validity of WISC–IV^{UK} factor index scores, H-MRA were used to first account for WIAT–II^{UK} achievement test score variance predicted by the WISC–IV^{UK} FSIQ and then account for additional proportions of achievement variance predicted by the WISC–IV^{UK} factor index scores.

Method

Participants

Participants were 1,014 children from the Republic of Ireland between the ages of 6–0 and 16–9. Students were most frequently referred to an educational psychologist by school personnel for evaluation of learning difficulties to determine eligibility for special education services or educational accommodations. Some were referred by their parents. Participants resided in the five major cities (Cork, Dublin, Galway, Limerick, and Waterford) in Ireland (20.1%), as well as in small towns and rural areas (79.9%). This compared closely to population estimates (five major cities [23.4%], small towns and rural areas [76.6%]) from the 2011 census for 5- to 19-year-olds (Central Statistics Office, 2013). The largest proportion of the sample was male (62.6%). The mean age was 10.77 (SD = 2.57), and the distribution was bimodal with peaks at 8 and 12 years of age.

This represented 3–4 years following entry into primary schools and entrance into post-primary schools, respectively. Unfortunately, agency practice and confidentiality standards allowed no other demographic information to be included in this archival data set.

Instruments

Wechsler Intelligence Scale for Children – Fourth UK Edition

The Wechsler Intelligence Scale for Children – Fourth Edition (WISC–IV; Wechsler, 2003a) is a test of general intelligence and is composed of 15 subtests (*M*s = 10, *SD*s = 3), 10 of which are mandatory and contribute to measurement of four factor-based index scores: VCI, PRI, WMI, and PSI. Each of the four indexes is expressed as a standard score (*M*s = 100, *SD*s = 15). The FSIQ is composed of 10 subtests (three Verbal Comprehension, three Perceptual Reasoning, two Working Memory, and two Processing Speed). The WISC–IV was anglicized and adapted for the United Kingdom (UK) in 2002 through item review and minor changes in items or language, spelling, and order of item difficulty (Wechsler, 2004b). The resulting WISC–IV^{UK} was standardized and normed on a sample of 780 children between the ages of 6–0 and 16–11, who were representative of the UK population stratified by geographic region, sex, race/ethnicity, and parent education level (Wechsler, 2004b). Of the 780 children in the standardization sample, 17 (2.2%) were from Northern Ireland. There are no separate norms for children in Ireland generally or the Republic of Ireland specifically. WISC–IV^{UK} reliability and validity data based on the UK standardization sample were not provided in the WISC–IV^{UK} manual, and standard errors of measurement were taken from the US version of the WISC–IV.

Wechsler Individual Achievement Test – Second UK Edition

The Wechsler Individual Achievement Test – Second UK Edition (WIAT–II^{UK}; Wechsler, 2005a) is an individually administered achievement test used for diagnostic purposes and includes nine subtests that measure a variety of academic skills in reading (Word Reading, Reading Comprehension, Pseudoword Decoding), mathematics (Numerical Operations, Mathematics Reasoning), written language (Spelling, Written Expression), and oral language (Listening Comprehension, Oral Expression). It also includes five composite scores (Reading, Mathematics, Written Language, Oral Language, and Total). Like the WISC–IV^{UK}, it was anglicized and adapted for use in the United Kingdom with what were described as minor modifications of spelling, illustrations, and item presentation order based on item difficulty estimates (Wechsler, 2005b).

The WIAT–II^{UK} was standardized on a demographically representative sample of 892 students (48.2% male) ages 4 through 16 between November 2003 and June 2004. The standardization sample was stratified across geographic region, sex, age, race/ethnicity, and parent education level and closely matched 2001 UK Census data (Wechsler, 2005b). The WIAT–II^{UK} was reportedly conormed with the WISC–IV^{UK}, but specific details regarding this sample are lacking, and some tables (i.e., E.4, E.5, E.6) in the WIAT–II^{UK} *UK Scoring and Normative Supplement* (Wechsler, 2005c) confusingly note WIAT–II^{UK} and WISC–IV^{UK} in titles but report WIAT–II and WISC–III in table notes. Subtest and composite scores are reported as commonly scaled standard scores (*M* = 100, *SD* = 15). Like the WISC–IV^{UK}, reliability estimates and standard errors of measurement reported in the WIAT–II^{UK} manual were based on US data, *not* the WIAT–II^{UK} standardization sample.

Much validity data reported in the WIAT–II^{UK} *UK Scoring and Normative Supplement* were likewise based on the US version and samples; however, Table H.6 presents zero-order Pearson correlations between the WISC–IV^{UK} and the WIAT–II^{UK}. Additional confusion in the WIAT–II^{UK} *Examiners Manual* (p. 80) is its description of WIAT–II linking samples pertaining to the US standardization, *not* the UK standardization.

Procedure

All WISC–IV^{UK} and WIAT–II^{UK} administrations were conducted by one of three educational psychologists according to standardized procedures. Only children with complete data for all 10 WISC–IV^{UK} core subtests were included in analyses. Children were administered WIAT–II^{UK} subtests appropriate to their referral for assessment resulting in sample size differences for WIAT–II^{UK} subtest and composite score analyses (see Table 1). Institutional review board approval was obtained, all data were de-identified, and no personal information was included.

Data analyses

WISC–IV^{UK} and WIAT–II^{UK} descriptive statistics and zero-order Pearson product-moment correlations were obtained for comparisons to the UK linking sample. Hierarchical multiple regression analyses (H-MRA) were conducted to assess proportions of WIAT–II^{UK} achievement subtest score and composite score variance accounted for by the observed WISC–IV^{UK} FSIQ and factor index scores. The WISC–IV^{UK} FSIQ was singularly entered into the first block, and the four WISC–IV^{UK} factor index scores were jointly

Table 1. Descriptive statistics for 1,014 Irish children administered the WISC–IV^{UK} and the WIAT–II^{UK}

Score	*n*	*M*	*SD*	Skewness	Kurtosis
WISC–IV^{UK}					
VCI	1,014	87.66	13.96	0	−0.10
PRI	1,014	88.25	13.95	−0.05	−0.01
WMI	1,014	86.99	13.02	−0.24	−0.06
PSI	1,014	90.65	13.36	0.07	0.24
Full Scale IQ	1,014	85.37	13.38	−0.02	0.10
WIAT–II^{UK}					
Word Reading	1,014	78.19	12.87	−0.04	0.35
Reading Comprehension	970	80.38	15.96	−0.09	−0.08
Pseudoword Decoding	970	80.08	12.47	0.32	0.53
Numerical Operations	1,010	84.19	13.67	0.02	0.33
Mathematics Reasoning	45	82.87	9.49	−0.02	0.30
Spelling	999	80.28	11.48	0.26	0.68
Reading Composite	909	77.46	12.86	−0.04	0.91
Mathematics Composite	42	81.00	10.65	−1.03	1.35

Note. WISC–IV^{UK}, Wechsler Intelligence Scale for Children – Fourth UK Edition; WIAT–II^{UK}, Wechsler Individual Achievement Test – Second UK Edition; VCI, Verbal Comprehension Index; PRI, Perceptual Reasoning Index; WMI, Working Memory Index; PSI, Processing Speed Index.
Other WIAT–II^{UK} subtests and composite scores (Written Expression, Oral Expression, Writing Composite, Listening Composite, and Oral Expression Composite) were available for fewer than 20 participants and thus excluded from analyses.

entered into the second block in the SPSS version 20 for Mac (IBM Corp. 2011) linear regression analysis. Separate analyses were conducted with WIAT–II^{UK} subtest scores (Word Reading, Reading Comprehension, Pseudoword Decoding, Numerical Operations, Mathematics Reasoning, and Spelling) and composite scores (Reading Composite and Mathematics Composite) serving as dependent variables. Fewer than 20 children were administered the WIAT–II^{UK} Written Expression, Listening Comprehension, and Oral Expression subtests, so these subtests and their related composite scores were not analysed because of the extremely small sample sizes. The change in WIAT–II^{UK} achievement variance provided by the four WISC–IV^{UK} factor index scores in the second block provided an estimate of the incremental prediction beyond the WISC–IV^{UK} FSIQ from the first block. As noted by Glutting *et al.* (2006), multiple regression analyses are appropriate due to the *predictive* nature of the study (Pedhazur, 1997). Cohen's (1988) guidelines 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

Table 1 presents WISC–IV^{UK} and WIAT–II^{UK} descriptive statistics. WISC–IV^{UK} factor index scores and the FSIQ were normally distributed with skewness ranging from −0.24 to 0.07 and kurtosis ranging from −0.10 to 0.24. WIAT–II^{UK} subtest and composite scores were also distributed normally with skewness ranging from −1.03 to 0.32 and kurtosis ranging from −0.08 to 1.35. Mean WISC–IV^{UK} scores were approximately 1 standard deviation below the population mean, and less variability in performance was also observed. Mean WIAT–II^{UK} scores were more than 1 standard deviation below population means and also evidenced reduced variability. Lower IQ and achievement scores are typically observed in referred samples (Canivez, 2014; Canivez & Watkins, 1998; Nelson *et al.*, 2013; Watkins, 2010).

Zero-order Pearson product-moment correlations between the WISC–IV^{UK} FSIQ and factor index score and WIAT–II^{UK} are presented in Table 2. Generally, the FSIQ had the highest correlations with WIAT–II^{UK} scores, but some correlations between WISC–IV^{UK} factor index scores and WIAT–II^{UK} scores were equally high. These were similar in magnitude to those reported in the WIAT–II^{UK} *UK Scoring and Normative Supplement* (Wechsler, 2005c). WISC–IV^{UK} FSIQ correlations with WIAT–II^{UK} subtests and composite scores ranged from .28 (*Pseudoword Decoding*) to .65 (*Reading Comprehension*, *Numerical Operations*). WISC–IV^{UK} VCI correlations with WIAT–II^{UK} subtests and composite scores ranged from .24 (*Pseudoword Decoding*) to .66 (*Reading Comprehension*). WISC–IV^{UK} PRI correlations with WIAT–II^{UK} subtests and composite scores ranged from .20 (Pseudoword Decoding) to .54 (Numerical Operations). WISC–IV^{UK} WMI correlations with WIAT–II^{UK} subtests and composite scores ranged from .28 (*Mathematics Composite*) to .49 (*Numerical Operations*). Consistently lower correlations were observed between the WISC–IV^{UK} PSI and WIAT–II^{UK} subtests and composite scores that ranged from −.04 (*Mathematics Reasoning*) to .48 (*Numerical Operations*). What is unknown from zero-order correlations is how much WIAT–II^{UK} subtest and composite score variance is uniquely related to the WISC–IV^{UK} factor index scores and how much is uniquely related to the FSIQ.

Table 3 presents results from H-MRA for WIAT–II^{UK} subtest and composite scores. The WISC–IV^{UK} FSIQ accounted for statistically significant (p < .0001) portions of the variance of each of the WIAT–II^{UK} subtests ranging from 7.7% (*Pseudoword Decoding*)

Table 2. Zero-order Pearson product-moment correlations between the WISC–IV^{UK} and the WIAT–II^{UK} for 1,014 referred Irish children

	WISC–IV^{UK}				
WIAT–II^{UK}	FSIQ	VCI	PRI	WMI	PSI
Word Reading	.48	.43	.35	.48	.25
Reading Comprehension	.65	.66	.51	.45	.35
Pseudoword Decoding	.28	.24	.20	.38	.10
Numerical Operations	.65	.54	.54	.49	.48
Mathematics Reasoning	.43	.39	.36	.33	–.04
Spelling	.38	.32	.26	.42	.23
Reading Composite	.56	.54	.43	.48	.28
Mathematics Composite	.48	.42	.35	.28	.09

Note. WISC–IV^{UK}, Wechsler Intelligence Scale for Children – Fourth UK Edition; WIAT–II^{UK}, Wechsler Individual Achievement Test – Second UK Edition; FSIQ, Full Scale IQ; VCI, Verbal Comprehension Index; PRI, Perceptual Reasoning Index; WMI, Working Memory Index; PSI, Processing Speed Index.
Not all 1,014 students were administered all WIAT–II^{UK} subtests, so correlations in this table are based on sample sizes noted in Table 1. Other WIAT–II^{UK} subtests and composite scores (Written Expression, Oral Expression, Writing Composite, Listening Composite, and Oral Expression Composite) were available for fewer than 20 participants and thus excluded from analyses.

to 42.5% (*Numerical Operations*). These portions of WIAT–II^{UK} subtest variance accounted for by the WISC–IV^{UK} FSIQ represented large effect sizes for Reading Comprehension and Numerical Operations; medium effect sizes for Word Reading, Mathematics Reasoning, and Spelling; and a small effect size for Pseudoword Decoding (Cohen, 1988). The WISC–IV^{UK} FSIQ accounted for statistically significant ($p < .0001$) portions of the variance of each WIAT–II^{UK} composite score ranging from 23.0% (*Mathematics Composite*) to 31.0% (*Reading Composite*). These portions of WIAT–II^{UK} achievement variance predicted from the WISC–IV^{UK} FSIQ represented medium and large effect sizes, respectively.

Also presented in Table 3 are the R^2 increases provided by the combined and unique effects of WISC–IV^{UK} factor index scores in predicting each of the six WIAT–II^{UK} subtests *after* achievement variance due to the WISC–IV^{UK} FSIQ was accounted for. Statistically significant ($p < .05$) portions of WIAT–II^{UK} subtest variance was incrementally accounted for by the combined WISC–IV^{UK} factor index scores for all subtests except Mathematics Reasoning ($n = 42$) and ranged from 1.4% (*Numerical Operations*) to 10.5% (*Mathematics Reasoning*). These increased variance portions represented small to medium effect sizes. Small effect sizes were observed for WISC–IV^{UK} factor index scores predicting WIAT–II^{UK} subtest scores for Word Reading, Reading Comprehension, Pseudoword Decoding, Numerical Operations, and Spelling, while a medium effect size was observed for Mathematics Reasoning. The *unique* contributions of WISC–IV^{UK} factor index scores in predicting each of the six WIAT–II^{UK} subtests (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (0.3–2.4%), PRI (0.1–3.0%), WMI (0.3–2.9%), and PSI (0–1.0%), and produced mostly small contributions. R^2 increases provided by the combined and unique effects of WISC–IV^{UK} factor index scores in predicting each of the two WIAT–II^{UK} composite scores *after* achievement variance due to the WISC–IV^{UK} FSIQ was accounted for are also presented in Table 3.

Table 3. Incremental contribution of observed WISC–IV^UK factor index scores in predicting WIAT–II^UK achievement subtest and composite scores in the referred Irish sample (N = 1,014)

Predictor	Word Reading (n = 1,014)		Reading Comprehension (n = 970)	
	Variance (%)	Increment[a] (%)	Variance (%)	Increment[a] (%)
FSIQ	22.6	22.6***	42.2	42.2***
Index scores (df = 4)[b]	28.2	5.6***	46.8	4.6***
VCI		0.8		1.6
PRI		0.2		0.1
WMI		2.4		0.3
PSI		0.1		0.1
	Pseudoword Decoding (n = 970)		Numerical Operations (n = 1,010)	
FSIQ	7.7	7.7***	42.5	42.5***
Index scores (df = 4)[b]	15.1	7.4***	43.9	1.4***
VCI		0.3		0.6
PRI		0.2		0.5
WMI		2.9		1.0
PSI		0		1.0
	Mathematics Reasoning (n = 45)		Spelling (n = 999)	
FSIQ	18.5	18.5*	14.1	14.1***
Index scores (df = 4)[b]	29.0	10.5	19.4	5.2***
VCI		2.4		0.4
PRI		3.0		0.1
WMI		2.6		2.4
PSI		0.6		0.3
	Reading Composite (n = 909)		Mathematics Composite (n = 42)	
FSIQ	31.0	31.0***	23.0	23.0**
Index scores (df = 4)[b]	35.7	4.7***	28.3	5.3
VCI		1.3		3.1
PRI		0.2		2.1
WMI		1.6		2.8
PSI		0.1		1.5

Note. WISC–IV^UK, Wechsler Intelligence Scale for Children – Fourth UK Edition; WIAT–II^UK, Wechsler Individual Achievement Test – Second UK 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 (semi-partial) correlations equivalent to changes in R^2 if this variable was entered last in block entry regression procedure.
[b]Partialling out FSIQ.
*p < .003, **p < .001, ***p < .0001.

A statistically significant ($p < .05$) portion of WIAT–II^{UK} composite score variance was incrementally accounted for by the combined WISC–IV^{UK} factor index scores for the Reading Composite (4.7%) but not the Mathematics Composite (5.3%), and both reflected small effect sizes. The *unique* contributions of WISC–IV^{UK} factor index scores in predicting the two WIAT–II^{UK} composite scores (based on squared part correlations from the predictor entered last in the block entry procedure) were as follows: VCI (1.3–3.1%), PRI (0.2–2.1%), WMI (1.6–2.8%), and PSI (0.1–1.5%), and were small for both the Reading and Mathematics composites.

Discussion

The present study assessed the WISC–IV^{UK} factor index score incremental validity in predicting academic achievement performance beyond the FSIQ in a large sample of Irish students referred for educational evaluations. Zero-order Pearson correlations between WISC–IV^{UK} and WIAT–II^{UK} scores were quite similar to (although slightly lower) than those from the WISC–IV^{UK} and WIAT–II^{UK} linking sample (Wechsler, 2005c). Lower correlations were expected in the present sample, given that many participants were students experiencing educational difficulties likely related to specific learning disability where achievement would be lower than expected. Lower correlations can also be partly the result of restriction of range as both WISC–IV^{UK} and WIAT–II^{UK} scores had smaller standard deviations than the population parameters (see Table 1). Some WISC–IV^{UK} factor index score correlations with WIAT–II^{UK} subtests and composites were nearly as high as for the FSIQ, but such correlations confound general intelligence variance and specific factor index variance. H-MRA addresses this by partialling out FSIQ (g) variance first to observe residual achievement variance uniquely accounted for by the WISC–IV^{UK} factor index scores.

H-MRA was used to determine the extent to which WISC–IV^{UK} factor index scores provided meaningful improvements in prediction of WIAT–II^{UK} scores *beyond* the WISC–IV^{UK} FSIQ. The WISC–IV^{UK} FSIQ provided statistically significant prediction with medium to large effect sizes for all WIAT–II^{UK} subtests examined (except Pseudoword Decoding which had a small effect size). The WISC–IV^{UK} FSIQ also provided statistically significant prediction with medium (*Mathematics Composite*) and large (*Reading Composite*) effect sizes for both WIAT–II^{UK} composite scores examined. These results are consistent with those from different intelligence tests, academic achievement measures, and samples (Canivez, 2013a; Freberg *et al.*, 2008; Glutting *et al.*, 1997, 2006; Kahana *et al.*, 2002; Nelson & Canivez, 2012; Nelson *et al.*, 2013; Ryan *et al.*, 2002; Watkins *et al.*, 2007; Youngstrom *et al.*, 1999). Statistically significant improvements in prediction of WIAT–II^{UK} subtests by the combined WISC–IV^{UK} factor index scores were observed for all subtests examined (except Mathematics Reasoning [$n = 45$]) and the Reading Composite (but not the Mathematics Composite [$n = 42$]). The lack of statistical significance for Mathematics Reasoning and the Mathematics Composite scores were likely due to the relatively small sample sizes for these variables. However, the effect sizes for incremental contribution of WISC–IV^{UK} factor index scores were mostly small and of little practical significance. To improve the incremental validity of WISC–IV^{UK}, it may be 'necessary to (1) increase the number of subtests estimating the factor scores to capture more variance, and/or (2) construct cognitive subtests that contain less g variance (and more Stratum II or broad-ability variance)' (Canivez, 2013b, p. 95).

Direct comparison of present results with those from Glutting *et al.* (2006) showed that in the present study, the combined WISC–IV^{UK} factor index scores provided somewhat greater incremental prediction of the WIAT–II^{UK} Reading Composite (4.7%) and WIAT–II^{UK} Mathematics Composite (5.3%) scores than that found in the US WISC–IV and WIAT–II linking sample (1.8% for Reading Composite and 0.3% for Mathematics Composite). Additional direct comparisons are not possible as Glutting *et al.* concentrated exclusively on the WIAT–II Reading and Mathematics composite scores. The present results are more similar to the somewhat larger portions of incremental prediction by the combined first-order factors (5–16%) reported by Glutting *et al.* (1997) using the WISC–III to predict WIAT scores with that linking sample. The present incremental prediction of WIAT–II^{UK} scores from combined WISC–IV^{UK} factor index scores was generally lower than that found with the WAIS–IV (Canivez, 2013a).

One question that might be raised is that of differential age effects. To address this issue, we divided the sample into two groups based on age (6–10 and 11–16), which produced fairly equal numbers across the eight achievement variables. Results indicated that there was generally *less* WIAT–II^{UK} achievement variance attributable to the WISC–IV^{UK} FSIQ for all subtests and composite scores for the younger group (6- to 10-year-olds) ranging from 2.9% less (*Pseudoword Decoding*) to 15.3% less (*Reading Composite*), except for the Spelling subtest (6- to 10-year-olds had 2.7% more FSIQ variance) and the Mathematics Composite (6- to 10-year-olds had 3.4% more FSIQ variance). Incremental prediction of WIAT–II^{UK} achievement variance by the combined WISC–IV^{UK} factor index scores was also lower for the younger age group (6–10) with smaller amounts ranging from 0.5% to 11.5% (*Mdn* = 7.25%) less. Thus, the incremental validity differences between the younger and older groups mostly represented small differences between the younger and older groups. Like results from the total sample, generally greater portions of WIAT–II^{UK} variance were attributed to the FSIQ. Comparisons of the older group (11- to 16-year-olds) with the WAIS–IV sample reported in Canivez (2013a) found the WISC–IV^{UK} FSIQ accounted for less WIAT–II^{UK} achievement variance than that observed in the WAIS–IV, ranging from 8.1% less (*Pseudoword Decoding*) to 51.8% less (*Mathematics Composite*) with a *Mdn* = 15.95% less achievement variance accounted for by the FSIQ. Incremental validity of factor index scores varied with the WISC–IV^{UK} factor index scores accounting for slightly more WIAT–II^{UK} variance in Reading Comprehension, Pseudoword Decoding, Numerical Operations, and Reading Composite (0.30–3.3% more), but the WAIS–IV factor index scores accounted for somewhat more WIAT–II variance in Word Reading, Math Reasoning, Spelling, and Mathematics Composite (0.60–12.8%). Speculation about why such differences were observed is difficult and may have as much to do with differences in samples as the specific tests. It would be preferable to compare the present results of WISC–IV^{UK} factor index score incremental prediction of WIAT–II^{UK} scores to WISC–IV^{UK} factor index score incremental prediction of WIAT–II^{UK} scores produced by the UK standardization linking sample, but such results are not available in a technical manual nor the extant literature, and Pearson UK denied access to the UK standardization sample raw data for such comparison. Thus, the present results must be considered in relation to replications with future WISC–IV^{UK} and WIAT–II^{UK} samples.

The WISC–IV^{UK} FSIQ typically accounted for substantially greater WIAT–II^{UK} achievement variance because it benefits from the aggregation of 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 all such investigations (i.e., Glutting *et al.*, 2006) due to the linear combination of subtests that produce both factor index scores *and* the FSIQ. However, this redundancy is precisely the problem practitioners must confront in their predictive (and explanatory) interpretations of both FSIQ *and* factor index scores in clinical assessments. Clinicians who interpret the FSIQ and then factor index scores will be counting some of the same subtest variance twice because subtest variance is in part due to hierarchical *g* and in part due to the first-order factor indexes. Such variance portions cannot be disaggregated for individuals, thus interpretation of observed FSIQ *and* factor index scores will result in redundancy. Multicollinearity does not invalidate the use of H-MRA in determining improvements in R^2 provided by WISC–IV^{UK} factor index scores and directly answers the question of improved prediction of an external criterion (Dana & Dawes, 2007; Schneider, 2008).

Incremental validity of multilevel intelligence tests is particularly important because it is based on the important scientific principle articulated by William of Ockham (alt. 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 are favoured over more complex explanations for phenomena, and in the case of intelligence test interpretation, the Full Scale score, an estimate of *g*, is a more parsimonious index than the multiple factor index scores 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.

Hale, Fiorello, Kavanagh, Hodnack, and Aloe (2007) argued that because variables entered into the multiple regression equation first capture greater portions of variance in the criterion scores than variables entered later, reversing entry of first-order factor index scores into Block 1 and the FSIQ into Block 2 would illustrate the incremental validity of the FSIQ *above and beyond the first-order factors.* Their illustration resulted in little incremental validity of the FSIQ in the second block. This is a result of the multicollinearity issue noted above. While this *can* be done, it *should not* be done and *was not* done in the present study because as Glutting *et al.* (2006) duly noted, such procedures reject the law of parsimony and thus 'repeal scientific law' (p. 106). Schneider (2008) also rejected the variable entry reversal by Hale *et al.* (2007) referring to it 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).

Due to the predictive nature of the present study and assessment of *observed* variables (WISC–IV^{UK} FSIQ and factor index scores), H-MRA were used. An alternative approach for analysis would be structural equation modelling (SEM) of latent WISC–IV^{UK} dimensions in *explaining* WIAT–II^{UK} achievement dimensional variance; however, there were too few participants with complete WIAT–II^{UK} subtests to estimate latent achievement constructs. While this is an interesting *theoretical* question (and one that might also examine rival CHC-based first-order factors when all 15 WISC–IV^{UK} subtests are administered), there are a number of problems with this approach as it applies to practitioner use of the observed scores that the WISC–IV^{UK} provides. Latent construct scores used in SEM analyses are not equivalent to the observed standard scores practitioners use, have different distributions and are not provided by the WISC–IV^{UK} (Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004). The present study was specifically interested in the incremental *prediction* provided by *observed* factor index scores beyond the FSIQ, scores that the WISC–IV^{UK} provides and that practitioners actually use.

Clinical utility of WISC–IV^{UK} factor index scores for individual diagnostic use is another important consideration and should be of interest. Incremental validity as examined in the present study pertains to *groups* and sets limits for clinical utility but does not assess clinical utility. Clinical utility is concerned with the ability of a test, comparison, or procedure to correctly predict an *individual's* membership in a clinical group and/or response to differential treatment (Meehl, 1959; Mullins-Sweatt & Widiger, 2009; Wiggins, 1988). Such diagnostic utility or diagnostic efficiency of WISC–IV^{UK} scores is as yet unknown. A number of distinct group differences comparisons are reported in the WISC–IV *Technical and Interpretive Manual* (Wechsler, 2003b), but these relate to the US samples and WISC–IV, *not* the WISC–IV^{UK} with UK samples. Examination of diagnostic utility or diagnostic efficiency of WISC–IV^{UK} factor index scores for use with these clinical groups has not yet been examined or reported. Weiner (2003) noted that even when group differences research showed statistically significant differences, it was rare that such results produced predictive power for individual diagnostic use. Given the small portions of *unique* WISC–IV^{UK} factor index score prediction of WIAT–II^{UK} achievement and the low latent factor reliabilities of WISC–IV^{UK} factor indexes (Watkins *et al.*, 2013), it is hard to imagine that the diagnostic utility of WISC–IV^{UK} factor index scores would be substantial.

Limitations

A significant limitation in the present study is that it is based on a non-random clinical sample of Irish children who were referred for educational evaluations of learning problems. As such, it would be expected that correlations between the WISC–IV^{UK} and WIAT–II^{UK} would be attenuated compared to those based on a more representative sample of the population. Such clinical samples also often have less variability, which can also attenuate correlations. Generalization to other populations is not recommended despite similar results to those obtained with demographically representative samples (Glutting *et al.*, 2006). Another limitation is that there are no other incremental validity studies of the WISC–IV^{UK} for comparison purposes. Direct comparisons to H-MRA based on the WISC–IV^{UK} and WIAT–II^{UK} standardization linking sample would be particularly informative, but the publisher denied access to the standardization data and linking samples for such analyses. Clearly, there is great need for publication of such critical psychometric information regarding the WISC–IV^{UK} standardization sample.

Conclusion

At this time, based on the present WISC–IV^{UK} incremental validity analyses, as well as WISC–IV^{UK} CFA results (Watkins *et al.*, 2013), primary interpretation of the WISC–IV^{UK} should focus on the WISC–IV^{UK} FSIQ and *not* the factor index scores. Given the low latent factor reliabilities of the four WISC–IV^{UK} factors (Watkins *et al.*) and the generally small portions of WIAT–II^{UK} achievement predicted beyond the WISC–IV^{UK} FSIQ, clinical interpretation of WISC–IV^{UK} factor index scores seems tenuous. Additionally, such low latent factor index reliability, low incremental validity, and poor longitudinal stability of factor index scores (Watkins & Canivez, 2004; Watkins & Smith, 2013) suggest that interpretation of WISC–IV^{UK} factor index score discrepancies would also be questionable. If interpretations of WISC–IV^{UK} factor index scores are made, they should be done with extreme caution and in the light of peer-reviewed empirical studies (AERA, APA, & NCME, 1999).

Acknowledgements

This research was partially supported by a 2013 Summer Research Grant from the Council on Faculty Research, Eastern Illinois University to the first author. Preliminary results were presented at the 2013 Annual Convention of the American Psychological Association, the 2014 Annual Convention of the National Association of School Psychologists, and the 9th Conference of the International Test Commission.

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: Author.

Binet, A., & Simon, T. (1904). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [New methods for the diagnosis of 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.

Brody, N. (2002). *g* and the one–many problem: Is one enough? In G. R. Bock, J. A. Goode & K. Webb (Eds.), *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 hierarchical 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. (2013a). Incremental validity of WAIS–IV factor index scores: Relationships with WIAT–II and WIAT-III subtest and composite scores. *Psychological Assessment*, *25*, 484–495. doi:10.1037/a0032092

Canivez, G. L. (2013b). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessments* (pp. 84–112). New York, NY: Oxford University Press.

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

Canivez, G. L., 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. doi:10.1037/1040-3590.10.3.285

Canivez, G. L., & Watkins, M. W. (2010a). 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

Canivez, G. L., & Watkins, M. W. (2010b). 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

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

Central Statistics Office (2013). Retrieved from http://www.cso.ie/px/pxeirestat/Statire/SelectVarVal/Define.asp?maintable=CD204

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

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21. doi:10.1016/j.intell.2006.02.001.

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly*, *28*, 154–169. doi:10.1037/spq0000010

Dombrowski, S. C. (2014a). Exploratory bifactor analysis of the WJ-III Cognitive in adulthood via the Schmid-Leiman procedure. *Journal of Psychoeducational Assessment*, *32*, 330–341. doi:10.1177/0734282913508243

Dombrowski, S. C. (2014b). Investigating the structure of the WJ-III Cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment*, *32*, 483–494. doi:10.1177/0734282914530838

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school aged analysis. *Psychological Assessment*, *25*, 442–455. doi:10.1037/a0031335

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-Second Edition: Introductory and technical handbook*. San Antonio, TX: The Psychological Corporation.

Ford, G., Andrews, R., Booth, A., Dibdin, J., Hardingham, S., & Kelly, T. P. (2008). Screening for learning disability in an adolescent forensic population. *The Journal of Forensic Psychiatry & Psychology*, *3*, 371–381. doi:10.1080/14789940802120944

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. (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., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639–662. doi:10.1080/00273171.2013.804398

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

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. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling (BSEM). *Psychological Assessment*, *25*, 496–508. doi:10.1037/a0030676

Gordon, S., Duff, S., Davidson, T., & Whitaker, S. (2010). Comparison of the WAIS-III and WISC-IV in 16-year-old special education students. *Journal of Applied Research in Intellectual Disabilities*, *23*, 197–200. doi:10.1111/j.1468-3148.2009.00538.x

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., & 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., Hodnack, 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

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

IBM Corp. (2011). *IBM SPSS statistics for Mac, version 20.0*. Armonk, NY: IBM Corp.

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

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-Second Edition*. Circle Pines, MN: AGS Publishing.

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

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 Psychology 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

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

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, *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

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

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. (1997). *Cognitive Assessment System: Interpretive handbook*. Itasca, IL: Riverside.

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

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV) with a clinical sample. *Psychological Assessment*, *25*, 618–630. doi:10.1037/a0032086

Niileksela, C. R., Reynolds, M. R., & Kaufman, A. S. (2013). An alternative Cattell–Horn–Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70–90. *Psychological Assessment*, *25*, 391–404. 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. *The 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.

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

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

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

Schneider, J. (2008). Playing statistical Ouija Board with commonality analysis (and other errors). *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

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

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., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition. *School Psychology Quarterly*, *29*, 52–63. doi:10.1037/spq0000038

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite strengths and weaknesses. *Psychological Assessment*, *16*, 133–138. doi:10.1037/1040-3590.16.2.133

Watkins, M. W., Canivez, G. L., James, T., Good, R., & James, K. (2013). Construct validity of the WISC–IV^{UK} with a large referred Irish sample. *International Journal of School and Educational Psychology*, *1*, 102–111. doi:10.1080/21683603.2013.794439

Watkins, M. W., 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., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*, 59–68. doi:10.1016/j.intell.2006.04.005

Watkins, M. W., & Smith, L. (2013). Long-term stability of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment*, *25*, 477–483. doi:10.1037/a0031653

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

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

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

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

Wechsler, D. (2005a). *Wechsler Individual Achievement Test-Second UK Edition*. London, UK: Pearson.

Wechsler, D. (2005b). *Wechsler Individual Achievement Test-Second UK Edition: Examiners manual*. London, UK: Pearson.

Wechsler, D. (2005c). *Wechsler Individual Achievement Test-Second UK Edition: UK scoring and normative supplement*. London, UK: Pearson.

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: The 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., & 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 Publishing Company.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

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 9 January 2014; revised version received 5 August 2014