Construct Validity of the WISC–IV UK With a Large Referred Irish Sample

Marley W. Watkins

Department of Educational Psychology, Baylor University, Waco, Texas, USA

Gary L. Canivez

Department of Psychology, Eastern Illinois University, Charleston, Illinois, USA

Trevor James, Kate James, and Rebecca Good

Éirim: The National Assessment Agency, Ltd., Dublin, Ireland

Irish educational psychologists frequently use the Wechsler Intelligence Scale for Children–Fourth U.K. Edition (WISC–IV UK) in clinical assessments of children with learning difficulties. Unfortunately, reliability and validity studies of the WISC–IV UK have not yet been reported. This study examined the construct validity of WISC–IV UK core subtest scores obtained from evaluations to assess learning difficulties in 794 Irish children (494 boys and 300 girls). One through four first-order factor models and indirect (higher-order) versus direct (bi-factor) hierarchical models were examined and compared using confirmatory factor analyses. The oblique four-factor Wechsler model provided the best fit to these data, but meaningful differences in fit statistics were not observed between this oblique four-factor model and rival indirect hierarchical and direct hierarchical models. For theoretical reasons, the direct (bi-factor) hierarchical model provided the best explanation of the WISC–IV UK factor structure. The general factor accounted for 63.7% of the common variance, whereas first-order factors each accounted for 8.6% to 9.6% of the common variance. Thus, the results with referred Irish children were similar to those from other investigations, further demonstrating the replication of the Wechsler Intelligence Scale for Children–Fourth Edition factor structure across cultures and the importance of focusing primary interpretation on the Full Scale Intelligence Quotient.

Keywords: Clinical assessment, cognitive assessment, intelligence, intelligence test, measurement/statistics

Wechsler intelligence scales have enjoyed immense popularity among educational (school) psychologists and clinical psychologists (Alfonso, Oakland, LaRocca, & Spanakos, 2000; Alfonso & Pratt, 1997; Belter & Piotrowski, 2001; Goh, Teslow, & Fuller, 1981; Hutton, Dubes, & Muir, 1992; Kaufman & Lichtenberger, 2000; Oakland & Hu, 1992; Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Stinnett, Havey, & Öehler-Stinnett, 1994; Watkins, Campbell, Nieberding, & Hallmark, 1995). Wechsler scales have been translated, adapted, and normed for use in other countries with different languages and cultures (Georgas, van de Vijver, Weiss, & Saklofske, 2003) and some evidence of factor invariance has been reported across cultures and between standardization and clinical samples (Chen, Keith, Weiss, Zhu, & Li, 2010; Chen & Zhu, 2012; Weiss, Keith, Zhu, & Chen, 2013a, 2013b).

During the U.S. revision of the Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991), the British version was simultaneously revised and normed for use in the United Kingdom. The Wechsler Intelligence Scale for Children–Fourth Edition (WISC–IV; Wechsler,
2003a) was a major revision that included the addition of new subtests (Picture Concepts, Letter–Number Sequencing, Matrix Reasoning, Cancellation, and Word Reasoning) and the deletion of others (Picture Arrangement, Object Assembly, and Mazes). Although the Full Scale Intelligence Quotient (FSIQ) was retained as an estimate of general intelligence, the Verbal and Performance IQs were deleted, and greater emphasis was placed on interpretation of factor index scores (Verbal Comprehension [VC], Perceptual Reasoning [PR], Working Memory [WM], and Processing Speed [PS]; Wechsler, 2003b; Weiss, Saklofske, & Prititera, 2005; Williams, Weiss, & Rolhus, 2003). The WISC–IV revision for use in the United Kingdom with U.K. norms was published one year later as the Wechsler Intelligence Scale for Children–Fourth U.K. Edition (WISC–IVUK, Wechsler, 2004).

The WISC–IVUK Administration and Scoring Manual provides a brief description of the standardization project including stratification and detailed information on administration, scoring, and analysis of index score and subtest score comparisons. Although raw score mean and standard deviation comparisons between the U.K. standardization sample and the U.S. standardization sample were provided in that manual, no further examinations of the U.K. standardization sample were reported. The WISC–IV Technical and Interpretive Manual (Wechsler, 2003b) provided with the WISC–IVUK is the version based on the U.S. standardization sample and supplemental validity samples. There is no mention in the WISC–IV Technical and Interpretive Manual of psychometric analyses with the U.K. sample.

Although the WISC–IVUK Administration and Scoring Manual states, “confidence in WISC–IVUK score interpretation is based on the extensive US standardization study” (Wechsler, 2004, p. 284), there are no reports of analyses beyond mean and standard deviation comparisons with the U.S. sample. Raw score means and standard deviations were similar between the U.K. and U.S. samples (Wechsler, 2004); however, reliability estimates and standard errors of measurement were based on the larger U.S. sample and no validity data were presented for the U.K. sample. Searches of the extant literature produced no studies reporting on the psychometric features of the WISC–IVUK with the standardization sample. Do other psychometric features of the WISC–IV based on U.S. samples generalize to children in the United Kingdom? Without extensive psychometric examination of the reliability, validity, and diagnostic efficiency/utility with the U.K. standardization sample much remains unknown regarding proper interpretation of WISC–IVUK scores.

Examination and reporting on the internal structure of the WISC–IV provided in the WISC–IV Technical and Interpretive Manual (Wechsler, 2003b) did not include a higher-order factor analysis to verify and describe the implied and theoretical structure of the WISC–IV. Three independent WISC–IV studies (Bodin, Pardini, Burns, & Stevens, 2009; Watkins, Wilson, Kotz, Carbone, & Babula, 2006) examined the higher-order structure of the WISC–IV and found that the majority of subtest variance was associated with the higher-order general intelligence dimension and substantially smaller amounts of variance were related to the first-order factors. This is a consistent finding among Wechsler scales, specifically, as also observed with the French WISC–IV (Golay, Reverte, Rossier, Favez, & Lecerf, 2012), the French Wechsler Adult Intelligence Scale–Third Edition (WAIS–III; Golay & Lecerf, 2011), and the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV; Canivez & Watkins, 2010a, 2010b), as well as intelligence tests, in general (Canivez, 2008, 2011; Canivez, Konold, Collins, & Wilson, 2009; Dombrowski & Watkins, 2013; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012, Nelson, Canivez, Lindstrom, & Hatt, 2007).

WISC–IV content and structure reflect current conceptualizations of intelligence articulated by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Cattell, 1966) and other WISC–IV internal structure studies have examined alternate structural models based on the Cattell–Horn–Carroll (CHC; McGrew, 1997, 2005) framework. Support for CHC-based structural models of the WISC–IV has been reported (Chen, Keith, Chen, & Chang, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010; Weiss et al., 2013b), where the basic Wechsler structure is retained for subtests and associations with the VC ($G_v$), WM ($G_m$)—except arithmetic—and PS ($G_s$); but, the PR dimension is divided into two CHC factors where Block Design and Picture Completion measure visual processing ($G_s$) and Matrix Reasoning and Picture Concepts measure fluid reasoning ($G_f$). However, standardized paths from $g$ to $G_f$ were 1.00 (Keith, 2005; Keith et al., 2006; Weiss et al., 2013b) with the U.S. standardization sample, 0.98 (Chen et al., 2009) with the Taiwan WISC–IV, and 1.00 (Lecerf et al., 2010) with the French WISC–IV basic CHC model patterned after Keith et al. (2006); but only 0.84 with the final modified six-factor CHC model of the French WISC–IV (Lecerf et al., 2010). This indicated that $G_f$ was isomorphic with the higher-order $g$ factor and not supportive of a CHC model. The exception was the modified six-factor CHC model of the French WISC–IV that suggested cultural differences (Lecerf et al., 2010). Isomorphism of $G_f$ with higher-order $g$ has also been observed in studies of various versions of the WAIS (Benson, Hulac, & Kranzler, 2010; Golay & Lecerf, 2011; Weiss et al., 2013a), but a recent study has suggested that isomorphism of $G_f$ with higher-order $g$ may be an artifact of confirmatory factor analysis (CFA) statistical procedures (Golay et al., 2012). It is also possible that results of Golay et al. may be unique to the French WISC–IV.
Educational psychologists in the Republic of Ireland frequently use the WISC–IV<sub>UK</sub> with the U.K. norms in clinical evaluations but there are no separate norms for Irish children. There are no equivalence or validity studies examining WISC–IV performance of Irish children compared to British or American children so, as with British children, proper interpretation of the WISC–IV<sub>UK</sub> scores is unknown. To examine the construct validity of the WISC–IV<sub>UK</sub> with an Irish sample, CFAs were used to test various theoretical models to determine the best fitting models identical to those examined by Watkins (2010). Based on results of Watkins (2010), Golay et al. (2012), and Gignac (2005, 2006), it was hypothesized that the direct hierarchical (bi-factor model as originally specified by Holzinger & Swineford, 1937) model allowing the general intelligence factor to directly influence WISC–IV<sub>UK</sub> subtest performance would best explain the WISC–IV<sub>UK</sub> structure with a sample of Irish children. Although some have examined a number of CHC-inspired theoretical structures, this requires all 15 WISC–IV subtests to be administered, which most clinicians rarely do (Watkins, 2010). Because data currently available for Irish children included only the 10 core WISC–IV<sub>UK</sub> subtests, CHC-based structures could not be examined.

METHOD

Participants and Procedures

Participants were 794 Irish children from the Republic of Ireland between the ages of 6 years, 0 months to 16 years, 9 months who were referred to an educational psychologist for evaluation of learning difficulties. Some children were referred for evaluation by their parents, but the vast majority of children were referred by their schools to determine eligibility for special education services or accommodations. Participants resided in the five major cities (Cork, Dublin, Galway, Limerick, and Waterford) in Ireland (19%), as well as in small towns and rural areas (81%). The largest portion of the sample were boys ($n = 494; 62.2\%$), as is typically observed in educational evaluation referrals. The mean age of the sample was 10.74 years ($SD = 2.56$) and bimodal in nature, with peaks at 8 and 12 years of age. This represented three to four years following entry into primary schools and entrance into post-primary school, respectively. Unfortunately, agency practice and confidentiality standards allowed no other demographic information to be included in this archival dataset.

All WISC–IV<sub>UK</sub> administrations were conducted by one of three educational psychologists according to the standardized procedure. Only children with complete data for all 10 core subtests were included in analyses. Institutional review board approval was obtained but all data were de-identified and no personal information included.

Instrument

The WISC–IV (Wechsler, 2003a) is a test of general intelligence and is composed of 15 subtests ($Ms = 10$, $SDs = 3$), 10 of which are mandatory and contribute to measurement of four factor-based index scores: Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. Each of the four indexes is expressed as a standard score ($Ms = 100$, $SDs = 15$). The FSIQ is composed of 10 core subtests (three VC, three PR, two WM, and two PS). The WISC–IV was anglicised and adapted for the United Kingdom in 2002 through item review and minor changes in items or language, spelling, and order of item difficulty (Wechsler, 2004). The resulting WISC–IV<sub>UK</sub> was standardized and normed on a sample of 780 children between the ages of 6 years, 0 months and 16 years, 11 months who were representative of the U.K. population stratified by geographic region, gender, race/ethnicity, and parent education level (Wechsler, 2004). 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. Reliability and validity data based on the WISC–IV<sub>UK</sub> standardization sample were not provided in the WISC–IV<sub>UK</sub> manual and standard errors of measurement were taken from the U.S. version of the WISC–IV.

Analyses

Mplus 7 for Macintosh (Muthén & Muthén, 2012) was used to conduct CFAs using maximum likelihood estimation. Consistent with previous WISC–IV structural analyses, four first-order models and two hierarchical models were specified and examined: (a) one factor; (b) two oblique verbal and nonverbal factors; (c) three oblique verbal, perceptual, and combined working memory/processing speed factors; (d) four oblique verbal, perceptual, working memory, and processing speed factors; (e) an indirect hierarchical (higher-order) model (as per Bodin et al., 2009) with four first-order factors; and (f) a direct hierarchical (bi-factor) model (as per Watkins, 2010) with four first-order factors. See Gignac (2008) for a detailed description of direct and indirect hierarchical models.

Although contentious (Marsh, Hau, & Wen, 2004), Hu and Bentler (1999, 1999) recommended a dual criterion to guard against both Type-1 and Type-2 errors with values of .95 for the comparative fit index (CFI) and .06 for the root mean square error of approximation (RMSEA). Higher CFI values and lower RMSEA values indicate better fit. These two indexes were supplemented with chi-square and Akaike information criterion (AIC) values. Nonsignificant chi-square values tend to indicate good model fit. Smaller AIC values indicate better fit after accounting for model complexity. Not all models were nested, so meaningful
differences between well-fitting models were evaluated using $\Delta \text{CFI} > .01$ (Cheung & Rensvold, 2002) and $\Delta \text{RMSEA} > .015$ (Chen, 2007) as standards.

Finally, latent factor reliabilities were estimated with coefficient omega ($\omega$) and omega hierarchical ($\omega_h$) as programmed by Watkins (2013). Omega estimated the reliability of the latent factor that combines the general and specific factor variance, whereas omega hierarchical (what Reise, 2012, termed the “omega subscale”) estimated the reliability of the latent factor with the general factor variance removed (Brunner, Nagy, & Wilhelm, 2012).

RESULTS

Descriptive statistics for participants’ mean WISC–IVUK subtest, factor index, and FSIQ scores are presented in Table 1 and illustrate univariate normality with the largest skewness index of $-0.52$ and the largest kurtosis index of $-0.33$. Mardia’s (1970) standardized multivariate kurtosis estimate for these data was $3.27$ and well under the criterion of $5.0$ for multivariate normality (Byrne, 2006). WISC–IVUK means for this sample were approximately $1 SD$ lower than the normative means and there was less variability observed among participants. Lower subtest, factor index, and FSIQ scores in referred samples are frequently observed (Canivez & Watkins, 1998; Watkins, 2010).

Model fit statistics presented in Table 2 illustrate the increasingly better fit from one to four factors; however, fit statistics indicated that the one-, two-, and three-factor models were inadequate. The correlated four-factor (VC, PR, WM, and PS) model provided the best fit to these data, but meaningful differences in fit statistics (CFI and RMSEA) were not observed between the four, first-order factor (see Figure 1), indirect hierarchical (see Figure 2), and direct hierarchical (see Figure 3) models. Because the four WISC–IVUK latent factors were highly correlated, a higher-order structure is implied (Gorsuch, 1988), making the correlated four-factor model an inadequate explanation of the factor structure. Both the direct and indirect hierarchical models exhibited good fit according to Hu and Bentler’s (1998, 1999) dual criteria. Neither was statistically superior to the other, but the direct hierarchical model offers several benefits (Brunner et al., 2012; Reise, 2012), so it was selected as the best explanation of the WISC–IVUK factor structure.

Table 3 presents decomposed WISC–IVUK subtest variance estimates based on the direct hierarchical model. The general factor accounted for $63.7\%$ of the common variance and $36.7\%$ of the total variance, the VC factor accounted for $9.4\%$ of the common variance and $5.4\%$ of total variance, the PR factor accounted for $8.6\%$ of the common variance and $5.0\%$ of total variance, the WM factor accounted for $8.7\%$ of the common variance and $5.0\%$ of the total variance, and the PS factor accounted for $9.6\%$ of the common variance and $5.5\%$ of the total variance (see Table 3). Thus, the higher-order $g$ factor accounted for substantially greater

### Table 1

<table>
<thead>
<tr>
<th>Scale</th>
<th>$M$</th>
<th>$SD$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Design</td>
<td>7.71</td>
<td>2.71</td>
<td>+.35</td>
<td>+.19</td>
</tr>
<tr>
<td>Similarities</td>
<td>8.46</td>
<td>2.74</td>
<td>+.22</td>
<td>-.33</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.62</td>
<td>2.63</td>
<td>+.17</td>
<td>+.31</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>9.15</td>
<td>2.81</td>
<td>-.30</td>
<td>+.31</td>
</tr>
<tr>
<td>Coding</td>
<td>8.15</td>
<td>2.76</td>
<td>+.31</td>
<td>+.21</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>6.86</td>
<td>2.65</td>
<td>+.33</td>
<td>+.12</td>
</tr>
<tr>
<td>Letter–Number Sequencing</td>
<td>7.80</td>
<td>2.63</td>
<td>-.52</td>
<td>-.27</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>7.32</td>
<td>2.70</td>
<td>+.11</td>
<td>-.26</td>
</tr>
<tr>
<td>Comprehension</td>
<td>8.07</td>
<td>2.85</td>
<td>-.12</td>
<td>-.04</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>8.41</td>
<td>2.73</td>
<td>-.34</td>
<td>+.01</td>
</tr>
<tr>
<td>Verbal Comprehension Index</td>
<td>87.19</td>
<td>13.90</td>
<td>+.08</td>
<td>+.02</td>
</tr>
<tr>
<td>Perceptual Reasoning Index</td>
<td>88.14</td>
<td>13.51</td>
<td>-.03</td>
<td>+.03</td>
</tr>
<tr>
<td>Working Memory Index</td>
<td>86.53</td>
<td>12.85</td>
<td>-.22</td>
<td>-.07</td>
</tr>
<tr>
<td>Perceptual Speed Index</td>
<td>90.40</td>
<td>13.30</td>
<td>+.03</td>
<td>+.19</td>
</tr>
<tr>
<td>FSIQ</td>
<td>84.92</td>
<td>13.06</td>
<td>+.03</td>
<td>+.06</td>
</tr>
</tbody>
</table>

Note. Mardia’s (1970) multivariate kurtosis was $3.27$. WISC–IVUK = Wechsler Intelligence Scale for Children–Fourth U.K. Edition; FSIQ = Full Scale Intelligence Quotient.

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>CFI</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>One factor</td>
<td>437.56</td>
<td>35</td>
<td>.867</td>
<td>.120</td>
<td>.110–.131</td>
<td>35,834.8</td>
</tr>
<tr>
<td>Two factors (V &amp; NV)</td>
<td>303.48</td>
<td>34</td>
<td>.911</td>
<td>.100</td>
<td>.090–.110</td>
<td>35,702.8</td>
</tr>
<tr>
<td>Three factors (VC, PR, &amp; WM + PS)</td>
<td>186.90</td>
<td>32</td>
<td>.949</td>
<td>.078</td>
<td>.067–.089</td>
<td>35,590.2</td>
</tr>
<tr>
<td>Four factors (VC, PR, WM, &amp; PS)</td>
<td>62.09</td>
<td>29</td>
<td>.989</td>
<td>.038</td>
<td>.025–.051</td>
<td>35,471.4</td>
</tr>
<tr>
<td>Indirect hierarchical</td>
<td>85.62</td>
<td>31</td>
<td>.982</td>
<td>.047</td>
<td>.035–.059</td>
<td>35,490.9</td>
</tr>
<tr>
<td>Direct hierarchical$^*$</td>
<td>78.14</td>
<td>27</td>
<td>.983</td>
<td>.049</td>
<td>.036–.062</td>
<td>35,491.4</td>
</tr>
</tbody>
</table>

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike information criterion; V = Verbal; NV = Nonverbal; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed. In the Wechsler first-order four-factor model, correlations are between the following: VC and PR = .81, VC and WM = .63, VC and PS = .49, PR and WM = .61, PR and PS = .63, and WM and PS = .55.

$^*$Two indicators of third and fourth factors were constrained to be equal to ensure identification.
portions of WISC–IV UK common and total variance relative to the factor index scores. Omega hierarchical (ωh) coefficients presented in Table 3 estimated the reliability of the latent constructs with the effects of other constructs removed. In the case of the four WISC–IV UK factor indexes, ωh coefficients estimated the scale reliabilities with the effects of the general factor removed and ranged from .143 (PR) to .376 (PS).

DISCUSSION

Factor analyses in this study of Irish children administered the WISC–IV UK in clinical evaluations provided strong replication of previous examinations of the internal structure of the WISC–IV (Bodin et al., 2009; Keith, 2005; Lecerf et al., 2006; Watkins, 2006, 2010; Wechsler, 2003b; Weiss et al., 2013b) with all 10 core subtests providing measurement of a broad general intelligence dimension and four specific first-order dimensions (VC, PR, WM, and PS). These results are also similar to those found with other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Gignac, 2005, 2006; Weiss et al., 2013a).

More specifically, these analyses supported the direct hierarchical (bi-factor) model, as have others (Gignac, 2005, 2006; Golay & Lecerf, 2011; Watkins, 2010). By specifying a direct hierarchical model, influences of g are direct to the
subtests as are influences of the four primary factors (VC, PR, WM, and PS), rather than subtest influences of g being mediated by the four specific factors prescribed in a higher-order model. The direct hierarchical (bi-factor) model allows g to be closer to the indicators (subtests) and g is conceptualized more as a breadth factor, rather than a superordinate factor (Gignac, 2008). This seems more consistent with Spearman’s (1904, 1927) conceptualization of general intelligence. By placing the general factor at the same level as the specific factors the direct hierarchical model is not really “hierarchical” as is the higher-order model that has dominated research on the structure of intelligence tests in the United States.

Decomposed variance estimates based on the direct hierarchical (bi-factor) model (see Figure 3) presented in Table 3 illustrate that the greatest portions of subtest variance were associated with the g factor and smaller portions of variance were associated with the four primary factors. Numerous studies of Wechsler scales and other intelligence tests have consistently found that the greatest portions of total and common variance are apportioned to the second-order g dimension (or bi-factor/direct hierarchical g), which is estimated by the FSIQ score, and much smaller portions of total and common variance are apportioned to the first-order or specific dimensions, estimated by the respective factor index scores. This has been documented for the WISC–IV (Bodin

et al., 2009; Watkins, 2006; Watkins et al., 2006), French WISC–IV (Golay et al., 2012), Stanford–Binet Intelligence Scales–Fifth Edition (Roid, 2003; see also Canivez, 2008), Wechsler Abbreviated Scale of Intelligence (Psychological Corporation, 1999) and Wide Range Intelligence Test (Glutting, Adams, & Sheslow, 2000; see also Canivez et al., 2009), Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003; see also Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007), Cognitive Assessment System (Naglieri & Das, 1997; see also Canivez, 2011), French WAIS–III (Golay & Lecerf, 2011), WAIS–IV (Canivez & Watkins, 2010a, 2010b; Niileksela et al., 2012), and the Woodcock–Johnson–Third Edition Psychoeduca-
tional Battery (Woodcock, McGrew, & Mather, 2001; see also Dombrowski & Watkins, 2013). The implication of these consistent findings is that the overall, omnibus FSIQ score should retain primary interpretive weight, rather than the first-order, specific, factor-based index scores.

Examination of reliability of the latent constructs indicated that the broad g factor had strong estimates allowing individual interpretation ($\omega = .904$, $\omega_h = .802$), but the $\omega_h$ estimates for the four WISC–IV UK narrow specific factors were very low (.143–.376) and extremely limited for measuring unique constructs (Brunner et al., 2012; Reise, 2012) and not high enough for individual interpretation. For comparison purposes, standardized path

---

coefficients from Watkins (2010) were used to calculate omega hierarchical and present results were quite similar. The $\omega_h$ estimates for the four WISC–IV narrow specific factors from Watkins (2010) were also very low (.112–.388). Canivez (in press) also reported very low $\omega_h$ coefficients for the four WISC–IV specific factors (.098–.330) in a sample of referred children demographically similar to Watkins (2010). In contrast to cross-battery (Flanagan, Alfonso, & Ortiz, 2012) and clinical (Weiss et al., 2005) interpretation approaches, these results further support primary interpretation of the FSIQ for the WISC–IV UK.

Limitations

Limitations of this study are primarily that of a restricted and nonrandom clinical sample of Irish students referred for evaluations of educational difficulties. Generalization to other populations is not recommended, despite the identical or similar results obtained with normative samples or large referred samples outside of Ireland. As no psychometric studies of the WISC–IV UK with British (normative or clinical) or Irish samples are presently available, it is impossible to know how the structure based on this sample compares to the British normative sample or to a normative Irish sample. Clearly there is great need for publication of such critical psychometric information for the WISC–IV UK normative sample.

CONCLUSION

Based on these results and strong replication of previous findings it seems prudent to focus WISC–IV UK interpretation at the FSIQ level and if going beyond the FSIQ to interpret factor index scores with extreme caution so as not to misinterpret or over-interpret scores given the small unique variance provided by the factor index scores. This conclusion is consistent with an extensive analysis of alternative methods of interpretation of intelligence tests, which recommended that “clinicians should restrain their clinical interpretations to the FSIQ score in most, if not all, instances” (Canivez, 2013, p. 96). Nevertheless, the WISC–IV UK structure should be examined in relation to external variables or criteria such as academic achievement to determine what, if any, reliable achievement variance is incrementally accounted for by the WISC–IV UK factor index scores beyond that accounted for by the FSIQ, as well as diagnostic utility studies.

TABLE 3

<table>
<thead>
<tr>
<th>Subtest</th>
<th>General b</th>
<th>VC Var b</th>
<th>PR Var b</th>
<th>WM Var b</th>
<th>PS Var b</th>
<th>$h^2$</th>
<th>$u^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarities</td>
<td>.725</td>
<td>.526</td>
<td>.411</td>
<td>.169</td>
<td>.695</td>
<td>.305</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.763</td>
<td>.582</td>
<td>.477</td>
<td>.228</td>
<td>.810</td>
<td>.190</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>.648</td>
<td>.420</td>
<td>.385</td>
<td>.148</td>
<td>.568</td>
<td>.432</td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>.654</td>
<td>.428</td>
<td>.152</td>
<td>.023</td>
<td>.451</td>
<td>.549</td>
<td></td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>.604</td>
<td>.365</td>
<td>.071</td>
<td>.005</td>
<td>.370</td>
<td>.630</td>
<td></td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>.652</td>
<td>.425</td>
<td>.684</td>
<td>.468</td>
<td>.893</td>
<td>.107</td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>.459</td>
<td>.211</td>
<td>.501</td>
<td>.251</td>
<td>.462</td>
<td>.538</td>
<td></td>
</tr>
<tr>
<td>Letter–Number Sequencing</td>
<td>.563</td>
<td>.317</td>
<td>.500</td>
<td>.250</td>
<td>.567</td>
<td>.433</td>
<td></td>
</tr>
<tr>
<td>Coding</td>
<td>.389</td>
<td>.151</td>
<td></td>
<td></td>
<td>.523</td>
<td>.274</td>
<td></td>
</tr>
<tr>
<td>Symbol Search</td>
<td>.500</td>
<td>.250</td>
<td></td>
<td></td>
<td>.528</td>
<td>.279</td>
<td></td>
</tr>
<tr>
<td>Total variance (%)</td>
<td>36.7</td>
<td>5.4</td>
<td>5.0</td>
<td>5.0</td>
<td>5.5</td>
<td>57.7</td>
<td></td>
</tr>
<tr>
<td>Common variance (%)</td>
<td>63.7</td>
<td>9.4</td>
<td>8.6</td>
<td>8.7</td>
<td>9.6</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>.904</td>
<td>.870</td>
<td>.777</td>
<td>.678</td>
<td>.644</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$\omega_h$</td>
<td>.802</td>
<td>.228</td>
<td>.143</td>
<td>.332</td>
<td>.376</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Note. $N = 794$. b = standardized loading of subtest on factor; Var = variance explained in the subtest; $h^2 =$ communality; $u^2 =$ uniqueness; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; $\omega = \omega$; $\omega_h = \omega$ hierarchical.

ACKNOWLEDGEMENTS

This research was partially supported by a 2012 Summer Research Grant to Gary L. Canivez from the Council on Faculty Research, Eastern Illinois University. Preliminary analyses were presented at the 2012 annual convention of the American Psychological Association and the 8th conference of the International Test Commission.

ABOUT THE AUTHORS

Marley W. Watkins is Professor and Chairman of the Department of Educational Psychology at Baylor University in Waco, TX. His research interests include the study of individual differences.

Gary L. Canivez is Professor of Psychology at Eastern Illinois University principally involved in training school psychologists.

Trevor James, psychologist at Éirim: The National Assessment Agency, who offers individual psycho-educational assessment and training in educational assessment in Ireland.
Kate James, psychologist at Éirim: The National Assessment Agency, who offers individual psycho-educational assessment and training in educational assessment in Ireland.

Rebecca Good, psychologist at Éirim: The National Assessment Agency, who offers individual psycho-educational assessment and training in educational assessment in Ireland.

REFERENCES


