Construct Validity of the WISC-IV With a Referred Sample:
Direct Versus Indirect Hierarchical Structures

Gary L. Canivez
Eastern Illinois University

The Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) is one of the most frequently used intelligence tests in clinical assessments of children with learning difficulties. Construct validity studies of the WISC-IV have generally supported the higher order structure with four correlated first-order factors and one higher-order general intelligence factor, but recent studies have supported an alternate model in which general intelligence is conceptualized as a breadth factor rather than a superordinate factor (M. W. Watkins, 2010, Structure of the Wechsler Intelligence Scale for Children–Fourth Edition among a national sample of referred students, Psychological Assessment, Vol. 22, pp. 782–787; M. W. Watkins, G. L. Canivez, T. James, K. & R. Good, in press, Construct validity of the WISC-IVUK with a large referred Irish sample, International Journal of School and Educational Psychology). WISC-IV core subtest data obtained from evaluations to assess learning difficulties in 345 children (224 boys, 121 girls) were examined. One through four, first order factor models and indirect versus direct hierarchical models were compared using confirmatory factor analyses. The correlated four-factor Wechsler model provided good fit to these data, but the direct hierarchical model showed statistically significant improvement over the indirect hierarchical model and correlated four-factor model. The direct hierarchical model was judged the best explanation of the WISC-IV factor structure, with the general factor accounting for 71.6% of the common variance while the first order factors accounted for 2.4–10.3% of the common variance. Thus, the results with the present sample of referred children were similar to those from other investigations (G. E. Gignac, 2005, Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling, Assessment, Vol. 12, pp. 320–329; G. E. Gignac, 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, Vol. 27, pp. 73–86; P. Golay, I. Reverte, J. Rossier, N. Pavez, & T. Lecerf, 2012, Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. Psychological Assessment, advance online publication; M. W. Watkins, 2010, Structure of the Wechsler Intelligence Scale for Children–Fourth Edition among a national sample of referred students, Psychological Assessment, Vol. 22, pp. 782–787; M. W. Watkins, G. L. Canivez, T. James, K. & R. Good, in press, Construct validity of the WISC-IVUK with a large referred Irish sample, International Journal of School and Educational Psychology) supporting primary interpretation of the Full Scale IQ rather than the factor index scores.

Keywords: confirmatory factor analysis, hierarchical CFA, WISC-IV, intelligence

Wechsler intelligence scales are repeatedly reported among the most frequently used intelligence tests among school 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; Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Stinnett, Havey, & Oehler-Stinnett, 1994; Watkins, Campbell, Nieberding, & Hallmark, 1995). Wechsler scales have been translated, adapted, and normed for use in other countries (Georgas, van de Vijver, Weiss, & Saklofske, 2003), and evidence of factor invariance across culture and between standardization and clinical samples...
has been reported (Chen, Keith, Weiss, Zhu, & Li, 2010; Chen & Zhu, 2012; Weiss, Keith, Zhu, & Chen, 2013a, 2013b).

The Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) was a major revision of the WISC-III (Wechsler, 1991) that included the deletion of some subtests (i.e., Picture Arrangement, Object Assembly, and Mazes) and the addition of new subtests (i.e., Picture Concepts, Letter–Number Sequencing, Matrix Reasoning, Cancellation, and Word Reasoning). The Full Scale IQ (FSIQ) was retained as an estimate of general intelligence, but the Verbal and Performance IQs were deleted, and emphasis was placed on interpretation of factor index scores (i.e., Verbal Comprehension [VC], Perceptual Reasoning [PR], Working Memory [WM], and Processing Speed [PS]; Prifitera, Saklofske, & Weiss, 2008; Wechsler, 2003b; Weiss, Saklofske, & Prifitera, 2005; Weiss, Saklofske, Prifitera, & Holdnack, 2006; Williams, Weiss, & Rolhus, 2003).

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 internal structure. Four studies have examined the higher order structure of the WISC-IV or WISC-IVUK (Wechsler, 2004), and all four found that the majority of subtest variance was associated with the higher order general intelligence dimension and considerably smaller amounts of variance were related to the first-order factors in U.S. and Irish samples (Bodin, Pardini, Burns, & Stevens, 2009; Watkins, 2006; Watkins, Canivez, James, James, & Good, in press; Watkins, Wilson, Kotz, Carbone, & Babula, 2006). This finding was consistently observed among various versions of Wechsler scales, including the French WISC-IV (Golay et al., 2012), the French Wechsler Adult Intelligence Scale–Third Edition (WAIS-III; Golay & Lecerf, 2011), and the WAIS-IV (Canivez & Lecerf, 2010a, 2010b; Niileksela, Reynolds, & Kaufman, 2012). Other intelligence tests also have larger portions of subtest variance associated with a higher order general factor than their first order factors, including the Stanford–Binet Intelligence Scales, Fifth Edition (SB-5; Roid, 2003; Canivez, 2008), Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler, 1999), Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000; Canivez, Konold, Collins, & Wilson, 2009), Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003; Dombrowski, Watkins, & Brogan, 2009; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007), the Cognitive Assessment System (CAS; Naglieri & Das, 1997; Canivez, 2011), and the Woodcock–Johnson Psycho-Educational Battery–III (WJ-III; Woodcock, McGrew, & Mather, 2001; McGrew & Woodcock, 2001; Dombrowski & Watkins, 2013).

The structure and content of the WISC-IV reflect current conceptualizations of intelligence articulated by Carroll, Cattell, and Horn (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Cattell, 1966), and some internal structure studies have examined alternate WISC-IV structural models based on the Cattell–Horn–Carroll (CHC; McGrew, 1997, 2005) framework. Some have reported support for CHC-based structural models of the WISC-IV (Chen, Keith, Chen, & Chang, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010; Weiss et al., 2013a), and although the basic Wechsler structure is retained for subtests and their associations within the VC (Gv), WM (Gsm) (except Arithmetic), and PS (Gs) factors, the PR dimension is divided into two CHC factors in which Block Design and Picture Completion are designated as visual processing (Gv) while Matrix Reasoning and Picture Concepts are designated as fluid reasoning (Gf). WISC-IV standardized paths from g to Gf, however, were 1.0 in the U.S. standardization sample (Keith, 2005; Keith et al., 2006; Weiss et al., 2013a), 0.98 with the Taiwan WISC-IV (Chen et al., 2009), and 1.0 with the French WISC-IV basic CHC model (Lecerf et al., 2010) that was patterned after Keith et al. (2006). The standardized path from g to Gf in the final modified six-factor CHC model of the French WISC-IV, however, was .84 (Lecerf et al., 2010). This indicated that Gf was often isomorphic with the higher order g factor and, thus, does not support a CHC-based model. The modified six-factor CHC model of the French WISC-IV was an exception, and it was suggested it could be related to cultural differences (Lecerf et al., 2010). Gf and higher order g isomorphism was also observed in studies of various versions of the WAIS (Benson, Hulac, & Kranzler, 2010;
Golay & Lecerf, 2011; Weiss et al., 2013b), but a recent study of the French WISC-IV has suggested that \( G_f \) and higher order \( g \) isomorphism may be an artifact of confirmatory factor analysis (CFA) statistical procedures (Golay et al., 2012), but could also be unique to the French WISC-IV.

Most hierarchical structure studies of the Wechsler scales have examined the higher order relationship of \( g \) with the first order factors, with effects of \( g \) going through first order factors with direct effects to the subtests. With hierarchical (higher order) models, it is important to note that such models obfuscate the role and influence of the second-order \( (g) \) factor on the subtests and may produce an illusion of strength of influence of the first order factors. Intelligence test subtests are influenced by both first order factors and the higher order \( g \) factor, which is why Carroll (1995) insisted on the use of the Schmid and Leiman (1957) transformation to apportion subtest variance to the first order and higher order dimensions. Interpretation of higher order models requires this partitioning of variance in order for the clinician to determine the relative value of the first order factors compared to the higher order factor. An alternate conceptualization of general intelligence was originally specified by Holzinger and Swineford (1937) as the bifactor model in which a hierarchy of \( g \) and the first order factors did not exist. Rather, both the general \( (g) \) and the specific factors had direct paths and influences on the subtests, allowing effects of \( g \) on the subtests to be direct rather than indirect. Gignac (2005, 2006, 2008) referred to the bifactor model as a direct hierarchical model (the \( g \) factor directly influenced subtests) in contrast to the indirect hierarchical model that is a higher order model (where the \( g \) factor influences subtests indirectly through the first order factors). Reise (2012) noted several advantages of the direct hierarchical (bifactor) model, including: the general factor having direct influences is easy to interpret, both general and specific influences on indicators (subtests) can be examined simultaneously, and examining the psychometric properties is necessary for determining scoring and interpretation of subscales. The direct hierarchical model can also be considered a more parsimonious model (Gignac, 2006) and more consistent with Spearman’s (1904, 1927) conceptualization of general intelligence.

Presently, the only CFA examination of the WISC-IV comparing direct hierarchical and indirect hierarchical structures with a sample of disabled youths is Watkins (2010), and although the sample was based on data from 355 students reported by 93 school psychologists from 34 states, replication is warranted. A similar examination of the WISC-IVUK (Watkins et al., in press) also compared direct versus indirect hierarchical models. Based on previous results (Gignac, 2005, 2006; Golay et al., 2012; Watkins, 2010; Watkins et al., in press), it was hypothesized that the direct hierarchical model (bifactor model as originally specified by Holzinger & Swineford, 1937) allowing the general intelligence factor to directly influence WISC-IV subtest performance would best explain the WISC-IV structure with a referred sample. Although CHC-inspired theoretical structures have been examined in a number of other studies, this requires data from administration of all 15 WISC-IV subtests, which most clinicians rarely administer (Watkins, 2010; Watkins et al., in press). Understanding the latent structure of the 15 core subtests is thus important, but because data available for referred students in the present study included only the 10 core WISC-IV subtests, CHC-based structures could not be examined. Understanding the WISC-IV latent structure of the 10 core subtests is important because it is that version that is most frequently used and interpreted and which has norms for interpretation of FSIQ and index scores.

**Method**

**Participants and Procedure**

Participants were 345 children between the ages of 6–0 to 16–1 who were referred for evaluation of learning difficulties and assessed with the WISC-IV by one of 30 school psychologists who worked in one medium-sized school district between 2003 and 2009. The number of WISC-IV assessments completed per school psychologist ranged from 1–52, with a mean of 10.73 (\( SD = 13.77 \)) and a median of 5. Students ranged in grade from kindergarten to Grade 10, with boys comprising a larger portion of the sample (\( n = 224, 65\% \)) as is often observed in educational disability referrals. The mean age of the sample was 10.25 years (\( SD = 2.25 \)). Of the
345, 124 (35.9%) were White and 182 (52.8%) were African American or Black. The two largest groups of remaining participants were bira
cial (n = 17, 4.9%) and Hispanic or Latino (n = 8, 2.3%). English was the primary language for 338 (98%) of the children, and those who spoke
another language, such as Spanish (n = 4, 1.2%) or Romanian (n = 1, 0.3%), were also English proficient for proper administration of the WISC-IV. It was reported that 51 (14.8%)
were retained in grade at least once and the most common disabilities were specific learning disabili
ity (n = 189, 54.8%), intellectual disability (n = 39, 11.3%), serious emotional disability (n = 24, 7.0%), and speech or language disabi
lity (n = 23, 6.7%), with 41 (11.9%) not being identified as disabled. Evaluations were fairly evenly distributed between initial evaluations
(n = 159, 46.1%) and reevaluations (n = 179, 51.9%), with seven cases not specified.

State-certified school psychologists provided all WISC-IV administrations according to the standardized procedures. Only children with complete data for all 10 core subtests were included in analyses. Analysis of WISC-IV core and supplemental subtests was considered, but only three (0.9%) cases included all 15 subtests. This appears to be common in clinical evaluations, in which frequently only the core subtests are administered. Institutional review board approval was obtained, and data analyzed were deidentified to maintain anonymity. Data were obtained through archival record review, and data were entered into the computer for analyses by pairs of trained under
graduate and graduate student research assistants (one read coding sheets while the other keyed data into the computer).

Instrument

The WISC-IV (Wechsler, 2003a) is a test of general intelligence and is composed of 15 sub
tests (Ms = 10, SDs = 3), 10 of which are mandatory and contribute to measurement of four factor-based index scores: Verbal Compre
hension 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 subtests (three VC, three PR, two WM, and two PS).

Analyses

EQS, Version 6.2 (Bentler & Wu, 2012) was used to conduct CFA 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 (see Figure 1); (e) an indirect hierarchical model (as per Bodin et al., 2009), with four first-order factors (see Figure 2); and (f) a direct hierarchical model (as per Watkins, 2010), with four first-order factors (see Figure 3). Because the 10 subtest WISC-IV configuration only has two subtest indicators for WM and PS, the two subtests were constrained to be equal in the direct hierarchical model to ensure specification as conducted by Watkins (2010) and Watkins et al. (in press). Gignac (2008) has provided a detailed description of direct and indirect hierarchical models, but the direct hierarchical model is the bifactor model described by Holzinger and Swineford (1937).

Hu and Bentler (1999, 1999) recommended a dual criterion of CFA fit statistic values of .95 for the comparative fit index (CFI) and .06 for the root mean square error of approximation (RMSEA) to guard against both Type I and Type II errors, but Marsh, Hau, and Wen (2004) warned about overgeneralizing these criteria. As Watkins (2010) noted, the Hu and Bentler criteria appear appropriate for studies of vari
ables included in the present study and were so applied. Higher CFI values and lower RMSEA values indicate better model fit, and these two indices were supplemented with chi-square and Akaike information criterion (AIC) values. Smaller AIC values indicate better fit after ac
counting for model complexity. Meaningful dif
terences between well fitting models were eval
uated using ΔCFI greater than .01 (Cheung & Rensvold, 2002) and ΔRMSEA greater than −.015 (Chen, 2007) as standards, because not all models were nested.

Finally, latent factor reliabilities were esti
mated with coefficient omega (ω) and omega hierarchical (ω_h), as programmed by Watkins (2013). Omega estimates the reliability of the latent factor combining the general and specific
factor variance, while omega hierarchical, what Reise (2012) termed the Omega subscale, estimates the reliability of the latent factor with all other latent construct variance removed (Brunner, Nagy, & Wilhelm, 2012).

Results

Descriptive statistics for participants’ mean WISC-IV subtest scores are presented in Table 1 and illustrate the univariate normality with the largest skewness index of .51 and the largest kurtosis index of −.55. Mardia’s standardized multivariate kurtosis estimate for these data was 1.17 and well under the criterion of 5.0 suggesting multivariate normality (Byrne, 2006). WISC-IV factor index scores and FSIQ were also reasonably normally distributed (see Table 1). WISC-IV subtests, factor indexes, and FSIQ means for this sample were approximately 1 SD lower than the normative means, and there was slightly less variability observed among participants on some scores than is typically observed.
in referred samples (Canivez & Watkins, 1998; Watkins, 2010; Watkins et al., in press).

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 (CFI < .95 and/or RMSEA > .06). The correlated four-factor (VC, PR, WM, PS) model (see Figure 1) provided the best fit to these data among the first order models, but, based on chi-square difference tests, the direct hierarchical model was a statistically better fit to these data than both the correlated four-factor model ($\Delta \chi^2 = 10.72$, $\Delta df = 2$, $p < .01$) and the indirect hierarchical model ($\Delta \chi^2 = 14.33$, $\Delta df = 4$, $p < .01$). Because the four WISC-IV latent factors were highly correlated, a higher order structure is implied (Gorsuch, 1988), rendering the correlated four-factor model an inadequate explanation of the WISC-IV factor structure. Based on Hu and Bentler’s (1998, 1999) dual criteria, both the direct hierarchical model and the indirect hierarchical model produced good fit. Meaningful differences in fit statistics based on criteria from Cheung and Rensvold (2002; $\Delta$CFI > .01) and Chen (2007; $\Delta$RMSEA > −.015) were not observed between the four correlated first-order factor (see Figure

Figure 2. Indirect hierarchical measurement model, with standardized coefficients, for the Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003a) for 345 referred children. SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter–Number Sequencing; CD = Coding; SS = Symbol Search; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; $g$ = general intelligence.
1), the indirect hierarchical (see Figure 2), and the direct hierarchical (see Figure 3) models. The direct hierarchical model produced the lowest AIC estimate, and, given the advantages (Brunner et al., 2012; Gignac, 2006; Reise, 2012), the direct hierarchical model was selected as the best explanation of the WISC-IV factor structure.

Table 3 presents the decomposed subtest variance estimates of the WISC-IV based on the direct hierarchical model, and the general factor accounted for 71.6% of the common variance and 43.6% of the total variance. The VC factor accounted for 10.3% of the common variance and 6.3% of total variance, the PR factor accounted for 7.4% of the common variance and 4.5% of total variance, the WM factor accounted for 2.4% of the common variance and 1.4% of the total variance, and the PS factor accounted for 8.3% of the common variance and 5.0% of the total variance (see Table 3). Thus, the higher order g factor accounted for substantially greater portions of WISC-IV common and total variance relative to the factor index scores. Omega hierarchical coefficients presented in Table 3 provide estimates of the reliability of the latent constructs with the effects of other constructs.
reduced. In the case of the four WISC-IV factor indexes, omega hierarchical coefficients estimated the scale reliabilities with the effects of the general factor removed, and ranged from .098 (WM) to .330 (PS).

Discussion

WISC-IV CFAs in the present study of referred children provide replication of previous examinations of the internal structure of the WISC-IV (Bodin et al., 2009; Keith, 2005; Lecerf et al., 2010; Watkins, 2006, 2010; Watkins et al., in press; Wechsler, 2003b; Weiss et al., 2013a), with all 10 core subtests contributing to the measurement of a broad general intelligence dimension and four specific dimensions (i.e., VC, PR, WM, and PS). These results are also similar to results found with other versions of Wechsler scales (Canivez & Watkins, 2010a, 2010b; Gignac, 2005, 2006; Weiss et al., 2013b).

The present analyses support the direct hierarchical (bifactor) model, as have others (Gig-
The direct hierarchical model (see Figure 3) specifies direct influences of \( g \) to the subtests (indicators) rather than subtest influences of \( g \) being mediated by the four specific factors prescribed in an indirect hierarchical (higher-order) model (Figure 2). Gignac (2006) argued that it is unreasonable or unrealistic to require full mediation of \( g \) through first order dimensions, and that “it is arguably more congruent and reasonable to specifically model the most significant factor of a battery of tests (i.e., ‘\( g \’) directly, rather than indirectly, through first-order factors” (p. 85). The direct hierarchical (bifactor) model allows \( g \) to be closer to the subtests, and \( g \) is conceptualized more as a breadth factor rather than as a superordinate factor (Gignac, 2008); and by specifying a unidimensional general factor, it can be considered more parsimonious than the indirect hierarchical model (Gignac, 2006). 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 intelligence test internal structure research. The direct hierarchical model is also more consistent with Spearman’s (1904, 1927) conceptualization of general intelligence as well as Wechsler’s (1939, 1958) definition of intelligence.

Decomposed variance estimates based on the direct hierarchical (bifactor) model (see Figure 3) presented in Table 3 illustrate that the greatest portions of subtest variance were associated with the \( g \) factor and that the smaller portions of variance were associated with the four specific 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 bifactor/direct hierarchical \( g \)), which is estimated by the Full Scale 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 observed in the WISC-IV (Bodin et al., 2009; Watkins, 2006; Watkins et al., in press; Watkins et al., 2006), French WISC-IV (Golay et al.,

Table 3

<table>
<thead>
<tr>
<th>Subtest</th>
<th>General b Variance</th>
<th>Verbal Comprehension b Variance</th>
<th>Perceptual Reasoning b Variance</th>
<th>Working Memory b Variance</th>
<th>Processing Speed b Variance</th>
<th>( h^2 )</th>
<th>( u^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarities</td>
<td>.691</td>
<td>.477</td>
<td>.417</td>
<td>.174</td>
<td>.651</td>
<td>.349</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.742</td>
<td>.551</td>
<td>.525</td>
<td>.276</td>
<td>.826</td>
<td>.174</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>.675</td>
<td>.456</td>
<td>.423</td>
<td>.179</td>
<td>.635</td>
<td>.365</td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>.708</td>
<td>.501</td>
<td></td>
<td></td>
<td></td>
<td>.867</td>
<td>.133</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>.663</td>
<td>.440</td>
<td>.052</td>
<td>.003</td>
<td>.442</td>
<td>.558</td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>.741</td>
<td>.549</td>
<td>.290</td>
<td>.084</td>
<td>.633</td>
<td>.367</td>
<td></td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>.561</td>
<td>.315</td>
<td></td>
<td></td>
<td></td>
<td>.394</td>
<td>.606</td>
</tr>
<tr>
<td>Letter–Number Sequencing</td>
<td>.692</td>
<td>.479</td>
<td>.254</td>
<td>.065</td>
<td></td>
<td>.543</td>
<td>.457</td>
</tr>
<tr>
<td>Coding</td>
<td>.405</td>
<td>.164</td>
<td></td>
<td></td>
<td></td>
<td>.545</td>
<td>.297</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>.652</td>
<td>.425</td>
<td></td>
<td></td>
<td></td>
<td>.454</td>
<td>.206</td>
</tr>
<tr>
<td>% Total variance</td>
<td>43.6</td>
<td>6.3</td>
<td>4.5</td>
<td>1.4</td>
<td>5.0</td>
<td>60.8</td>
<td>39.2</td>
</tr>
<tr>
<td>% Common variance</td>
<td>71.6</td>
<td>10.3</td>
<td>7.4</td>
<td>2.4</td>
<td>8.3</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>( \omega )</td>
<td>.923</td>
<td>.877</td>
<td>.835</td>
<td>.636</td>
<td>.700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \omega_n )</td>
<td>.843</td>
<td>.259</td>
<td>.140</td>
<td>.098</td>
<td>.330</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. WISC-IV = Wechsler Intelligence Scale for Children–Fourth Edition; Variance = variance explained in subtest; \( h^2 \) = communality; \( u^2 \) = uniqueness.
2012), SB-5 (Canivez, 2008), WASI and WRIT (Canivez et al., 2009), RIAS (Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007), CAS (Canivez, 2011), French WAIS-III (Golay & Lecerf, 2011), WAIS-IV (Canivez & Watkins, 2010a, 2010b; Niileksela et al., 2012), and the WJ-III (Dombrowski & Watkins, 2013). These consistent findings imply that the overall omnibus FSIQ should retain primary interpretive weight rather than the specific factor-based index scores. The particularly low variance estimates of PR on Picture Concepts and Matrix Reasoning and of WM on Digit Span and Letter–Number Sequencing (see Table 3) were also quite low in the WISC-IV standardization sample (Watkins, 2006) and another referred sample (Watkins et al., 2006).

Examination of reliability of the latent WISC-IV constructs indicated that although the broad g factor had strong estimates allowing individual interpretation (ω = .923, ωh = .843), the omega hierarchical estimates for the four WISC-IV narrow specific factors were very low (.098–.330), 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 the Watkins (2010) direct hierarchical model were used to calculate omega hierarchical estimates, and the present results were quite similar. The omega hierarchical estimates for the four WISC-IV narrow specific factors from Watkins (2010) were also low (.112–.388), as were reported omega hierarchical estimates for the four WISC-IVUK specific factors (.143–.376) with a large sample of referred Irish children (Watkins et al., in press). Thus the dominance of the general factor supports primary interpretation of the FSIQ for the WISC-IV rather than the four factor index scores.

Study Limitations

The limitations of the present study are primarily those of a geographically restricted and nonrandom clinical sample of 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 with large referred samples. Another limitation is that use of archival data restricted analyses to only the 10 core WISC-IV subtests, because the majority (99.1%) of children evaluated by school psychologists in the present study did not administer the supplemental subtests that would have permitted examination of alternate CHC-based configurations. Although this is a limitation, the present analyses address the structure of the WISC-IV based on the subtests that are most likely to be administered in clinical evaluations, and there are noted problems related to CHC models when applied to the WISC-IV and Wechsler scales in general (Canivez & Kush, 2013).

Conclusion

Based on the present results and replication of previous findings, it is appropriate to continue the recommendation of concentrating primary WISC-IV interpretation on the FSIQ and, when going beyond the FSIQ in interpreting factor index scores, clinicians must exercise extreme caution to guard against misinterpretation or overinterpretation of scores, given the poor reliability of the latent specific WISC-IV factors. Although recommendations for clinical interpretation of the WISC-IV often focus on the specific factor index scores (Prifitera et al., 2008; Wechsler, 2003b; Weiss et al., 2005; Weiss et al., 2006; Williams et al., 2003), data from the present study and those of others ( Bodin et al., 2009; Golay et al., 2012; Watkins, 2006, 2010; Watkins et al., in press; Watkins et al., 2006) do not support such recommendations. WISC-IV validity should continue to be examined in relation to external variables and criteria, such as academic achievement, to determine what reliable achievement variance is incrementally accounted for by the factor index scores, beyond that accounted for by the FSIQ. Diagnostic utility studies of WISC-IV scores would also be helpful (see Canivez, 2013).

References

Canivez, G. L., Konold, T. R., Collins, J. M., Structural equation modeling
Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, EQS for
Bentler, P. M., & Wu, E. J. C. (2012). EQS for
Bentler, P. M., & Wu, E. J. C. (2012). EQS for
Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, EQS for
Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school aged anal-
CONSTRUCT VALIDITY OF THE WISC-IV

ysis. Psychological Assessment. Advance online publication. doi:10.1037/a0031335


Received March 12, 2013
Revision received April 17, 2013
Accepted May 16, 2013