Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests

CrossMark

Stefan C. Dombrowski [a,*], Gary L. Canivez [b], Marley W. Watkins [c], A. Alexander Beaujean [c]

[a] Department of Graduate Education, Leadership and Counseling, Rider University, United States
[b] Department of Psychology, Eastern Illinois University, United States
[c] Department of Educational Psychology, Baylor University, United States

A B S T R A C T

Despite substantial revisions involved in creating the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014a), the test publisher relied exclusively upon confirmatory factor analytic procedures to determine the instrument's structure and failed to apportion the variance among factors and subtests. To fill this lacuna, the factor structure of the 16 primary and secondary subtests of the WISC-V standardization sample was examined with exploratory bifactor analysis (EBFA). EBFA results provided strong support for a general intelligence (g) factor, but nominal evidence for three group factors (i.e., Processing Speed, Working Memory, and Perceptual Reasoning). There was no evidence for distinct verbal, fluid reasoning or visual–spatial factors. The g factor accounted for large portions of total and common subtest variance while the group factors accounted for negligible portions of total and common variance. These results suggest that clinical interpretation of the WISC-V should reside primarily at the global level (i.e., Full Scale IQ).

© 2015 Published by Elsevier Inc.

1. Introduction

The developers of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler, 2014a) stated that they not only used current cognitive, intellectual, and neuropsychological theories (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Horn & Blankson, 2012; Horn & Cattell, 1966; McCloskey, Whitaker, Murphy, & Rogers, 2012; Miller & Maricle, 2012) to guide its creation, but also retained its long-standing linkage to Spearman's (1904) notion of general intelligence. Evidence of structural validity was established via confirmatory factor analyses (CFA) and reported in the WISC-V *Technical and Interpretive Manual* (Wechsler, 2014b), which included the specification of higher-order factor models with a single second-order general intelligence (g) factor *indirectly* influencing subtests through five first-order factors. However, scholars have raised a number of concerns regarding that structure (Canivez & Watkins, in press; Canivez, Watkins, & Dombrowski, 2015). Canivez and Watkins (in press) and Canivez, Watkins, James, James, and Good (2014) noted that there was insufficient detail in describing how the factors were defined and why weighted least squares estimation was used. For example, WLS estimation is typically used with categorical or non-normal data, requires much larger sample sizes, and can lead to model misspecification more readily

than maximum likelihood estimation (Hu, Bentler, & Kano, 1992; Olsson, Foss, Troye, & Howell, 2000; Yuan & Chan, 2005). Canivez and colleagues also indicated that the preferred CFA model abandoned the parsimony of simple structure by allowing cross-loadings of the Arithmetic subtest. Further, there was a standardized path coefficient of 1.00 between the higher-order general intelligence factor and the first-order Fluid Reasoning (FR) factor, suggesting that g and FR were empirically redundant (Le, Schmidt, Harter, & Lauver, 2010). Canivez et al. also expressed concern about the use of chi-square difference tests of nested models to identify the five-factor model because this approach has been shown to be misleading when the base model is misspecified (Yuan & Bentler, 2004) and is overly powerful with large sample sizes (Millsap, 2007).

There are five additional issues with the test publisher's approach to documenting the WISC-V structure. First, the test publisher did not examine rival models, such as a bifactor model. Bifactor models are sometimes preferred over higher-order models (Canivez, in press; Reise, 2012) and have been recommended for tests of cognitive ability because they allow for partitioning of general and group factor variance (Beaujean, Parkin, & Parker, 2014; Canivez, 2014b; Canivez et al., 2015, 2014; Carroll, 1997; Gignac, 2005, 2006; Gignac & Watkins, 2013; Nelson, Canivez, & Watkins, 2013; Watkins, 2010; Watkins & Beaujean, 2014; Watkins, Canivez, James, James, & Good, 2013; Brunner, Nagy, & Wilhelm, 2012) and are more in line with Carroll's three-stratum theory cognitive ability (Beaujean, 2015). This inclusion would aid clinicians and researchers in determining the interpretability

* Corresponding author at: School Psychology Program, 2083 Lawrenceville Road, Lawrenceville, NJ 08648, United States.
E-mail address: sdombrowski@rider.edu (S.C. Dombrowski).

of group factors (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Gustafsson & Aberg-Bengtsson, 2010).

Second, model-based reliability estimates including omega-hierarchical (ω_h) and omega-subscale (ω_s) (Gignac & Watkins, 2013; Reise, 2012; Reise, Bonifay, & Haviland, 2013; Shrout & Lane, 2012; Zinbarg, Revelle, Yovel, & Li, 2005, 2009) were not included in the *Technical and Interpretive Manual.* Several researchers (e.g. Canivez, 2010; Canivez, 2014a; Canivez & Kush, 2013) as well as the *Standards for educational and psychological testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), have emphasized the need for these statistics in IQ test manuals that recommend the interpretation of subscores. Along with the measurement of total and common variance for general- and group/specific, ω estimates can aid in determining how much interpretive emphasis should be placed upon scores designed to measure primary and secondary factors.

Third, the WISC-V authors did not furnish EFA results; instead they relied exclusively upon CFA procedures when providing structural validity evidence. Gorsuch (1983) and others (e.g. Brown, 2015; Carroll, 1993; Reise, 2012) indicated that EFA and CFA are complementary, and test users can have greater confidence in an instrument's structure when both procedures are in agreement, particularly when an instrument has been revised and reformulated. For instance, elimination of the Word Reasoning and Picture Completion subtests and addition of Visual Puzzles, Figure Weights, and Picture Span subtests could have caused unexpected changes to the WISC-V factor structure that would benefit from EFA prior to the use of CFA (Strauss, Spreen, & Hunter, 2000).

Fourth, previous independent investigations of intelligence test factor structures using EFA methods have produced divergent results from those offered by CFA-based models of extant IQ subtests (e.g. Canivez, 2008; Canivez & Watkins, 2010a, 2010b; DiStefano & Dombrowski, 2006; Dombrowski, 2013; Dombrowski, 2014a, 2014b; Dombrowski & Watkins, 2013; Dombrowski, Watkins, & Brogan, 2009; Watkins, 2006). In fact, some researchers contend that present day IQ tests are overfactored (Frazier & Youngstrom, 2007).

Finally, Canivez et al. (2015, 2014) recently subjected the WISC-V total sample correlation matrix to EFA using the Schmid–Leiman (SL) procedure. The SL procedure mathematically transforms a second-order factor solution into an orthogonal first-order structure where general and group factors both directly influence indicator variables. Schmid and Leiman (1957) argued that this process "preserves the desired characteristics of the oblique solution" and "discloses the hierarchical structure of the variables" (p. 53). Carroll (1995) also emphasized that orthogonal factors are appropriate only when produced in the context of a Schmid–Leiman solution. Canivez et al.'s SL analysis resulted in a four-factor solution where the fluid reasoning and visual spatial subtests combined to form the WISC-IV's previously identified perceptual reasoning factor. Additionally, their results revealed the preeminence of the higher-order *g* factor and prompted them to recommend that primary interpretative emphasis should be placed on the FSIQ with possible secondary interpretive emphasis on the processing speed index score.

Although useful, the SL procedure is simply a re-parameterization of the higher-order model to show how the measured variables relate to the second-order factor and residualized versions of the first-order factors. As with higher-order models in general, loading values from the SL transformation may be biased if there are cross-loadings (Reise, 2012). Likewise, the loadings of all measured variables on a group factor are constrained to be proportional (Schmiedek & Li, 2004). Given these issues, Jennrich and Bentler (2011) developed an alternative to the SL procedure for EFA: exploratory bifactor analysis (EBFA). They described EBFA as "simply exploratory factor analysis using a bi-factor rotation criterion" (p. 2). EBFA is designed to estimate loadings from bifactor models directly, which Jennrich and Bentler contend can be better

than the SL transformation in some cases. The only independently published article comparing the two procedures on cognitive ability data, however, found consistent results between EBFA and the SL (Dombrowski, 2014b).

2. Method

2.1. Participants

Participants included the entire WISC-V standardization sample ($N = 2200$), ranging in age from 6 to 16 years. Detailed demographic characteristics are available in the WISC-V *Technical and Interpretive Manual* (Wechsler, 2014b). The standardization sample was obtained using stratified proportional sampling across variables of age, sex, race/ethnicity, parental education level, and geographic region. Education level was used as a proxy for socioeconomic status. Examination of tables in the *Technical and Interpretive Manual* (Wechsler, 2014b) revealed a close match to the U.S. census across stratification variables.

2.2. Instrument

The WISC-V is an individually administered test of cognitive ability for children aged 6–16 years. The Full Scale IQ (FSIQ) is composed of seven primary subtests across five domains [Verbal Comprehension (VC), Visual Spatial (VS), Fluid Reasoning (FR), Working Memory (WM), and Processing Speed (PS)]. The Primary Index Scale level is composed of 10 WISC-V subtests (primary subtests) that are used to estimate the five WISC-V factor index scores (VCI, VSI, FRI, WMI, PSI). Fig. 1 illustrates the publisher's proposed latent factor structure.

2.3. Procedure and analyses

The WISC-V subtest correlation matrix for the total standardization sample was obtained from the *Technical and Interpretive Manual* (Table 5.1; Wechsler, 2014b). In addition to purposefully extracting 2 to 5 factors to map onto the CFA models posited in the WISC-V *Technical and Interpretive Manual*, multiple empirical factor extraction criteria were also examined (Gorsuch, 1983) as well as factor interpretability and compliance with simple structure (Thurstone, 1947). Specifically, eigenvalues > 1 (Kaiser, 1960), the scree test (Cattell, 1966), standard error of scree (SE_{Scree}; Zoski & Jurs, 1996), Horn's parallel analysis (HPA; Horn, 1965), and minimum average partials (MAP; Velicer, 1976) were examined. After determining the number of factors to extract, factors were then rotated using the bifactor rotation with orthogonal group factors (Jennrich & Bentler, 2011). All EBFA analyses were conducted using the R statistical programming language (R Development Core Team, 2015) using the BaylorEdPsych and Psych packages (cf. Beaujean, 2013, 2014; Revelle, 2012). Omega estimates (ω_h and ω_s) were produced using the *Omega* program developed by Watkins (2013).

3. Results

3.1. Factor extraction criteria comparisons

MAP suggested one factor; eigenvalue > 1, scree, and HPA suggested 2–3 factors; whereas the test publisher recommended five factors. Given that it is better to overextract than underextract (Wood, Tataryn, & Gorsuch, 1996), and to attempt a replication of the publisher's five-factor model, we began by extracting five factors. Models with four, three, and two factors were also successively examined for adequacy.