Assessment
1–15

DOI: 10.1177/1073191116677799
asm.sagepub.com
SAGE

What Does the WRAML2 Core Battery Measure? Utilizing Exploratory and Confirmatory Techniques to Disclose Higher Order Structure

Ryan J. McGill[1] and Stefan C. Dombrowski[2]

Abstract

The present study examined the factor structure of the Wide Range Assessment of Memory and Learning–Second Edition (WRAML2) core battery with participants from the normative sample aged 9 to 90 years (n = 880) using higher order exploratory and confirmatory factor analytic techniques that were not reported in the in the WRAML2 *Administration and Technical Manual*. Exploratory factor analysis results suggested only one factor, whereas confirmatory factor analysis results favored the three factors posited by the test authors. Although model fit statistics were equivalent for the oblique, indirect hierarchical, and direct hierarchical measurement models, it was determined that the bifactor model best disclosed the influence of latent dimensions on WRAML2 manifest variables. In the three-factor bifactor model, the general factor accounted for 31% of the total variance and 69% of the common variance, whereas the three first-order factors combined accounted for 41% of the total variance and 31% of the common variance. Latent factor reliability coefficients (as estimated by ω_h) indicated that only the general factor was measured with enough precision to warrant confident clinical interpretation. Implications for clinical interpretation of WRAML2 scores and the procedures utilized in the development of related measures are discussed.

Keywords

WRAML2, factor analysis, structural validity, bifactor model

The Wide Range Assessment of Memory and Learning–Second Edition (WRAML2; Sheslow & Adams, 2003a) is an individually administered test battery designed to assess memory ability in children and adults aged 5 to 90 years and is widely utilized by neuropsychologists, clinical psychologists, and school psychologists. The WRAML2 is a major structural and conceptual revision of the WRAML (Sheslow & Adams, 1990); specifically, the age range for the instrument was extended to be inclusive of adult participants, the core battery was reduced from nine subtests to six, and the theoretical foundation for the battery was updated.

The WRAML2 is composed of six core subtests that combine to yield three index scores (Verbal Memory, Visual Memory, and Attention/Concentration) as well as a full-scale global memory composite (General Memory Index [GMI]). Additionally, users can elect to administer up to 11 optional subtests that can yield additional clinical index scores though it should be noted that these indicators were not included in the WRAML2 structural validation studies reported in the *Administration and Technical Manual* (Sheslow & Adams, 2003b). Thus, their potential relationship to the aforementioned first-order indexes and second-order global composite is presently unknown. In terms of clinical interpretation, the Technical Manual suggests that users should interpret the scores obtained from the WRAML2 in a stepwise fashion beginning with the GMI and then proceeding to more specific measures (e.g., indexes and subtests).

Although users are advised to focus most of their interpretive weight on the GMI, additional consideration of performance on the first-order dimensions is encouraged. For instance, the Technical Manual suggests that discrepant performance across the scales may be clinically noteworthy (p. 67) and base rates for observed differences are reported in supplementary tables (Tables A.8-A.9, pp. 217-219). Although no evidence is provided to support the use of these procedures in clinical practice, it should be noted that discrepant performance on memory indexes such as those provided

[1]College of William & Mary, Williamsburg, VA, USA
[2]Rider University, Lawrenceville, NJ, USA

Corresponding Author:
Ryan J. McGill, School of Education, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187, USA.
Email: rmcgill@wm.edu

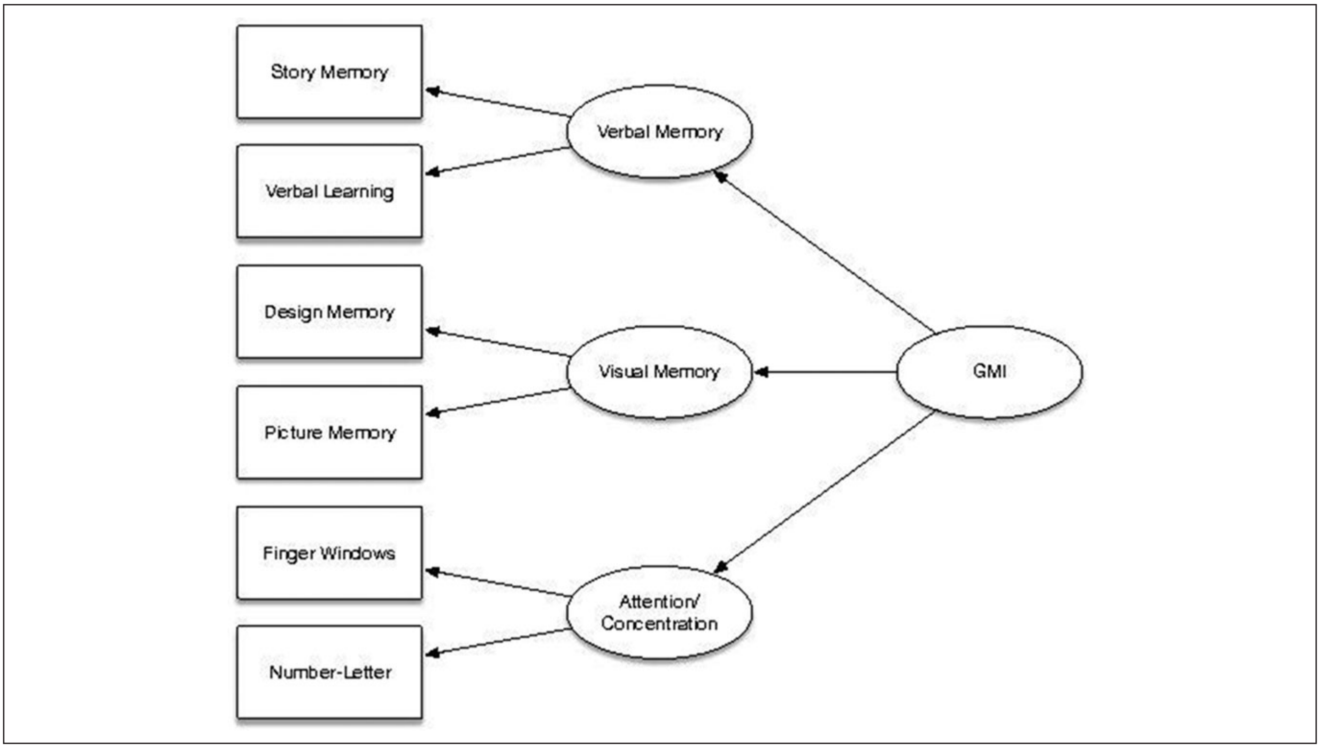

Figure 1. Implied hierarchical model for the Wide Range Assessment of Memory and Learning–Second Edition (Sheslow & Adams, 2003a) core battery.

on the WRAML2 has long been noted as a questionable diagnostic sign for lateralization difficulties and more focal cortical insults in clinical populations (e.g., Engle & Smith, 2010; Loring, Lee, Martin, & Meador, 1989).

Although a series of exploratory and confirmatory factor analytic techniques were used to evaluate the internal structure of the measurement instrument, considerable problems remain. Several of these concerns involve the choice of the procedures employed for exploratory factor analysis (EFA).

For EFA, the subtest intercorrelation matrix for the entire WRAML2 normative sample (N = 1,200) was subjected to a principal components analysis (PCA) with an oblique rotation (rotation not specified) with a forced extraction based on a *predicted* three-factor structure for the measurement instrument, in spite of the fact that the WRAML2 was designed to be atheoretical (Hartman, 2007). Initial results indicated that all WRAML2 subtest alignments were salient and consistent with the predicted model thus demonstrating desired simple structure. These results were later replicated when applying the forced three-factor extraction procedure to five age groups (5-10, 11-20, 21-40, 41-60, and 61-90) spanning the entire normative sample (Tables 7.19 to 7.23). Nevertheless, the use of a constrained analytical approach in which factor extraction is based on subjective inference rather than more robust empirical criteria (e.g., parallel analysis, minimum average partials [MAP]) is problematic as it is essentially using EFA in a confirmatory context, a practice that has been critiqued extensively within the empirical literature (Costello & Osborne, 2005; Haig, 2005; Mulaik, 1987; Thompson, 2004).

Further degrading the utility of the EFA analyses undertaken was the choice to examine latent structure via PCA. Although these validation procedures have been referred to as a *factor analysis* in both the Technical Manual and professional literature (e.g., Adams, 2013; Maricle, Miller, & Mortimer, 2011; Strauss, Sherman, & Spreen, 2006), PCA is not considered to be a factor analytic procedure due to the fact that its algorithm differs mathematically from the assumptions of the common factor model (Fabrigar & Wegener, 2012; Gorsuch, 1983). Although it has been argued that there are negligible differences between principal components and common factor analysis (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Velicer & Jackson, 1990), components analysis is computed without regard for the influence of latent variables and does not discriminate between different dimensions of variance in the manifest variables (e.g., shared and unique variance). As a result, the components derived from PCA should not be interpreted as reflecting latent dimensions such as memory and learning abilities (Bentler & Kano, 1990; Preacher & MacCallum, 2003).

It is also unknown why the test authors decided to forgo an explication of higher order structure in any meaningful way given the implied structure of the test (see Figure 1). The test authors correctly employed an oblique rotation under the assumption of correlated factors. Although an oblique rotation is necessary, it may not be singularly sufficient and an additional step is required. According to Gorsuch (1983), higher order factors are implicit in all oblique rotations; consequently, Gorsuch recommended that second-order factors be extracted and examined. Unfortunately, Sheslow and Adams (2003b) did not use second-order factor analysis in both the EFA procedures they employed.

According to Carroll (1995), all cognitive measures are composed of reliable variance that is attributable to a higher order general factor, reliable variance that is attributable to first-order group factors, and error variance. Because of this, Carroll argued that variance from the higher order factor must be extracted first to residualize the lower order factors, leaving them orthogonal to the higher order dimension. Thus, variability associated with a higher order factor is accounted for before interpreting variability associated with lower order factors, resulting in variance being apportioned to higher order and lower order dimensions. To accomplish this task, Carroll (1993, 1995) recommended second-order factor analysis of first-order factor correlations followed by a Schmid–Leiman transformation (SL; Schmid & Leiman, 1957). The SL technique allows for the orthogonalization of higher order variance from lower order factors and results in an approximate bifactor solution (Beaujean, 2015). The bifactor procedure mathematically separates variance attributable to the general factor and to group-specific factors by directly residualizing variance assigned to each level (e.g., see Figure 4). In contrast, an indirect hierarchical model has general factor variance mediated through group factors (e.g., see Figure 1) and may not present as lucid an accounting of variance apportionment making clinical interpretation more difficult. A bifactor solution permits the practitioner to more clearly understand how the variance attributable to the general factor (i.e., the WRAML-2 GMI score) and to group factors (i.e., WRAML-2 Index scores) should be assigned and how much interpretative emphasis should be placed on each level of the instrument (e.g., full scale vs. index; Canivez, 2016). If the variance attributable to group factors is low, then interpretive emphasis should reside at the full-scale level. The SL technique has been used by many researchers to investigate the approximate bifactor structure of tests of cognitive ability (e.g., Canivez, 2011; Canivez & Watkins, 2010; Canivez, Watkins, & Dombrowski, 2016; Dombrowski, 2013; Dombrowski, McGill, & Canivez, 2016).

EFA results in the Technical Manual were additionally supported by confirmatory factor analysis (CFA) results that compared competing models with one, two, and three correlated factors with the core subtests to the entire normative sample. Of the three first-order models, the three oblique factor model was the best fitting with factor correlations ranging from .51 to .73. Because these three factors are highly correlated, a higher order or hierarchical structure is implied and should be explicated (Gorsuch, 1983; Thompson & Daniel, 1996). However, the oblique three-factor model did not include a higher order dimension despite the fact that the Technical Manual (Sheslow & Adams, 2003b) posits a hierarchical structure with a global composite (serving as a proxy for a higher order general memory dimension) at the apex of the model (Kranzler & Keith, 1999). According to McClain (1996), it is a mistake to interpret a second-order factor on the basis of first-order dimensions, as it can lead to overinterpretation of lower order factors in clinical practice.

Although independent structural validity investigations of the WRAML2 have been scarce, several studies provided inconsistent support for a similar oblique three-factor structure posited in its predecessor in both clinical and normative samples (e.g., Aylward, Gioia, Verhulst, & Bell, 1995; Burton, Donders, & Mittenberg, 1996; Dewey, Kaplan, & Crawford, 1997; Gioia, 1998; Phelps, 1995). The EFA results produced by Gioia (1998) are particularly noteworthy as he was the only researcher to attempt to explicate second-order structure using the SL procedure, as insisted by Carroll (1993, 1995). In doing so, it was found that the vast majority of WRAML subtest variance was attributable to a higher order general memory dimension and that the residual variance afforded to the three first-order group factors was consistently weak, calling into question their potential clinical utility. Additionally, he critiqued the use of PCA as a proxy for common factor analysis and the failure to explicate higher order structure in the original WRAML validation analyses. Thus, it is surprising that the SL procedure, or any form of higher order analyses, was overlooked when developing the WRAML2 as the authors cited the Gioia (1998) study within the Technical Manual and appeared to be aware of the serious questions raised about the structure of the WRAML. Also missing from the WRAML2 Technical Manual were proportions of variance accounted for by the hypothesized second-order factor and the three first-order factors, second-order subtest loadings, subtest specificity estimates, and model-based reliability estimates including omega coefficients (ω; Canivez, 2016; Rodriguez, Reise, & Haviland, 2016). The body of literature on factor analysis methodology (e.g., Carroll, 1993, 1995; Gorsuch, 1983; Thompson, 2004) and model-based reliability (e.g., Reise, 2012; Reise, Bonifay, & Haviland, 2013) recommends the inclusion of this information because it assists test users in determining how the instrument should be interpreted.

Since its publication, WRAML2 validity studies have largely focused on examining the diagnostic utility of the measurement instrument with different clinical populations (e.g., Atkinson, Konold, & Glutting, 2008; Pham & Hasson, 2014; Sowerby, Seal, & Tripp, 2011). Despite the information provided by these researchers, these procedures are inadequate for validating the internal structure of a measurement instrument (Clark & Watson, 1995; Cronbach & Meehl, 1955). Of greater concern, the WRAML2 has yet to be subjected to higher order analyses, suggesting that our understanding of the structuring of these variables is presently unknown.

Purpose of the Current Study

Over the past decade, unrestricted EFA methods have been eclipsed by more restrictive CFA methods when examining the structural validity of psychological measures (Reynolds & Keith, 2013). However, it has been suggested (e.g., Canivez, 2013; Frazier & Youngstrom, 2007) that use of CFA has resulted in overfactoring of ability measures such as the WRAML2. Although EFA and CFA are considered to be complementary procedures, they provide answers to different empirical questions and that when the results from these procedures are in agreement, greater confidence can be placed in the internal structure of a test (Gerbing & Hamilton, 1996; Schmitt, 2011). As a consequence, Carroll (1995) recommended that "a confirmatory analysis of a dataset should not be published without an accompanying statement or report on one or more appropriate exploratory analyses" (p. 437). Consequently, the current study examined the applicability of a hierarchical measurement model for the WRAML2 normative sample data using exploratory and confirmatory methodologies. Although commonly applied to intelligence tests, use of higher order variance partitioning procedures on other types of psychological assessment tools has increased in recent years to account for the ever increasing specification of multilevel models and related interpretive procedures for these measures (i.e., Dombrowski, 2015; Reise, Moore, & Haviland, 2010). Therefore, it is important to ascertain the degree to which tools such as the WRAML2 measure multiple lower order constructs well if at all (Glutting, Watkins, & Youngstrom, 2003). It is believed that the results that are obtained will be instructive for correct interpretation of the WRAML2 and other related measures of memory and learning.

Method

Participants

Participants were children and adults aged 9 to 90 years (n = 880) drawn from the WRAML2 standardization sample. Demographic characteristics are provided in detail in the WRAML2 Technical Manual (Sheslow & Adams, 2003b). The standardization sample (N = 1,200) was obtained using stratified proportional sampling across demographic variables of age, sex, race/ethnicity, parent educational level, and geographic region. Examination of the tables provided in the Technical Manual revealed a close correspondence to the 2001 U.S. Census estimates across the stratification variables. The present sample was selected for analyses so as to provide results that could be compared with the structural analyses reported in the Technical Manual with the total sample without completing redundant analyses. Additionally, normative participants aged 5 to 8 years were excluded because of previous research (e.g., Putzke, Williams, Adams, & Boll, 1998; Putzke, Williams, Glutting, Konold, & Boll, 2001) suggesting that lower order WRAML abilities failed to emerge consistently at those ages.

Measurement Instrument

The WRAML2 is a multidimensional test of memory abilities for children and adults aged 5 to 90 years. The measure is composed of 17 subtests, 6 of which contribute to the measurement of three first-order index scores: Verbal Memory, Visual Memory, and Attention/Concentration. The core subtests are linearly combined to form the second-order full-scale GMI composite. It should be noted that supplementary measures can be combined to form an additional Working Memory Index score, although performance on these measures does not contribute to the core battery indexes or GMI scores, thus, they were excluded from the present study. All index and composite variables on the WRAML2 are expressed as standard scores with a mean of 100 and a standard deviation of 15. Extensive normative and psychometric data can be found in the WRAML2 Technical Manual (Sheslow & Adams, 2003b).

Procedure and Data Analyses

Exploratory Factor Analysis. Following best practice guidelines (e.g., Fabrigar et al., 1999; Preacher & MacCallum, 2003; Thompson, 2004), EFA was implemented using SPSS version 23 (IBM Corp., 2014) using data extracted from the intercorrelation matrix of the six WRAML2 core subtests for ages 9 to 90 (Sheslow & Adams, 2003b, p. 113). Bartlett's test of sphericity was used to ensure that the correlation matrix was not random, and the Keiser–Meyer–Olkin statistic was required to be above a minimum standard of .60 to ensure that the matrix was suitable for factor analysis (Kaiser, 1974). The principal axis factoring (PAF) method was used because of its ability to recover weak factors as well its ability to disclose latent structure with measurement models that have few indicators per factor or that may be just identified (de Winter & Dodou, 2012). Iterations in first-order PAF extraction were limited to two in estimating final communality estimates. According to Gorsuch (2003), limiting iterations to two provides an optimal balance between sampling and measurement error in estimating communality. As recommended by Velicer, Eaton, and Fava (2000), multiple criteria for determining the number of factors to retain were examined. The procedures used to determine the appropriate number of factors for retention and rotation included Horn's parallel

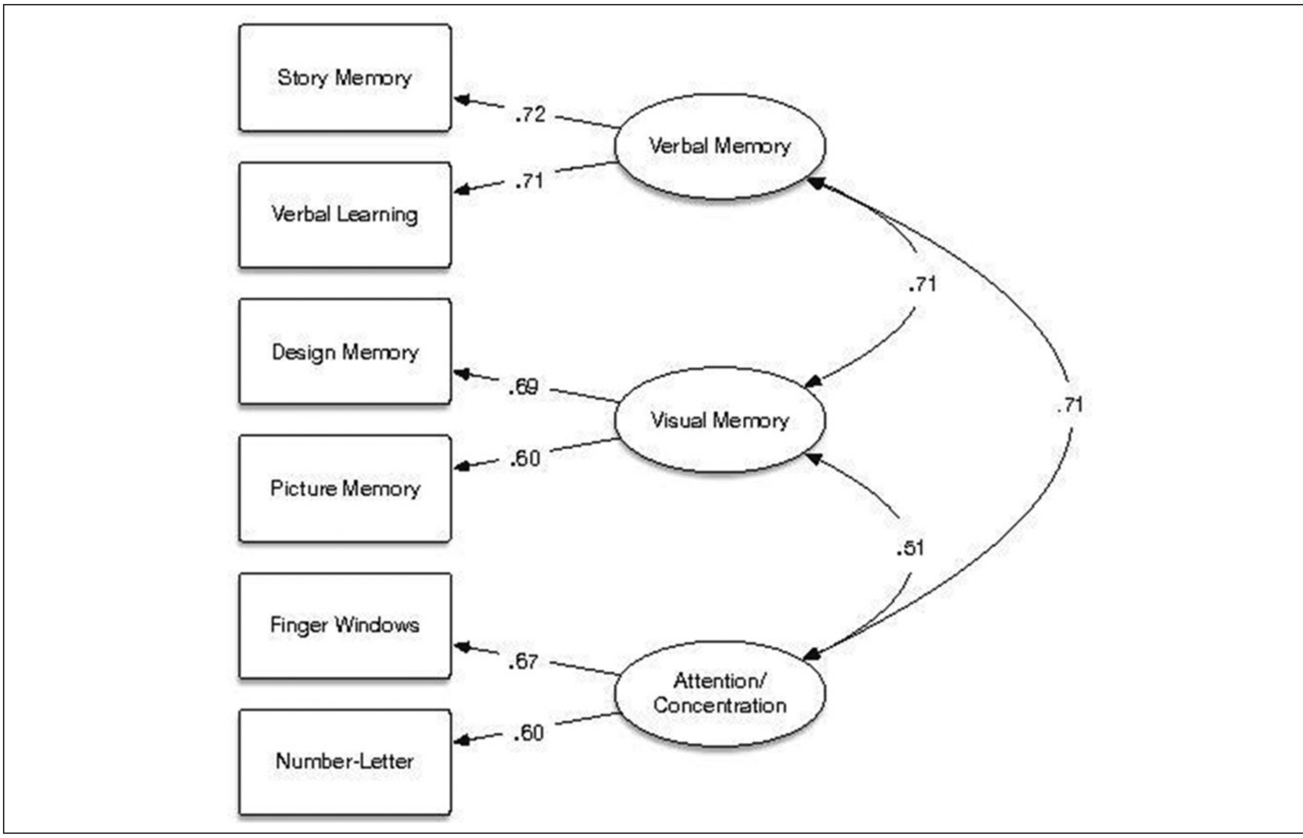

Figure 2. Correlated three-factor first-order measurement model, with standardized coefficients, for the Wide Range Assessment of Memory and Learning–Second Edition ages 9 to 90 (n = 880).

analysis (HPA; Horn, 1965), and MAP (Velicer, 1976), to compliment a visual scree test (Cattell, 1966). While the scree test was used to visually determine the optimum number of factors to retain, it is a subjective methodology. As recommended by Frazier and Youngstrom (2007), HPA and MAP were also included as they potentially protect against the threat of overfactoring in EFA. Random data for HPA analyses were generated using Watkins's (2000) parallel analysis program with 100 replications to produce stable estimates. MAP procedures were conducted using O'Connor's (2000) SPSS syntax program. Finally, additional consideration was given to theoretical convergence.

Initial PAF extraction was followed by a promax (oblique) rotation (k = 4; Gorsuch, 2003). Because an indirect hierarchical measurement model was implied for the WRAML2, higher order factor analysis using the SL (Schmid & Leiman, 1957) procedure was applied to oblique first-order factors to elucidate the hierarchical structuring of variables using the MacOrtho program by Watkins (2004). This procedure allows for the extraction of a second-order factor from a first-order factor correlation matrix. According to Schmid and Leiman (1957), this transforms "an oblique solution containing a hierarchy of higher-order factors into an orthogonal solution which not only preserves the desired interpretation characteristics of the oblique solution, but also discloses the hierarchical structuring of the variables" (p. 53). Criteria for determining factor adequacy were established a priori. In accord with Dombrowski (2013), salient coefficients were defined as those ≥.30, for the oblique solution, and ≥.20, for the orthogonalized solution.

Confirmatory Factor Analysis. EQS, Version 6.2 (Bentler & Wu, 2012) was used to conduct CFA using maximum likelihood estimation. To align with the CFA analyses reported in the WRAML2 Technical Manual (Sheslow & Adams, 2003b), three first-order models were specified and examined at ages 9 to 90: (a) one factor; (b) two oblique Verbal and Visual Memory factors; and (c) three oblique Verbal Memory, Visual Memory, and Attention/Concentration factors (see Figure 2). Additionally, two higher order models were explicated: an indirect hierarchical model and a direct hierarchical model, with three first-order group factors. Because the six subtest WRAML2 model configuration only has two subtest indicators for the three resulting group-specific factors, subtest indicators were constrained to be equal in the direct hierarchical (i.e., bifactor) model to ensure specification as conducted by Watkins and Beaujean (2014). Beaujean (2015) has provided a detailed description of the salient differences between direct and indirect hierarchical models and assumptions regarding the appropriate structuring and influence of latent cognitive dimensions, but the direct hierarchical model is a variant of the so-called *bifactor* model originally described by Holzinger and Swineford (1937).

To comport with best practice (e.g., Brown, 2016; Marsh, Hau, & Grayson, 2005), multiple indices were examined to evaluate the adequacy of model fit. Specifically, the (a) chi-square, (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), (d) standardized root mean square residual (SRMR), and (e) Akaike information criterion (AIC). Although there are no golden rules for evaluating model fit indices (i.e., Markland, 2007), the following guidelines were used for good model fit criteria: (a) CFI $\geq$ 0.95; (b) SRMR and RMSEA $\leq$ 0.06 (Hu & Bentler, 1999). Higher CFI values and lower RMSEA values indicate better model fit, and these two indices were supplemented with chi-square and AIC values. There are no specific criteria for information-based indices like the AIC, but smaller values may indicate better approximations of the true measurement model after accounting for model complexity (Vrieze, 2012). Meaningful differences between well-fitting models were evaluated based on the following criteria: (a) exhibit good fit according to CFI, RMSEA, and SRMR indices; (b) demonstrate a ΔCFI value $\leq$ 0.01 for nested models (Dimitrov, 2012); and/or (c) display the smallest AIC value (Burnham & Anderson, 2004).

Finally, the bifactor model hypothesizes that each WRAML2 subtest is influenced simultaneously by two orthogonal latent constructs: a general memory factor (*g*) and a first-order domain-specific group factor (e.g., Verbal Memory, Attention/Concentration, etc.). As a consequence, Omega (ω) and omega-hierarchical/hierarchical subscale (ω_h/ω_{hs}) were estimated as model-based reliability estimates of the latent factors (Gignac & Watkins, 2013). Whereas ω estimates the variance accounted for by both of the constructs in a given domain, ω_h estimates the variance accounted for by a single target construct. Chen, Hayes, Carver, Laurenceau, and Zhang (2012) stressed that "for multidimensional constructs, the alpha coefficient is complexly determined, and McDonald's omega-hierarchical (ω_h; 1999) provides a better estimate for the composite score and thus should be used" (p. 228). Omega estimates were produced using the Omega program (Watkins, 2013). Albeit subjective, omega coefficients should at a minimum exceed .50, but .75 would be preferred (Reise, 2012; Reise et al., 2013).

Results

Exploratory Factor Analysis

Factor Extraction Criteria. Whereas parallel analysis (Horn, 1965) and the MAP (Velicer, 1976) criterion recommended retention of one factor, visual scree suggested that three factors be retained. Given that, it is better to overfactor than underfactor (Wood, Tataryn, & Gorsuch, 1996), three factors were extracted to accord with the theoretical structure delineated in the WRAML2 Technical Manual (Sheslow & Adams, 2003b).

Oblique Solution. The results of Bartlett's test of sphericity indicated that the correlation matrix was not random, $\chi^2(15)$ = 1000.11, p < .001, and the Kaiser–Meyer–Olkin measure of sampling adequacy coefficient of .77, was well above the minimum standard for conducting factor analysis (Kaiser, 1974). Communality estimates ranged from .331 (Number–Letter) to .481 (Story Memory). On the basis of these values, it was determined that the correlation matrix was appropriate for the EFA procedures that were employed. Table 1 presents results from extracting three WRAML2 factors with promax (k = 4) rotation. All WRAML2 subtests were saliently and properly associated with their theoretical factor demonstrating desirable simple structure. Correlations between the factors ranged from .53 and .74 implying the possible presence of a higher order dimension requiring explication (Gorsuch, 1983; Thompson, 2004).

Orthogonalized Solution. The three first-order oblique EFA factor solution was transformed with the SL orthogonalization procedure. Results for the second-order factor analysis of three first-order WRAML2 factors are also presented in Table 1. All subtests were properly associated (higher residual variance) with their theoretically proposed factor after removing variance associated with a general memory dimension. The general factor accounted for 28.9% of the total variance and 72% of the common variance. The general factor also accounted for between 18.4% (Picture Memory) and 45.4% (Story Memory) of individual subtest variability. At the first-order level, the Verbal Memory factor accounted for an additional 0.6% of the total variance and 1.5% of the common variance, the Visual Memory factor accounted for an additional 5.4% of the total variance and 13.5% of the common variance, and the Attention/Concentration factor accounted for an additional 5.2% of the total variance and 13% of the common variance. The general and group factors combined to measure 40% of the variance in WRAML2 scores resulting in 59% unique variance (combination of specific and error variance). Subtest specificity (reliable variance unique to the individual measures) ranged from .38 to .50.

Confirmatory Factor Analysis

Model fit statistics presented in Table 2 illustrate the increasingly better fit from one to three factors; however, fit statistics indicated that the one- and two-factor models were inadequate (CFI < 0.95 and/or RMSEA > 0.06). Consistent

Table 1. Exploratory Factor Analysis With Oblique and Orthogonalized Pattern Coefficients of the Wide Range Assessment of Memory and Learning–Second Edition Core Subtests Ages 9 to 90 (n = 880).

	Oblique solution			Orthogonalized solution					
Subtest	I	II	III	General	I	II	III	h^2	u^2
Story Memory	**.604**	.118	−.002	.674	.132	.080	.080	.485	.516
Verbal Learning	**.634**	−.022	.083	.664	.139	−.015	.055	.463	.537
Design Memory	.021	**.565**	.067	.487	.005	**.381**	.045	.384	.616
Picture Memory	.007	**.616**	−.043	.429	.002	**.416**	−.029	.357	.643
Finger Windows	.054	−.011	**.585**	.480	.012	−.007	**.390**	.384	.616
Number–Letter	−.013	.016	**.577**	.430	−.003	.011	**.384**	.332	.668
Total variance (%)				28.9	0.6	5.4	5.2	40.1	59.9
Common variance (%)				72.0	1.5	13.5	13.0		

Note. h^2 = communality; u^2 = uniqueness. As per Dombrowski (2013), salient loadings ≥.30, for the oblique solution, and ≥.20, for the orthogonalized solution are denoted in bold.

Table 2. Fit Statistics for Competing Structural Models of the Wide Range Assessment of Memory and Learning–Second Edition Core Subtests Ages 9 to 90 (n = 880).

Model	χ^2	df	CFI	SRMR	RMSEA	90% CI RMSEA	AIC
First-order models							
One factor	119.24	9	0.888	0.058	0.118	[0.099, 0.137]	101.24
Two oblique factors	113.41*	8	0.893	0.058	0.122	[0.103, 0.143]	97.41
Three oblique factors	21.91**	6	0.984	0.021	0.055	[0.031, 0.080]	9.91
Hierarchical models							
Indirect hierarchical[a]	21.91	6	0.984	0.021	0.055	[0.031, 0.080]	9.91
Direct hierarchical[b]	21.91	6	0.984	0.021	0.055	[0.031, 0.080]	9.91

Note. df = degrees of freedom; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike information criterion.
[a]Identical goodness-of-fit and fit statistics with previous model due to just identification. [b]Identical goodness-of-fit and fit statistics with previous two models due to model constraints.
*Statistically different (p < .01) from previous model. **Statistically different (p < .01) from previous two models.

with the CFA results reported in the Technical Manual for the total normative sample, the correlated three-factor model (see Figure 2) provided the best fit to these data among the first-order models and was statistically a better fit to these data than the rival two-factor model, $\Delta\chi^2(2)$ = 91.50, p < .001. However, because the latent factors were highly correlated, a higher order structure is implied (Gorsuch, 1983), rendering the correlated three-factor model a less than optimal explanatory model of the WRAML2 core battery structure (Canivez, 2016; Gignac, 2016; Thompson, 2004). Since the three-factor first-order model is underidentified, a higher order solution will yield the same goodness of fit as the first-order model; thus, meaningful differences in fit statistics cannot be assessed. Nevertheless, Brown (2016) suggests that it may be "substantively meaningful to evaluate such a solution in order to examine the magnitude of (and statistical significance) of the higher-order factor loadings and relationships of the higher-order factors to the observed measures" (p. 292). Accordingly, an indirect hierarchical solution was estimated to better examine the structural relationships between latent dimensions and measured variables to aide clinical interpretation of these constructs.

In the indirect hierarchical model (Figure 3), the standardized path from the second-order general factor to Verbal Memory was .99, indicating that Verbal Memory was difficult to disentangle from the higher order general factor (Canivez & Kush, 2013). As a consequence, an orthogonal version of the theoretical three-factor model was also examined (Figure 4). In contrast to the indirect hierarchical model, the bifactor model (Reise, 2012) which has also been called a *direct hierarchical model* (Gignac, 2007) estimates the influences of first-order factors and the general factor on measured variables simultaneously. This model has been recommended for hierarchically structured constructs such as memory and other related cognitive abilities (Brunner, Nagy, & Wilhelm, 2012) and has been applied to other cognitive scales (see Canivez, 2016; Gignac & Watkins, 2013; Watkins & Beaujean, 2014). Of note, a bifactor model was explicitly recommended by John Carroll

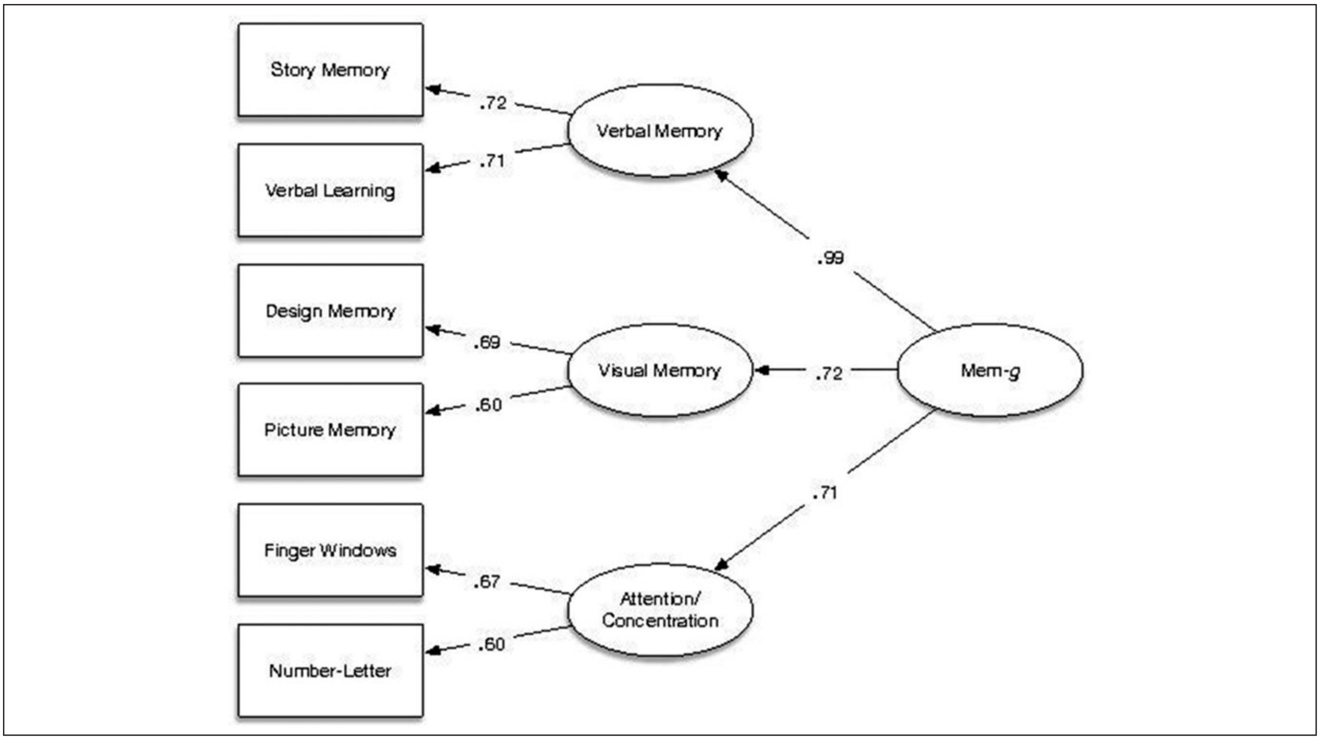

Figure 3. Indirect hierarchical measurement model, with standardized coefficients, for the Wide Range Assessment of Memory and Learning–Second Edition ages 9 to 90 (n = 880).
Note. Mem-*g* = general memory factor.

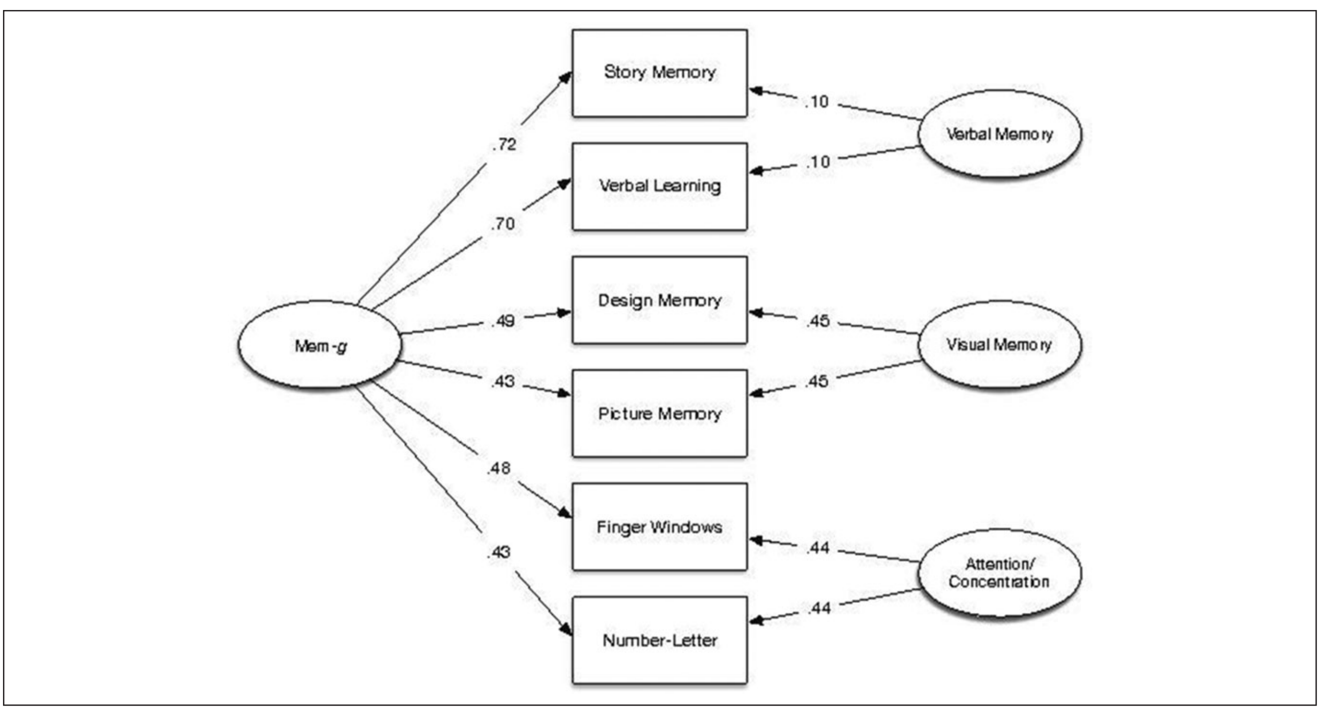

Figure 4. Direct hierarchical (bifactor) measurement model, with standardized coefficients, for the Wide Range Assessment of Memory and Learning–Second Edition ages 9 to 90 (n = 880).
Note. Mem-*g* = general memory factor.

Table 3. Sources of Variance in the WRAML-2 According to a Three-Factor Direct Hierarchical Model.

	General		Verbal		Visual		A/C					
Subtest	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2	Error	s^2
Story Memory	.715	.511	.097*	.009					.521	.479	.080	.399
Verbal Learning	.700	.490	.097*	.009					.499	.501	.160	.341
Design Memory	.493	.243			.448	.201			.444	.556	.140	.416
Picture Memory	.425	.181			.448	.201			.381	.619	.170	.449
Finger Windows	.478	.228					.443	.196	.425	.575	.190	.385
Number–Letter	.425	.181					.443	.196	.377	.623	.170	.453
% Total variance	30.6		0.9		20.1		19.6		44.1	55.9	15.2	40.7
% Common variance	69.3		0.7		15.2		14.8			100.0		
	ω = .783		ω = .675		ω = .583		ω = .572					
	ω_h = .678		ω_{hs} = .012		ω_{hs} = .285		ω_{hs} = .280					

Note. WRAML-2 = Wide Range Assessment of Memory and Learning–Second Edition; A/C = Attention/Concentration factor; b = standardized loading of subtest on factor; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; Error = 1 − reliability from Sheslow and Adams (2003b); $s^2 = u^2$ − Error; ω = omega; ω_h = omega hierarchical; ω_{hs} = omega hierarchical subscale.
*p < .05.

in his explorations of the structure of human cognitive abilities such as memory (Beaujean, 2015). Given the underidentification of the WRAML2 factors, test loadings within each factor were constrained to equality.

Table 3 presents decomposed core battery subtest variance estimates of the WRAML2 based on a direct hierarchical model. In this orthogonal model, the general factor accounted for 30.6% of the total variance and 69.3% of the common variance. Among the group factors, the Verbal Memory factor accounted for 0.9% of the total variance and 0.7% of the common variance, the Visual Memory for 20.1% of the total variance and 15.2% of the common variance, and the Attention/Concentration factor accounted for 19.6% of the total variance and 14.8% of the common variance. Thus, the higher order general factor accounted for greater portions of WRAML2 common and total variance relative to the individual factor index scores. Omega hierarchical and hierarchical subscale coefficients presented in Table 3 provide estimates of the reliability of the latent constructs with the effects of other constructs removed. In the case of the three WRAML2 factor indexes, omega hierarchical subscale coefficients estimated the scale reliabilities with the effects of the general factor removed, and ranged from .012 (Verbal Memory) to .285 (Visual Memory), well below recommend levels for confidant clinical interpretation.

Discussion

The WRAML2 is a multidimensional battery of memory and learning abilities that is routinely utilized by assessment psychologists in a variety of clinical settings. As its predecessor was the first memory assessment tool designed for use with children, the WRAML2 is especially popular in school and pediatric psychology for appraising these dimensions (Homack, 2013). Despite providing users with a variety of scores that imply a multilevel structure, hierarchical structure was not explicated in the Technical Manual (Sheslow & Adams, 2003b) and has yet to be investigated since its publication. Instead, subtests were grouped into several first-order indexes in part on the results obtained from components analysis and the resulting three-factor oblique structure was supported in a CFA with the total normative sample.

Whereas users are encouraged to focus most of their interpretive weight on the second-order GMI composite, the EFA and CFA results reported in the Technical Manual tend to provide evidence only for the interpretability of the first-order index scores. Although oblique measurement models are widely utilized to validate the structure of cognitive measures, it is a mistake to extrapolate higher order structure from first-order dimensions (Carretta & Ree, 2001; McClain, 1996). Consequently, this study applied both EFA and CFA methods to the WRAML2 core battery subtests for participants in the normative sample aged 9 to 90 years in order to better disclose higher order structure.

EFA results suggested the retention of only one factor, in contrast to the theoretical three factors posited by Sheslow and Adams (2003b). The extraction of three factors resulted in highly correlated first-order dimensions, suggesting the presence of second-order general factor. Gignac (2007) has encouraged researchers to always perform orthogonalization procedures when examining higher order model solutions. Thus, in order to better understand the underlying structure of the WRAML2, we utilized recommended procedures (e.g., Carroll, 1993, 1995; Schmid & Leiman, 1957) in order to correctly apportion subtest variance appropriately to higher and lower order dimensions. Similar to Gioia (1998), we found that the oblique loading coefficients produced from PAF were significantly discrepant and attenuated from the

PCA coefficients reported in the Technical Manual. Because common and specific factor variances are conflated in PCA, these solutions have a tendency to inflate loadings and provide the illusion of a strong component structure (Snook & Gorsuch, 1989). According to Fabrigar and Wegener (2012), differences between PCA and PAF are more likely to emerge with low communalities such as those observed in the present analyses with the WRAML2.

When the theoretical model was subjected to higher order EFA via the Schmid and Leiman (1957) procedure, a strong general factor accounted for 29% of the total variance and 72% of the common variance, whereas all three group factors combined accounted for 6% of the total variance and 28% of the common variance. Although WRAML2 subtests were consistently aligned with their theoretically assigned factors, the three-factor model appears to be overfactored. Whereas desired simple structure was observed for the Visual Memory and Attention/Concentration factors, residual loadings on the Verbal Memory factor were weak (<.20), calling into question its viability within the WRAML2 measurement model. These EFA results replicate the important role of the general factor found in related EFA investigations of intelligence tests using similar methods (Canivez et al., 2016; Dombrowski, 2013; McGill, 2016; McGill & Spurgin, 2015) and provide little support for interpretation of the WRAML2 beyond the GMI composite.

CFA results were more ambiguous. Consistent with the CFA results reported in the Technical Manual with the total normative sample, fit statistics suggested that a three-factor oblique correlated factors model best fit the WRAML2 data. Similar to EFA results, correlations between first-order dimensions were moderate to strong, indicating the presence of second-order factor that required explication. The application of an indirect hierarchical model with a general memory dimension at the apex resulted in a near perfect loading (.99) between Verbal Memory and the general factor. When the first- and second-order loadings on the WRAML2 subtests were residualized using the procedures described in Reynolds and Keith (2013), the general factor accounted for 43% to 72% of subtest variance, whereas the first-order factors accounted for 0% to 23% of the residual variance. Results from the application of a bifactor measurement model in which the effects of the general and first-order group factors on the WRAML2 subtests were directly estimated, diverged significantly from EFA results and provided stronger support for interpretation beyond the GMI. Whereas the general factor accounted for 31% of the total variance and 69% of the common variance, the three first-order factors combined accounted for 41% of the total variance and 31% of the common variance. Five of the six WRAML2 tests displayed uniqueness values that that exceeded their communality, indicating that much of the variability in these measures is attributable to test-specific and error variances. Nevertheless, omega coefficients indicated that only the general factor was measured with enough precision to warrant confident clinical interpretation.

As stated by Gorsuch (2003), "the ultimate arbiter in science is well established: replication" (p. 153). As EFA and CFA provide answers to different empirical questions, contradictory results are commonly reported within the cognitive assessment literature (Reynolds & Keith, 2013). Whereas strong support was found for the GMI in the present study, the strength and consistency of the three first-orders group factors varied across EFA and CFA analyses. Despite this variation, the Verbal Memory factor was consistently weak across all of the measurement models, indicating that dimension accounted for trivial proportions of variance in tests assigned to that factor apart from the general factor.

Although users of the WRAML2 have been encouraged (e.g., Adams, 2013; Miller, 2013) to forgo administration of the core battery at the expense of more selective assessment of first-order dimensions, results from the present study suggest that specific approach and other related cross-battery assessment procedures (e.g., Flanagan, Ortiz, & Alfonso, 2013) should be employed with the WRAML2 with extreme caution, if at all, until more consistent evidence is provided to support interpretation of the group-specific factors. These results illustrate well that nontrivial variance attributable to the general factor is endemic at all levels of the measurement instrument and must be accounted for when interpreting group-specific WRAML2 dimensions in isolation (Gignac, 2007; Gustaffson & Åberg-Bengtsen, 2010).

Implications for Test Development

Since general and more specific psychological constructs cannot be observed directly, assessment researchers must choose from a variety of measurement models that link latent constructs to measured variables when validating an assessment tool. This decision is crucial as it provides the statistical rationale for the standardized scores that are computed for that measure that are thought to reflect these constructs (Brunner et al., 2012). According to the *Standards for Educational and Psychological Testing*, "When a test provides more than one score, the distinctiveness of the separate scores should be demonstrated" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 26). Furthermore, Standard 1.14 explicitly requires that a rationale be provided for the development of composite scores. As a consequence, it is difficult to understand why the authors of the WRAML2 and other related ability measures continue to rely on correlated factors models to validate internal structure in which a second-order model is consistent with the constructs being measured and implied in the scores that are later developed for that measure and presented as capable of being interpreted. First-order measurement models focus only on specific abilities; as a consequence, extrapolating second-order structure from the relationships observed between first-order dimensions may result in the retention of spurious constructs (Gignac, 2007, 2016).

The EFA/CFA results for the WRAML2 presented herein illustrate well that the correlated factors model presented in the Technical Manual overestimates the importance of the Verbal Memory factor when the hierarchical structuring of the variables are disclosed. As stated by Gustaffson and Åberg-Bengtsen (2010),

> if our intention is to measure one or more narrow factors, it is necessary to partial out the influence of the more general factors that exert influence on the observed variables . . . but this cannot be done unless information is available that allows estimation of the more general dimensions. (p. 109)

It should be noted that estimation of a hierarchical model is possible in most CFA programs with data sets that yield two or more first-order group factors and procedures for doing so are commonly provided in accompanying program guides and manuals (e.g., Bentler, 2006; Byrne, 2006).

Limitations

The present results were derived from a single intercorrelation matrix representing a large proportion of the WRAML2 age span. Although these results provide a relevant comparison with the psychometric analyses presented in the Technical Manual that utilized the total normative sample, additional independent examinations of the construct validity of the measurement instrument at different points of the age span would be beneficial. Also, given the neuropsychological foci of the instrument, it is possible that the factor structure of the WRAML2 may emerge more consistently in clinical populations such as those suspected of having more focal cortical injuries. Furthermore, the consistently low communalities produced for WRAML2 indicators suggest that the weak and inconsistent first-order effects may be the product of the way in which these dimensions were sampled and/or measured (Gioia, 1998). Therefore, additional research to examine the potential generalizability of these findings would be of benefit to WRAML2 users.

Finally, as the WRAML2 core battery measurement model is underidentified, it is not possible to determine which manifestation of the hierarchical model (i.e., direct vs. indirect) is most appropriate for the data via CFA (MacCallum, Wegener, Uchino, & Fabrigar, 1993). As noted by Beaujean (2015), "While the higher-order model [indirect hierarchical] is technically nested within the bifactor model, they provide very different conceptualizations of *g* and other factors in the model" (p. 122). In the indirect hierarchical model, the general factor's influence is mediated by the first-order factors and the general factor is produced from the correlations between the group factors. As a result, this model produces a constraint that restricts the general and specific variance within a group factor to be proportional (Gignac, 2016; Yung, McLeod, & Thissen, 1999). As a result, it has been argued that the bifactor model, in which the effects of the general and specific factors on measured variables are estimated directly, is preferred as the effects of the latent variables are easier to interpret (Canivez, 2016; Gignac, 2007, 2016).

Nevertheless, some researchers have questioned whether the bifactor model is a tenable structure for human cognitive abilities (e.g., Murray & Johnson, 2013; Reynolds & Keith, 2013). While adjudication of this issue is beyond the scope of the present discussion, CFA results indicate that, regardless of one's preference in terms of a hierarchical model, the effects of the general factor are superior to the individual group-specific factors on the WRAML2.

Conclusion and Implications for Interpretation

The present study provides clinicians with important information for interpreting the WRAML2 (Sheslow & Adams, 2003a). Whereas our results suggest that users can be reasonably confident in their interpretations of the GMI composite, the contribution of the group-specific factors was less consistent. As a consequence, it is recommended that users of the WRAML2, focus most of their interpretive weight on the GMI score, with additional consideration of select factor scores (Visual Memory and Attention/Concentration) to be employed cautiously. As an example, if users elect to interpret the Visual Memory score, they must bear in mind that Visual Memory tasks have long been known to be factorially complex. According to Carroll (1993), it is common for examinees to encode information on these tasks in a nonvisual manner. He noted that if a Visual Memory factor exists it "would appear in tests that emphasize the person's ability to form and remember . . . a mental image or representation of a visual shape or configuration that does not represent some easily recognized object" (p. 282). As both of the subtests that combine to form the Visual factor on the WRAML2 contain stimuli that is not novel, additional information is needed to determine the sensory modality that is sampled by these tasks.

As a result of these deficiencies, we encourage users of the WRAML2 to additionally forgo the recommended procedures for index score discrepancy analyses nor use that information to make inferences regarding cortical lateralization (i.e., Adams & Reynolds, 2009. As "the ultimate responsibility for appropriate test use and interpretation lies predominantly with the test user" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 141), clinicians using the WRAML2 in clinical evaluations must seriously consider the present information to make informed decisions about which scores have satisfactory reliability, validity, and utility.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Adams, W. (2013). Memory assessment. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 494-525). New York, NY: Oxford University Press.

Adams, W., & Reynolds, C. R. (2009). *Essentials of WRAML2 and TOMAL-2 assessment*. Hoboken, NJ: John Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: Author.

Atkinson, T. M., Konold, T. R., & Glutting, J. J. (2008). Patterns of memory: A normative taxonomy of the Wide Range Assessment of Memory and Learning–Second Edition (WRAML-2). *Journal of the International Neuropsychological Society*, *14*, 869-877. doi:10.1017/S1355617708081137

Aylward, G. P., Gioia, G., Verhulst, S. J., & Bell, S. (1995). Factor structure of the Wide Range Assessment of Memory and Learning in a clinical population. *Journal of Psychoeducational Assessment*, *13*, 132-142. doi:10.1177/073428299501300203

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence*, *3*, 121-136. doi:10.3390/jintelligence3040121

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Retrieved from http://84.89.132.1/~satorra/CourseSEMVienna2010/EQSManual.pdf

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*, 67-74. doi:10.1207/s15327906mbr2501_8

Bentler, P. M., & Wu, E. J. C. (2012). *EQS for Windows*. Encino, CA: Multivariate Software.

Brown, T. A. (2016). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796-846. doi:10.1111/j.1467-6494.2011.00749.x

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261-304. doi:10.1177/0049124104268644

Burton, D. B., Donders, J., & Mittenberg, W. (1996). A structural equation analysis of the Wide Range Assessment of Memory and Learning in the standardization sample. *Child Neuropsychology*, *2*, 39-47. doi:10.1080/09297049608401349

Byrne, B. M. (2006). *Structural equation modeling with EQS* (2nd ed.). New York, NY: Psychology Press.

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. (2013). 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 assessment* (pp. 84-112). New York, NY: Oxford University Press.

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Gottingen, Germany: Hogrefe.

Canivez, G. L., & Kush, J. C. (2013). WAIS-IV and WISC-IV structural validity: Alternative methods, alternate results: Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). *Journal of Psychoeducational Assessment*, *31*, 157-169. doi:10.1177/0734282913478036

Canivez, G. L., & Watkins, M. W. (2010). 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., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *28*, 975-986. doi:10.1037/pas0000238

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

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

Cattell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioral Research*, *1*, 245-276. doi: 10.1207/s1532790mbr0102_10

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*, 219-251. doi:10.1111/j.1467-6494.2011.00739.x

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319. doi:10.1037/1040-3590.7.3.309

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1-9. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302. doi:10.1037/h0040957

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principle axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*, 695-710. doi:10.1080/02664763.2011.610445

Dewey, D., Kaplan, B. J., & Crawford, S. G. (1997). Factor structure of the WRAML in children with ADHD or reading disabilities: Further evidence of an attention/concentration factor. *Developmental Neuropsychology*, *13*, 501-506. doi:10.1080/87565649709540691

Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.

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. (2015). Exploratory bifactor analysis of the WJ-III Achievement at school age via the Schmid–Leiman procedure. *Canadian Journal of School Psychology*, *30*, 34-50. doi:10.1177/0829573514560529

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2016). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000350

Engle, J. A., & Smith, M. L. (2010). Attention and material-specific memory in children with lateralized epilepsy. *Neuropsychologia*, *48*, 38-42. doi:10.1016/j.neuropsychologia.2009.08.005

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York, NY: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299. doi:10.1037/1082-989X.4.3.272

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: John Wiley.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, *35*, 169-182. doi:10.1016/j.intell.2006.07.002

Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling*, *3*, 62-72. doi:10.1080/10705519609540030

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences*, *42*, 37-48. doi:10.1016/j.paid.2006.06.019

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68. doi:10.1016/j.intell.2016.01.006

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

Gioia, G. A. (1998). Re-examining the factor structure of the Wide Range Assessment of Memory and Learning: Implications for clinical interpretation. *Assessment*, *5*, 127-139. doi:10.1177/107319119800500204

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 343-374). New York, NY: Guilford Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 143-164). Hoboken, NJ: John Wiley.

Gustaffson, J. E., & Åberg-Bengtsen, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 97-122). Washington, DC: American Psychological Association.

Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*, 303-329. doi:10.1207/s15327906mbr4003_2

Hartman, D. E. (2007). Test review: Wide Range Assessment of Memory and Learning-2 (WRAML-2): WRedesigned and WReally improved. *Applied Neuropsychology*, *14*, 138-140. doi:10.1080/09084280701322908

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

Homack, S. (2013). Methods of neuropsychological assessment. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 474-493). New York, NY: Oxford University Press.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185. doi:10.1007/BF02289447

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

IBM Corp. (2014). IBM SPSS Statistics for Windows (Version 23.0) [Computer software]. Armonk, NY: Author.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36. doi:10.1007/BF02291575

Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review*, *28*, 117-144.

Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1989). Verbal and visual memory index discrepancies from the Wechsler Memory Scale—Revised: Cautions in interpretation. *Psychological Assessment*, *1*, 198-202. doi:10.1037/1040-3590.1.3.198

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199. doi:10.1037//0033-2909.114.1.185

Maricle, D. E., Miller, D. C., & Mortimer, J. (2011). Memory tests in pediatric neuropsychology. In A. S. Davis (Ed.), *Handbook* *of pediatric neuropsychology* (pp. 275-292). New York, NY: Springer.

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, *42*, 851-858. doi:10.1016/j.paid.2006.09.023

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah, NJ: Lawrence Erlbaum.

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

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

McGill, R. J. (2016). Investigation of the factor structure of the Comprehensive Test of Nonverbal Intelligence–Second Edition (CTONI-2) using exploratory factor analysis. *Journal of Psychoeducational Assessment*, *34*, 339-350. doi:10.1177/0734282915610717

McGill, R. J., & Spurgin, A. R. (2015). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children–Second Edition (KABC-II) school-age battery. *Assessment*. Advanced online publication. doi:10.1177/1073191115614081

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken, NJ: John Wiley.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267-305. doi:10.1207/s15327906mbr2203_3

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*, *37*, 487-494. doi:10.1016/j.intell.2013.06.004

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavioral Research Methods*, *32*, 396-402. doi:10.3758/BF03200807

Pham, A. V., & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, *29*, 467-477. doi:10.1093/arclin/acu024

Phelps, L. (1995). Exploratory factor analysis of the WRAML with academically at-risk students. *Journal of Psychoeducational Assessment*, *13*, 384-390. doi:10.1177/073428299501300405

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13-43. doi:10.1207/S15328031US0201_02

Putzke, J. D., Williams, M. A., Adams, W., & Boll, T. J. (1998). Does memory test performance in children become more consistent with age? Cross-sectional comparisons using the WRAML. *Journal of Clinical and Experimental Neuropsychology*, *20*, 835-845. doi:10.1076/jcen.20.6.835.1115

Putzke, J. D., Williams, M. A., Glutting, J. J., Konold, T. R., & Boll, T. J. (2001). Developmental memory performance: Inter-task consistency and base-rate variability on the WRAML. *Journal of Clinical and Experimental Neuropsychology*, *23*, 253-264. doi:10.1076/jcen.23.3.253.1189

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667-696. doi:10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129-140. doi:10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544-559. doi:10.1080/00223891.2010.496477

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 48-83). New York, NY: Oxford University Press.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 223-237. doi:10.1080/00223891.2015.1089249

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

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*, 304-321. doi:10.1177/0734282911406653

Sheslow, D., & Adams, W. (1990). *Wide Range Assessment of Memory and Learning (WRAML)*. Wilmington, DE: Wide Range.

Sheslow, D., & Adams, W. (2003a). *Wide Range Assessment of Memory and Learning* (2nd ed.). Wilmington, DE: Wide Range.

Sheslow, D., & Adams, W. (2003b). *Wide Range Assessment of Memory and Learning, Second Edition (WRAML2): Administration and Technical Manual*. Wilmington, DE: Wide Range.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*, 148-154. doi:10.1037/0033-2909.106.1.148

Sowerby, P., Seal, S., & Tripp, G. (2011). Working memory deficits in ADHD: The contributions of age, learning/language difficulties, and task parameters. *Journal of Attention Disorders*, *15*, 461-472. doi:10.1177/1087054710370674

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, *56*, 197-208. doi:10.1177/0013164496056002001

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *31*, 321-327. doi:10.1007/BF02293557

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A view and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: A* *festschrift to Douglas Jackson at seventy* (pp. 41-71). Norwell, MA: Kluwer Academic.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research*, *25*, 89-95. doi:10.1207/s15327906mbr2501_12

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228-243. doi:10.1037/a0027127

Watkins, M. W. (2000). Monte Carlo PCA for Parallel Analysis [Computer software]. State College, PA: Ed & Psych.

Watkins, M. W. (2004). MacOrtho [Computer software]. State College, PA: Ed & Psych.

Watkins, M. W. (2013). Omega [Computer software]. Phoenix, AZ: Ed & Psych.

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

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*, 354-365. doi:10.1037/1082-989X.1.4.354

Yung, Y. F., McLeod, L. D., & Thissen, D. (1999). On the relationship between the higher-order model and the hierarchical factor model. *Psychometrika*, *64*, 113-128. doi:10.1007/BF02294531