This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Structural Equation Modeling: A Multidisciplinary Journal* on 27/06/2019, available online: https://doi.org/10.1080/10705511.2019.1622421.

Please use the following citation when referencing this work:

Dombrowski, S. C., Beaujean, A. A., McGill, R. J., Benson, N. F., & Schneider, W. J. (in press). Using exploratory bifactor analysis to understand the latent structure of multidimensional psychological measures: An example featuring the WISC-V. *Structural Equation Modeling*. Advance online publication. doi: 10.1080/10705511.2019.1622421

# Using Exploratory Bifactor Analysis to Understand the Latent Structure of Multidimensional Psychological Measures: An Example Featuring the WISC-V

Stefan C. Dombrowski

Rider University

A. Alexander Beaujean

Baylor University

Ryan J. McGill

William & Mary

Nicholas F. Benson

Baylor University

W. Joel Schneider

Temple University

Author Note

Correspondence concerning this article should be addressed to Stefan C. Dombrowski, Ph.D., Professor, Department of Graduate Education, Leadership & Counseling. Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648. E-Mail: sdombrowski@rider.edu

**Abstract**

Exploratory bifactor analysis (EBFA) represents a methodological advancement for implementing a bifactor model in exploratory factor analysis (EFA). It can be useful for understanding the multidimensional structure of measures of cognitive ability, achievement, behavior and personality. However, little is known about how to properly employ the procedure. The current rotation criteria available for EBFA make it more likely to "get stuck" in local minima, contributing to possible group factor collapse, than more traditional EFA rotations. Thus, getting a proper solution is a more complex and involved process than typical EFA and may require a sensitivity analysis. This article examines EBFA through a sensitivity analysis and subsequent simulation of parameters thought to contribute to group factor collapse. Results support the use of sensitivity analysis, as the problematic variable was shown to greatly increase the likelihood of factor collapse. The hypothesis that estimation start values contribute to factor collapse was not supported. Accompanying R syntax for all analyses is provided to facilitate reproducibility.

Keywords: Exploratory bifactor analysis; Sensitivity analysis; Monte Carlo Simulation; Wechsler Intelligence Scale for Children–Fifth Edition

**Using Exploratory Bifactor Analysis to Understand the Latent Structure of Multidimensional Psychological Measures: An Example Featuring the WISC-V**

Bifactor models can be a useful way to represent the theories concerning some psychological structures (Canivez, 2016; Reise, 2012). This model represents theories that posit two main sources of influence on observable phenomena: (1) a single, general factor that is common to all phenomena; and (2) one or more non-general (i.e., "group") factors that represent what is common to only a subset of the phenomena independent of the general factor. If there are two or more group factors they can be correlated or uncorrelated, although traditionally they are uncorrelated so that each factor represents a unique contribution on the phenomena.

A bifactor model has a factor loading matrix ($\Lambda_B$) of the form

$$\Lambda_B = \begin{bmatrix} * & * & - \\ * & * & - \\ * & * & - \\ * & - & * \\ * & - & * \\ * & - & * \end{bmatrix}. \tag{1}$$

In Equation (1), * denotes an estimated parameter and – either denote parameters constrained to be zero (in confirmatory factor analysis [CFA]) or estimated parameters thought to be close to zero (exploratory factor analysis [EFA]).

The procedures for implementing a bifactor model in CFA are well established, but implementing a bifactor model in EFA has historically been more difficult. Karl Holzinger and colleagues (e.g., Holzinger & Swineford, 1937) developed the bifactor technique as an extension of Spearman's two-factor model, but it bears little resemblance to modern conceptualizations of EFA. For example, it requires *a priori* specification of what measured variables relate to what group factors, involves no factor rotation, and does not use an iterative analytic method to extract factors. Consequently, while the notion that a general factor should be of prime importance has been continued, Holzinger's bifactor technique has not. Instead, researchers wanting to specify a bifactor model in EFA had to rely on the use of anapproximation of the technique (Reise, Moore, & Haviland, 2010). One of the most common EFA approximations has been to conduct a second-order factor analysis using some iterative method for factor extraction and rotation, and then calculate both the indirect effects of the general factor and the direct effects of the residualized group factors (e.g., Schmid-Leiman or Cattell-White procedures). [1]

The use of bifactor models in EFA changed when Jennrich and Bentler (2011) developed an analytic factor rotation with a bifactor criterion. Their rotation criterion specifies that all measured variables load on the first general factor and encourages a perfect cluster structure for the loadings on the remaining group factors (i.e., $\Lambda_B$ in Equation 1). They called the process of conducting EFA using a bi-factor rotation criterion *exploratory bifactor analysis* (EBFA).

**Higher-Order vs. Bifactor Models**

As previously mentioned, conducting a higher-order EFA followed by an "orthogonalization" of higher-order factor loadings has been used by researchers to approximate a bifactor solution in EFA for decades (Carroll, 1983). In particular, use of this technique grew dramatically in popularity in the 1990s after Carroll (1993) used the procedure developed by Schmid and Leiman (SL; 1957) throughout his masterwork on human intelligence. Historically, the SL technique has been implemented primarily in studies of cognitive ability (e.g., Canivez & Watkins, 2010a, 2010b; Carroll, 1993; Dombrowski, Watkins, & Brogan, 2009; Dombrowski & Watkins, 2013; Dombrowski, 2013; Dombrowski, Canivez & Watkins, 2017; Dombrowski, McGill, & Canivez, 2017; Gignac, 2005; Gignac & Watkins, 2013), but it has also been used to

---

[1] Throughout this paper, we assume all high-order factor models include multiple first-order factors and a single second-order factor, and all bifactor models include only first-order orthogonal group factors.

investigate other psychological constructs (e.g., Ackerman, Donnellan, & Robins, 2012; Brouwer, Meijer, & Zevalkink, 2013; Chen, West, & Sousa, 2006).

Because of its popularity, it is not uncommon to believe that the factor loadings from a bifactor model are functionally equivalent to those from the SL procedure. This will not always be the case, however, since higher-order and bifactor models are generally not equivalent and represent competing representations of the structure that is thought to underlie the data. The SL procedure is just a transformation of a higher order model's loadings—nothing new is estimated in the process (Loehlin & Beaujean, 2016b). Thus, the SL procedure will only duplicate the results of a bifactor model if the measured phenomena actually came from a higher-order process; otherwise, the SL procedure can provide misleading results in some cases where it is implemented as a proxy for EBFA (Jennrich & Bentler, 2011; Reise, 2012; Yung, Thissen, & McLeod, 1999).

Conceptually, a higher-order model (with or without applying the SL procedure) represents something distinct from a bifactor model (Beaujean, 2015). In a typical higher-order model, the general factor has no direct influence on the measured variables—only the first-order (group) variables have a direct effect. In a bifactor model, however, both general and group factors directly affect the measured variables.

In addition, group and general factors take on different meanings in the two models. In a bifactor model, the general factor represents what is in common with all the measured variables and is prioritized over group factors. Group factors represent things (e.g., abilities) that can explain individual differences in the measured phenomena over and above the general factor. This interpretation of both the general factor in group factors is aligned with Spearman's theory of noëgenesis, which was the original basis of the model.

By contrast, in a higher-order model group factors represent all that is in common with a subset of the measured variables. They are prioritized over a general factor, so the general factor represents all that is in common with the first-order factors (i.e., it is a superordinate variable).[2] Such interpretations are consistent with L. L. Thurstone's Primary Mental Abilities theory, which was the original basis of this model.

While there are other differences as well (e.g., Gignac, 2008), the main point is that the two models represent distinct theories of psychological phenomena. For example, if an intervention was discovered by which a general factor of intelligence could be manipulated directly, the two models would predict different treatment outcomes. A higher-order model implies that any influence on the general factor would necessarily alter the group factors as well. A bifactor model implies that the general factor is causally independent of the group factors, and thus selective manipulation of the general factor would have no downstream effects on the group factors.

**Exploratory Bifactor Analysis**

The EBFA process is similar to typical EFA: extract a set of common factors and then rotate the factors (Loehlin & Beaujean, 2016a). Yet, there are two key differences (Mansolf & Reise, 2016). First, in EBFA the number of factors to extract has to include all hypothesized group factors plus a general factor. Thus, EBFA always requires extracting one additional factor than the number of group factors from a commensurate higher-order EFA. The extraction of an additional factor is not an issue for data if they represent phenomena consistent with being generated from a bifactor structure. However, extracting additional factors can sometimes produce parameter estimation problems especially when used with data consistent with being generated from a higher-order or non-hierarchical structure. In these situations, extracting the

[2] In theory, the general factor could also have both direct and indirect effects on the measured variables (e.g., Yung et al., 1999), but there is not a method for examining such a model in EFA.

additional factor is tantamount to over-extraction which can result in biased or nonsensical parameter estimates (for more detail, see Mansolf & Reise, 2016; Wood, Tataryn, & Gorsuch, 1996).

Second, multiple random starting values need to be used for the bifactor rotation. Although any analytic EFA rotation can get stuck in local minima (Browne, 2001), the optimization algorithm for the bifactor rotations currently available are especially prone to this. Using multiple random starting values increases the likelihood of finding results that come from a global minimum (Mansolf & Reise, 2016). As a consequence, EBFA requires examining the resulting optimization values from the random starts as well as loadings from multiple unique solutions.

There are currently two bifactor rotations available: bi-quartimin and bi-geomin. They are adaptions of the quartimin and geomin rotation criteria, respectively, for use with a bifactor criterion (for the actual criteria, see Jennrich & Bentler, 2011, 2012). Both rotations have oblique and orthogonal options for the group factors, but most implementations in substantive research use the orthogonal rotation. The rotations are currently available as options in R, EQS, and Mplus, but can be implemented in other software with some additional programming.

The two bifactor rotations are similar in that they only penalize row complexity for group factor loadings and only the group factors are rotated explicitly; the general factor is rotated implicitly after rotating the group factors. Thus, optimizing either bifactor rotation’s criterion is tantamount to minimizing row complexity for the group factors. They will produce similar results except when there is rank complexity in the measured variables such as cross-loadings. In this situation, the bi-geomin produces superior results to bi-quartimin since the geomin rotation works better than quartimin rotation in the presence of departures from perfect cluster structure (Sass & Schmitt, 2010).

Although EBFA is relatively new, it is gaining in popularity. Beaujean (2013, 2014, Loehlin & Beaujean, 2016b) discussed how to implement the procedure in R, but did so only within the context of general EFA procedures. Others have used EBFA to understand psychological constructs (Revelle & Wilt, 2013) and investigate the structure of certain tests, such as the Anxiety Sensitivity Index-3 (Ebesutani, McLeish, Luberto, Young, & Maack, 2014), Psychological Need Thwarting Scale (Myers, Martin, Ntoumanis, Celimli, & Bartholomew, 2014), the Woodcock-Johnson III (Dombrowski, 2014a; 2014b), and the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; Dombrowski, Beaujean, Canivez & Watkins, 2015).

**Cautions in Using Exploratory Bifactor Analysis**

Having an analytic bifactor rotation is a considerable methodological step forward. It allows for a direct application of a bifactor model within an EFA framework. Thus, if the data come from a bifactor structure, using a bifactor rotation should provide a more accurate estimate of the influence of general and group factors than using models or rotations that only approximate a bifactor model. As with any new method, it needs to be implemented cautiously. A recent application of EBFA serves as an example.

Dombrowski, Beaujean, Canivez, and Watkins (2015) used EBFA to analyze the normative data provided in the WISC-V Technical and Interpretive Manual (Wechsler, 2014b). They found evidence for a factor solution that differed considerably from the theoretical model posited by the test publisher. Specifically, Dombrowski et al. found evidence for three group factors (Perceptual Reasoning, Working Memory, Processing Speed), which was in stark contrast to the publisher's five-factor structure (Fluid Reasoning, Processing Speed, Verbal Comprehension, Visual Spatial, and Working Memory).[3] Of particular note in Dombrowski et

[3] Perceptual Reasoning is a combination of the Fluid Reasoning and Visual Spatial factors.

al.'s (2015) factor solution was the lack of a Verbal Comprehension factor; the subtests designed to measure verbal comprehension only loaded onto the general factor. Dombrowski et al. ascribed a plausible theoretical rationale for the resulting structure: the WISC-V is a verbally dominated instrument and therefore the general factor explains most of the verbal subtests' variance and leaves nominal residual variance to form a separate group factor. Others have found similar results (e.g., cognitive dedifferentiation; Baltes, Cornelius, Spiro, Nesselroade, & Willis, 1980), but such findings have typically been restricted to samples of elderly individuals. Moreover, other investigations of the WISC-V have generally supported the viability of a solution with four group factors that included the verbal factor (i.e., Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed; Canivez, Watkins, & Dombrowski, 2016, 2017; Dombrowski, Canivez, & Watkins, 2017; Watkins, Dombrowski, & Canivez, 2018).

Beyond the rationale offered by Dombrowski et al. (2015), it is possible that there is a methodological explanation for their results. When data come from a complex factor structure where both the loadings and factor correlations are strong, then some oblique rotations are prone to produce factor collapse (i.e., the correlation between two factors tend toward unity; Sass & Schmitt, 2010). Measures of cognitive ability have historically been strongly correlated and produce strong factor loadings—especially Wechsler scales. In the WISC-V, for example, the group factor correlations range from .64–.85 and subtests' factor loadings range from .64–.85. A notable exception is the Cancellation subtest, which has a loading of .41 (WISC-V Technical Manual, p. 83).[4]

Although bifactor rotations are orthogonal (at least between the general and group factors), they are susceptible to factor collapse. This is because of the extraction of an additional

[4] The technical manual does not provide the actual factor correlations, so we used the implied correlations from the publisher's preferred five-factor model.

factor and their proneness to get stuck in local minima (Mansolf & Reise, 2015). In EBFA, factor collapse occurs when the general factor absorbs too much variance from a subset of the measured variables. This causes those variables to have loadings on the general factor that are too large, which either makes their group factor loadings too small or removes them altogether. Since the bifactor rotations' criteria only penalize row complexity for the group factors' loadings, a solution with a collapsed factor could provide values for either rotation's criterion that are as small as a true bifactor solution. Thus, the current bi-factor rotations available require using multiple starting values and examining the interpretability of multiple solutions to ensure that problematic solutions are not retained.

Because of the estimation quirks with the bifactor rotations, Dombrowski et al.'s (2015) EBFA may support the cognitive dedifferentiation idea in younger samples or may have arisen from methodological issues. One way to aid in differentiating between these two hypotheses is to conduct a sensitivity analysis. A sensitivity analysis is a study of whether, and to what extent, the results of a given analysis are sensitive to plausible changes in the main assumptions and numeric inputs (Office for the Management and Budget of the White House, 2003; Saltelli, Tarantola, Campolongo, & Ratto, 2004). In other words, do the results substantially change when changing some aspect involved in parameter estimation? If so, a particular solution may be an artifact of an arbitrary input parameter employed by a researcher and may not represent actual relationships among psychological variables.

The aspects varied in a sensitivity analysis can include any part of the parameter estimation process, ranging from sample composition (e.g., inclusion and exclusion of influential observations) to technical aspects in the estimation program (e.g., differing starting values). If the method is robust, then the conclusions from the differing analyses should remain relatively constant. If the conclusions differ, then the results may—at least in part—stem from methodological decisions made throughout the analysis process raising concern about the potential impact of the garden of forking paths (i.e., Gelman & Loken, 2013). Thus, it is important to examine how the results differ and determine if there are particular assumptions or inputs that are responsible for the differing conclusions. If results differ, then it is important to further investigate the possible causes for differing results.

In EFA, sensitivity analysis can be applied to any aspect of the analysis, including both the factor extraction and factor rotation steps (Flora, LaBrish, & Chalmers, 2012). In the factor extraction step, the initial un-rotated factor loadings are usually not interpretable, so sensitivity studies focus on the indicators' communalities. In the factor rotation step, the focus is often on the factor loadings and their subsequent interpretation.

**Purpose of the Study**

The present study consists of two experiments. The first offers an applied example of EBFA using data from the WISC-V normative sample and examines the influence of some input decisions through a sensitivity analysis. The WISC-V data were used for a few reasons. First, it has some features that make it ripe for possible EBFA problems such as strong factor loadings, strong factor correlations, and problematic variables (e.g., Cancellation subtest). Second, it uses the same data that Dombrowski et al. (2015). Thus, this analysis can serve as a re-investigation of the resulting structure of the WISC-V when both four and five group factors are extracted.

The second experiment is a Monte Carlo study designed to examine the prevalence of factor collapse. Using conditions similar to those from the first experiment, we simulated data in multiple conditions concerning factor loadings and commonality start values to determine the effect on factor collapse.

## Experiment 1 Method

**Participants**

Participants included the entire WISC-V standardization sample (*n* = 2200), ranging in age from 6 to 16 years. The test publisher obtained the standardization sample using stratified proportional sampling across variables of age, sex, race/ethnicity, parental education level, and geographic region. Other demographic characteristics are available in the WISC-V technical and interpretive manual (Wechsler, 2014b).

**Instrument**

The WISC-V is an individually administered test of cognitive ability for children aged 6–16 years. The instrument contains 16 subtests thought to measure cognitive ability in five unique domains: Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed. Table 1 summarizes the publisher's proposed subtest alignment. A more thorough description of the instrument can be found in the Technical Manual and Interpretive Manual (Wechsler, 2014b; cf. Canivez & Watkins, 2016).

**Procedure and Analyses**

All analyses were conducted in R (R Development Core Team, 2019), using the *psych* (Revelle, 2012) package for factor extraction and the *GPArotation* (Bernaards & Jennrich, 2005) package for rotation. To facilitate replication, the R syntax used in the present study is furnished through the Open Science Framework at (https://osf.io/ne795/). For all analyses, the total sample correlation matrix (Wechsler, 2014b, p. 74) was used.

First, the number of factors to extract were examined by conducting the minimum average partial (MAP) values, parallel analysis (PA; using both the mean and 95$^{th}$ percentile criteria), and Bayesian information criterion (BIC). Second, the factors were extracted using principal axis factoring. Starting values for the communality estimates that ranged from .1 to .9 (using increments of .1), as well as each variable's squared multiple correlation were utilized. Principal axis factoring was used because it is the only extraction method in the program that allows for user-specified starting values. Although the starting values should not affect communality estimates if the number of iterations is large and minimization criterion is small (Widaman & Herringer, 1985), the starting values were investigated because some of the WISC-V subtests have the potential to be problematic because of small inter-subtest correlations. For comparison, maximum likelihood (ML) extraction was also utilized. For all extractions, this study allowed for 100 iterations and used a convergence criterion (i.e., difference in communality estimates) of .001. Third, the quality of the initial extractions was examined by evaluating the degree to which the model reproduced the sample correlation matrix, the stability of the communality estimates, and the value of the ML objective function.[5] Although both principal axis and ML were used, values from this objective function can still be utilized to compare solutions.

Fourth, the optimal solutions from step three were identified, selected, and subsequently rotated using the orthogonal bi-quartimin and bi-geomin rotations. For each rotation, 1000 random starts were used and loadings from up to 10 of the best unique solutions were examined as recommended by Mansolf and Reise (2016). If the solutions differed in their interpretability, the solution that most aligned with the WISC-V's publisher's intended factor structure was selected.

## Experiment 1 Results

### Factor Extraction

**Number of factors.** The MAP criterion indicated extracting a single factor, but the BIC and PA converged to suggest four group factors. Consequently, four group factors (five factors

---

[5] The ML objective function is

$$\log(\mathrm{tr}\left(\left(\mathbf{\Lambda\Lambda}' + \mathbf{U}^2\right)^{-1}\mathbf{R}\right) - \log(|\left(\mathbf{\Lambda\Lambda}' + \mathbf{U}^2\right)^{-1}\mathbf{R}|) - p,$$

where $\mathbf{U}^2$ are the variable uniqueness's, $\mathbf{R}$ is the sample correlation matrix, tr() the matrix trace function, and $p$ is the number of measured variables.

total) were extracted for the EBFA and much of this study focuses on these results. Although no index indicated extracting five group factors (six factors total), the results from such a solution were examined since that is the number of group factors indicated by the WISC-V publisher.

**Communality estimates.** All five-factor solutions converged. The final communality estimates from each starting value are provided in Table 2. The difference in communality estimates for a given subtest was ≤ .20 across all start values except for Cancellation. For Cancellation, the final communality estimates ranged from .18 to .82—a difference of .63. This degree of communality fluctuation across start values suggests that this subtest might be a problematic variable.

Inspection of the fit and objective functions indicated that starting values of .3 and the variables' squared multiple correlations (SMCs) produced the best fit, both of which were equivalent to the ML extraction. Of note, the communality estimates are identical within .04 units for all subtests across all three start values (SMC, .3 & ML) solutions.

For the six-factor solutions, there were multiple indications of factor instability. First, the results did not converge when using SMCs as start values. Second, although the difference in communality estimates for a given indicator were ≤ .20 for most subtests, there were large differences for three of the subtests: Coding (difference of .33), Picture Concepts (difference of .70), and Cancellation (difference of .76). Communality estimates are provided in Table 3. The communality differences are noticeably larger for the six-factor solutions than the five-factor solutions, both in number of problematic subtests as well as actual communality differences. Third, the fit and the objective functions indicated that .30 starting value had the best fit, and these values were very similar to those for the ML extraction. While the communality estimates were similar for both solutions, there were some noticeable differences for the Information (.32), Coding (.34), and Figure Weights (.38) subtests. Thus, even though extraction methods produce almost identical values for the fit and objective function, the resulting factor loadings are far from identical for some of the subtests.

**Bifactor Rotation**

The loadings from the five-factor solution using the .30 and SMC starting values were rotated using the bi-quartimin and bi-geomin rotations. The results were identical, so the results from the .30 start value received interpretive emphasis. In addition, the six-factor solution for the .30 starting value was rotated.

In the five-factor solution, there were four unique optimization values for the bi-quartimin rotation and three for the bi-geomin rotation. The smallest optimization value for both rotations produced results with identical factor interpretations and the factors align with some of the constructs the WISC-V publishers intended to measure: Working Memory, Verbal Comprehension, Processing Speed, and Visual Spatial abilities. Thus, it is likely that both solutions result from reaching a global minimum.

The bi-quartimin and bi-geomin solutions are presented in Table 4. The factor loadings are relatively large on the general factors for all subtests except for those measuring processing speed: Coding, Symbol Search, and Cancellation. Conversely, factor loadings are relatively small for the group factors except for Processing Speed. The exception is with the Cancellation subtest, which has low loadings on both the general factor (.20) and Processing Speed factor (.37–.38)

For both rotations, a group factor emerged that represents Verbal Comprehension. Yet, the subtests that comprise this factor have relatively small loadings, indicating that the Verbal Comprehension contains little reliable variance that is unique from the general factor. In fact, the $\omega_{\text{Subscale}}$ value (Reise, Bonifay, & Haviland, 2013) is only approximately .20 across both rotations.

For the six-factor solution, there were multiple indications that the results were problematic (see Table 5). First, there was 1 unique optimization value for the bi-quartimin rotation, but 7 for the bi-geomin rotation. While the bi-geomin rotations produced somewhat similar interpretations, these differed from interpretation of the bi-quartimin rotation. The bi-quartimin's group factors represent Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed abilities. In the bi-geomin rotation there are two district Processing Speed factors—and this was consistent across all seven solutions. The additional group factors from the bi-geomin solution represent Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed abilities. Second, in both rotations the Fluid Reasoning factor was only comprised of one subtest, Figure Weights. Moreover, in the bi-geomin rotation two of the Processing Speed subtests (Cancellation and Symbol Search) had non-negligible cross loadings.

In summary, there were multiple problems with the six-factor solution, both in the extraction and rotation aspects of EBFA. Thus, the optimal five-factor solution appears to provide the best—if not only—interpretable solution.

## Experiment 2 Method

### Procedure and Analyses

All analyses were conducted in R (R Development Core Team, 2019). Factor extraction was done using the *psych* (Revelle, 2012) package and factor extraction was done using the *GPArotation* (Bernaards & Jennrich, 2005) package. The R syntax used in the present study is furnished through the Open Science Framework at (https://osf.io/ne795/).

We conducted a Monte Carlo (MC) study (Paxton, Curran, Bollen, Kirby, & Chen, 2001) to see the effect on factor collapse from EBFA solutions. We manipulated two conditions: (a) presence of a problematic variable and (b) estimation start values. We simulated the data using a model with bifactor cluster structure that had one general factor and five group factors. We selected most of the loadings in such a way that (a) they made theoretical sense based on previous WISC-V factor analyses, and (b) the resulting correlations would approximate those from the WISC-V data in Experiment 1 (see Table 6). For the problematic variable (V16) we set the loadings to .30, .60, and .80. We simulated $m$ =1000 data sets for all three V16 loadings, each using $n$ = 500 observations. For each of the datasets, we extracted six factors using principal axis extraction with either a communality start value of .3 or .5, since these levels reflected non-collapse and collapse, respectively, in Experiment 1. This produced 6000 sets of unrotated factor loadings (i.e., 3 problematic variable loadings x 2 start values x 1000 replications). For each EBFA, we used 50 random starts and selected the single solution that produced the smallest value for the bi-quartamin criterion.

There are not currently any quantitative criteria for determining when factor collapse occurs. Consequently, to examine collapse we did the following for each EFBA solution. First, we took the absolute value of the loadings to make sure that the loadings that defined each factor were positive. Second, we removed the general factor loadings. Third, we subset the loadings to only include the block of variables representing the Verbal Comprehension Index (VCI. Fourth, we selected the group factor that contained the maximum loading to represent the VCI factor and summed the loadings. This sum of loadings was the value we used to examine collapse. Fifth, we examined the proportion of the summed loadings that were below (i.e., collapsed) or above (i.e., not collapsed) threshold levels ranging from .80 to 2.0, in increments of .20.

## Experiment 2 Results

The loading sums are displayed in Figure 1 for each of the six conditions. For all conditions, the distribution of loadings sums is bimodal. Most of the sums are relatively large (i.e., not collapsed) and center around 1.5, which is an average loading of approximately .38 (i.e., 1.5/4). Nonetheles, in all conditions there is also a small proportion of values that are relatively small (i.e., centered around 0.5), which likely indicate that the factor had collapsed.

Although we originally simulated $m = 1000$ replications in each condition, some of the factor solutions did not converge. The number of converged solutions is provided in Table 7. Table 7 also contains the proportion of collapsed factors for a given threshold value, communality start value, and population loadings of the problematic variable. The table indicates that across all thresholds and start values, collapse was more likely to occur for smaller population values.

Table 8 has the ANOVA source table for the MC experiment. There was no interaction between the communality start values and problematic variable population values, nor was there a main effect of start value. Consequently, the start values we used did not typically influence the sum of the loadings. There was a small, yet statistically significant, effect of the population start value for the problematic variable ($F = 87.477$, $p < .001$, Partial $\omega^2$=.03).

## Discussion

Although EBFA has only been available a short period of time, researchers and practitioners have used it to examine the structure of multidimensional data, especially when there are questions about the use of scores representing general versus more specific constructs (Rodriguez, Reise, & Haviland, 2016). EBFA procedures that are currently available to researchers are a little more quirky than more traditional EFA since they require using multiple starting values and examining multiple solutions. To paraphrase Mansolf and Reise (2016): EBFA works but requires more from users than traditional EFA for it to work consistently. The tradeoff for this extra effort is that it produces more accurate factor loadings estimates than more approximate methods when the underlying data come from a bifactor structure.

Previously, Dombrowski and colleagues (2015) conducted an EBFA to investigate the structure of the WISC-V. Their results were consistent with some previous cognitive ability research, but inconsistent with the WISC-V theoretical structure as well as other WISC-V studies that have been produced over the course of the last five years. This raised questions about the tenability of the findings, so the present study re-analyzed the WISC-V data using a sensitivity analysis with EBFA to shed insight on the source of these incongruent results. Specifically, this study examined whether the initial starting values for the factor communalities or method of bi-factor rotation could have influenced Dombrowski et al.'s findings. The results of the current study suggest that both can affect the results.

**Sensitivity Analysis Findings**

The initial communality estimate/starting value made a substantial difference for some WISC-V subtests, with this effect being more potent for the six-factor solution (one general and five group factors) than the five-factor solution (one general and four group factors). For the five-factor solution, there were minimal differences in any of the subtests' communality values across start values except for one subtest: Cancellation. For this subtest, the final communality values ranged from .18 to .82 depending on the starting value. Differences this large for a single variable would influence the results from any factor rotation, including bi-factor rotations. For the six-factor solution, there were multiple problems with the initial factor extraction, including non-convergence for some start values and sizable communality differences for multiple variables across both start values and extraction methods.

There are two implications from this study. First, a six-factor solution of the WISC-V does not work very well. Although contrary to the preferred solution of the WISC-V publishers, this is consistent with Dombrowski et al.'s (2015) findings as well as other WISC-V studies. Unlike Dombrowski et al., the current study indicated that a five-factor solution (1 general 4 group) was both optimal and interpretable. Although Dombrowski et al. examined a five-factor solution, the Processing Speed subtests formed two Processing Speed factors, with one of them being comprised solely of the Cancellation subtest.

Second, the start values used to estimate the communalities for the extracted factors can make a difference—at least with the WISC-V data. On the one hand, this was unexpected given that other research has indicated that the start values used tend to be irrelevant (Widaman & Herringer, 1985). On the other hand, this makes sense given the low communality for the Cancellation subtest (.18 for the best five-factor solution in the current study). In other words, when there are variables that have such little in common with other variables, the initial communality estimate becomes very important. Consequently, it is possible that the starting values for the communality estimates contributed to Dombrowski et al.'s results. Dombrowski et al. used a starting value of .50 for all their factor solutions because the default start values (i.e., SMCs) produced non-convergence for some of the factor solutions. In the current study, a variety of start values were used, but the .30 start value produced the best results. The differences in the communality estimates for Dombrowski et al.'s five-factor solution (Table 3, p. 198) and the best five-factor solution in the current study are relatively small ($\leq .14$) except for the Cancellation subtest (difference of .31).[6]

As with starting values, the effect of the bifactor rotation procedure depended on the initial factor solution. In the current study, both bi-quartimin and bi-geomin rotations were implemented. Based on Mansolf and Reise's (2016) EBFA recommendations, 1000 random starts were used and up to 10 of the best unique solutions for each rotation were examined. With the best five-factor solution, the optimum bi-quartimin and bi-geomin solutions were determined to be functionally equivalent and produced identical factor interpretations. With the best six-

[6] Based on the factor loadings, the communality reported by Dombrowski et al. (2015, Table 3, p. 198) for Cancellation in the five-factor solution should be .49.

factor solution, however, the optimum bi-quartimin and bi-geomin loadings produced considerably different factor interpretations (see Table 5). The bi-quartimin rotation produced factors representing Visual Spatial, Fluid Reasoning, Working Memory, Processing Speed, and Verbal Comprehension abilities. The bi-geomin rotation also had factors representing Visual Spatial, Fluid Reasoning, and Working Memory abilities; however, instead of a Verbal Comprehension factor it produced two factors representing Processing Speed—results similar to those reported by Dombrowski et al. (2015) for their six-factor solution.

Dombrowski et al. (2015) only used bi-quartimin rotation, 10 random starts, and examined the single best solution—what Jennrich and Bentler (2011) initially recommended for EBFA modeling. Mansolf and Reise (2016) found that using so few random starts and solutions increases the likelihood of the factor loadings produced from the bifactor rotation getting stuck in a local minimum. Thus, it is possible Dombrowski and colleagues (2015) inability to locate a Verbal Comprehension factor with their five-factor solution is an artifact obtaining results from a local minimum (and subsequent factor collapse). In addition to communality starting values and factor rotation procedures, the unique nature of the WISC-V norming data likely also contributed to to these results. As a whole, the WISC-V data are factorable—the correlation matrix is not spherical, only 8% of the elements in anti-image correlation matrix were not zero to the first decimal place, and the measure of sampling adequacy value of .94 makes it "marvelous" for factor analysis (Dziuban & Shirkey, 1974). Yet, there are aspects of the WISC-V data that make factor analysis—especially EBFA—challenging.

First, the Cancellation subtest seems to have little in common with other subtests—even those measuring processing speed—making its communality very low. Likely this is because cancellation-type tasks are traditionally used to measure visual attention or perceptual speed, which is distinct from general processing speed (Carroll, 1993). Having subtests with little in common with other subtests makes the whole enterprise of EFA difficult. These variables make little contribution to any common factor, so they hinder the process of correctly determining the number of factors to extract as well as their interpretation. As demonstrated in the current study, these variables can also contribute to specious communality estimates.

Second, there is no theoretical consensus about the number of distinct abilities that the WISC-V measures (Canivez & Kush, 2013). While the WISC-V technical manual says it measures five abilities in addition to *g*, other factor solutions are also common. For example, some analyses have shown it best to combine Visual Spatial and Fluid Reasoning factors into a single factor measuring Perceptual Reasoning (Canivez, Watkins & Dombrowski, 2016, 2017; Dombrowski, Canivez & Watkins, 2017; Watkins, Dombrowski & Canivez, 2017). In addition, the subtests designed to measure Fluid Reasoning—and to a lesser extent those measuring Verbal Comprehension and Visual Spatial abilities, too—are primarily measuring *g*. Thus, it is difficult to separate factors that measure these abilities from *g* very reliability. In the WISC-V technical manual, the best model had correlations between a higher-order *g* and lower-order Fluid Reasoning, Visual Spatial, and Verbal Comprehension factors that were 1.00, .88, and .85, respectively (Wechsler, 2014b, p. 83). In the current study, our best EBFA solution did not even include a separate Fluid Reasoning factor and the loadings on the Verbal Comprehension and Visual Spatial factors were all relatively low (average value of .39; See Table 4). Consequently, while Dombrowski et al.'s (2015) favored five-factor model is distinct from the five-factor model favored in the current study, the actual difference between these models is minimal. In fact, their major conclusion that "the WISC-V is primarily a measure of *g*, as it accounts for a majority of the subtests' total and common variance" applies to this study as much as it does to theirs.

There is a final point regarding EBFA that deserves comment. Research has discussed the bifactor model's propensity to fit any data, including randomly generated data (Bonifay & Cai, 2017; Reise et al., 2016). Bonifay and Cai (2017) contend that the bifactor is nearly as flexible as EFA, which suggests that the likelihood of extracting a well-fitting solution, even if totally contrived, may well be quite high. Of course, this is the case with many models and theory should be the ultimate arbiter of which model is preferred or best "fits" the data.

**Monte Carlo Study Findings**

The hypothesis that a problematic variable (e.g., Cancellation) could cause factor collapse was examined through Monte Carlo simulatoin. We simulated data under various three population start values (.3, .6, and .8) and examined the EFBA solutions using communality start values of .30 and .50. We then examined collapse using a range of thresholds for the sum of absolute loadings. The results of the MC study indicated that the problematic valuable was the only contributor to factor collapse. It appears that the observations from the sensitivity analysis portion of this study (i.e., Cancellation caused the optimization algorithm to get stuck in local minima subsequently leading to group factor collapse) is further supported by this MC simulation. The hypothesis that the communality start value also contributed to collapse was not supported by MC study.

The results of both the sensitivity and the MC studies suggests that Cancellation is a problematic variable. It has little in common with any other WISC-V subtest and likely caused the verbal group factor to collapse. This may well lead to the suggestion that the Cancellation subtest should be eliminated from future editions of the WISC.

It might be tempting to blame the collapse on the bifactor rotation—after all, there was no ostensible evidence for group factor collapse when extracting using maximum likelihood or when using the SL orthogonalization. This would lead to an inaccurate conclusion about the tenability of EBFA. Our results indicate that this issue is not with EBFA; it is with the inclusion of a variable that likely should have been eliminated from the dataset prior to conducting factor analysis. EBFA appears to be sensitive to problematic indicators that have less in common with others.

**Conclusion**

The present study has some direct implications concerning EBFA. First, when using datasets containing problematic indicator variables, both the factor extraction and factor rotation steps need to be thoroughly investigated for possible methodological problems. Overreliance on the default options available in conventional statistical programs may lead to specious or unstable outcomes. A sensitivity analysis investigation of the effect of communality starting values and rotation starting values does not take very long, but can prove to be very insightful. The accompanying R syntax that is furnished in this study can aid in this endeavor.

Second, EBFA may not always work well directly "out of the box"—especially when there are problematic variables (e.g., small factor loadings, non-negligible cross-loadings). This means that the initial results from a bi-factor rotation—and perhaps factor extraction, too—cannot automatically be trusted. Instead, EBFA requires completing and examining multiple solutions. As Cliff (1983) noted many years ago, just because a computer was used to complete a complex statistical procedure it should not suspend people's normal critical faculties when interpreting such results.

Third, although the SL procedure has historically been the procedure of choice for approximating a bifactor model in EFA, consideration should be given to using EBFA if an exploratory bifactor structure is desired. The SL will still be useful for aiding with interpretation of results from a higher-order model. While in some situations the results from the SL procedure and EBFA will be similar, in general they represent competing theories about the influence and composition of general and group factors and the results should not *a priori* be expected to be the same (citation removed for masked review). Thus, EBFA should be strongly considered for directly implementing a bifactor model within EFA with the caveat that simulation studies will further support this contention.

Although the implementation of EBFA requires more steps than the SL procedure, the additional work it is not cost prohibitive. Mplus already has built-in options to use multiple starting values and extract results from multiple unique solutions, and Loehlin and Beaujean (2016b) provide similar capabilities for R. While processing time will differ among computers and datasets, the time required for each set of extractions and rotations in the sensitivity analysis took less than one minute using either R or Mplus.

In sum, EBFA is a viable option for implementing a bifactor model within the EFA framework; however, it should not be implemented without employing additional analytical procedures for investigating whether the resulting solutions are accurate. As demonstrated in this study, a sensitivity analysis can in aid in this investigation by examining the influence of communality start values as well as any influence of the implementation process for the bi-factor rotation.

**References**

Ackerman, R. A., Donnellan, M. B., & Robins, R. W. (2012). An Item Response Theory Analysis of the Narcissistic Personality Inventory. *Journal of Personality Assessment, 94*, 141-155. doi: 10.1080/00223891.2011.645934

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. (1980). Integration versus differentiation of fluid/crystallized intelligence in old age. *Developmental Psychology*, *16*, 625–635. doi: 10.1037/0012-1649.16.6.625

Beaujean, A. A. (2013). Factor analysis using R. *Practical Assessment, Research, and Evaluation, 18 (4),* 1–11. Retrieved from http://pareonline.net/pdf/v18n4.pdf

Beaujean, A. A. (2014). R syntax to accompany *Best practices in exploratory factor analysis* (2014) by Jason Osborne. Retrieved from: https://dl.dropboxusercontent.com/u/18489687/EFAbook/BestPracticesEFA_Rcode.pdf

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

Bernaards, C.A. and Jennrich, R.I. (2005) Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*, 676–696. doi: 10.1177/0013164404272507

Bonifay, W. & Cai, L. (2017). On the Complexity of Item Response Theory Models, *Multivariate Behavioral Research, 52*, 465-484. doi: 10.1080/00273171.2017.1309262

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology. *Clinical Psychological Science*, *5*, 184–186. doi: 10.1177/2167702616657069

Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). On the factor structure of the Beck

Depression Inventory–II: G is the key. *Psychological Assessment, 25*, 136-145. doi: 10.1037/a0029228

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150. doi:10.1207/s15327906mbr3601_05

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: Alternate 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. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV) adolescent subsample. *School Psychology Quarterly, 25,* 223-235. doi: 10.1037/a0022046

Canivez, G. L**.,** & Watkins, M. W. (2010b). 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. (2016). Review of the Wechsler Intelligence Scale for Children–Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 683–702). Hoboken, NJ: Wiley.

Canivez, G. L., Watkins, M. W. & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29*, 458–472. doi: 10.1037/pas0000358

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

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 1–33). New York, NY: Academic Press.

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

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189–225. doi: 10.1207/s15327906mbr4102_5

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115–126. doi: 10.1207/s15327906mbr1801_7

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

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

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

Dombrowski, S. C., Canivez, G. L. & Watkins, M. W. (2018). Factor structure of the 10 WISC–V primary subtests in four standardization age groups. *Contemporary School Psychology, 22,* 90-104. doi: 10.1007/s40688-017-0125-2

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

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment,* 27, 494–507. doi: 10.1177/0734282909333179

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

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53*, 194–201. Doi: 10.1016/j.intell.2015.10.009

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*, 358–361. doi:10.1037/h0036316

Ebesutani, C., McLeish, A. C., Luberto, C. M., Young, J., & Maack, D. J. (2014). A bifactor model of anxiety sensitivity: Analysis of the Anxiety Sensitivity Index–3. *Journal of Psychopathology and Behavioral Assessment*, *36*, 452–464. doi: 10.1007/s10862-013-9400-3

Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, *3* (55), 1–21. doi: 10.3389/fpsyg.2012.00055

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited a head of time.* Unpublished manuscript available at http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment, 12,* 320–329. doi: 10.1177/1073191105278118

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly, 50*, 21–43.

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

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41–54.
` doi: l 0.1007 /BF02287965

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 6*, 537–549. doi: 10.1007/s11336-011-9218-4

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*, 442–454. doi: 10.1007/s11336-012-9269-1

Loehlin, J. C., & Beaujean, A. A. (2016a). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). New York, NY: Routledge.

Loehlin, J. C., & Beaujean, A. A. (2016b). *Syntax companion for Latent variable models: An introduction to factor, path, and structural equation analysis (5th ed.)*. New York: NY. Retrieved from: sites.baylor.edu/lvm5

Mansolf, M. & Reise, S. P. (2015). Local minima in exploratory bifactor analysis. *Multivariate Behavioral Research, 50*, 738. doi: 10.1080/00273171.2015.1121127

Mansolf, M. & Reise, S. P. (2016) Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations, *Multivariate Behavioral Research, 51*, 698–717, doi: 10.1080/00273171.2016.1215898

Myers, N. D., Martin, J. J., Ntoumanis, N., Celimli, S., & Bartholomew, K. J. (2014). Exploratory bifactor analysis in sport, exercise, and performance psychology: A substantive-methodological synergy. *Sport, Exercise, and Performance Psychology*, *3*, 258–272. doi: 10.1037/spy0000015

Office for the Management and Budget of the White House (2003), Circular A4. Washington, DC: Author.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 287-312. doi:10.1207/S15328007SEM0802_7

R Development Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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 scores. *Journal of Personality Assessment, 92*, 544–559. doi: 10.1080/00223891.2010.496477

Revelle, W. (2012). *psych: Procedures for psychological, psychometric, and personality research* (version 1.2.4) [computer software]. Evanston, IL: Northwestern University.

Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, *47*, 493–504. doi: 10.1016/j.jrp.2013.04.012

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

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Hoboken, NJ: Wiley.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*, 73–103. doi: 10.1080/00273170903504810

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

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–Fifth Edition. *International Journal of School and Educational Psychology, 6*, 252–265. doi: 10.1080/21683603.2017.1342580

Wechsler, D., (2014b). *Wechsler Intelligence Scale for Children–Fifth Edition technical and interpretive manual.* Bloomington, MN: Pearson.

Widaman, K. F., & Herringer, L. G. (1985). Iterative least squares estimates of communality: Initial estimate need not affect stabilized value. *Psychometrika, 50*, 469–477. doi: 10.1007/BF02296264

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

**Conflicts of Interest: None**

Table 1

*WISC-V Publisher Proposed Subtest Alignment*

| Verbal Comprehension | Visual Spatial | Fluid Reasoning |
|---|---|---|
| Similarities | Block Design | Matrix Reasoning |
| Vocabulary | Visual Puzzles | Figure Weights |
| Information | | Picture Concepts |
| Comprehension | | Arithmetic |
| | | |
| Working Memory | Processing Speed | |
| Digit Span | Coding | |
| Picture Span | Symbol Search | |
| Letter-Number Sequencing | Cancellation | |

Table 2

*Communality Estimates Following Initial Five-Factor Extraction*

| | Initial Principal Axis Communality Start Values | | | | | | | | Maximum | |
|---|---|---|---|---|---|---|---|---|---|---|
| Subtest | SMC | .30 | .40 | .50 | .60 | .70 | .80 | .90 | Difference | ML |
| Similarities | .65 | .65 | .65 | .65 | .65 | .65 | .65 | .65 | .00 | .64 |
| Vocabulary | .74 | .74 | .74 | .74 | .72 | .73 | .73 | .74 | .02 | .74 |
| Information | .67 | .67 | .67 | .67 | .66 | .67 | .67 | .67 | .01 | .67 |
| Comprehension | .51 | .51 | .51 | .51 | .52 | .51 | .51 | .51 | .01 | .51 |
| Block Design | .55 | .55 | .57 | .57 | .57 | .57 | .57 | .57 | .02 | .54 |
| Visual Puzzles | .69 | .69 | .61 | .61 | .60 | .61 | .61 | .61 | .09 | .73 |
| Matrix Reasoning | .43 | .43 | .43 | .43 | .43 | .43 | .43 | .43 | .00 | .43 |
| Figure Weights | .56 | .56 | .45 | .45 | .46 | .45 | .45 | .45 | .11 | .55 |
| Picture Concepts | .28 | .28 | .27 | .27 | .28 | .27 | .27 | .27 | .00 | .27 |
| Arithmetic | .55 | .55 | .53 | .53 | .54 | .53 | .53 | .53 | .02 | .55 |
| Digit Span | .66 | .66 | .67 | .67 | .64 | .66 | .66 | .66 | .03 | .66 |
| Picture Span | .40 | .40 | .40 | .40 | .41 | .40 | .40 | .40 | .02 | .40 |
| Letter-Number Sequencing | .63 | .63 | .63 | .63 | .62 | .64 | .64 | .64 | .01 | .64 |
| Coding | .56 | .56 | .73 | .70 | .59 | .61 | .62 | .63 | .17 | .57 |
| Symbol Search | .60 | .60 | .51 | .51 | .58 | .57 | .56 | .55 | .10 | .60 |
| Cancellation | .18 | .18 | .40 | .49 | .56 | .65 | .73 | .82 | .63 | .18 |
| | | | | | | | | | | |
| Fit | .931 | .931 | .934 | .936 | .937 | .938 | .939 | .940 | .02 | .931 |
| Objective | .045 | .045 | .058 | .059 | .065 | .061 | .060 | .060 | .01 | .044 |

*Note*. SMC = Squared multiple correlation, where SMC = Initial communality estimate. ML = Maximum Likelihood Extraction with start value = 1. Fit: sum of actual squared correlations minus the sum of the implied squared correlations, divided by sum of actual squared correlations. Objective: value of objective function.

Table 3

*Communality estimates following initial six factor extraction*

| | Initial Principal Axis Communality Start Values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subtest** | **SMC** | **.30** | **.40** | **.50** | **.60** | **.70** | **.80** | **.90** | **1.0** | Maximum Difference | **ML** |
| Similarities | NC | .65 | .65 | .65 | .65 | .65 | .65 | .65 | .65 | .00 | .66 |
| Vocabulary | NC | .74 | .73 | .73 | .73 | .73 | .73 | .73 | .73 | .01 | .71 |
| Information | NC | .67 | .67 | .67 | .67 | .67 | .67 | .67 | .67 | .00 | 1.0 |
| Comprehension | NC | .51 | .51 | .51 | .51 | .51 | .51 | .51 | .51 | .00 | .53 |
| Block Design | NC | .56 | .58 | .59 | .59 | .58 | .58 | .58 | .58 | .03 | .57 |
| Visual Puzzles | NC | .68 | .61 | .61 | .60 | .60 | .60 | .60 | .60 | .08 | .66 |
| Matrix Reasoning | NC | .43 | .43 | .43 | .43 | .43 | .43 | .43 | .43 | .00 | .42 |
| Figure Weights | ***NC*** | ***.61*** | ***.46*** | ***.46*** | ***.46*** | ***.46*** | ***.46*** | ***.46*** | ***.46*** | ***.15*** | ***1.0*** |
| Picture Concepts | ***NC*** | ***.28*** | ***.38*** | ***.50*** | ***.59*** | ***.69*** | ***.79*** | ***.88*** | ***.98*** | ***.70*** | ***.28*** |
| Arithmetic | NC | .54 | .54 | .54 | .54 | .54 | .53 | .53 | .53 | .01 | .53 |
| Digit Span | NC | .67 | .66 | .66 | .66 | .66 | .66 | .66 | .66 | .01 | .67 |
| Picture Span | NC | .40 | .40 | .40 | .40 | .40 | .39 | .39 | .39 | .01 | .39 |
| Letter-Number Sequencing | NC | .63 | .64 | .64 | .64 | .64 | .64 | .64 | .64 | .01 | .64 |
| Coding | NC | ***.92*** | ***.60*** | ***.61*** | ***.62*** | ***.61*** | ***.63*** | ***.64*** | ***.66*** | ***.34*** | ***.58*** |
| Symbol Search | NC | .53 | .58 | .56 | .56 | .56 | .55 | .54 | .53 | .05 | .59 |
| Cancellation | ***NC*** | ***.23*** | ***.40*** | ***.50*** | ***.60*** | ***.70*** | ***.80*** | ***.90*** | ***1.0*** | ***.77*** | ***.18*** |
| Fit | NA | .936 | .937 | .941 | .944 | .947 | .949 | .950 | .950 | .14 | .937 |
| Objective | NA | .031 | .055 | .055 | .055 | .055 | .055 | .054 | .054 | .24 | .026 |

*Note*. SMC=Squared multiple correlation, where SMC=Initial communality estimate. ML=Maximum Likelihood Extraction with start value = 1.0. Fit: sum of actual squared correlations minus the sum of the implied squared correlations, divided by sum of actual squared correlations. Objective: value of objective function. NC=Non-Convergence. NA=Not available.

Table 4

*Five factor principal axis rotated solutions (start = .30 and SMC)*

| | Bi-quartimin rotation | | | | | Bi-geomin rotation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | g | Gc | Gv | Gwm | Gs | g | Gc | Gv | Gwm | Gs |
| Similarities | **.71** | **.37** | .01 | .03 | –.03 | **.72** | **.37** | –.00 | .02 | –.02 |
| Vocabulary | **.74** | **.43** | .02 | –.02 | –.09 | **.74** | **.42** | .01 | –.03 | –.07 |
| Information | **.73** | **.35** | –.02 | –.04 | –.05 | **.74** | **.35** | –.03 | –.05 | –.04 |
| Comprehension | **.61** | **.37** | –.01 | .06 | .03 | **.61** | **.36** | –.02 | .06 | .04 |
| Block Design | **.66** | -.02 | **.33** | –.05 | .11 | **.66** | –.03 | **.32** | –.05 | .12 |
| Visual Puzzles | **.66** | .01 | **.49** | –.03 | –.03 | **.67** | .00 | **.48** | –.04 | –.02 |
| Matrix Reasoning | **.64** | -.03 | .12 | .05 | .02 | **.64** | –.04 | .11 | .04 | .03 |
| Figure Weights | **.72** | -.14 | .03 | –.10 | –.11 | **.72** | –.15 | .02 | –.11 | –.09 |
| Picture Concepts | **.50** | .12 | .11 | .04 | .03 | **.50** | .11 | .11 | .04 | .03 |
| Arithmetic | **.73** | .04 | –.04 | .14 | .04 | **.73** | .03 | –.05 | .13 | .05 |
| Digit Span | **.66** | -.03 | –.01 | **.47** | .04 | **.67** | –.04 | –.02 | **.47** | .03 |
| Picture Span | **.54** | .00 | .04 | **.33** | .06 | **.54** | –.01 | .03 | **.32** | .05 |
| Letter-Number Sequencing | **.64** | .04 | –.06 | **.46** | .03 | **.65** | .04 | –.06 | **.46** | .02 |
| Coding | **.37** | -.03 | –.03 | .05 | **.65** | **.35** | –.03 | –.03 | .08 | **.65** |
| Symbol Search | **.42** | .00 | .03 | .00 | **.65** | **.41** | .00 | .03 | .03 | **.66** |
| Cancellation | .20 | -.01 | .02 | –.08 | **.37** | .19 | –.01 | .03 | –.06 | **.38** |

*Note*. *g* = General Factor. Gwm = Working memory. Gc = Crystallized ability. Gv = Visualization. Gs = Processing speed. SMC = Squared multiple correlation.

Table 5

*Six factor principal axis rotated solutions (start=.30)*

| | Bi-Quartimin Rotation | | | | | | Bi-Geomin Rotation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | Gc | Gv | Gf | Gwm | Gv | g | Gv | Gf | Gwm | Gs1 | Gs2 |
| Similarities | **.69** | **.41** | .01 | –.00 | .05 | –.04 | **.79** | –.04 | –.05 | –.12 | –.04 | –.05 |
| Vocabulary | **.70** | **.49** | .04 | .03 | .02 | –.07 | **.83** | –.03 | .05 | –.19 | –.08 | –.10 |
| Information | **.72** | **.38** | –.03 | .02 | –.02 | –.08 | **.80** | –.04 | .01 | –.17 | –.07 | .02 |
| Comprehension | **.61** | **.36** | –.03 | –.06 | .05 | .01 | **.70** | –.08 | –.08 | –.10 | –.00 | –.00 |
| Block Design | **.66** | –.01 | **.32** | .05 | –.05 | .06 | **.61** | **.41** | .05 | –.00 | .12 | .08 |
| Visual Puzzles | **.67** | .03 | **.48** | .04 | –.03 | –.06 | **.63** | **.53** | .01 | –.01 | –.01 | .01 |
| Matrix Reasoning | **.62** | .02 | .14 | .12 | .08 | –.00 | **.60** | .20 | .12 | .09 | .04 | .04 |
| Figure Weights | **.65** | .02 | .09 | **.42** | –.00 | –.07 | **.62** | .21 | **.43** | .02 | –.02 | –.02 |
| Picture Concepts | **.50** | .12 | .10 | –.00 | .04 | –.00 | **.51** | .11 | –.01 | .01 | .02 | .03 |
| Arithmetic | **.69** | .11 | –.01 | .12 | .19 | .04 | **.71** | .03 | .13 | .15 | .06 | .02 |
| Digit Span | **.66** | .02 | –.02 | –.00 | **.48** | .02 | **.67** | –.00 | –.02 | **.47** | .03 | .03 |
| Picture Span | **.53** | .01 | .04 | –.00 | **.33** | .05 | **.54** | .05 | –.00 | **.31** | .06 | –.00 |
| Letter-Number Sequencing | **.63** | .07 | –.04 | –.00 | **.48** | .04 | **.67** | –.07 | –.01 | **.41** | .04 | –.04 |
| Coding | **.38** | .00 | .00 | .01 | .02 | **.88** | **.35** | –.02 | .00 | .05 | **.90** | –.01 |
| Symbol Search | **.53** | –.17 | –.08 | –.16 | –.10 | **.43** | **.40** | .05 | –.02 | .04 | **.49** | **.36** |
| Cancellation | **.30** | –.17 | –.10 | –.13 | –.19 | .22 | .18 | .03 | –.01 | –.05 | .27 | **.35** |

*Note*. g=General Factor. Gwm=Working memory. Gc=Crystallized ability. Gv=Visualization. Gf=Fluid reasoning. Gs=Processing speed.

Table 6

*Population loadings for Monte Carlo Study*

| Variable Cluster | Manifest Variable | Factor | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | General | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | $u^2$ |
| VCI | V1 | .70 | .40 | | | | | 0.35 |
| | V2 | .70 | .40 | | | | | 0.35 |
| | V3 | .70 | .40 | | | | | 0.35 |
| | V4 | .70 | .40 | | | | | 0.35 |
| VSI | V5 | .60 | | .40 | | | | 0.48 |
| | V6 | .60 | | .40 | | | | 0.48 |
| FRI | V7 | .60 | | | .10 | | | 0.63 |
| | V8 | .60 | | | .40 | | | 0.48 |
| | V9 | .60 | | | .00 | | | 0.64 |
| | V10 | .70 | | | .00 | | | 0.51 |
| WMI | V11 | .70 | | | | .40 | | 0.35 |
| | V12 | .60 | | | | .40 | | 0.48 |
| | V13 | .60 | | | | .40 | | 0.48 |
| PSI | V14 | .40 | | | | | .60 | 0.48 |
| | V15 | .40 | | | | | .60 | 0.48 |
| | V16 | .30, .60, .80 | | | | | .40 | .75, .48, .20 |

Table 7

*Simulation Results*

| | | | Collapse Threshold (Sum of Absolute Loadings) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Communality Starting Value | V16 Population Value | *m* | 0.8 | | 1.0 | | 1.2 | | 1.4 | | 1.6 | | 1.8 | | 2.0 | |
| | | | # | *p* | # | *p* | # | *p* | # | *p* | # | *p* | # | *p* | # | *p* |
| .3 | 0.3 | 969 | 211 | 0.22 | 264 | 0.27 | 330 | 0.34 | 470 | 0.49 | 741 | 0.76 | 947 | 0.98 | 969 | 1.00 |
| | 0.6 | 875 | 126 | 0.14 | 153 | 0.17 | 191 | 0.22 | 319 | 0.36 | 638 | 0.73 | 855 | 0.98 | 875 | 1.00 |
| | 0.8 | 874 | 70 | 0.08 | 103 | 0.12 | 145 | 0.17 | 280 | 0.32 | 629 | 0.72 | 860 | 0.98 | 874 | 1.00 |
| .5 | | | | | | | | | | | | | | | | |
| | 0.3 | 999 | 217 | 0.22 | 273 | 0.27 | 348 | 0.35 | 502 | 0.50 | 781 | 0.78 | 976 | 0.98 | 999 | 1.00 |
| | 0.6 | 1000 | 160 | 0.16 | 208 | 0.21 | 273 | 0.27 | 439 | 0.44 | 755 | 0.76 | 982 | 0.98 | 1000 | 1.00 |
| | 0.8 | 1000 | 89 | 0.09 | 122 | 0.12 | 169 | 0.17 | 308 | 0.31 | 710 | 0.71 | 978 | 0.98 | 1000 | 1.00 |

*Note*. p=Proportion of collapse. m=Number of successful replications. #=Number of cases of collapse. Originally there were *m=1000* replications in each condition, but we removed factor solutions that did not converge.

Table 8

*ANOVA Source table for Monte Carlo*

| Factor | SS | *df* | MS | *F* | Partial $\omega^2$ |
|---|---|---|---|---|---|
| Start Value | 0.32 | 1 | 0.32 | 2.15 | < 0.00 |
| Loading | 26.05 | 2 | 13.03 | 87.48 | 0.03 |
| Start Value x Loading | 0.52 | 2 | 0.26 | 1.74 | < 0.00 |
| Error | 850.56 | 5712 | 0.15 | | |

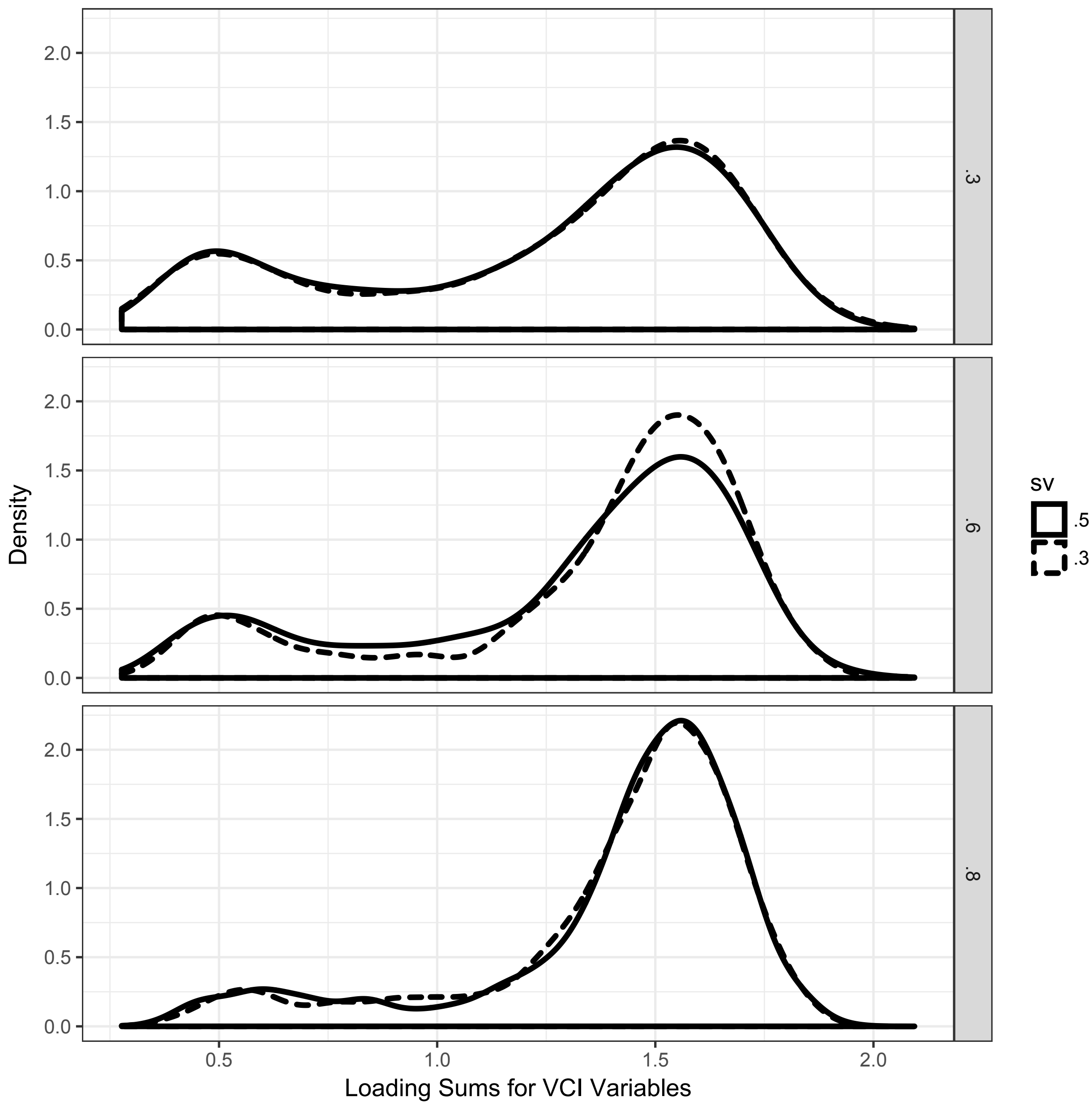


*Figure 1*. Density plots for loading sums for each of the six conditions.