Factor Analysis and Variance Partitioning in Intelligence Test Research: Clarifying Misconceptions

Stefan C. Dombrowski[1] ⓘ**, Ryan J. McGill**[2]**,
Gary L. Canivez**[3] ⓘ**, Marley W. Watkins**[4]**,
and A. Alexander Beaujean**[4]

Abstract

This article addresses conceptual and methodological shortcomings regarding conducting and interpreting intelligence test factor analytic research that appeared in the Decker, S. L., Bridges, R. M., Luedke, J. C., & Eason, M. J. (2020). Dimensional evaluation of cognitive measures: Methodological confounds and theoretical concerns. *Journal of Psychoeducational Assessment*. Advance online publication article.

Keywords

factor analysis, Schmid–Leiman procedure, bifactor model, intelligence testing, exploratory factor analysis, confirmatory factor analysis

Decker, Bridges, Luedke, & Eason's (2020) article "Dimensional Evaluation of Cognitive Measures: Methodological Confounds and Theoretical Concerns" was recently published online in the *Journal of Psychoeducational Assessment*. We read this article with great interest as it promised to present an applied demonstration of how two methodological approaches, the bifactor method (BF; see Brown, 2015), and the Schmid–Leiman procedure (SL; Schmid & Leiman, 1957), are problematic for structural validity research on intelligence tests. In our view, the article suffers from numerous conceptual and methodological shortcomings which we articulate below to better inform ongoing debates about these topics.

Conceptual Misunderstandings

Beginning in the abstract and continuing throughout the article, Decker et al. (2020) mis-characterized the BF and SL methods and misrepresented the results from studies using these

[1]Rider University, NJ, USA
[2]William & Mary, VA, USA
[3]Eastern Illinois University, IL, USA
[4]Baylor University, TX, USA

Corresponding Author:
Stefan C. Dombrowski, Department of Graduate Education, Leadership, and Counseling, Rider University, 2083 Lawrenceville Road, Lawrenceville 08648, NJ, USA.
Email: sdombrowski@rider.edu

procedures. These mischaracterizations are central to the philosophical narrative of the article and deserve correction. As one example, Decker et al. indicated that "subtest loadings on a general factor for a BF model (their Figure 2) are approximately equivalent to loadings for a unidimensional model (their Figure 1). This result occurs because a BF model estimates parameters for a general factor 'as if' it is a unidimensional model" (p. 19). This statement is not correct. Neither the BF method nor the SL procedure estimates parameters in this fashion. The SL procedure[1] and BF modeling, or the results produced by these procedures, should not be construed as representing a unidimensional model of intelligence. Simply put, one cannot use a BF method, or its approximation (i.e., SL), without assuming a multidimensional structure (Bonifay, Lane, & Reise, 2017).

Additionally, Decker and colleagues incorrectly contended that the SL procedure "forces" indicators to load on a general factor. With the SL procedure, correlated group factors are prioritized in the first-order solution (with oblique rotation) and allowed to be freely estimated independent of the influence of *g*. The ability of the SL procedure to recover group factors is well established (Giordano & Waller, 2020). Decker and colleagues neither acknowledged nor discussed this mathematical capability. Brown (2015) and Wolff and Preising (2005) provided an accessible explication for applied users of how parameters are estimated using both procedures.

Further, Decker and colleagues' literature review claimed to make the case that the SL and the BF are equivalent with the SL characterized as an exploratory version of the BF procedure. This characterization is incorrect but understandable, given how loosely terminology has been used in the literature (Beaujean, 2015). While Reise (2012) referred to the SL procedure as an *approximate* exploratory BF model, the SL procedure technically represents a mathematical transformation of the higher-order model using an elegant procedure to apportion the variance to the higher- and lower-order factors (Carroll, 1993; Gorsuch, 1983; Loehlin & Beaujean, 2016; Schmid & Leiman, 1957; Wolff & Preising, 2005). However, the SL procedure is not a *true* BF procedure but rather a reparameterization of the higher-order model. The BF model and SL procedure originate from different assumptions and carry with them their own set of strengths and limitations. Some of the researchers cited in Decker et al.'s Table 1 have clarified this distinction and moved away from labeling the SL as a BF procedure since it assumes a *higher-order* structure.

With this understanding, it is important to note that a true exploratory bifactor analysis (EBFA) procedure—using a BF rotation—was created by Jennrich and Bentler (2011). Whereas the SL procedure extracts *p* rotated oblique factors, applies a second-order analysis, and transforms the variance to be apportioned to the general and group factors, the EBFA procedure simply extracts *p* + 1 rotated factors and apportions variance to those dimensions. No acknowledgment or discussion of the EBFA procedure and how it differs from the SL method was provided in the Decker and colleague's article.

Framing of the Literature

Decker et al. cited several studies (e.g., Chen, West, & Sousa, 2006; Maydeau-Okivares & Coffman, 2006; Schmiedek & Li, 2004) to make the case that the BF model erroneously fits better than rival models in simulations where the model is implausible for the data. The authors also claimed that the SL procedure and BF model suffer from additional methodological issues such as proportionality constraints, failure to account for unmodeled dimensional complexity, overinflation of *g* variance, and underestimation of variance to group factors. In contrast, the article did not provide a discussion of the potential strengths and utility of SL and BF procedures. From their framing of the literature, Decker and colleagues appear to be opposed to the use of either the BF model or SL procedure for investigating the multidimensional structure of intelligence tests and for making interpretive recommendations for those measures. We respect Decker et al.'s right to

maintain this position; however, their article provided a one-sided presentation to frame an argument that the SL/BF procedures are methodologically flawed.

While all modeling procedures have limitations, it is important to frame the limitations accurately and provide a balanced presentation of the literature and conclusions from research studies. For instance, Decker et al. cited Chen et al. (2006) as evidence that a BF model is problematic, but the conclusions from Chen et al. (2006) plainly provided the opposite conclusion as reflected in that article's abstract:

> "The bifactor model allowed for easier interpretation of the relationship between the domain specific factors and external variables, over and above the general factor. Contrary to the literature, sufficient power existed to distinguish between bifactor and corresponding second-order models in one actual and one simulated example, given reasonable sample sizes. Advantages of bifactor models over second-order models are discussed" (p. 181).

Additionally, the proportionality constraint issue is only attributable to the SL procedure (and the higher-order model from which it is derived; Gignac, 2016). EBFA overcomes this issue, but it has been shown to "get stuck" in local minima, which can result in group factor collapse particularly with problematic indicators. These are all issues that several researchers on this commentary have uncovered and disclosed in the methodological literature (see Dombrowski, Beaujean, McGill, Benson, & Schneider, 2019).

Studies suggesting that model fit statistics are biased in favor of the BF model were cited throughout the article and leveraged as a central critique of the BF model (e.g., Mansolf & Reise, 2017; Murray & Johnson, 2013). Although this potential limitation is framed as a fatal flaw, some of the authors cited by Decker et al. to support this contention offer more circumspect conclusions. For example, Maydeu-Olivares and Coffman (2006) discussed limitations of the BF model yet noted that the BF model should be preferred when "when researchers are interested in domain-specific factors over and above the general factor and, particularly, when researchers are interested in their differential predictive validity" (p. 359). As another example, Murray and Johnson (2013) expressed a preference for the BF model when interpreting domain specific factors. This is the explicitly stated goal of most, if not all, of the studies listed in Decker et al.'s Table 1, as well as the interpretive strategies articulated in intelligence test technical and interpretation manuals. Finally, Decker et al. did not explicate the results from simulation studies that have countered their own claims that the BF model is biased (e.g., Chen et al., 2006; Morgan, Hodge, Wells, & Watkins, 2015). In fact, the Morgan et al. citation is referenced in a way (p. 2) that supports the Decker et al. position on these matters; however, our reading of the Morgan et al. simulation study does not support this contention. On the contrary, the Morgan et al. results clearly indicate that model fit indices are not biased in favor of the BF model when a BF model reflects the "true"[2] model underlying the data.

Building a Narrative Around Interpretation of Research

Decker et al. reported a review of the factor analytic methods utilized by a particular group of researchers (see their Table 1). From this review, they made several unsubstantiated claims that are used to support their premises. For example, Decker and colleagues asserted that "no justification was given for the preference of a BF/SL model" (p. 4) for the Differential Abilities Scales, Second Edition (DAS-II; Elliot, 2007) exploratory factor analysis (EFA) conducted by Dombrowski, McGill, Canivez, & Peterson (2019). This characterization is inaccurate. Dombrowski and colleagues provided ample justification for the SL procedures used in their study (pp. 92–93). Similarly, in a later section of the article, Decker et al. claimed that McGill, Dombrowski, and Canivez (2018) committed what they regard as "a clear methodological oversight[3]" because they

did "not take into consideration that 100% of the reviewed studies used BF/SL models" (p. 19). However, the authors failed to disclose that the table in question from that article was located in a section of the article dedicated *specifically* to the discussion of BF modeling relative to the estimation of coefficient omega in intelligence test research.

Decker and colleagues then alleged that the studies they evaluated demonstrated evidence of methodological bias because all employed the SL procedure or BF model estimation. The only thing that Decker et al. (2020) have *actually* demonstrated is that researchers who are well versed in the use of appropriate factor analytic methods have properly utilized recommended procedures when conducting these analyses. For example, Keith and Reynolds (2018) described higher-order and BF models as well as the SL procedure in their tutorial on the use of confirmatory factor analysis (CFA) with intelligence tests. It is unclear what alternative procedures Decker et al. would prefer to have been employed other than to eschew use of SL/BF methods altogether.

We also draw attention to Decker and colleagues omission of research that has supported the retention of a higher-order (HO) model at the expense of the BF model (e.g., Dombrowski, McGill, & Morgan, 2019; McGill, 2020) and other research that has impartially presented both the HO *and* BF models to readers when they could not be distinguished statistically (e.g., Canivez, Watkins, & Dombrowski, 2017; Canivez, Watkins, Good, James, & James, 2017; Dombrowski, Golay, McGill, & Canivez, 2018; Strickland, Watkins, & Caterino, 2015). Further, Decker et al. also overlooked studies that have acknowledged the existence of broad abilities beyond *g* (e.g., Benson, Beaujean, McGill, & Dombrowski, 2018). Results furnished in these studies cast substantial doubt on Decker et al.'s claims of systemic bias in the literature.

Based in part on their fundamental misunderstanding of the SL procedure and BF modeling, construing them as "equivalent to a unidimensional model of intelligence," Decker et al. (p. 19) appear to have misinterpreted the conclusions of the findings reported in their Table 1 as evidence that this body of work has argued for a "unidimensional model of intelligence" (p. 4). This represents a logical fallacy in the form of a false equivalence. Instead, every one of the studies in Decker et al.'s Table 1 (and others published by some of the authors listed in that table) evaluated the tenability of a single-factor unidimensional *g* model, found it empirically lacking, and rejected it as inadequate. A review of Decker et al.'s Table 1 shows that each study supported the presence of 3–7 group factors (across multiple IQ tests) that is made clear in the *actual* studies themselves. The conclusion from these studies that primary emphasis should be placed on interpreting the Full Scale Intelligence Quotient (FSIQ) should not be conflated as suggesting that an instrument is unidimensional or that broad abilities *do not* exist. None of the studies listed in Decker et al.'s Table 1 have ever made that contention. Rather, the following statement by Schneider and Kaufman (2016) summarize well the philosophical position of most of the studies' authors listed in Decker and colleagues' Table 1: "Although no scholars believe that intelligence consists solely of the *g* factor, some believe that our intelligence tests are simply too crude to effectively isolate the smaller factors of ability" (p. 286).

It is important to highlight that the discussion of group factor interpretability involves more than just simple variance partitioning. The alignment of subtests with theoretically posited group factors—a requirement not discussed but tacitly acknowledged by Decker et al. when they disclosed that their EFA results suggested retention of a five-factor model, which diverges from the publisher theory—should also be considered when deciding to interpret indices derived from group factors. Additionally, metrics of interpretive relevance (e.g., omega-hierarchical [ω_H] and omega-hierarchical subscale [ω_{HS}] coefficients, construct reliability or replicability [H], and percentage of uncontaminated correlations) should also be considered. These metrics offer an additional vantage from which to view whether an instrument's group factors are empirically suitable for interpretation (Dombrowski, 2020; Reise, Bonifay, & Haviland, 2013; Rodriguez et al., 2016a, 2016b; Selbom & Tellegen, 2019).

Questionable Methodological Procedures

Although the theoretical and conceptual errors within Decker et al.'s article are sufficient to give one pause, there are numerous methodological shortcomings (not all are presented in this commentary) that deserve highlighting as they were further used by Decker et al. to generate incorrect conclusions.

Varimax Rotation in EFA

Decker et al. did not disclose the rotation method applied in their EFAs; however, the summary data reported in their article were used to determine that an orthogonal rotation method, varimax, was employed to produce their findings.[4] The use of varimax rotation[5] was methodologically inappropriate, given the aims of the Decker et al. study as well as the nature of tests of intelligence (Gorsuch, 1983; Watkins, 2018). Jensen (1998) noted that an orthogonal rotation "expressly precludes the appearance of a *g* factor. With orthogonal rotation, the *g* variance remains in the factor matrix but is dispersed among all of the group (or primary) factors. This method of factor analysis is not appropriate for any domain of variables, such as mental abilities, in which substantial positive correlations among all the variables reveal a large general factor" (p. 73).[6]

Incorrect Interpretation of "g" Loadings

The most substantive flaw in the article is the misinterpretation/misrepresentation of their own results reported in Tables 5 and 6 of their article. Decker et al. examined a series of increasingly complex first-order models ranging from one to eight factors[7] (1F–8F). Consistent with established practice (e.g., Jensen, 1998), the 1F solution was used to obtain initial *g* loadings from the first *unrotated* factor.[8] Then, only the loadings from the first *rotated* factor extracted for Models 2F–7F were reported in their Table 5 and erroneously interpreted as representing a general factor. However, once a first-order factor solution is rotated, the loadings from the first factor no longer represent *g* loadings unless the solution is inherently unidimensional, which the authors acknowledge is not the case.[9] "When rotation occurs, the variance associated with the first factor seems to disappear, but in reality, it has become the dominant source of variance in the now rotated factors" (Carretta & Ree, 2001, p. 332).

It is well-known that *g* loadings are represented by the first *unrotated* factor coefficients and remain relatively stable, regardless of the number of factors that are extracted (Jensen, 1998). Thus, the rotated factor loadings in Table 5 of the Decker et al. article do *not* measure the general factor. The *correct* general factor loadings for models 1F–7F are presented in our Table 1 based on our replication of the results produced by Decker et al. using SPSS. The percentage of variance explained by the first factor in each solution did *not* decrease but remained relatively stable at 32%–33% in the solutions that converged. Inspection of these results illustrates that the conclusions reached from their EFA are incorrect. While Decker et al. correctly suggested that there are multiple ways to obtain *g* loadings in the Limitations section of their article, they failed to disclose that their approach is not recognized within the psychometric literature as one of them (Jensen, 1998).[10]

It is not clear why Decker et al. overlooked the results of hundreds of studies spanning more than a century that consistently found a strong general factor, regardless of factor analytic methodology (e.g., Carroll, 1993; Jensen, 1998; Lubinski, 2004; Warne & Burningham, 2019). As summarized by Sternberg (2003), "the evidence in favor of a general factor of intelligence is, in one sense, overwhelming....One would have to be blind or intransigent not to give this evidence its due" (p. 374).

Table 1. First Unrotated Factor Coefficients (g loadings) using Maximum Likelihood Estimation across Various WJ III Cognitive Factor Extractions.

WJ III Cognitive Subtest	Number of factor extracted						
	1F	2F	3F	4F	5F	6F*	7F
Verbal Comprehension	.844	.870	.877	.876	.891	.365	.849
General Information	.786	.806	.825	.815	.811	.293	.794
Concept Formation	.737	.716	.728	.728	.721	.999	.808
Analysis-Synthesis	.623	.600	.611	.613	.606	.273	.627
Visual Auditory Learning	.622	.609	.607	.605	.604	.414	.619
Sound Blending	.558	.547	.544	.550	.583	.222	.568
Memory for Words	.514	.499	.494	.527	.507	.273	.510
Numbers Reversed	.512	.500	.494	.501	.496	.334	.504
Retrieval Fluency	.485	.496	.489	.491	.486	.133	.479
Spatial Relations	.458	.444	.447	.444	.443	.141	.470
Visual Matching	.448	.474	.458	.462	.453	.173	.455
Auditory Attention	.438	.449	.440	.446	.449	.383	.438
Decision Speed	.373	.405	.400	.408	.404	.152	.399
Picture Recognition	.274	.274	.271	.274	.269	.183	.272
Variance (%)	32.40	32.57	32.70	33.05	33.05	14.08	33.56

*Improper solution because the extracted communality estimate for spatial relations was >1.00 indicating a Heywood case. A Heywood case warning was encountered when estimating Models 4F–7F.

Lack of Disclosure in EFA

In addition to not disclosing which rotation method was used in EFA, factor loadings (coefficients) for the remaining extracted first-order factors in Models 2F–7F were not reported. This makes it difficult to determine whether the models are free from local identification issues which is a critical component of model evaluation in EFA (Gorsuch, 1983; Watkins, 2018). For example, a Heywood case (Spatial Relations) was encountered in Model 6F (see Table 1). Additionally, extracted factors that contain too few measured variables for identification, measured variables that cross-load, or measured variables that migrate to theoretically different factors were apparently not considered. The authors did not report encountering these issues in their analyses. It is certainly possible, in the models that they examined, that the Heywood case evaded extraction, but it does suggest that there is instability in these analyses.[11]

Lack of Disclosure in CFA

All of the models with group factors examined in the CFA analyses reported by Decker and colleagues are just identified which renders the models statistically indistinguishable. However, the degrees of freedom reported for their Model 5 are not consistent with what was expected (i.e., $df = 70$). Instead, the authors reported, "Aside from adding a degree of freedom to the analysis, the effect of the additional specification for the hierarchical model was likely to be *minimal* [emphasis added] since only one additional parameter was required" (p. 18). The specific parameter that was modified and the rationale for its specification were not disclosed, a clear violation of research reporting standards (Appelbaum et al., 2018).

Another major problem is apparent in Decker et al.'s Figure 2. There are no standardized path coefficients reported from Long-Term Storage and Retrieval (Glr) to Visual Auditory Learning or Retrieval Fluency, suggesting that these paths were near zero, and thus, Glr is *not* a viable group

factor. Further, there is no indication whether the standardized path coefficients from Visual-Spatial Processing (Gv) to Spatial Relations (.10) and Picture Recognition (.10) were statistically significant, but even if so, they are trivial and indicate that Gv in this model is likely also untenable. Such omissions suggest disregard of *local fit* problems as an important aspect of judgment of model adequacy because CFA models should never be retained "solely on global fit testing" (Kline, 2016, p. 461). Recall that the EFA analyses reported by Decker et al. supported a five-factor model.

As it appears to have been necessary to modify the Woodcock Johnson III Tests of Cognitive Abilities (WJ III Cognitive; Woodcock, McGrew & Mather, 2001) theoretical model, we used the summary data reported by Decker and colleagues to examine the publisher proposed theoretically derived HO model. Specification of the HO model produced clear evidence of model misspecification in the form of a Heywood case (the second-order loading between Glr and *g* was 1.05 and Glr had a negative residual variance estimate [-.094])[12] indicating an impermissible solution.[13] We can only speculate as to what constraint Decker et al. applied to Model 5 to rectify the offending estimate as this was not disclosed, nor was the additional parameter made evident in their Figure 5. Thus, we examined the most plausible option and constrained the variance in Glr to 1.0 to restrict its ability to produce an out of bounds estimate. With the addition of this constraint (which added a degree of freedom), the resulting standardized loadings were identical to those reported in their Figure 5 and the global fit statistics approximately the same as those reported in their Table 7. Ironically, the HO model appears to have required the use of a constraint to be identified, a major source of contention by Decker et al. with respect to the BF model.

Finally, it remains unclear what the results reported in their Table 8 represent. Based on Decker et al.'s description, these appear to be standardized loadings (coefficients) on the *g* factor based on their Models 1 and 2 that should correspond to the loading coefficients reported in their corresponding Figures 1 and 2. However, the loadings for Verbal Comprehension are absent from their Table 8 without explanation.[14] Of even greater concern, none of the values reported in their Table 8 are consistent with the "*g*" loadings reported for the corresponding Models 1 and 2. As a result, it remains unclear how these results were obtained. Even so, the variance explained by *g* is 48.7% and 39.9%, respectively, for their Models 1 and 2. Thus, the BF produced a general factor that is approximately 20% weaker than the unidimensional model. This loss of power is not trivial, illustrating that the models should not be regarded as equivalent. This is also evident in the fit statistics reported in their Table 7, which indicated that Model 1, in their analyses, is structurally deficient and should not be retained. Inadequacy of their Model 1 (*g* only) was also reported in every CFA study reported in Decker et al.'s Table 1 and studies conspicuously absent from that systematic review (e.g., Canivez, Watkins, & Dombrowski, 2017; Canivez, Watkins, Good, et al., 2017; Canivez, Watkins, & McGill, 2019; Dombrowski, McGill & Morgan, 2019; Fenollar-Cortés & Watkins, 2019; Lecerf & Canivez, 2018). In sum, it is unclear what the Decker et al. CFA results prove relative to determining whether the BF model is biased or the psychometric adequacy of the variance partitioning procedures that emanate from that model. It may be the case that "sometimes a method or concept is baffling not because it is profound but because it is wrong" (Sasso, 2001, p. 188).

Conclusion

The Decker et al. study is predicated on theoretical, philosophical, and statistical misunderstandings and contains serious methodological flaws. In our view, their conclusions are not supported by the analyses reported. In fact, the analyses were not designed to falsify their contention that the SL procedure and BF model are biased methodological choices. Accordingly, readers are strongly urged to take our rebuttal into account when evaluating the commentary and speculation stemming from their analyses and when citing Decker, et al. in future discussions on these matters.

Nevertheless, the theoretical and empirical questions raised by Decker et al. remain important points for future research. Researchers (e.g., Horn, 1968; Johnson & Bouchard, 2005; Kovacs & Conway, 2016) have presented alternative views of the structure of human intelligence that should be investigated. Additionally, "a plethora of fundamental questions about cognitive abilities— their structure, sources, and meanings" (Carroll, 1998, p. 22) remain unanswered for even the most widely accepted models of intelligence.

Declaration of Conflicting Interests

Funding

ORCID iDs

Stefan C. Dombrowski https://orcid.org/0000-0002-8057-3751
Gary L. Canivez https://orcid.org/0000-0002-5347-6534

Notes

1. Throughout Decker et al. it is implied and in some cases stated outright that the SL procedure is a "model." This is not entirely correct. It is a variance partitioning procedure that is most commonly applied to existing higher-order EFA solutions after those solutions have been estimated.
2. It is understood that modeling efforts are approximations and should not be construed as an exact representation of reality.
3. It is not clear how a methodological error is possible in a literature review.
4. In email correspondence, the lead author confirmed the use of varimax rotation and SPSS for their EFA.
5. Varimax is the default rotation in SPSS (IBM Corp., 2020).
6. One exception is EBFA (Jennrich & Bentler, 2011), which uses orthogonal rotation to generate general factor loadings.
7. It is unclear why an eight-factor model was deemed plausible, given identification limitations.
8. Jensen (1987) discussed preference for the SL procedure in obtaining a cognitive ability measures' *g* loadings: "hierarchical analysis can be accomplished by means of the Schmid and Leiman (1957) procedure, which yields the factor loadings of all the tests on each of the orthogonal factors at every level of the hierarchy. Is there a preference among these methods of extracting a *g* factor?"(p. 91). "Yes, although each method has certain advantages and disadvantages" (p. 91). "In short, the hierarchical *g* is more stable than the first principal factor across variations in psychometric sampling" (p. 92).
9. Inspection of the loadings (coefficients) reported in Decker et al.'s Table 5 reveal that the first *rotated* factor is likely a complexly determined verbal or language ability dimension, which is often the case with IQ tests.
10. The authors cite Yung, Thissen, and McLeod (1999) as supporting their approach; however, we are unable to locate any relevant discussion of this issue in that article.
11. A known shortcoming of maximum likelihood (ML) estimation is that it is prone to produce Heywood cases and sometimes recovers unstable solutions in EFA (Gorsuch, 1983).
12. Disclosure of this finding would have left the authors no choice but to retain the BF model as their Model 4 (HO) is impermissible due to a perfect correlation between visual-spatial processing (Gv) and fluid reasoning (Gf), clear evidence of over-factoring. However, consideration of local fit problems for all models containing seven group factors would lead to rejection (see Decker et al.'s Figures 2, 4, and 5) and evidence that the WJ III Cognitive does not appear to adequately measure seven group factors.
13. MPLUS (Muthén & Muthén, 1998–2017) was used for these analyses as opposed to AMOS (Arbuckle, 2019); however, the results were cross-validated and replicated using EQS (Bentler & Wu, 2016) suggesting they are not an artifact of the program being used.

14.   This is possibly an oversight as is the misspelling of several names throughout the article including Bonifay and Dombrowski.

References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, *73*(1), 3-25. doi:10.1037/amp0000191

Arbuckle, J. L. (2019). *Amos (Version 26.0) [Computer Program]*. Chicago: IBM SPSS.

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence*, *3*(4), 121-136. doi:10.3390/jintelligence3040121

Benson, N. F., Beaujean, A. A., McGill, R. J., & Dombrowski, S. C. (2018). Revisiting Carroll's survey of factor-analytic studies: Implications for the clinical assessment of intelligence. *Psychological Assessment*, *30*(8), 1028-1038. doi:10.1037/pas0000556

Bentler, P. M., & Wu, E. J. C. (2016). EQS for windows. Multivariate software.

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*(1), 184-186. doi:10.1177/2167702616657069

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford.

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*(4), 458-472. doi:10.1037/pas0000358

Canivez, G. L., Watkins, M. W., Good, R., James, K., & James, T. (2017). Construct validity of the WISC–IV[UK] with a referred Irish sample: Wechsler and CHC model comparisons with 15 subtests. *British Journal of Educational Psychology*, *87*(3), 383-407. doi:10.1111/bjep.12155

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler intelligence scale for children-fifth UK edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *The British Journal of Educational Psychology*, *89*(2), 195-224. doi:10.1111/bjep.12230

Carretta, T. R., & Ree, M. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment*, *9*(4), 325-335. doi:10.1111/1468-2389.00184

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511571312

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5-23). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189-225. doi:10.1207/s15327906mbr4102_5

Decker, S. L., Bridges, R. M., Luedke, J. C., & Eason, M. J. (2020). Dimensional evaluation of cognitive measures: Methodological confounds and theoretical concerns. *Journal of Psychoeducational Assessment*. Advance online publication. doi:10.1177/0734282920940879

Dombrowski S. C. (2020). A newly proposed framework and a clarion call to improve practice. In S. C. Dombrowski (Ed.), *Psychoeducational assessment and report writing* (2nd ed., pp. 9-59). Cham Switzerland: Springer Nature. doi:10.1007/978-3-030-44641-3_2

Dombrowski, S. C., Beaujean, A. A., McGill, R. J., Benson, N. F., & Schneider, W. J. (2019). Using exploratory bifactor analysis to understand the latent structure of multidimensional psychological measures: An example featuring the WISC-V. *Structural Equation Modeling*, *26*(6), 847-860. doi:10.1080/10705511.2019.1622421

Dombrowski, S. C., Golay, P., McGill, R. J., & Canivez, G. L. (2018). Investigating the theoretical structure of the DAS-II core battery at school age using Bayesian structural equation modeling. *Psychology in the Schools*, *55*(2), 190-207. doi:10.1002/pits.22096

Dombrowski, S. C., McGill, R. J., Canivez, G. L., & Peterson, C. H. (2019). Investigating the theoretical structure of the differential ability scales-second edition through hierarchical exploratory factor analysis. *Journal of Psychoeducational Assessment*, *37*(1), 91-104. doi:10.1177/0734282918760724

Dombrowski, S. C., McGill, R. J. & Morgan, G. W. (2019). Monte Carlo modeling of contemporary intelligence test (IQ) factor structure: Implications for IQ assessment, interpretation and theory. *Assessment*. Advance online publication. doi:10.1177/1073191119869828

Elliott, C. D. (2007). *Differential ability scales* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Fenollar-Cortés, J., & Watkins, M. W. (2019). Construct validity of the Spanish version of the Wechsler intelligence scale for children fifth edition (WISC-V[Spain]). *International Journal of School & Educational Psychology*, 7(3), 150-164. doi:10.1080/21683603.2017.1414006

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

Giordano, C., & Waller, N. G. (2020). Recovering bifactor models: A comparison of seven methods. *Psychological Methods*, 25(2), 143-156. doi:10.1037/met0000227

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). New Jersey: Erlbaum.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75(3), 242-259. doi:10.1037/h0025662

IBM Corp (2020). *IBM SPSS statistics for windows, version 27.0*. Armonk, NY: IBM Corp.

Jennrich, RI, & Bentler, PM (2011). Exploratory bi-factor analysis. *Psychometrika*, 76(4), 537-549. doi:10.1007/BF02287965

Jensen, A. R. (1987). The g beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87-142). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Santa Barbara, CA: Praeger.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393-416. doi:10.1016/j.intell2004.12.002

Keith, T. Z., & Reynolds, M. R. (2018). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 853-900). New York: Guilford.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford.

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27(3), 151-177. doi:10.1080/1047840X.2016.1153946

Lecerf, T, & Canivez, GL (2018). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment*, 30(6), 1009-1808. doi:10.1037/pas0000526

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

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured." *Journal of Personality and Social Psychology*, 86(1), 96-111. doi:10.1037/0022-3514.86.1.96

McGill, R. J. (2020). An instrument in search of a theory: Structural validity of the Kaufman assessment battery for children-second edition normative update at school-age. *Psychology in the Schools*, 57(2), 247-264. doi:10.1002/pits.22304

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, 71, 108-121. doi:10.1016/j.jsp.2018.10.007

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61, 120-129. doi:10.1016/j.intell.2017.01.012

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344-362. doi:10.1037/1082-989X.11.4.344

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3(1), 2-20. doi:10.3390/jintelligence3010002

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.).

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*, *41*(5), 407-422. doi:10.1016/j.intell.2013.06.004

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 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*(2), 129-140. doi:10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150. doi:10.1037/met0000045

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237. doi:10.1080/00223891.2015.1089249

Sasso, G. M. (2001). The retreat from inquiry and knowledge in special education. *Journal of Special Education*, *34*(4), 178-193. doi:10.1177/002246690103400401

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

Schmiedek, F., & Li, S.-C. (2004). Toward an alternative representation for disentangling age- associated differences in general and specific cognitive abilities. *Psychology and Aging, 19*(1), 40-56. doi:10.1037/0882-7974.19.1.40

Schneider, W. J., & Kaufman, A. S. (2016). Commentary on current practices and future directions for the assessment of child and adolescent intelligence in schools around the world. *International Journal of School & Educational Psychology*, *4*(4), 283-288. doi:10.1080/21683603.2016.1206383

Selbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428-1441. doi:10.1037/pas0000623

Sternberg, R. J. (2003). "My house is a very very very fine house" –But it is not the only house. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 373-395). Oxford, UK: Elsevier Science, LTD.

Strickland, T, Watkins, MW, & Caterino, LC (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment*, *27*(2), 689-697. doi:10.1037/pas0000052

Warne, RT, & Burningham, C (2019). Spearman's g found in 31 non-Western nations: Strong evidence that *g* is a universal phenomenon. *Psychological Bulletin*, *145*(3), 237-272. doi:10.1037/bul0000184

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219-246. doi:10.1177/0095798418771807

Wolff, HG, & Preising, K (2005). Exploring item and higher order factor structure with the Schmid–Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods*, *37*, 48-58. doi:10.3758/BF03206397

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113-128. doi:10.1007/bf02294531