serval
serveur académique lausannois

Unil
UNIL | Université de Lausanne
Faculty of Biology and Medicine

CHUV

Investigating the Theoretical Structure of the DAS-II Core Battery at School Age using

Bayesian Structural Equation Modeling

Stefan C. Dombrowski

Rider University

Philippe Golay

Lausanne University Hospital

Ryan J. McGill

College of William & Mary

Gary L. Canivez

Eastern Illinois University

Correspondence: Stefan C. Dombrowski, Ph.D., Professor & Director, School Psychology Program, Department of Graduate Education, Leadership & Counseling. Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648. sdombrowski@rider.edu

Abstract

Bayesian structural equation modeling (BSEM) was used to investigate the latent structure of the DAS–II core battery using the standardization sample normative data for ages 7 to 17. Results revealed plausibility of a three-factor model, consistent with publisher theory, expressed as either a higher-order (HO) or a bifactor (BF) model. The results also revealed an alternative structure with the best model fit, a two-factor bifactor model with Matrices (MA) and Sequential & Quantitative Reasoning (SQ) loading on g only with no respective group factor loading. This was only the second study to use BSEM to investigate the structure of a commercial ability test and the first to use a large normative sample and the specification of both approximate zero cross-loadings and correlated residuals terms.

Keywords: Bayesian structural equation modeling; Differential Abilities Scales, Second Edition; confirmatory factor analysis, bifactor model, Structural Validity, intelligence

Investigating the Theoretical Structure of the DAS-II Core Battery at School Age

using Bayesian Structural Equation Modeling

The Differential Ability Scales-Second Edition (DAS–II; Elliott, 2007a) is an

individually administered battery of cognitive tests for children and adolescents ages 2-17 years.

The DAS–II is a complex instrument that can be divided into three levels: Lower Early Years

(ages 2:6 through 3:5), Upper Early Years (3:6 through 6:11), and School Age (7:0 through

7:11). At school age, the DAS-II contains six core subtests that combine to yield three first-order

composite scores referred to as *cluster* scores (Verbal Ability, Nonverbal Reasoning Ability, and

Spatial Ability) as well as an omnibus full scale General Conceptual Ability (GCA) score

thought to reflect psychometric *g* (Spearman, 1927). There are also 10 diagnostic subtests which

contribute the measurement of two additional cluster scores (Working Memory and Processing

Speed) which can be used by examiners to supplement the core battery. However, none of these

supplemental measures contribute to the measurement of the GCA or the three primary clusters,

nor can they be exchanged for any of the core battery measures. It should also be noted that the

Early Years battery features different core and diagnostic subtest configurations and not all

school-age clusters are available[1]. According to the *Introductory and Technical Handbook*

(Elliott, 2007b; hereafter referred to as the 'Technical Handbook'), this is the result of not being

able to measure certain constructs well (e.g., Processing Speed, Working Memory) at younger

ages.

---

[1] Several Early Years measures have restricted age bands that preclude them from being
administered at school-age. However, both batteries are co-normed at ages 5:0 through 8:11,
permitting "out of level" testing for examinees in that age bracket.

Factor Structure of the DAS–II

To validate the DAS–II at school-age, the test publisher relied exclusively on confirmatory factor analysis (CFA) using maximum likelihood (ML) estimation to appraise the six subtest core battery and structure for normative participants ages 7 to 17[2]. Four oblique (correlated) factors models ranging from 1-3 factors (one model was a variant of the two-factors model with cross-loading permitted) were specified and evaluated for adequacy. Fit statistics reported in the Technical Handbook indicate that a three-factors model consistent with publisher theory fit the standardization sample data well though the factor loadings for this model were not presented.

Similar analyses were also conducted to evaluate different configurations of the core and diagnostic measures at school-age. For these analyses, the normative sample was split into two groups (6:0-12:11 and 6:0-17:11) with a 14 subtest configuration used at ages 6-12 and a 12 subtest configuration used at ages 6-17. Whereas a seven-factors model was retained for ages 6-12, it was suggested that a six-factors model best fit the normative data for ages 6-17. Although the Technical Handbook indicates that both structural models are *likely* consistent with the Cattell-Horn-Carroll theory of cognitive abilities (CHC; Schneider & McGrew, 2012), several first-order factors not available in the actual DAS–II were specified (e.g., Auditory Processing, Visual–Verbal Memory, and Verbal Short-Term Memory). In addition, the Auditory Processing and Visual-Verbal Memory factors in the final validation models for ages 6-17 were each produced from a single indicator reflecting variables that are underidentified. Although the

---

[2] It appears additional exploratory analyses were conducted by the project team (see p. 157, Elliot, 2007b); however, description and results of these procedures are not presented in the Technical Handbook.

inclusion of singlet variables is possible in CFA, they should not be interpreted as latent factors became they do not contain any shared common variance (Brown, 2015).

Since its publication, independent factor analytic investigations of the DAS–II structure have been scarce. In one of the two studies that could be located, Keith et al. (2010) used CFA to investigate the age invariance of the DAS–II full test battery (20 subtests). The measurement model was derived from the normative data from participants ages 5-8. As previously mentioned, this is the only age-bracket at which the Early Years and School-Age batteries are co-normed. Rival models were evaluated, containing different mixtures of correlated errors ($n = 10$), cross-loadings ($n = 10$), and additional post-hoc modifications with a separate validation sample ($n = 5$). In spite of these modifications, the fit statistics for many of the models were indistinguishable. Nevertheless, a measurement model was selected and tested. The subtests not administered in other age groups were treated as latent variables using the "reference variable" approach suggested by McArdle (1994). As described by Keith et al. (2010), "This method allows the researcher to keep the full model as the comparison model," (p. 688). In this procedure, subtests that are not administered at an age level, are treated as latent variables, while constraining their parameters and loadings to be equal to the values obtained for the age group at which they are administered (ages 5-8). Ultimately, a six-factor, CHC-based, higher-order model (Crystallized Ability, Fluid Reasoning, Visual Processing, Long-Term Retrieval, Short-Term Memory, and Processing Speed) was found to be invariant across the instrument. Although it should be noted that the final validation model for ages 4-17 required the specification of additional parameters, including correlated residual terms for both subtests (Copying and Recall of Designs) and group-factors (Visual Processing [Spatial Ability] and Fluid Reasoning [Nonverbal Reasoning]), and a theoretically inconsistent cross-loading (Verbal Comprehension

was found to loading on Crystallized Ability and Fluid Reasoning). Given these departures from desired simple structure and the incorporation of "out of range" measures across the age span, the practical implications of these findings are unclear.

Given the fact that the models produced from the core battery CFA analyses were not presented in the Technical Handbook, users of the DAS–II electing to administer and interpret the core battery may be tempted to extrapolate from the CFA analyses from the full DAS–II battery. However, results furnished by a recent exploratory factor analytic (EFA) of the DAS–II core battery structure suggest this practice may be problematic. Canivez and McGill (2016) used principal axis factoring with promax rotation followed by the Schmid-Leiman Orthogonalization (Schmid & Leiman, 1957) to disclose an approximate exploratory bifactor structure of the DAS–II core battery. Whereas empirical extraction criteria suggested that DAS–II was a one-factor test, a forced three-factors extraction produced subtest alignment consistent with that proposed within the Technical Handbook. Nevertheless, the reliable variance accounted for by the three group factors (Verbal, Nonverbal, and Spatial) was consistently small suggesting the DAS–II may be overfactored (Frazier & Youngstrom, 2007). Specifically, once variance was apportioned to higher- and lower-order constructs, as recommended by Carroll (1993, 1995), most of the reliable variance in the DAS–II subtests was sourced to g, rendering the Nonverbal factor ill-defined (i.e., contained less than two salient subtest loadings).

Historically, two basic factor analytic techniques have been used to evaluate the internal structure of intelligence tests: EFA and CFA. Although, EFA and CFA have been used to shed insight on the DAS–II structure, the results of these investigations have not clarified what the DAS–II core battery measures. Whereas, EFA results suggest that the core battery may be overfactored and mostly reflects general intelligence, CFA investigations using various

combinations of the core and diagnostic subtests, have provided evidence to support the three

group-factors posited for the core battery model. The invariance results produced by Keith et al.

(2010) suggest that the relationships between DAS–II variables may be more complex than the

simple structure portrayed in the CFAs reported in the Technical Handbook (i.e., no cross-

loading or correlated residuals). Gorsuch (1983) and others (Carroll, 1985) suggest that when

different methods of factor analysis converge upon the same solution then greater confidence

may be engendered in the instrument's factor structure. These discrepant results suggest that

additional analyses of the DAS–II factor structure may be worthwhile.

It is worth pointing out that there are important differences between EFA and CFA.

Whereas EFA models are weakly specified, CFA models are more flexible, requiring the

research to specify all relevant aspects of the model *a priori*. Within the factor analytic literature,

it is frequently suggested that EFA is preferred when the relations between variables is less

understood and CFA is a better method for formal model testing[3]. Nevertheless, both methods

have limitations. Although EFA procedures are easier to implement, those methods can

underestimate the number of factors and may produce solutions that oversimplify data (Mulaik,

2010). On the other hand, CFA may be able to detect previously omitted variance however, as

models become more complex, there is a threat of capitalizing on chance and retaining a model

that may not generalize to other samples (MacCallum, Roznowski, & Necowitz, 1992). As a

result, "researchers are often left with the dilemma of whether to keep meaningful alternatives

untested or to risk overfitting their model to the data" (Golay, Reverte, Favez, & Lecerf, 2012, p.

---

[3] In practice, the line between EFA and CFA is less clear. For instance, one can use EFA in a
confirmatory context and CFA in an exploratory fashion. Thus, it is better to think of EFA and
CFA more generally as techniques for conducting factor analysis. Whether an approach is
exploratory or *confirmatory* depends on its application.

498). Bayesian Structural Equation Modeling (BSEM) can be a useful solution for researchers faced with this dilemma.

Bayesian Structural Equation Modeling

BSEM is based upon Bayes' Theorem, a mathematical proof created by Thomas Bayes, an 18th century cleric, that has been recently re-discovered by applied measurement researchers following the arrival of microcomputers with sufficient processing capabilities, the creation of statistical software capable of performing complex Bayesian modeling, and greater confidence in Bayesian estimation which challenges many assumptions of traditional frequentist statistics (Brown, 2015; Kaplan & Depaoli, 2013). One of the most famous–but until recently, secret–uses of Bayesian methodology was to decipher the German enigma code during the Second World War (Stone, 2013). However, the application of Bayesian methodology to understand applied cognitive measurement issues is in its infancy. To date, within the fields of psychometrics and intelligence research there has been only one application of Bayesian estimation. Golay, Reverte, Rossier, Favez, and Lecerf (2013) used the procedure to acquire further insight into the French WISC–IV theoretical structure. Application of BSEM revealed that a five-factor CHC-based direct hierarchical (bifactor) model best fit the data produced from a clinical sample ($N =$ 249) of French-speaking Swiss children. However, in their application of BSEM, Golay et al. (2013) did not include estimation of small variance correlated residuals, potentially important features of the BSEM technology (Muthén & Asparouhov, 2012). The present study seeks to extend use of BSEM to a different measure of cognitive ability to help understand its factor structure and apply other aspects of the BSEM model not included in previous analyses (e.g., simultaneous estimation of approximate zero cross-loadings and approximate zero correlated error terms).

BSEM holds promise for the understanding of the latent structure of assessment instruments used within many fields including psychology, health, business, and education (Muthén & Asparouhov, 2012). It portends to better reflect substantive theory and overcome some of the limitations of traditional (i.e., termed frequentist) exploratory and confirmatory factor analytic procedures (Brown, 2015). In many ways BSEM attempts to strike a balance between exploratory and confirmatory factor analytic methods (Golay et al., 2013) but is still more confirmatory in nature.

Purpose of the Current Study

The present investigation sought to apply BSEM to the DAS–II standardization sample data to understand better the core battery factor structure for ages 7-17. The application of BSEM to the DAS–II factor structure presents an opportunity to compare the procedure across different types of structural models (oblique, higher-order, bifactor). The present study will also serve as a comparative test of BSEM relative to results produced from frequentist exploratory (i.e., Canivez & McGill, 2016) and confirmatory factor analytic methods (i.e., Technical Handbook, Elliot, 2007b; Keith et al., 2010) for the measurement instrument. This is also the first BSEM study of a cognitive ability test taking advantage of a large sample size and the use of correlated residuals; thus, it is believed that the results produced from the current study will be instructive for advancing the field's understanding of not only the factor structure of the DAS-II core battery but also the potential utility of BSEM in psychometric investigations of intelligence test structures.

Method

Participants

Participants were drawn from the DAS–II standardization sample and included a total of 2,188 individuals ranging in age from 7 to 17:11 years. The standardization sample was obtained using stratified proportional sampling across demographic variables of age, sex, race/ethnicity, parent educational level, and geographic region. Details of demographic characteristics and close approximation to population characteristics are provided in the Technical Handbook (Elliott, 2007b).

Instrument

The DAS–II is an individually administered test of intelligence that includes 6 core subtests across the 7 to 17:11 age range and a mixture of 10 supplemental diagnostic subtests. At this age range the DAS–II core subtests combine to form a General Conceptual Ability (GCA) score as well as three primary cognitive clusters at the first-order level, each composed of two subtests. The clusters include Verbal, Nonverbal, and Spatial. Supplemental diagnostic subtests are also available, which can be combined to form additional first-order clusters (e.g., working memory, processing speed) but these measures are not utilized to calculate the higher-order GCA or the three primary cognitive clusters. As previously noted, the Early Years battery contains different combinations of core subtests and cluster scores. For the sake of parsimony, the present study is focused specifically on the core battery at school-age as it is at that age that the DAS–II structure is most consistent.

Procedure and Analyses

The DAS–II standardization sample participant raw data for the 6 core, age 7 to 17:11 subtests were obtained from the test publisher. Bayesian structural equation modeling was used

to investigate two- and three-factor (oblique, higher-order, bifactor) models from the DAS–II, which included a test of the three-factor higher-order structure furnished in the Technical Handbook. Additionally, a derivation of a two-factor bifactor structure was investigated where the two nonverbal subtests (Matrices [MA] and Sequential and Quantitative Reasoning [SQ]) loaded only on *g*. This model was tested post hoc and after observing that the two- and three-bifactor structures had subtests (MA and SQ) with approximate zero loadings on the nonverbal group factor.

Mplus 8.0 (Muthén and Muthén, 1998-2017) was used for Bayesian estimation. The code for undertaking BSEM with explanation is offered in the appendix. Three different BSEM procedures were invoked to test each of the models: (1) an analysis *without* cross-loadings or correlated residuals; (2) an analysis where all cross-loading are simultaneously estimated; and (3) and an analysis where all cross-loadings *and* correlated residuals are simultaneously estimated.

A prior mean of 0 and variance of .01 was established for cross-loadings.   For the cross-loadings this resulted in a range of -.20 to .20 for the resulting cross-loading estimates. If the model failed to converge then a prior cross loading variance of .001 was specified. This reduced the range of the cross loadings estimates from -.06 to .06. An Inverse-Wishart prior variance of .01 was selected for specification of residual prior variance (Asparouhov & Muthén, 2010).

Three Markov Chain Monte Carlo (MCMC) chains were utilized and iterations were established at 150,000 with the first 75,000 being discarded as the burn-in phase. A model was determined to have attained convergence under two conditions: (1) a potential scale reduction (PSR) value stabilizing on a value less than 1.10; and (2) a satisfactory Kolmogorov-Smirnov Distribution (i.e., no discrepant posterior distributions in the different MCMC chains; Muthén & Muthén, 1998-2017). In cases where the model failed to converge using 150,000 iterations then

the number of iterations was increased to 250,000. Generally, it is appropriate to increase

iterations (I) by a factor of two (i.e., I, I2, then I4; Muthén & Muthén, 2017) but this is dependent

upon computing power. If the model converged then the next step was to investigate the

Posterior Predictive *p*-value (PPp). As previously noted, a perfect fit of the model to the data is a

PPp of .50 with values < .10 or > .90 considered poor model fit meriting model rejection.

Following acceptable model fit with these data via the PPp, the Deviance Information Criterion

(DIC) was referenced as the main index to compare competing models. The DIC was used in a

similar fashion to other frequentist fit indices with lower values generally considered better

values. Finally, models were examined in relation to theoretical plausibility as guided by the

prevailing literature base.

Omega–hierarchical (ω_H) and omega–hierarchical subscale (ω_{HS}) coefficients (Reise,

2012; Rodriguez et al., 2015) were estimated as model–based reliability estimates of the latent

factors (Gignac & Watkins, 2013) for both the bifactor and higher-order models. The ω_H

coefficient is the model–based reliability estimate for the hierarchical general intelligence factor

independent of the variance of group factors. The ω_{HS} coefficient is the model–based reliability

estimate of a group factor with all other group *and* general factors removed (Reise, 2012;

Rodriguez et al., 2015). Although omega coefficients have been referred to as model-based

reliability estimates they may also be conceived of as validity estimates as they present data

regarding the plausibility of interpreting general and group factors (Gustafsson & Aberg-

Bengtsson, 2010). Omega coefficients should at a minimum exceed .50, but .75 would be

preferred (Reise, 2012; Reise, Bonifay, & Haviland, 2013). Additionally, Hancock and Mueller

(2001) suggested use of an index of construct reliability or replicability (called H) that reflects

the proportion of variability in the construct that is explained by its own indicators and furnishes

an estimate of the reliability of the underlying factor. High H-values (> .80) suggest a well-defined latent variable which portends to be stable across studies. Rodriquez et al. (2016a) indicated that it is difficult to specify group factors within a single instrument and it should only be done when H-values are higher than .70. Further, when H-values are large it might be useful to utilize a weighted composite score instead of unit-weighted composite score. The percentage of uncontaminated correlations (PUC) was also referenced. PUC determines the potential bias associated with forcing multidimensional data into a unidimensional model. When explained common variance (ECV) and PUC are both greater than .70 then the relative bias will be slight and the common variance might best be considered unidimensional (Rodriquez et al., 2016a). Omega–hierarchical and omega–hierarchical subscale coefficients, PUC, and H were estimated using Watkin's (2013) Omega program. To estimate these values in the higher-order models, the group factors needed to be residualized of general factor variance. This was accomplished through the following formula from Reynolds and Keith (2013):

$$\sqrt{R^2 - (g\ loading)^2}$$

Results

Table 2 presents the results of BSEM of the DAS–II investigating the two- and three-factor oblique, higher-order, and bifactor models under three conditions: 1) without small variance priors as identified by the 'a' model versions; 2) with small variance priors for cross loadings only, as identified by the 'b' model versions; and 3) with small variance priors for cross loadings and correlated residuals, as identified by the 'c' model versions. A single factor (*g*) model was also investigated. When BSEM does not utilize small variance, informative priors for cross loadings or correlated residuals (i.e., all 'a' models from Table 1) then the model is said to be akin to a frequentist maximum likelihood CFA.

All of the models examined, except for models 1 (single factor), 2a, and 2b (two-factor oblique), and 3a and 3b (two-factor higher-order), fit these data well according to an examination of the posterior predictive *p*-value (PPp) (PPp > .10). When investigating the PPp, it is further noted that several of the models displayed near perfect fit with these data (0.500; see models 2c, 3c, 5c, 6c, and 7c). This was most commonly found when both cross-loadings and correlated residuals were specified [two exceptions were the bifactor models (i.e., 5b and 8b) in which only cross-loadings were specified]. These latter two models failed to converge when correlated residuals were estimated.

While the PPp value should be used to determine how well the data fit the model, DIC along with theoretical considerations should be used to compare models and determine which model is preferred (Asparouhov et al., 2015; Brown, 2015). Improvements in model fit both within (i.e. models 'a' to 'c') and between (i.e., 1 through 8) models was determined by examining the DIC (with models that had a PPp > .10). All models with PPp > .10 demonstrated a slightly lower DIC when cross-loadings were incorporated, except for models 3b and 5b. In those two cases the 'a' version (that did not incorporate cross-loadings or correlated residuals) was preferred to the models that incorporated small variance cross-loadings.

When correlated residuals, along with cross-loadings, were incorporated, five of the models (2c, two-factor oblique; 3c, two-factor HO; 5c, two-factor BF with MA & SQ on *g* only; 6c, three-factor oblique; 7c three-factor HO) then demonstrated perfect fit with these data (PPp = .499 or .500). However, the two- and three-factor bifactor (BF) models (4c and 8c) failed to converge when specifying correlated residuals. Additionally, the three-factor HO (model 7c) demonstrated a slightly higher DIC when all residuals were correlated. The three remaining models [2 oblique (2c); 3 HO (3c); 2 BF plus MA and SQ on *g* only (5c); and 3 factor oblique

(6c)] demonstrated a lower DIC, indicative of improved model fit, when both correlated

residuals and cross-loadings were specified. Additionally, examining the publisher's proposed

three-factor higher-order model versus a three-factor bifactor model revealed nearly identical

DIC when no cross-loadings (model 'a' versions) or when cross-loadings (model 'b' versions)

were specified. This is consistent with ML CFA research that suggests that just identified models

have nearly identical fit whether a higher-order or bifactor model is specified (Brown, 2015; also

see McGill & Dombrowski, 2017 for an applied example).

Pattern of Subtest Loadings

An investigation of the pattern of subtest loadings was informative. Within the two- and

three-factor BF models (4b & 8b) the group nonverbal factor loadings were near zero for all BF

models (Tables A1 and 2) suggesting that once the two subtests were residualized of their

general factor variance the two subtests had negligible group factor variance. This finding

similarly occurred when the 'a' model versions without informative cross-loadings or correlated

residuals were included with the bifactor models, although the 'a' model version had lower g

loadings for MA and SQ compared to the 'b' model version. Thus, the decision was made to test

a derivation of the two-factor BF model where MA and SQ loaded only on g [2 BF plus MA and

SQ on g only (model 5 (a to c); Tables 3 and A2))]. With the exception of the 3 oblique factors

model (6c; Table A3), the 2 BF plus SQ and MA on g only model (model 5c; Table 3) had the

lowest DIC when both cross-loadings and correlated errors were specified. Although the oblique

model (Table A3) had a lower DIC it was deemed to be theoretically inferior as tests of cognitive

ability are generally presumed to have a hierarchical latent ability factor, presumably general

intelligence (Carroll, 1993; Gorsuch, 1983). An examination of the three-factor higher-order

model (Table 4), which included cross-loadings, suggested that all subtests were aligned with

theoretically proposed factors. This did not occur with its three-factor bifactor counterpart (Table 2) wherein MA and SQ had approximate zero loadings on the nonverbal group factor once general ability was residualized. The two-factor higher-order model (Table A4) with both cross-loadings and correlated residuals produced loadings consistent with theoretically proposed factors.

Examination of variance apportionment along with omega statistics, H and PUC— presented at the bottom of Tables 2 through 4 and A1 to A4— all converge to suggest that the general factor absorbed a considerable proportion of both total and common variance across all higher-order and bifactor models. Across all BF and HO models investigated, the ECV of the general factor ranged from .663 to .823. Individual group ECV ranged from .000 to .218. The general factor similarly accounted for a considerably higher proportion of total variance ranging from .442 to .508 than did the group factors. Group factor total variance ranged from .000 to .145.

Omega hierarchical and omega hierarchical subscale coefficients suggested that interpretation of the DAS–II should reside primarily at the higher-order or general (GCA) level, whether a BF or HO was referenced, with omega hierarchical ranging from .711 to .838. Omega hierarchical subscale ranged from .000 to .274, again supporting primary emphasis on general factor interpretation. When looking at PUC in combination with the ECV of the general factor, it is evident that the DAS–II is dominated by a general factor. Similarly, the high H values ($>.80$) also suggests a dominant general factor that portends to be stable across studies. Thus, consistent with other frequentist EFA and CFA studies (e.g., Bodin, Pardini, Burns & Stevens, 2009; Canivez, 2014; Canivez & McGill, 2016; Canivez, Watkins & Dombrowski, 2016, 2017; DiStefano & Dombrowski, 2006; Dombrowski, 2013, 2014a, 2014b; Dombrowski, Brogan & Watkins, 2009; Dombrowski, Canivez, Watkins & Beaujean, 2015; Dombrowski, McGill & Canivez, 2016; Watkins &

Beaujean, 2014) and consistent with Frazier and Youngstrom (2007), the DAS–II appears to be an instrument dominated by a general factor.

Discussion

The present study permitted a comparison of BSEM across different types of structural models (oblique, higher-order, bifactor). It also furnished information about possible alternative structures (i.e., 2 BF plus MA and SQ on *g* only; Model 5, Tables 3 and A2) for the DAS–II which were not described in the Technical Handbook nor observed within Canivez and McGill's (2016) EFA-SL study.

One of the more potentially useful capabilities of BSEM (Muthén & Muthén, 2015; Asparouhov et al., 2015) is that it permits the simultaneous estimation of cross-loadings and correlated error terms using small variance priors. This would not be possible on a six subtest instrument, such as the DAS–II, using classical ML CFA estimation. The attempt to estimate this many parameters in frequentist CFA would simply lead to an unidentified model. With ML CFA most cross-loadings have to be fixed at zero to achieve model identification and most error terms remain uncorrelated for that same reason. But, this may not reflect the researcher's hypothesis or even the structural reality of a cognitive ability instrument that often has overlapping, highly correlated constructs (Carroll, 1993; Gorsuch, 1983). Unnecessarily strict models and inappropriate zero cross-loadings could contribute to poor model fit, distorted factors, inflated loadings, and biased correlations (Asparouhov & Muthén, 2009; Brown, 2015; Marsh et al., 2009). McCrae et al. (2008) recognized this concern within the personality structural validity research literature and posited that ML CFA was overly restrictive (i.e., independent cluster assumption requiring an indicator to load only one factor and disregard cross-loadings) leading to correlations among the factors that tend to be overestimated.

BSEM offers the potential of an elegant solution to this problem that accounts for both cross-loadings and correlated residuals through simultaneous estimation. It may also be considered a hybrid estimation procedure in between EFA and CFA. It is noted, however, that the use of correlated residual terms represents a novel approach to structural modeling that is not yet fully embraced by the statistical community (Rindskopf, 2012; Stromeyer et al., 2015). The incorporation of all correlated residuals terms within BSEM deserves further study and debate but has potential to help clarify more complex elements of an instrument's internal structure (Asparouhov et al. 2015).

Within this study, the inclusion of correlated residuals improved model fit in some cases, (e.g., models 2c, 3c & 5c) as determined by PPp values suggesting that the model nearly perfectly fit these data, and produced lower DIC scores. However, there were also cases where incorporation of correlated errors produced models that failed to converge (the two- and three-factor BF models, 4c & 8c; Tables 2 and A1), failed to yield a lower DIC (model 7c, three factor HO), or did not enhance structural clarity based on the patterns of loadings, as the loadings were essentially the same whether or not correlated residuals were incorporated. In those cases, the more parsimonious model (cross-loadings only or no incorporation of cross-loadings) may be favored. For instance, model (5c; Table 3) produced a DIC that was lower than all models except model 6c (three-factor oblique model; Table A3) but it is unknown whether any structural clarity or theoretical gains could be made by choosing the correlated errors version (Model 5c; Table 3) over its cross-loading only counterpart (Model 5b).

Also, theoretical considerations must be accounted for. Although model 6c (three factor oblique) produced the lowest DIC, and one could indeed offer a statistical defense for an oblique model, but, at present, oblique models do not reflect the consensual theoretical conceptualization

for measures of cognitive abilities (Carroll, 1993; Gignac, 2016; Gignac & Watkins, 2013; Dombrowski, 2015). Therefore, oblique models were incorporated for pedagogical reasons since BSEM has only been used once before in the professional literature to understand cognitive ability instruments (i.e., Golay et al., 2013).

Moving next to an understanding of the three factor structure posited in the Technical Handbook, the three-factor HO (cross-loadings only; Table A4) and the three-factor BF (cross-loadings only; Table 2) models demonstrated a nearly identical DIC. This is not surprising. As occurs with ML CFA estimation, in BSEM estimation when a just identified model is investigated, fit indices are virtually identical (Brown, 2015). When correlated residuals where incorporated both the two- and three-factor bifactor models failed to converge. When correlated residuals were specified for the higher-order models, the two-factor HO model saw improved model fit while the three-factor HO model evidenced a reduced model fit as noted by an increase in DIC. In the case of model 5c (Two BF plus MA & SQ on g only; Table 3) the incorporation of correlated residuals improved model fit with these data and lowered DIC. However, the pattern of subtest loadings was essentially the same as when cross-loadings only approach was specified (see Tables A2 and 3). Across all models investigated, parameter estimates for correlated residuals were not statistically significant. This information is important in its own right and along with inclusion of cross-loadings (all were non-significant) suggested that the subtests may be statistically homogenous.

The results of this study indicate that the DAS–II six core subtest battery may be conceptualized not only as a three-factor higher-order model, as indicated in the Technical Handbook (although the standardized loadings associated with this model were not reported), but also as a three-factor bifactor model. With both models, the incorporation of cross-loadings

improved model fit with these data. However, the incorporation of correlated residuals caused

the three-factor bifactor model (and two-factor BF model) to fail to converge.

In addition to being conceptualized as a three-factor higher-order model (Table 4) or

three-factor bifactor model (Table 2), the DAS–II may be conceptualized as a two-factor bifactor

model with two of its subtests (Matrices and Sequential and Quantitative Reasoning; Tables A2

and 4) loading on g only. If one ascribes to a bifactor conceptualization of intelligence then this

hybrid bifactor model appears plausible: the three-factor bifactor model produced loadings (MA

and SQ close to zero on their theoretically posited group factor. Whether a two- (Tables A2 & 3)

or three-factor BF (Table 2) model is investigated MA and SQ load on their respective group

factors at close to zero, but have high general ability loadings. If the choice is for a BF model

then the hybrid approach (i.e., Model 5a-c; Tables A2 and 3) is viable as having MA and SQ

load on g only improves structural clarity. When correlated residuals were included, this model

produced the lowest DIC and affirmed a lack of relationship among the error terms for the DAS–

II, a finding that is important in its own right.

Regardless of whether a BF or HO model is adopted omega statistics suggest that the

DAS–II is an instrument dominated by general ability. This was similarly supported by H and

PUC. The finding is also consistent with prior findings from Canivez and McGill (2016) who

cautioned about moving beyond interpretation of the general factor even though they found

evidence for three group factors consistent with that posited in the Technical Handbook when

force extracting that model in their EFA analyses.

Similar to Golay et al. (2013), the present results suggest the use of BSEM appears to be

a viable option for the investigation of the structure of cognitive ability instruments. With the

DAS–II it produced results that appear theoretically plausible and in fact offered an alternative

structure (2 BF plus MA and SQ on *g* only; Model 5 a-c) that was not described within the

Technical Handbook nor described within the extant DAS-II factor analytic research (Canivez &

McGill, 2016; Keith, Low, Reynolds, Patel & Ridley, 2010). Within the present study, the

inclusion of small variance cross-loadings appeared to aide in theoretical interpretation of the

DAS–II structure. The inclusion of correlated residuals did not necessarily improve the structural

clarity of the model beyond the use of cross-loadings, lowered DIC in some cases, and failed to

permit the model to converge in others. But, it did offer additional insight into the DAS–II

structure by demonstrating that subtests were not confounded by error terms that were correlated

and that cross-loadings do not detract from the core battery's structural clarity as none were

statistically significant.

Whereas cross-loadings are familiar to the structural validity researcher who encounters

them when using exploratory factor analysis, the use of correlated residuals may well be a

procedure that requires further explication, scrutiny, and debate. Questions remain about

whether it improves structural clarity, whether it introduces statistical noise, or whether it may be

exploited for the sole purpose of improving model fit. Because of this it is suggested that

guidelines be established. However, the specification of correlated residuals may be of benefit.

Unlike with ML CFA which permits only the specification of just a few correlated residuals

often guided by theory, with BSEM the model identification issues are less of a concern and

portend to uncover relationships that were not specified. Keep in mind, however, that BSEM is

not a panacea for model identification issues, and is not the only option to the structural validity

mountain top. This study demonstrated that bifactor models still experience identification

problems when correlated residuals were specified quite possibility due to the inclusion of

additional parameters that had to be estimated. This study's findings regarding the DAS–II

support either a three-factor higher-order or three-factor bifactor structure. This study also lends

support for an alternative two factor BF structure where MA and SQ load only on the general

factor.

Limitations include the need for further research on the use of BSEM. There has only

been one prior study using BSEM for cognitive ability and just a handful investigating

psychology, health, and management (De Bondt, Van Petegem, 2015; Fong & Ho, 2013, 2014;

Stromeyer et al., 2015; Zyphur & Oswald, 2015). Although proponents of BSEM may claim that

BSEM is devoid of statistical fishing expeditions, this is not entirely true. One still needs to

specify in advance the selection of a prior and avoid the temptation to search for improved model

fit just for its own sake. The results of this study showed that it was indeed possible to

simultaneously estimate all cross-loadings to evaluate the nature of the constructs measured by

each subtest scores. Thus BSEM avoided resorting to many comparisons and potentially biasing

the estimation of the model parameters. The most controversial aspect of BSEM is the use of

correlated residuals. There are researchers who raise concerns about their use (Stromeyer et al.,

2015). On the other hand, Muthén and Asparouhov (2012) and Asparouhov et al. (2015) contend

that if used appropriately then the specification of correlated residuals may enhance the

understanding of an instrument's structure. Additional discussion and debate into this topic is

necessary.

In totality, the use of BSEM on the six core subtest DAS–II structure offered additional

insight into the structure of the DAS–II not previously uncovered by the use of ML CFA within

the Technical Handbook nor within the exploratory and Schmid-Leiman procedures used by

Canivez and McGill (2016). As a result, a follow-up ML CFA study comparing the various two-

and three-factor structures, including the 2 BF plus MA and SQ on g only, appears worthwhile.

References

Asparouhov, T., & Muthen, B. (2009). Exploratory structure equation modeling. *Structural Equation Modeling, 16,* 397-438. doi: 10.1080/10705510903008204

Asparouhov, T. & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Retrieved from https://www.statmodel.com/download/Bayes3.pdf

Asparouhov, T., Muthén, B. & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al*., Journal of Management, 41,* 1561-1577.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC–IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417–424.

Brown, T. A. (2015). *Confirmatory factor analysis for the applied researcher*. New York: Guilford Press.

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales— Second Edition: Exploratory and hierarchical factor analysis with the core subtests. *Psychological Assessment, 28,* 1475-1488. doi: 10.1037/pas0000279

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

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*. Advance online publication. doi: 10.1037/pas0000358

Carroll, J. B. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence, Vol. 1: Research methodology* (pp. 25-58). Norwood, NJ: Ablex.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge, England: Cambridge University Press.

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

De Bondt, N., & Van Petegem, P. (2015). Psychometric evaluation of the Overexcitability Questionnaire-Two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology, 6,* 1-17, doi: 10.3389/fpsyg.2015.01963

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford–Binet–Fifth Edition. *Journal of Psychoeducational Assessment, 24,* 123–136.

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. (2015). *Psychoeducational assessment and report writing*. New York:

Springer Science.

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

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2016). Exploratory and hierarchical factor

analysis of the WJ IV Cognitive at school age. *Psychological Assessment*. Advance

online publication. http://dx.doi.org/10.1037/pas0000350

Elliott, C. D. (2007a). Differential Ability Scales(2nd ed.). San Antonio, TX: Harcourt

Assessment.

Elliott, C. D. (2007b). Differential Ability Scales–2nd ed.: Introductory and technical handbook.

San Antonio, TX: Harcourt Assessment.

Fong, T. C., & Ho, R. T. (2013). Factor analyses of the Hospital Anxiety and Depression Scale:

A Bayesian structural equation modeling approach. *Quality of Life Research,* 22: 2857-

2863.

Fong, T. C., & Ho, R. T. (2014). Testing gender invariance of the Hospital Anxiety and

Depression Scale using the classical approach and Bayesian approach. *Qual. Life Res.* 23,

1421–1426. doi: 10.1007/s11136-013-0594-3

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors

measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence,*

35, 169−182. doi:10.1016/j.intell.2006.07.002

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

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48,* 639-662.

Golay, P., Reverte, I., Rossier, J., Favez, N., and Lecerf, T. (2013). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling. *Psychological Assessment,* 25, 496–508. doi: 10.1037/a0030676

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

Gustafsson, J. –E., & Åberg-Bengtsson,L.(2010). Unidimensionality and interpretability of psychological instruments. In S. Embretson (Ed.), Measuring psychological constructs: Advances in model-based approaches (pp. 97–121). Washington, DC: American Psychological Association.

Kaplan, D. & Depaoli, S. (2013). Bayesian statistical methods. In T. Little (Ed.), The Oxford handbook of quantitative methods (Vol. 1, pp. 407-437). New York: Oxford University Press.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), Structural equation modeling: Present and Future (pp. 195-216). Lincolnwood, IL: Scientific Software International.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales-II: Consistency across 4 to 17. *Psychology in the Schools, 47,* 676-697. doi: 10.1002/pits.20498

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in

    covariance structures analysis: The problem of capitalizing on chance. *Psychological*

    Bulletin, 111, 490-504. doi: 10.1037//0033-2909.111.3.490

Marsh, H. W., Muthén, B., Asparouhov, A., Ldtke, O., Robitzsch, A., Morin, A. J. S. &

    Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and

    EFA: Application to students' evaluations of university teaching. *Structural Equation*

    Modeling, 16, 439-476. doi: 10.1080/10705510903008220

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate*

    Behavioral Research, 29, 409-454. doi: 10.1207/s15327906mbr2904_5

McCrae, R. R., Yamagata, S., Jang, K. L., Riemann, R., Ando, J., Ono, Y.,...Spinath, F. M.

    (2008). Substance and artifact in the higher-order factors of the Big Five. Journal of

    Personality and Social Psychology, 95, 442–455. doi:10.1037/0022-3514.95.2.442

McGill, R. J., & Dombrowski, S. C. (2017). What does the WRAML2 core battery measure?

    Utilizing exploratory and confirmatory techniques to disclose higher-order structure.

    Assessment. Advance online publication. doi: 10.1177/1073191116677799

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Muthén, L.K. and Muthén, B.O. (1998-2015). Mplus User's Guide. Seventh Edition. Los

    Angeles, CA: Muthén & Muthén

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible

    representation of substantive theory. *Psychological Method, 17,* 313–335. doi:

    10.1037/a0026802

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

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment

research. In D. H. Soklofske, V. L. Schwean & C. R. Reynolds (Eds.),

Oxford handbook of child psychological assessment (pp. 48-83). New York: Oxford

University Press.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in

the evaluation of psychological measures. *Journal of Personality Assessment, 98,* 223-

237.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions.

Psychometrika, 22, 53– 61. doi: 10.1007/BF02289209

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D.

P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories,*

tests, and issues (3rd ed., pp. 99-144). New York: Guilford Press.

Spearman, C. (1927). *The abilities of man*. New York, NY: Cambridge University Press.

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The

prowess and pitfalls of Bayesian structural equation modeling: Important considerations

for management research. *Journal of Management,* 41: 491–520.

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

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary

Scale of Intelligence–Fourth edition. *School Psychology Quarterly, 29,* 52–63.

doi:10.1037/spq0000038

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management,* 41: 390-420. doi: 10.1177/0149206313501200

Table 1

Comparison of Model Fit for the DAS-II Core Battery Ages 7 to 17 using Bayesian Structural Equation Modeling

Models	Number of free parameters	Posterior Predictive P-Value (PPP)	Difference between observed & replicated χ^2 95% CrI		DIC	pD
			Lower 2.5%	Upper 2.5%		
1. Single Factor (g)	18	0.000	327.114	364.872	31902.352	18.116
2a. Two Factor Oblique	19	0.000	32.459	70.882	31608.850	18.994
2b. Two Factor Oblique (with Cross Loadings .01)	25	0.001	50.473	103.480	31635.589	20.912
2c. Two Factor Oblique (Cross Loadings & Correlated Residuals .01)	40	0.500	-20.558	20.566	31553.384	15.302
3a. Two Factor Higher Order	20	0.000	16.227	55.776	31594.712	20.354
3b. Two Factor Higher Order (Cross Loadings .001)	26	0.000	22.503	63.462	31601.397	19.686
3c. Two Factor Higher Order (Xload/Corr Resd .001; I=250K; Table A4)	41	0.500	-20.702	20.445	31554.280	16.227
4a. Two Factor Bifactor	23	0.475	-19.705	19.802	31557.689	19.196
4b. Two Factor Bifactor (Xload .001)	29	0.539	-21.158	18.862	31554.227	17.146
4c. Two Factor Bifactor (Xload/Resd Corr .001; I=250k)*		Model	Did not	Converge*		
5a. Two Factor Bifactor MA & SQ on g only	20	0.385	-16.952	21.401	31560.490	19.897
5b. Two Factor Bifactor MA & SQ on g only (Cross Loadings .01; Table A2)	28	0.507	-20.894	19.514	31561.036	23.211
5c. Two Factor Bifactor MA & SQ on g only (Xload & Corr Res .01; Table 3)	43	0.499	-20.532	20.655	31546.422	8.318
6a. Three Factor Oblique	21	0.358	-16.651	23.073	31562.399	20.854
6b. Three Factor Oblique (with Cross Loadings .01)	33	0.458	-19.659	20.276	31560.556	21.537
6c. Three Factor Oblique (Xload & Corr Residual .01; Table A3)**	48	0.499	-20.620	20.289	31543.577	5.483
7a. Three Factor Higher Order	21	0.353	-16.289	23.101	31562.034	20.266
7b. Three Factor Higher Order (Cross Loadings .001; Table 4)	33	0.464	-19.530	19.080	31559.694	20.926
7c. Three Factor Higher Order (Xload & Corr Residual .001)	48	0.501	-20.672	20.668	31561.587	23.546
8a. Three Factor Bifactor	21	0.354	-16.812	22.217	31561.803	20.504
8b. Three Factor Bifactor (with Cross Loadings .001; Table 2)	33	0.508	-20.115	18.786	31559.564	21.828
8c. Three Factor Bifactor (Xload & Corr Resid .001; I=250k)*		Model	Did not	Converge*		

Note. PPP = posterior predictive p-value; CrI = credibility interval; DIC = deviance information criterion; pD = Estimated number of parameters, MA = Matrices, SQ = Sequential & Quantitative Reasoning, *g* = general intelligence. *Kolmogorov-Smirnov Distribution test indicates model noncovergence. **CI for individual subtests contains values >1.00 for all subtest on respective factors. I=Iterations

Table 2

Three Factor Bifactor BSEM with Cross-Loadings and Small Variance (.001) Priors

	General		Verbal		Nonverbal		Spatial			
Loading estimates (median)	*g*									
Subtest	*b* [95% CI]	S^2	*b* [95% CI]	S^2	*b* [95% CI]	S^2	*b* [95% CI]	S^2	h^2	u^2
Word Definitions	**.658** **[.621**	**.433** **.694]**	**.467** **[.419**	**.218** **.511]**	.000 [-.061	.000 .061]	-.004 [-.056	.000 .047]	.653	.347
Verbal Similarities	**.665** **[.628**	**.442** **.701]**	**.467** **[.419**	**.218** **.511]**	.000 [-.062	.000 .060]	-.002 [-.054	.000 .049]	.662	.338
Matrices	**.766** **[.731**	**.587** **.796]**	-.004 [-.055	.000 .045]	**-.020** **[-.215**	**.000** **.196]**	.014 [-.045	.000 .071]	.601	.399
Sequential & Quantitative	**.810** **[.777**	**.656** **.840]**	.017 [-.036	.000 .067]	**-.020** **[-.215**	**.000** **.197]**	-.007 [-.064	.000 .050]	.671	.329
Pattern Construction	**.714** **[.681**	**.510** **.748]**	-.027 [-.075	.001 .020]	.000 [-.063	.000 .063]	**.302** **[.222**	**.091** **.363]**	.604	.396
Recall of Designs	**.639** **[.602**	**.408** **.676]**	.016 [-.032	.000 .063]	.000 [-.060	.000 .062]	**.302** **[.223**	**.091** **.363]**	.501	.499
									.615	.385
ECV*		.822		.118		.000		.049	.991*	
Total Variance		.506		.073		.000		.030		
ω_H / ω_{HS}		.835		.263		.000		.118		
H		.869		.358		.001		.167		
PUC		.800								

Note. *b* = standardized loading of subtest on factor, S^2 = variance explained in the subtest, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω_H = Omega-hierarchical (general factor), ω_{HS} = Omega-hierarchical subscale (group factors). BSEM=Bayesian Structural Equation Modeling, CI=Confidence Interval, *g* = general intelligence. *Does not total to 100% due to use of median parameter estimates. Loadings in bold were freely estimated. Other loadings were estimated with small (0.001) variance priors.

Table 3

Two Factor Bifactor BSEM MA and SQ on g only with Cross-Loadings and Correlated Residuals (.01).

| | General | | | | | | | |
| | *G* | | Verbal | | Nonverbal | | | |
Subtest	*b* [95% CI]	S^2	*b* [95% CI]	S^2	*b* [95% CI]	S^2	h^2	u^2
Word Definitions	**.607** **[.441**	**.368** **.733]**	**.612** **[.464**	**.375** **.703]**	.037 [-.144	.001 .209]	.754	.246
Verbal Similarities	**.621** **[.454**	**.386** **.747]**	**.612** **[.463**	**.375** **.703]**	.028 [-.156	.001 .203]	.771	.229
Matrices	**.839** **[.652**	**.704** **.930]**	-.029 [-.212	.001 .157]	-.018 [-.200	.000 .161]	.722	.278
Sequential & Quantitative	**.866** **[.697**	**.750** **.940]**	-.012 [-.200	.000 .168]	-.021 [-.209	.003 .161]	.767	.233
Pattern Construction	**.658** **[.501**	**.433** **.782]**	.037 [-.140	.001 .212]	**.550** **[-.046**	**.330** **.646]**	.748	.252
Recall of Designs	**.590** **[.395**	**.348** **.740]**	.042 [-.144	.002 .217]	**.550** **[-.046**	**.330** **.647]**	.664	.336
ECV*		675		.170		.137	.983	
Total Variance		.498		.126		.101	.725	.275
ω_H / ω_{HS}		.800		.428		.358		
H		.887		.545		.464		
PUC		.800						

Note. *b* = standardized loading of subtest on factor, S^2 = variance explained in the subtest, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω_H = Omega-hierarchical (general factor), ω_{HS} = Omega-hierarchical subscale (group factors). BSEM=Bayesian Structural Equation Modeling, CI=Confidence Interval, MA=Matrices, SQ= Sequential & Quantitative Reasoning, *g* = general intelligence. *Does not total to 100% due to use of median parameter estimates. Loadings in bold were freely estimated. Other loadings were estimated with small (0.01) variance priors.

Table 4
Three Factor Higher Order BSEM with Cross Loadings and Small Variance (.001) Priors

Loading estimates (median)	General* g		Verbal		Residualized Verbal		Nonverbal		Residualized Nonverbal		Spatial		Residualized Spatial			
Subtest	b [95% CI]	S^2	b [95% CI]	S^2	b	S^2	b [95% CI]	S^2	b	S^2	b [95% CI]	S^2	b	S^2	h^2	u^2
Word Definitions	**.654**	**.428**	**.804** [.740	**.646** .878]	*.470*	*.221*	.002 [-.046	.000 .048]			-.001 [-.048	.000 .045]			.649	.351
Verbal Similarities	**.663**	**.439**	**.814** [.742	**.663** .885]	*.474*	*.225*	.002 [-.047	.000 .050]			.000 [-.048	.000 .047]			.664	.336
Matrices	**.769**	**.591**	-.006 [-.052	.000 .038]			**.774** [.705	**.599** .841]	*.101*	*.010*	.007 [-.041	.000 .057]			.601	.399
Sequential & Quantitative	**.801**	**.642**	.016 [-.030	.000 .062]			**.807** [.739	**.651** .875]	*.170*	*.029*	-.001 [-.050	.000 .048]			.671	.329
Pattern Construction	**.728**	**.530**	-.021 [-.066	.000 .022]			.000 [-.048	.000 .047]			**.800** [.735	**.640** .869]	*.293*	*.086*	.616	.384
Recall of Designs	**.625**	**.391**	.015 [-.029	.000 .060]			.001 [-.047	.000 .049]			**.687** [.619	**.472** .755]	*.312*	*.097*	.488	.512
ECV		.819				.121				.011				.050		
Total Variance		.504				.074				.007				.031	.615	.385
ω_H\ ω_{HS}**		.833				.269				.005				.118		
H		.867				.364				.011				.168		
PUC		.800														

Second Order Loadings (median)

Verbal	**.814** [.776 .847]
Nonverbal	**.993** [.972 1.00]
Spatial	**.910** [.875 .941]

Note. b = standardized loading of subtest on factor, S^2 = variance explained in the subtest, h^2 = communality, u^2 = uniqueness, ECV = explained common variance, ω_H = Omega-hierarchical (general factor), ω_{HS} = Omega-hierarchical subscale (group factors), g = general intelligence. Omega estimates based on residualized group factor loadings. Loadings in bold were freely estimated. Other loadings were estimated with small (0.001) variance priors. Residualized using the following formula: $\sqrt{R^2 - (g\ loading)^2}$ *Calculated using the path tracing rules. **Used residualized estimates to calculate omega.