Embracing cross-loading to improve latent variable models fit: A comparison of exploratory and Bayesian structural equation modeling

Kevin Gittner[1], Niloofar Ramezani[2], Kate Mobley[1]
[1]Kennesaw State University, 275 Kennesaw State University Road NW, Kennesaw, GA 30144
[2]George Mason University, 4511 Patriot Cir, Fairfax, VA 22030

Abstract
Traditional methods of latent variable modeling fix all non primary factor cross-loading values to zero, but the methods have evolved to allow non-zero cross-loading values. These new methods allow for improved modeling of latent constructs by enabling researchers to achieve better model fit while preserving the realistic relationship between items and factors. Exploratory structural equation modeling (ESEM) and Bayesian structural equation modeling (BSEM) both allow relaxing the restrictive assumption that force all minor cross-loading values to be constrained to zero. While some studies have compared BSEM and ESEM models to maximum likelihood confirmatory factor analysis models (Gucciardi & Zyphur, 2016; Guo et al., 2019; Wei et al., 2022), these comparisons have not been extended to include estimation across the different available software that have incorporated both methodologies. Considering the flexibility in specification of ESEM and BSEM in the presence of cross-loading, with the limited computational algorithms developed for estimating parameters of these models, a comparison of these models across different software packages is vital to understand the full potential that ESEM and BSEM methods provide to improve model fit. Using real data on mental ability test scores (Holzinger & Swineford, 1939), this study compares ESEM and BSEM estimation in Mplus and R. The results provide an appropriate application of these methods and an evaluation of the consistency of parameter estimates across software packages.

Key Words: Bayesian structural equation modeling, exploratory structural equation modeling, latent variable analysis, factor cross loading, multi-group measurement models

1. Background Literature

Structural equation modeling (SEM) involves the estimation of latent variables in a model. While estimating latent variables is imperfect by nature, methodological advances in structural equation modeling have helped to improve model fit in models that would otherwise have been found unacceptable with traditional techniques. Confirmatory factor analysis (CFA) was first improved by the development of exploratory structural equation modeling (ESEM) which allowed for cross-loading values to be included in the models, rather than being forced to zero, as is done in traditional CFA (Asparouhov & Muthén, 2009). ESEM allows statistically significant primary cross-loadings to be estimated, while non-primary cross-loadings can still be estimated with a restriction that they are close to zero rather than being fixed to zero. More recently, Bayesian structural equation modeling (BSEM) has sought to further improve model fit by allowing small, non-zero values for minor cross-loadings, rather than fixing them to zero (Muthén & Asparouhov, 2012). In other words, researchers can investigate non-primary factor-loadings without leading to model identification issues accompanied by estimating all minor cross-loadings in traditional estimation techniques (Muthén & Asparouhov, 2012; Steiger, 2002). ESEM and BSEM both offer significant advantages over traditional CFA by resulting in more realistic models that may have better model fit (Koizumi & In'nami, 2020; Muthén & Asparouhov, 2012). While there are many software options for conducting CFA, few have been updated to include both ESEM and BSEM in their capabilities, and there has been little investigation into the accuracy and dependability of the software options that are capable of ESEM and BSEM. The purpose of this study is to compare the estimation of ESEM and BSEM across the software packages that have adopted both methodologies and learn more about the accuracy and precision of estimations provided by these software packages.

The emerging field of structural equation modeling is valuable because it allows researchers to distinguish between observed and latent variables in their modes while also directly estimating latent variables (Guo et al., 2019). Structural Equation Models are commonly used in the social sciences to measure abstract concepts and variables that cannot be directly measured, and therefore typically require some prior domain knowledge to create successful models of latent concepts (Guo et al., 2019). In traditional structural equation models, non-significant factor loadings are fixed to zero and only hypothesized primary factor loadings have non-zero parameters that are freely estimated (Muthén & Asparouhov, 2012).

ESEM, a technique developed over the past two decades, improved model fit and flexibility by allowing cross-loading factors into the models. More recently, BSEM started allowing all factors to have non-zero loading values, with non-significant factor loadings being near zero. BSEM models also specify informative prior distributions (Asparouhov & Muthén, 2021; Muthén & Asparouhov, 2012). Since, it is unlikely that two variables in a model have no association of any kind, especially when measuring abstract concepts in the social sciences, BSEM may create models which better fit the data (Koizumi & In'nami, 2020).

There have been studies, using both simulated and real-world data, that have demonstrated the ability of BSEM and ESEM methods to improve SEM/CFA (Craig, 2017; Gucciardi & Zyphur, 2016; Guo et al., 2019; Liang et al., 2020; Xiao et al., 2019). Most commonly, BSEM is compared to ICM-CFA and ESEM. Previous research comparing Bayesian and traditional SEM indicated that a Bayesian model with correctly specified informative priors outperforms frequentist models as well as naïve and non-informative Bayesian models (Smid et al., 2019). Although it is not surprising that the use of more accurate prior information outperforms less informative priors, the accessibility of known prior information in latent variable modeling is quite common. “Prior dependence” is a known issue in parameter estimate bias that practitioners need to consider when defining Bayesian priors (Asparouhov & Muthén, 2021; Muthén & Asparouhov, 2012). Bayesian models have performed better in estimating more complex models because they could incorporate model uncertainty more successfully than traditional SEM methods that involve no prior information (Smid et al., 2019). BSEM generally has better model fit compared to traditional CFA models, partly due to the Bayesian models being able to determine which covariates should vary and which covariates should remain fixed (Guo et al., 2019). Correctly specifying prior information is crucial to maximize the potential of BSEM in providing more realistic estimates of model parameters (Liang, 2020; Smid et al., 2019). It is known that correctly specifying target values is just as crucial to maximizing the potential of ESEM, but BSEM still provides additional model specifications over ESEM (Guo et al., 2019).

There is a growing number of studies that support the notion that ESEM and Bayesian methods applied to SEM improve model fit by allowing for more flexible models. However, this is contingent on the prior information being correctly specified for factor loadings. The primary reason for improvement in model fit is allowing for major and minor, non-zero cross-loadings, rather than only allowing for major non-zero cross-loadings and forcing non primary cross-loadings to be equal to zero, as is done in traditional SEM (Asparouhov & Muthén, 2009; Marsh et al., 2009; Muthén & Asparouhov, 2012). However, with increased model complexity and model fit may come increased estimation error. There is some concern that the more complex BSEM, especially those with poor prior specification or non-informative prior distributions, learn sample-specific aspects of the data, and as a result become noisier and less generalizable. Some studies have found that BSEM has better model fit, but also has higher estimation bias than ESEM (Guo et al., 2019).

Applied studies using real-world data have similarly found that BSEM and ESEM have improved model fit over traditional CFA models. Yet comparisons of analyses in Mplus provided varying degrees of model improvement for BSEM over ESEM (Koizumi & In'nami, 2020; Reis, 2017). Improved model fit of BSEM is often attributed to the strength of the known prior information available to researchers. Through all of this research, general software selection by researchers has been relegated to Mplus. There is a need to compare analysis results from Mplus to those from other software packages for traditional, ESEM, and BSEM latent construct estimation/validation methods.

The only statistical analysis software that has incorporated both ESEM and BSEM are R and Mplus. BSEM methodology was introduced to Mplus around 2012 and R, through the *blavaan* package, around 2015 (Merkle & Rosseel, 2015; Muthén & Asparouhov, 2012). ESEM with minor cross-loadings methodology was introduced to Mplus around 2010 and to R, through various packages, around 2013 (Asparouhov & Muthén, 2009; De Beer & Van Zyl, 2019; Guàrdia-Olmos et al., 2013; Muthén & Asparouhov, 2012; Revelle, 2015).

Since the adoption of both BSEM and ESEM methods in software, numerous applied and simulation studies have compared ESEM and BSEM performance, yet they have only utilized Mplus for the analysis (Gucciardi & Zyphur, 2016; Guo et al., 2019; Liang et al., 2020; Morin, 2020; Wei et al., 2022; Xiao et al., 2019). Previous studies have compared parameter estimates for various types of analysis methods across software, but not within SEM framework, resulting in mixed conclusions regarding the consistency of parameter estimate. While some studies found consistent parameter estimates across software packages for non-SEM analyses, others found unexplainable differences in parameter estimates among different software for non-SEM analysis (Chang et al., 2020; Harper et al., 2011). While consistent parameter estimates across different software is expected, the only way to validate this is to estimate and evaluate models with multiple software. As the literature shows, the previous studies that compared BSEM and ESEM model parameter estimates used only Mplus for model validation and parameter estimation. None of the studies compared the results of model estimation in Mplus to parameter estimates obtained in R. An investigation of the available literature performed in August 2022 on *blavaan* R package, the only package available in R to estimate BSEM, and exploratory structural equation modeling returned only four relevant publications that discussed both analysis methods, yet none of the publications compared ESEM and BSEM estimates across software. To the best of our knowledge, no comparisons have been done across the software and methods.

There is some documentation of model convergence issues and lengthy computation time when building Bayesian models using the *blavaan* package in R (Merkle & Jorgensen, 2022). The *blavaan* package developers were made aware of a potential issue regarding lengthy estimation time and convergence issues in early 2022; however, no response to improve estimation time has been provided yet. While any individual can develop an R package, inherent issues are commonly recorded and must be updated by package authors/developers. Authors often become aware of issues through user reported issues and may attempt to fix them.

Few studies have compared the parameter estimates of traditional SEM measurement across different software. No study to our knowledge has compared parameter estimates of BSEM and ESEM across different software. With the recent adoption of ESEM and BSEM techniques into statistical analysis software, research is needed to test the consistency of parameter estimates across different software. The primary purpose of this study was to explore parameter estimates and global fit indices across the different software that include both ESEM and BSEM techniques, thus allowing for estimation of all minimal cross-loadings. To date, only Mplus and R include these features. We expect to find similar parameter estimates between Mplus and R, as well as similar improvements in model fit in ESEM and BSEM models with minimal cross-loadings, regardless of software choice. A secondary hypothesis of interest in this study was to explore whether estimation speed was improved with R, as previous Mplus simulations studies have been restricted by convergence times (Gittner, 2021). We hope this study provides continued support for alternatives to traditional measurement models and the practice of fixing non primary cross-loadings to zero. The results of this study will provide practitioners with knowledge regarding consistency of parameter estimates and global fit indices when using ESEM and BSEM methodology, as well as information on R and Mplus software specifications for these modeling techniques.

2. Methods

2.1 Data

The data used for this methodological comparison was obtained from the 1939 Holzinger and Swineford study measuring mental abilities in 301 middle school aged children in 7th and 8th grades at two schools (Holzinger & Swineford, 1939). Their dataset has been used throughout the development of modern measurement techniques. Within their study, 26 different tests were conducted to measure various mental abilities of these children. Each variable was measured on a continuous scale based on the outcome of the various tests conducted by Holzinger and Swineford, with the original hypothesized model measuring a general factor and five other primary factors (Holzinger & Swineford, 1939). Over time, the general factor indicators and two additional indicators have been dropped due to poor fit, leaving 19 of the original indicators to be used in different measurement models (Gustafsson, 2001; Harman, 1976). The same 19, out of the original 26, items, were used for analysis in our study, replicating the original development of BSEM techniques (Muthén & Asparouhov, 2012). Due to the known small sample size convergence issues and possible parameter estimate bias in SEM, all sample participants (n=301) were analyzed in this study. School variable was not necessary in providing substantive context to provide evidence of improved model fit with ESEM and BSEM for this methodological study, so unlike the original analysis by Muthén and Asparouhov, the school that a participant attended was not of interest in this study. The small sample size, while estimating independent models through simulation studies, have been shown to increase bias in parameter estimates (Gittner, 2021; Ma, 2020; Muthén & Asparouhov, 2012).

2.2 Software

Analyses were completed in Mplus version 8.6 (Muthén & Muthén, 2017) and R version 4.1.0 (R Core Team, 2022). Within R, the model estimations were completed using the *lavaan* package version 0.6.12 and the *blavaan* package version 0.4.3, along with each package's various dependencies (Merkle et al., 2021; Rosseel, 2012). As there are various Markov chain Monte Carlo (MCMC) algorithms in Bayesian estimation, all BSEM estimations in *blavaan* were completed using the package default MCMC algorithm, *Stan*, through *rstan* version 2.21.5 (Stan Development Team, 2022). Recent advances in ESEM literature have provided tools to assist practitioners in coding ESEM options in both R and Mplus. For ease of specifying models with target values for ESEM in R, the *esemComp* package version 0.2 was used (Silverstrin & De Beer, 2022). For ease of specifying models with target values for ESEM in Mplus, researchers developed the ESEM Mplus code generator (De Beer & Van Zyl, 2019). These are some tools researchers could benefit from to reduce the coding burden required to specify ESEM models.

2.3 Analysis

A new analysis was completed of the Holzinger-Swineford mental abilities dataset, where a simple structure with non primary factor loadings constrained to zero does not fit well. (Muthén & Asparouhov, 2012). The original measurement model was modified and proposed to compose four primary factor domains (Gustafsson, 2002; Harman, 1976). Factors/latent constructs measuring spatial ability, verbal ability, speed, and memory were composed of the 19 items in Table 1. The model outlined in Table 1 provides a traditional measurement model that will be used as the baseline estimation method for this study. All non primary factor loadings "NP" were constrained by being fixed to 0, while primary factor loadings "X" were freely estimated. This model has been found to provide unacceptable model fit and require post hoc modifications (Muthén & Asparouhov, 2012). For models that provide flexible approaches to handling minor cross-loadings, the "NP" factor loadings were restricted to small factor loadings with an expectation of being close to zero. Details of this restriction are explained below.

For clarity purposes, packages in R and estimators in Mplus are outlined. All four models in both software were estimated with a standardized latent variable by fixing the latent variable variance to 1.0 (Brown, 2015; Kline, 2015). Because all indicators were measured on different scales, all observations were standardized before model estimation (Brown, 2015; Kline, 2015). All models were estimated using a local hard drive on a computer with an Intel i7-9850H processor and 16 GB of RAM.

Table 1: Primary Factor Loading Pattern w/ Minor Cross-Loadings Fixed to Zero

	Spatial	Verbal	Speed	Memory
visual	X	NP	NP	NP
cubes	X	NP	NP	NP
paper	X	NP	NP	NP
flags	X	NP	NP	NP
general	NP	X	NP	NP
paragrap	NP	X	NP	NP
sentence	NP	X	NP	NP
wordc	NP	X	NP	NP
wordm	NP	X	NP	NP
addition	NP	NP	X	NP
code	NP	NP	X	NP
counting	NP	NP	X	NP
straight	NP	NP	X	NP
wordr	NP	NP	NP	X
numberr	NP	NP	NP	X
figurer	NP	NP	NP	X
object	NP	NP	NP	X
numberf	NP	NP	NP	X
figurew	NP	NP	NP	X

This study considered the previous hypothesized model to be a baseline for comparison purposes, to provide readers with an understanding of potential improvement in model fit of using ESEM and BSEM methodology. Four different measurement models were estimated using both Mplus and R.

Model 1, maximum likelihood (ML), estimated the traditional measurement model with fixed non-primary loadings "NP" constrained to zero with a maximum likelihood estimator. Model 1 utilized the previous simple structure model which fixed all non primary cross-loadings to zero as outlined in Table 1. No further estimation of non-primary cross loadings with ML techniques was completed due to model identification issues. Model 1 in R was estimated using the *lavaan* package *cfa()* function with the maximum likelihood estimator, and all other package arguments set to the default settings. Model 1 in Mplus used the ML estimator with all other software arguments set to the default settings.

Model 2, Bayesian structural equation modeling with non primary cross-loadings fixed to zero (BSEM-NOCL), estimated the traditional measurement model with fixed non primary loadings of zero with a Bayesian estimator using the software's default and non-informative prior specifications. Model 2 utilized the previous simple structure model outline in Table 1 with no modifications. Model 2 in R used the *blavaan* package using the *bcfa()* function with the *Stan* MCMC sampling method, while Mplus utilized the Bayes estimator with Mplus' default MCMC sampling method. All other package arguments were set to the default settings, including the *blavaan*, *rstan* and Mplus default's non-informative priors. In R, an additional null model was estimated with the same specification to allow for the estimation of additional Bayesian-adjusted specific global fit indices using the *blavFitIndices()* function.

Model 3, Bayesian structural equation modeling with non primary cross-loadings being estimated (BSEM-CL), adapted the traditional measurement model outlined in Table 1 by eliminating the restriction of fixing the minimal cross-loadings to zero. Instead, informative priors for non-primary factor loadings were defined with a normal distribution with mean of 0 and variance of 0.01. This model is a more realistic approach compared to the estimated models 1 and 2 because it provides an expectation of a minimal relationship, even if close to zero, for non-primary factors with a 95% credibility interval of (-.2, .2). All other prior information were specified by the software default settings for BSEM-CL model 3. Model 3 in R was estimated using the same specifications outlined for model 2, with additional informative priors specified for all non-primary loadings defined to be normally distributed with mean 0 and standard deviation 0.1. All other package options were set to the default settings including additional non-informative priors for other various model parameters. The same null model used for model 2 was used in model 3 to estimate the additional Bayesian-adjusted global fit indices. Model 3 was estimated in Mplus using the same specifications outlined for model 2, with additional informative priors specified for all non-primary loadings defined to be normally distributed with mean 0 and variance deviation 0.01. All other Mplus options including additional non-informative priors for other various model parameters were kept at the default settings. It is worth noting that *Stan* requires prior specification to be defined as a mean-standard deviation relationship while Mplus requires prior specification to be defined as a mean-variance relationship.

Model 4, Exploratory structural equation model with non-primary cross-loadings being estimated (ESEM), is comprised of the two-part estimation ESEM-within-CFA approach (Asparouhov & Muthén, 2009; Marsh et al., 2009). Part one utilized the traditional measurement model from model 1 as the target model for primary factors versus non-primary factors that were expected to have small factor loadings close to zero. A target rotation, within exploratory factor analysis (EFA) was used to estimate factor loadings for items across all factors "X" and "NP" (Table 1) (Zhang et al., 2019). Part 2 of this model consisted of using the EFA estimates for all primary and non-primary loadings as starting values to estimate the CFA model (Asparouhov & Muthén, 2009; Marsh et al., 2009). Model 4 in R was estimated using the *esemComp* package, which has a *lavaan* dependency. The *esemComp* package provided a user-friendly method for defining the target rotation matrix from part 1 of the ESEM-within-CFA analysis method, for use in the part 2 CFA analysis. For model 4, Part 1 was specified the same in both R and Mplus by using the loading matrix outlined in Table 1. The ML estimator was used for part 1 and part 2 in both R and Mplus. This included ensuring that fixed starting values of some non-primary factor loadings in model identification for part 2 were the same across software. The authors confirmed results in the *esemComp* package with simply estimating the models in the *lavaan* package for the model estimated in R to ensure consistency.

Models 3 and 4 estimated all parameters outlined in Table 1, including providing estimates for all non-primary "NP" loadings that were fixed to zero in models 1 and 2. All four proceeding model estimations were estimated in both Mplus and R, resulting in eight total models estimated in this study.

2.3.1 *Bayesian estimation considerations*
Default non-informative priors are specified slightly differently in Mplus compared to *blavaan*'s *Stan* estimator. Because the purpose of this study was to compare software, software defaults were utilized unless otherwise noted. The default MCMC technique used for estimation in the *blavaan* package used *Stan*, thus the *Stan* MCMC sampling technique was used for BSEM estimation in R as opposed to Just Another Gibbs Sampler (JAGS). It has been indicated that the *Stan* MCMC is more efficient and the optimal option for model estimation in the *blavaan* package (Bølstad, 2019; Hecht et al., 2021; Merkle et al., 2021; Smeets & van de Schoot, 2019). The number of MCMC iterations was set to 10,000 for both software for the BSEM estimations. The first half of the iterations (n=5,000) were designated as burn-in iterations and the last half of the samples (n=5,000) were used for estimation. Two MCMC chains were used for the BSEM models.

While a more appropriate technique may have used the potential scale reduction (PSR) value to provide evidence of convergence towards equilibrium of Bayesian chains, using this method would likely result in different iterations being completed across software. Instead, the PSR values of each Bayesian model were inspected after the 10,000 total iterations and confirmed for values below 0.05, suggesting that all chains converged for BSEM (Gelman, 2013).

Using *Stan*'s default starting values for the BSEM estimates presented an issue with model convergence in R for the BSEM-CL models. Therefore, the starting values in the estimation algorithm were specified as similar starting values that Mplus uses in its estimation algorithm. The initial starting values in both *blavaan*'s *Stan* and Mplus were chosen based on the ML estimated starting values (Jöreskog & Sörbom, 1982; Muthén & Muthén, 2017). As this study was focusing on comparisons of default options, the starting values were only adjusted if the attempted analysis had a non-convergence issue.

3. Results

The results of this study are presented as a comparison of the standardized factor loading estimates between Mplus and R for each of the four models explained above, and results can be found in Tables 2 through 5. Global fit indices were estimated next and compared between these two software for each of the four models (see Table 6). Additional global fit indices, which are only available for BSEM in R, and required estimation of the additional null model discussed above were reported in Table 7. Finally, the estimation times were reported in Table 8 for all four models in both R and Mplus.

3.1 Standardized Factor Loadings
The standardized factor loadings and their associated parameter estimate standard errors for ML estimation, as well as the parameter estimate posterior standard deviations, are compared in Tables 2 through 5. The primary expected factor loadings based on the original hypothesized model are highlighted for each of the four factors to assist readers in comparing the results to Table 1. Generally, only the trivial differences in parameter estimates were found when comparing estimates between the two software.

3.1.1 Model 1 ML standardized factor loadings
The standardized factor loadings were similar across the two software packages when ML estimation was used in model 1. Only trivial differences were noted between software for model 1, as is shown in Table 2. Standardized factor loadings were within .001 of the corresponding factor/item relationship parameter estimates for both software. Standardized

standard error estimates were within .002 of their corresponding parameter estimates for the specific factor/item relationship across both software.

3.1.2 Model 2 BSEM-NOCL standardized factor loadings

The standardized factor loadings for the BSEM-NOCL model 2, were similar for both software. The standardized factor loading parameter estimates were found to be within .003 of their corresponding estimates between the two software. Standardized posterior standard deviations were found to be within .002 of their corresponding estimates between the two software. While the differences between software were slightly larger for the BSEM-NOCL model compared to ML model, the difference is likely due to the differences in the specification of non-informative priors between these two software, as well as variations in the random seeds that each MCMC started with and default starting values within each software. The trivial differences in standardized parameter estimates would still result in the same conclusions about primary loadings for the specified model.

3.1.3 Model 3 BSEM-CL standardized factor loadings

The standardized factor loadings were similar across the two packages for the BSEM-CL model 3. The standardized factor loading parameter estimates were within .008 of their corresponding estimates between these software. The standardized posterior standard deviations were within .005 of their corresponding estimates between the two software. These differences are larger than what we observed for models 1 and 2, yet they are still considered trivial differences, and are likely due to variations in the random seed and slight differences in the other default non-informative priors between the software. The largest differences in the standardized factor loadings and the posterior standard deviations in the BSEM-CL model still did not change the practical interpretations of the models.

3.1.4 Model 4 ESEM standardized factor loadings

The standardized factor loadings were similar across the two software for the ESEM model 4. The standardized factor loadings were within .001 of each parameter estimate for both packages. The standardized standard error estimates were within .005 of their corresponding parameter estimates between the two software packages. The difference in parameter estimates between the two software is seen as trivial round differences. The ESEM model primary factor standardized loadings were on average smaller than the BSEM model estimates in both software, with the difference in standardized loadings between ESEM and BSEM being visually noticeable. These smaller estimates did not change any contextual interpretation of the measurement models.

Table 2: Model 1 factor loading parameter estimates with ML approach

	Mplus ML estimation				R ML estimation (*lavaan*)			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	0.750 (0.046)	0	0	0	0.749 (0.045)	0	0	0
cubes	0.434 (0.058)	0	0	0	0.434 (0.057)	0	0	0
paper	0.491 (0.055)	0	0	0	0.491 (0.055)	0	0	0
flags	0.605 (0.051)	0	0	0	0.605 (0.050)	0	0	0
general	0	0.836 (0.021)	0	0	0	0.836 (0.021)	0	0
paragrap	0	0.821 (0.022)	0	0	0	0.821 (0.022)	0	0
sentence	0	0.867 (0.018)	0	0	0	0.867 (0.018)	0	0
wordc	0	0.741 (0.029)	0	0	0	0.741 (0.029)	0	0
wordm	0	0.847 (0.020)	0	0	0	0.847 (0.020)	0	0
addition	0	0	0.585 (0.049)	0	0	0	0.585 (0.047)	0
code	0	0	0.718 (0.042)	0	0	0	0.718 (0.040)	0
counting	0	0	0.626 (0.047)	0	0	0	0.626 (0.045)	0
straight	0	0	0.678 (0.044)	0	0	0	0.678 (0.042)	0
wordr	0	0	0	0.578 (0.051)	0	0	0	0.578 (0.050)
numberr	0	0	0	0.517 (0.054)	0	0	0	0.517 (0.053)
figurer	0	0	0	0.604 (0.050)	0	0	0	0.604 (0.048)
object	0	0	0	0.556 (0.052)	0	0	0	0.556 (0.051)
numberf	0	0	0	0.548 (0.053)	0	0	0	0.548 (0.051)
figurew	0	0	0	0.454 (0.057)	0	0	0	0.454 (0.056)

Note. Standardized factor loadings (SE)

Table 3: Model 2 factor loading parameter estimates with BSEM-NOCL approach

	Mplus BSEM-NOCL estimation				R BSEM-NOCL estimation (*Stan*)			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	0.746 (0.047)	0	0	0	0.749 (0.047)	0	0	0
cubes	0.436 (0.060)	0	0	0	0.436 (0.058)	0	0	0
paper	0.492 (0.056)	0	0	0	0.492 (0.056)	0	0	0
flags	0.608 (0.052)	0	0	0	0.607 (0.052)	0	0	0
general	0	0.837 (0.021)	0	0	0	0.838 (0.021)	0	0
paragrap	0	0.822 (0.022)	0	0	0	0.823 (0.022)	0	0
sentence	0	0.868 (0.019)	0	0	0	0.869 (0.018)	0	0
wordc	0	0.743 (0.029)	0	0	0	0.743 (0.030)	0	0
wordm	0	0.848 (0.020)	0	0	0	0.849 (0.020)	0	0
addition	0	0	0.586 (0.050)	0	0	0	0.587 (0.050)	0
code	0	0	0.715 (0.042)	0	0	0	0.718 (0.043)	0
counting	0	0	0.626 (0.048)	0	0	0	0.628 (0.048)	0
straight	0	0	0.679 (0.044)	0	0	0	0.679 (0.044)	0
wordr	0	0	0	0.580 (0.052)	0	0	0	0.581 (0.052)
numberr	0	0	0	0.519 (0.054)	0	0	0	0.521 (0.055)
figurer	0	0	0	0.606 (0.050)	0	0	0	0.606 (0.051)
object	0	0	0	0.558 (0.054)	0	0	0	0.559 (0.054)
numberf	0	0	0	0.551 (0.053)	0	0	0	0.549 (0.053)
figurew	0	0	0	0.459 (0.057)	0	0	0	0.456 (0.057)

Note. Standardized factor loadings (posterior SD)

Table 4: Model 3 factor loading parameter estimates with BSEM-CL approach

	Mplus BSEM-CL estimation				R BSEM-CL estimation (*Stan*)			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	0.637 (0.074)	0.089 (0.066)	0.050 (0.070)	0.038 (0.071)	0.631 (0.070)	0.097 (0.066)	0.055 (0.069)	0.045 (0.070)
cubes	0.510 (0.076)	0.011 (0.063)	-0.052 (0.066)	-0.023 (0.068)	0.505 (0.073)	0.017 (0.063)	-0.049 (0.067)	-0.021 (0.068)
paper	0.483 (0.074)	0.049 (0.063)	0.025 (0.066)	-0.055 (0.067)	0.480 (0.073)	0.054 (0.062)	0.029 (0.066)	-0.050 (0.068)
flags	0.639 (0.073)	-0.109 (0.066)	0.036 (0.069)	0.077 (0.072)	0.635 (0.070)	-0.103 (0.067)	0.041 (0.070)	0.084 (0.070)
general	-0.030 (0.060)	0.860 (0.041)	0.040 (0.058)	-0.074 (0.061)	-0.026 (0.060)	0.858 (0.040)	0.039 (0.058)	-0.073 (0.058)
paragrap	0.001 (0.059)	0.809 (0.041)	-0.014 (0.058)	0.057 (0.059)	0.005 (0.058)	0.808 (0.040)	-0.016 (0.058)	0.058 (0.059)
sentence	-0.068 (0.061)	0.925 (0.042)	-0.006 (0.060)	-0.056 (0.062)	-0.063 (0.060)	0.922 (0.038)	-0.007 (0.059)	-0.055 (0.060)
wordc	0.041 (0.059)	0.708 (0.045)	0.021 (0.058)	0.041 (0.059)	0.044 (0.058)	0.707 (0.044)	0.018 (0.058)	0.045 (0.058)
wordm	0.046 (0.059)	0.831 (0.040)	-0.036 (0.057)	0.029 (0.058)	0.048 (0.058)	0.830 (0.039)	-0.038 (0.057)	0.031 (0.058)
addition	-0.192 (0.073)	-0.020 (0.071)	0.756 (0.071)	0.014 (0.075)	-0.190 (0.072)	-0.023 (0.070)	0.755 (0.069)	0.011 (0.073)
code	-0.005 (0.065)	0.114 (0.063)	0.585 (0.067)	0.113 (0.067)	-0.003 (0.064)	0.115 (0.061)	0.584 (0.065)	0.113 (0.067)
counting	0.041 (0.071)	-0.052 (0.066)	0.716 (0.068)	-0.072 (0.070)	0.044 (0.071)	-0.055 (0.067)	0.720 (0.067)	-0.075 (0.071)
straight	0.226 (0.066)	-0.013 (0.063)	0.581 (0.073)	-0.027 (0.069)	0.228 (0.067)	-0.012 (0.063)	0.585 (0.071)	-0.029 (0.068)
wordr	-0.082 (0.072)	0.040 (0.067)	-0.105 (0.071)	0.697 (0.074)	-0.078 (0.070)	0.038 (0.067)	-0.109 (0.069)	0.698 (0.072)
numberr	0.015 (0.070)	-0.110 (0.064)	-0.081 (0.070)	0.628 (0.073)	0.015 (0.070)	-0.111 (0.065)	-0.085 (0.068)	0.634 (0.071)
figurer	0.186 (0.067)	0.026 (0.059)	-0.018 (0.066)	0.519 (0.072)	0.188 (0.065)	0.027 (0.060)	-0.019 (0.065)	0.522 (0.071)
object	-0.138 (0.068)	-0.049 (0.065)	0.160 (0.069)	0.588 (0.077)	-0.136 (0.067)	-0.050 (0.063)	0.158 (0.068)	0.589 (0.072)
numberf	0.036 (0.066)	-0.002 (0.060)	0.078 (0.066)	0.472 (0.075)	0.036 (0.065)	-0.003 (0.060)	0.078 (0.066)	0.475 (0.074)
figurew	0.030 (0.065)	0.118 (0.058)	0.018 (0.065)	0.367 (0.077)	0.032 (0.064)	0.120 (0.058)	0.017 (0.064)	0.369 (0.076)

Note. Standardized factor loadings (posterior SD)

Table 5: Model 4 factor loading parameter estimates with ESEM approach

	Mplus ESEM				R ESEM			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	0.587 (0.063)	0.147 (0.071)	0.089 (0.080)	0.082 (0.084)	0.587 (0.062)	0.146 (0.071)	0.089 (0.080)	0.082 (0.084)
cubes	0.499 (0.068)	0.046 (0.071)	-0.052 (0.082)	0.000 (0.084)	0.499 (0.068)	0.046 (0.072)	-0.052 (0.083)	0.001 (0.084)
paper	0.449 (0.069)	0.096 (0.071)	0.073 (0.081)	-0.058 (0.085)	0.449 (0.069)	0.096 (0.071)	0.073 (0.081)	-0.058 (0.085)
flags	0.599 (0.051)	-0.105 (0.004)	0.081 (0.003)	0.147 (0.006)	0.598 (0.051)	-0.105 (0.004)	0.081 (0.003)	0.147 (0.006)
general	-0.017 (0.057)	0.847 (0.033)	0.058 (0.056)	-0.083 (0.059)	-0.017 (0.055)	0.847 (0.032)	0.059 (0.054)	-0.083 (0.057)
paragrap	-0.001 (0.056)	0.801 (0.033)	-0.013 (0.056)	0.084 (0.057)	-0.001 (0.056)	0.801 (0.033)	-0.013 (0.056)	0.083 (0.057)
sentence	-0.055 (0.002)	0.911 (0.019)	0.000 (0.000)	-0.057 (0.002)	-0.055 (0.002)	0.911 (0.018)	0.000 (0.000)	-0.057 (0.002)
wordc	0.067 (0.059)	0.695 (0.040)	0.029 (0.060)	0.052 (0.062)	0.067 (0.059)	0.695 (0.040)	0.029 (0.060)	0.052 (0.062)
wordm	0.059 (0.056)	0.819 (0.032)	-0.028 (0.056)	0.040 (0.058)	0.059 (0.054)	0.819 (0.031)	-0.027 (0.055)	0.040 (0.056)
addition	-0.241 (0.010)	-0.004 (0.000)	0.746 (0.051)	0.046 (0.002)	-0.241 (0.010)	-0.004 (0.000)	0.746 (0.049)	0.046 (0.002)
code	0.000 (0.073)	0.158 (0.062)	0.530 (0.061)	0.169 (0.073)	0.000 (0.072)	0.158 (0.062)	0.530 (0.060)	0.169 (0.071)
counting	0.091 (0.079)	-0.052 (0.069)	0.716 (0.063)	-0.087 (0.084)	0.090 (0.078)	-0.053 (0.069)	0.716 (0.063)	-0.087 (0.083)
straight	0.327 (0.077)	-0.008 (0.072)	0.555 (0.072)	-0.029 (0.089)	0.327 (0.073)	-0.008 (0.071)	0.555 (0.067)	-0.028 (0.085)
wordr	-0.084 (0.003)	0.069 (0.003)	-0.144 (0.006)	0.702 (0.050)	-0.084 (0.003)	0.069 (0.003)	-0.144 (0.006)	0.702 (0.050)
numberr	0.047 (0.082)	-0.135 (0.071)	-0.095 (0.083)	0.634 (0.070)	0.047 (0.082)	-0.135 (0.071)	-0.095 (0.083)	0.634 (0.070)
figurer	0.275 (0.073)	0.035 (0.067)	-0.024 (0.077)	0.497 (0.069)	0.275 (0.073)	0.035 (0.067)	-0.024 (0.077)	0.497 (0.069)
object	-0.184 (0.083)	-0.054 (0.069)	0.238 (0.080)	0.565 (0.072)	-0.185 (0.081)	-0.054 (0.069)	0.238 (0.079)	0.565 (0.070)
numberf	0.057 (0.080)	-0.002 (0.068)	0.137 (0.078)	0.432 (0.074)	0.057 (0.077)	-0.002 (0.067)	0.137 (0.075)	0.433 (0.070)
figurew	0.057 (0.077)	0.168 (0.065)	0.022 (0.076)	0.332 (0.074)	0.057 (0.075)	0.168 (0.065)	0.022 (0.075)	0.332 (0.072)

Note. Standardized factor loadings (SE)

3.2 Global Fit Indices

Evaluation of the global fit indices included comparisons between the two software and different methods, which are included in Table 6. Mplus and R both estimate the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the standardized root mean squared residual (SRMR) values. Both Mplus and the *blavaan* package in R have adopted several Bayesian adjusted fit indices along with the posterior predictive p-value (PPP) (Hoofs et al., 2018). The *blavaan* package in R provides additional Bayesian adjusted global fit indices. Currently, the only way to estimate these additional Bayesian adjusted global fit indices is in R. Additionally, the information criteria and log-likelihood tests were computed by both software for each of the various models but were not reported in this study.

Models 1 and 4, were both developed using ML estimators, which provided consistent values for RMSEA, CFI, and TLI global fit indices between the two software. The default settings of R and Mplus differ in how the number of freely estimated parameters are specified, but this only impacted the number of degrees of freedom during model estimation. The difference in degrees of freedom did not impact the conclusions drawn from the two software. Still, practitioners should be aware of the potential impact of that varying degrees of freedom may have when using default model specifications. There were trivial differences in the SRMR estimates, which is a direct result of the impact degrees of freedom has on estimating absolute measures of fit like SRMR between the estimated models (Bentler, 1995; Taasoobshirazi & Wang, 2016). Both ML and ESEM models provided consistent global fit estimates across the two software.

The BSEM-NOCL model 2, provided consistent Bayesian adjusted global fit indices across the two software. Both the PPP and Bayesian CFI (BCFI) estimates were similar across the two software and the Bayesian adjusted RMSEA (BRMSEA) and TLI (BTLI) differed by only .002. The highest PSR value for each software was within the acceptable values, and the values differed by only .003 between the two software. This slight difference between software results is likely attributed to the sampling seeds and different non-informative prior specification. The conclusions drawn from model 2 were consistent across R and Mplus, despite the trivial model differences.

The BSEM-CL model 3, provided consistent BRMSEA, BCFI, and BTLI values as seen in the model 2 comparison. However, the PPP values were considerably different between the two software, with the R *blavaan* estimate being .025 larger than the Mplus estimate. Both BSEM-CL models met the criteria for acceptable model fit, but seeing such a large discrepancy between estimates of PPP is concerning. Additionally, the highest PSR value for the model was .011 larger in Mplus than in R. Some of the difference is likely attributed to the difference in prior specifications between software for other parameters in the model. Regardless, with heavy reliance being placed on the PPP value in Mplus, this could lead to differing conclusions being drawn from the PPP value. While rejection of model's goodness of fit should not rely on one global fit index, this discrepancy could result in a borderline model meeting the model fit criteria in one software while having poor fit in another software. While the other default non-informative prior information is different between software, further research should explore and address the root of this discrepancy between software.

3.2.1 Flexible approaches allowing cross-loading

The results of this study support both previously published research and the hypothesis of this study; allowing minor cross-loadings for more flexible and realistic models improves the model fit. Both BSEM-CL and ESEM with minor cross-loading improved model fit with acceptable global fit indices compared to their BSEM-NOCL and ML counterparts, respectively. The improvement in model fit was consistent across software.

By removing the restriction that forces minor cross-loadings to zero in the BSEM-CL and ESEM models, indicators can share variance with the non-primary factors. This allowance had considerable impact on the standardized factor loading estimates. Differences in the standardized factor loadings of primary loadings of up to .170 were seen when estimating with BSEM-NOCL compared to BSEM-CL. Differences in the standardized factor loading of primary loadings of up to .188 were seen when estimating with ESEM compared to ML. With these considerable differences in the standardized factor loadings, contextual interpretation of the entire measurement model changed when including estimation of minor minor-cross loadings. The overall model fit was acceptable when allowing minor cross-loading within ESEM and BSEM-CL. Some indicators that were highly related to primary factors using the traditional approach, were then found to be only moderately related when non-primary cross-loadings were allowed to be estimated instead of being fixed to zero. To see this shift in standardized factor loading estimates, review the items 'code' and 'addition' across estimation methods for the 'Speed' factor. There is an overall improvement in model fit when allowing non-primary loadings to be estimated, yet the standardized factor loading will shift from a high loading seen in traditional fixed zero cross-loading ML and BSEM-NOCL to a moderate loading in ESEM and BSEM-CL. This supports the ESEM and BSEM-CL approach's benefit which is providing a more realistic model of shared variance with non-primary factors.

3.2.2 blavaan additional Bayesian adjusted global fit indices

R's *blavaan* package provided additional Bayesian adjusted global fit indices, including the Bayesian normed fit index (BNFI), Bayesian McDonald's centrality index (BMc), Bayesian gamma-hat, and Bayesian adjusted gamma-hat (Garnier-Villarreal & Jorgensen, 2020). These added Bayesian adjusted global fit indices provide information for researchers seeking additional evidence regarding model fit. These adjusted global fit indices may be of interest to practitioner and more appropriate given certain applications, such as gamma hat's appropriate use in studies with smaller sample sizes.

Table 6: Software default global fit indices for each model

	Number of freely estimated parameters	Chi-square (d.f.) [p-value]	RMSEA	CFI	TLI	SRMR	Highest PSR	PPP	BRMSEA	BCFI	BTLI
Mplus ML estimation	63	314 (171) [<0.0001]	0.062	0.915	0.900	0.064					
R ML estimation (*lavaan*)	44	314 (146) [<0.0001]	0.062	0.915	0.900	0.067					
Mplus BSEM-NOCL estimation	63						1.004	<0.001	0.062	0.914	0.899
R BSEM-NOCL estimation (*Stan*)**							1.001	<0.001	0.064	0.914	0.897
Mplus BSEM-CL estimation	120						1.012	0.093	0.033	0.982	0.972
R BSEM-CL estimation (*Stan*)**							1.001	0.118	0.032	0.982	0.972
Mplus ESEM	101	130 (101) [0.0251]	0.031	0.985	0.975	0.023					
R ESEM	101	130 (101) [0.0251]	0.031	0.985	0.975	0.024					

**Default starting values adjusted in R

Table 7: Additional global fit indices available in R blavaan

	Additional global fit indices			
	BNFI	BMc	B Gamma Hat	Adjusted B Gamma Hat
R BSEM-NOCL estimation (*Stan*)**	0.857	0.755	0.944	0.916
R BSEM-CL estimation (*Stan*)**	0.933	0.943	0.988	0.976

**Default starting values adjusted in R

3.3 Comparison of Software Estimation Time

Also compared in this study were the computational times that each software took to estimate each model. Table 8 outlines the estimation time required to complete models 1 through 4 in both software. All time estimates were collected from the same computer used to complete all the analyses. All four models in Mplus as well as estimation of ML and ESEM with R *blavaan* package required very little computational time to complete estimations. However, the Bayesian estimation in R *blavaan* had considerably longer computation times, taking up to fifteen minutes depending on model complexity. The additional global fit indices available in R *blavaan* come with an additional computational cost when estimating the null models. The known dependency of *Stan* to estimate Bayesian models is the likely cause of this issue but will surely impact a researcher's decision to use *blavaan* to estimate BSEM models.

Table 8: Estimation times for model convergence in R and Mplus

Model & Software	Estimation of primary analysis	Estimation of null model for additional global fit indices
Mplus ML estimation	1 second	
R ML estimation (*lavaan*)	1.9 second	
Mplus BSEM-NOCL estimation	4 seconds	
R BSEM-NOCL (*Stan*)	10.1 minutes	6.3 minutes
Mplus BSEM-CL estimation	9 seconds	
R BSEM-CL estimation (*Stan*)	14.5 minutes	6.5 minutes
Mplus ESEM	7 seconds	
R ESEM	5 seconds	

Note: Time estimates using Intel i7-9850H processor with 16 GB of RAM

4. Discussion

The result of this methodological study provide evidence for practitioners that both ESEM and BSEM-CL techniques in R and Mplus provide consistent estimates across software. Researcher should feel comfortable in selecting R or Mplus to estimate ESEM or BSEM models. ML and ESEM approaches across the two software, resulted in only trivial differences in parameter estimates. BSEM approaches across the two software, resulted in observing slightly more variation in estimates. However, these variations were not surprising since there were differences in the software default settings, including default non-informative priors as well as variations in sampling seeds and starting points. The only concerning discrepancies between software were the PPP and PSR value estimated for the BSEM-CL model. These two measures, PPP and PSR, differed between the software by up to .025 and 0.011 respectively. While the differences did not change the contextual interpretation of the models, further research is needed to determine the root cause and potential solutions of the problem. Further methodological studies could define all non-informative prior information to be consistent across different software, and test to determine if this resolves the discrepancies.

This study utilized data from a commonly known study with 301 subjects, but there are known parameter estimate bias issues when using sample sizes less than 400. While parameter bias was outside the scope of this study, future simulation studies could explore the general smaller standardized factor loadings of ESEM estimates when compared to BSEM-CL estimates. This would help determine if the difference is due to sample size, other unexplored software defaults, or is attributed to software algorithm differences. While there was some variation between model estimates, ESEM and BSEM-CL provide an improved model fit that mirror realistic measurement models across both software.

As the true population parameters are unknown, the deviation from the true population parameters for ESEM varied in comparison to BSEM with dependence on data conditions (Wei et al., 2022; Xiao et al., 2019). Thus, the deviation in parameter estimates between ESEM and BSEM in this analysis is not surprising. Ultimately, the smaller ESEM estimates did not change any contextual interpretation of the measurement models and would corroborate the study findings with BSEM as previously suggested (Xiao et al., 2019). Yet continued simulation research can explore the measurable difference of bias between ESEM and BSEM under varying conditions.

ESEM techniques have been adopted in other various packages including *psych* and *sem* (Guàrdia-Olmos et al., 2013; Revelle, 2015). Given that anyone can develop a package, and that there is a documented method to reporting package bugs to authors, comparing parameter estimates across packages is necessary. While this study utilized *lavaan* for model estimation, future research could confirm consistency of parameter estimates across the various R packages for ESEM. Additionally, further exploration of consistency of parameter estimates across other packages that have only adopted ESEM methods including AMOS, STATA, SAS, EQS, etc.

An unsatisfactory conclusion in this study was the observed substantial increase in the computational time when estimating the BSEM model using R's *blavaan*. The package authors are aware of the computational issues but may not be aware of the considerably higher computation time compared to Mplus (Merkle & Jorgensen, 2022). Future studies could explore parallel computing in R to see if computational time is reduced. This study also estimated the BSEM-NOCL model using the Just Another Gibbs Sampler (*JAGS*) non-default option in *blavaan* but found the computation time to be even longer and as a result, authors terminated the analyses. Based on this issue and suggestions provided by the package authors to use *Stan* for optimum efficiency, the authors did not continue using *JAGS*. Further research may be able to confirm whether *Stan* is more efficient, by estimating increased computational time if using *JAGS*, and document if *Stan* provides consistent parameter estimates compared with *JAGS* when using *blavaan*. Based on estimation times alone, the results of this study suggest avoiding estimating BSEM models with *blavaan* until the computational times are reduced. The additional Bayesian adjusted global fit indices may be of importance to some researchers, but the additional computational time for more complex models to estimate the null model may outweigh the benefit. While Mplus has adopted some Bayesian adjusted fit indices, there is hope that Mplus will consider adopting the additional Bayesian adjusted fit indices.

5. Conclusions

Both BSEM-CL and ESEM provided an improved approach to measurement model fit with similar parameter estimates in both R and Mplus. These methods support a more flexible and realistic model generation and provide similar contextual conclusions regarding the measurement models across the two software, R and Mplus. For researchers looking to estimate BSEM models, they will need to weigh the potential estimation time due to higher computation times, with the cost of an Mplus license. The package dependencies required to estimate a BSEM model in R might seem excessive or challenging when compared to the clear documentation of analyses provided by Mplus. As software continue to adopt ESEM and BSEM, comparison of model estimates in R and Mplus should be completed.

Acknowledgements

We would like to thank Kennesaw State University's financial assistance in obtaining an Mplus license.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438.

Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(1), 1-14. http://www.statmodel.com/download/BayesAdvantages18.pdf

Bentler, P. M. (1995). EQS. Structural equations program manual.

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research. Guilford publications.

Bølstad, J. 2019. How Efficient is Stan Compared to JAGS? Conjugacy, Pooling, Centering, and Posterior Correlations. *Playing with Numbers: Notes on Bayesian Statistics*. www.boelstad.net/post/stan_vs_jags_speed/

Chang, C., Gardiner, J., Houang, R., & Yu, Y. L. (2020). Comparing multiple statistical software for multiple-indicator, multiple-cause modeling: an application of gender disparity in adult cognitive functioning using MIDUS II dataset. *BMC Medical Research Methodology*, 20(1), 1-14.

De Beer, L.T. & Van Zyl, L.E. (2019). ESEM code generator for Mplus. Retrieved from https://www.surveyhost.co.za/esem/ doi: 10.6084/m9.figshare.8320250

Craig, J. F. (2017). Employee Empowerment, Self-determination Theory, and Employee Engagement: A Mediation Model. Doctoral dissertation, University of Oklahoma. https://shareok.org/handle/11244/52759

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, 25(1), 46.

Gelman, A. (2013). Two simple examples for understanding posterior p-values whose distributions are far from uniform. *Electronic Journal of Statistics*, 7, 2595-2602.

Gittner, K. B. (2021). Small sample performance of Bayesian confirmatory factor analysis in the presence of missing data. Doctoral dissertation, University of Northern Colorado. Retrieved from https://digscholarship.unco.edu/dissertations/818/

Guàrdia-Olmos, J., Peró-Cebollero, M., Benítez-Borrego, S., & Fox, J. (2013). Using sem library in r software to analyze exploratory structural equation models. *Proceedings of the 59th ISI World Statistics Congress* (pp. 25-30).

Gucciardi, D. F., & Zyphur, M. J. (2016). Exploratory structural equation modelling and Bayesian estimation. *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists*, 172-194.

Gustafsson, J.-E. (2001). On the hierarchical structure of ability and personality. *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 25–42). Lawrence Erlbaum Associates Publishers.

Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Lüdtke, O., & Diallo, T. M. (2019). A systematic evaluation and comparison between exploratory structural equation modeling and Bayesian structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(4), 529-556.

Harman, H. H. (1976). Modern Factor Analysis. University of Chicago press.

Harper, W. V., Eschenbach, T. G., & James, T. R. (2011). Concerns about maximum likelihood estimation for the three-parameter Weibull distribution: Case study of statistical software. *The American Statistician*, 65:1, 44-54, DOI: 10.1198/tast.2011.09103

Hecht, M., Weirich, S., & Zitzmann, S. (2021). Comparing the MCMC efficiency of JAGS and Stan for the multi-level intercept-only model in the covariance-and mean-based and classic parametrization. *Psych*, 3(4), 751-779.

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: the stability of a bi-factor solution. *Supplementary Educational Monographs*, 48, xi + 91.

Hoofs, H., van de Schoot, R., Jansen, N. W., & Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement*, 78(4), 537-568.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of marketing research*, 19(4), 404-416.

Kline, R. B. (2015). Principles and practice of structural equation modeling. Guilford publications.

Koizumi, R., & In'nami, Y. (2020). Structural equation modeling of vocabulary size and depth using conventional and Bayesian methods. *Frontiers in Psychology*, 11, 618.

Liang, X., Yang, Y., & Cao, C. (2020). The performance of ESEM and BSEM in structural equation models with ordinal indicators. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(6), 874-887.

Ma, H. (2020). Evaluation of the utility of informative priors in Bayesian structural equation modeling with small samples. Doctoral dissertation, Southern Methodist University. https://scholar.smu.edu/cgi/viewcontent.cgi?article=1004&context=simmons_depl_etds

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, 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: A Multidisciplinary Journal*, 16(3), 439–476. doi:10.1080/10705510903008220

Merkle EC, Fitzsimmons E, Uanhoro J, & Goodrich B (2021). Efficient Bayesian Structural Equation Modeling in Stan. *Journal of Statistical Software*, 100(6), 1–22. doi:10.18637/jss.v100.i06.

Merkle, E., & Jorgensen, T., (April, 2022). Blavaan incredible slow...speed-up solutions?. *Discussion group for blavaan, an R package for Bayesian latent variable analysis*. https://groups.google.com/g/blavaan/c/jiBuYc1inYE?pli=1

Merkle, E. C., & Rosseel, Y. (2015). blavaan: Bayesian structural equation models via parameter expansion. *arXiv*, 1511.05604.

Morin, A. J., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. *Handbook of sport psychology*, 1044-1073.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313.

Muthén, L.K., & Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA.

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Reis, D. (2017). Further insights into the German version of the multidimensional assessment of interoceptive awareness (MAIA). *European Journal of Psychological Assessment*.

Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, 337, 338.

Rosseel Y (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. doi:10.18637/jss.v048.i02.

Silvestrin M, & T. de Beer L (2022). esemComp: ESEM-within-CFA syntax composer. R package version 0.2, https://mateuspsi.github.io/esemComp.

Smeets, L., & van de Schoot, R. (October, 2019) Bayesian Regression in Blavaan (using Stan). https://www.rensvandeschoot.com/tutorials/bayesian-regression-in-blavaan-using-stan/

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 131-161.

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.7, https://mc-stan.org/.

Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, 7(2), 210.

Taasoobshirazi, G., & Wang, S. (2016). The performance of the SRMR, RMSEA, CFI, and TLI: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods*, 11(3), 31-39.

Wei, X., Huang, J., Zhang, L., Pan, D., & Pan, J. (2022). Evaluation and Comparison of SEM, ESEM, and BSEM in Estimating Structural Models with Potentially Unknown Cross-loadings. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(3), 327-338.

Xiao, Y., Liu, H., & Hau, K. T. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5), 665-677.

Zhang, G., Hattori, M., Trichtinger, L. A., & Wang, X. (2019). Target rotation with both factor loadings and factor correlations. *Psychological methods*, 24(3), 390.