Systematic Evaluation and Comparison between Exploratory Structural Equation Modeling and Bayesian Structural Equation Modeling

Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Lüdtke, O., & Diallo, T. M. O. (2019). A Systematic Evaluation and Comparison Between Exploratory Structural Equation Modeling and Bayesian Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, Advance Online Publication. https://doi.org/10.1080/10705511.2018.1554999

Acknowledgments

The authors would like to acknowledge David Kaplan, Bengt Muthén, and Tihomir Asparouhov for their comments on earlier versions of this paper.

Abstract

In this study, we contrast two competing approaches, not previously compared, that balance the rigor of CFA/SEM with the flexibility to fit realistically complex data. Exploratory SEM (ESEM) is claimed to provide an optimal compromise between EFA and CFA/SEM. Alternatively, a family of three Bayesian SEMs (BSEMs) replace fixed-zero estimates with informative, small-variance priors for different subsets of parameters: cross-loadings (CL), residual covariances (RC) or CLs and RCs (CLRC). In Study 1, using three simulation studies, results showed that (1) BSEM-CL performed more closely to ESEM; (2) BSEM-CLRC did not provide more accurate model estimation compared with BSEM-CL; (3) BSEM-RC provided unstable estimation; (4) different specifications of targeted values in ESEM and informative priors in BSEM have significant impacts on model estimation. The real data analysis (Study 2) showed that the differences in estimation between different models were largely consistent with those in Study1 but somewhat smaller.

Key words: factor analysis, Bayesian statistics, exploratory structural equation modeling, informative priors.

Factor analysis is a mainstream statistical technique for multivariate data analysis. Typically, confirmatory factor analysis (CFA) models are used to formalize measurement hypotheses and develop measurement instruments. However, in CFA, unnecessarily strict constraints with inappropriate exact zero cross-loadings and residual covariances can result in poor model fit; substantial parameter biases in estimation of factor loadings and correlations; and a series of model modifications capitalizing on chance features of the data (Cole, Ciesla, & Steiger, 2007; MacCallum, Roznowski, & Necowitz, 1992; Marsh et al., 2009; 2017). In recent years, competing approaches to structural equation modeling (SEM) have been developed which aim to balance CFA/SEM rigor with the flexibility to fit realistically complex data. These include various specifications for Bayesian Structural Equation Modeling (BSEM; Muthén & Asparouhov, 2012; see Van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017 for a review) and Exploratory Structural Equation Modeling (ESEM) with a reliance on target rotation (Asparouhov & Muthén, 2009, Marsh et al., 2009; Marsh, Morin, Parker, & Kaur, 2014). Although ESEM and BSEM approaches are based on different methodological frameworks (Maximum Likelihood [ML] and Bayesian respectively), both allow researchers to freely estimate inappropriate exact zero cross-loadings or residual covariances and still have a priori control on the expected factor structure, thus better representing substantive theory. However, few studies to our knowledge have directly compared the two approaches in estimation of factor structure. In particular, the great modeling flexibility of BSEM allows researchers to assess and estimate measurement models in the various ways (e.g., using informative cross-loading priors or/and residual covariance priors, see below for more discussion). Nevertheless, the performance of BSEM models incorporating different subsets of priors has not been systematically examined. To fill this gap, this study is among the first to provide a comprehensive comparison of CFA, ESEM, and alternative BSEM approaches based on both real and simulated data.

Factor Analysis: EFA vs. CFA

In the SEM framework, factor analysis is a dimensional reduction procedure that extracts information from high-dimensional observed indicators to an underlying set of latent variables of lower dimensionality through the following equations:

$$y_i = \mu + \Lambda\eta_i + \varepsilon_i \tag{1}$$

$$V(y_i) = \Lambda\Psi\Lambda' + \Theta \tag{2}$$

where $i = 1, \ldots, N$; N is the sample size; μ is a $p \times 1$ vector of intercepts; y_i is a $p \times 1$ vector of observed indicators; η_i is a $q \times 1$ vector of latent variables; ε_i is a $p \times 1$ vector of measurement errors; Λ is a $p \times q$ loading matrix, reflecting the relations between observed indicators and latent factors; Ψ is a factor covariance matrix; and Θ is a residual covariance matrix. Standard assumptions of this model are that η_i and ε_i are normally distributed and independent.

Historically, exploratory factor analysis (EFA; Jennrich & Sampson, 1966) and confirmatory factor analysis (CFA; Jöreskog, 1969) are the two key variants of factor analysis, each following different approaches and assumptions to estimate Λ. EFA and CFA have certain advantages and disadvantages. EFA is an important precursor of CFA that is used to identify and distinguish between key psychological constructs (Cudeck & MacCallum, 2007). EFA first optimizes a target function for the parameters based on a minimally identified version of the model in Equation 1 to generate preliminary estimates. These preliminary estimates are then applied with rotation to produce a parsimonious Λ that optimizes a specific simplicity function. Analytic rotation of the factor pattern matrix involves the postmultiplication of the pattern matrix by the inverse of an optimal transformation matrix:

$$\Lambda^* = \Lambda(\mathrm{T}^*)^{-1} \tag{3}$$

where T^* is an $q \times q$ optimal transformation matrix, determined by minimizing a continuous complexity function, $f(\Lambda)$, of the elements in the pattern matrix. Various rotation procedures define $f(\Lambda)$ differently and yield different rotated matrices Λ^* with a simple pattern of loadings. A mechanical rotation criterion (e.g., geomin[1]) is thought to be relatively easy to implement. However, in the mechanical approach "the factors are extracted from the data without specifying the number and pattern of loadings between the observed variables and the latent factor

[1] The rotation function for the Geomin rotation criterion is $f(\Lambda) = \sum_{i=1}^{p}\left(\prod_{j=1}^{m}(\lambda_{ij}^2 + \varepsilon)\right)^{1/m}$

Where ε is a small positive constant added by Browne (2001) to reduce the problem of indeterminacy. Geomin has performed relatively well when numbers of non-zero cross-loadings for each latent variables are greater than 1 in both simulation and empirical examples, when compared with other mechanical rotation criteria (Marsh et al., 2009; McDonald, 2005).

variables” (Bollen, 2002, p. 615), thus providing little to no opportunity to incorporate a priori factor structure into the $f(\Lambda)$. In contrast, CFA starts with stronger theoretical assumptions by specifying numbers, associations, and the pattern of free parameters in Λ (Jöreskog, 1969). The basic independent cluster model of CFA (ICM-CFA) posits each observed indicator is only allowed to load on one latent factor (McDonald, 1985). In this regard, all cross-loadings that are freely estimated in EFA are constrained to be zero in ICM-CFA. These constraints mean that many psychological measures with well-defined EFA factor structures are not supported in ICM-CFA (Marsh et al., 2009, 2014). Marsh and his colleagues (Marsh et al., 2009, 2013, 2014; also see Asparouhov, Muthén, and Morin, 2015) argue that ICM-CFA is too restrictive for psychological and applied research, because most psychological items have multiple determinants and small cross-loadings which are logically justifiable in terms of substantive theory or item content (e.g., method effects). The inappropriate imposition of zero factor loadings usually results in systematically inflated factor correlations associated with poor discriminant validity, poor model fit to item-level factor structures, and biased structural parameter estimates in SEMs (Marsh et al., 2009, 2013, 2014). Furthermore, the strategies often used to compensate for ICM-CFA model’s inadequacies (e.g., a stepwise relaxation of parameters in relation to cross-loadings and residual covariances using model modification indices) can be misleading (Asparouhov et al., 2015; Marsh et al., 2014, 2017; Muthén & Asparouhov, 2012).

Conceptually, EFA with target rotation (Browne, 2001) can be assumed to lie in-between the mechanical approach of EFA rotation (weak a priori factor structure) and ICM-CFA model specification (strong a priori factor structure; Asparouhov & Muthén, 2009; Marsh et al., 2014). The target rotation criterion is designed to find a rotated solution $f(\Lambda)$ that is closest to a targeted pattern matrix. In the early versions of target rotation in EFA, a fully specified target matrix was indirectly used (Horst, 1941; Tucker, 1944), whereas its later versions were direct and could be based on only a partially specified target matrix (Browne, 1972a, 1972b; Gruvaeus, 1970). For identification purposes at least $q-1$ entries must be specified in each column for oblique rotation and $(q-1)/2$ entries must be specified in each column for orthogonal rotation. The rotation function is:

$$f(\Lambda)=\sum_{i=1}^{p}\sum_{j=1}^{q}a_{ij}(\lambda_{ij}-b_{ij})^2 \tag{4}$$

where $a_{ij} = 1$ if λ_{ij} is a target and 0 if λ_{ij} is not a target, and b_{ij} is the targeted value. Note that the user must provide a_{ij} and b_{ij}, to define $f(\Lambda)$. Supposed a structure of the loading matrix ($p = 9, q = 3$) with population values Λ, which includes main loadings (.8), major cross-loadings (.2) and minor cross-loadings (.01) (see Figure 1). Two matrices were provided in EFA with target rotation: a matrix A that designates whether each pattern coefficient was (1) or was not (0) a target, and a matrix B that provides values that targeted elements will be rotated toward and denotes nontargeted elements with a ? sign. As shown in Figure 1, 0 was chosen for target values b_{ij} in B, which is the most common specification in practice (see Marsh et al., 2014 for a review). In such cases, the cross-loadings of the rotated factor pattern matrix are only made as close to the specified zeros as possible (Browne, 2001), whereas in CFA cross-loadings are constrained to be the specified values of zero. Thus, target rotation allows researchers to have more a priori control on the expected factor structure and have approximately fixed-to-zero cross-loadings estimates.

Recently researchers examined how the number of targets and target error (i.e., $b_{ij} \neq \lambda_{ij}$) influence the accuracy and stability in relation to a rotated pattern matrix in EFA with target rotation (e.g., Myers, Ahn, & Jin, 2013, Myers, Jin, Ahn, Celimli, & Zopluoglu, 2015). Myers et al. (2013, 2015) found that the effects of target error on both accuracy (bias) and stability (variability) in relation to the rotated pattern matrix were negligible, but a small positive effect of (increasing) the number of targets specified was evident. In comparison with an easier-to-use mechanical rotation criterion (i.e., geomin rotation), target rotation has been shown to perform better in terms of accuracy, particularly when factor structures were more complex, whereas geomin rotation produced more stable factor solutions (Asparouhov & Muthén, 2009; Myers et al., 2015).

In the EFA framework, however, the new and evolving methodologies associated with CFA and SEM cannot be appropriately evaluated and applied. For example, in EFA it is not feasible to test measurement invariance (relating to groups, time, and covariates) and evaluate relations between latent variables with other constructs (Marsh et al., 2009, 2014, see below). Recently, in order to resolve these dilemmas between CFA and EFA, researchers have developed ESEM and BSEM approaches that allow researchers to define more

appropriately the underlying factor structure and still apply the advanced statistical methods relating to CFAs and SEMs (Marsh et al., 2014; Muthén & Asparouhov, 2012).

The most basic ESEM model is equivalent to EFA. Nevertheless, ESEM offers greater flexibility as it can accommodate residual covariances, covariates, and measurement invariance test in an EFA model (Asparouhov & Muthén, 2009; Marsh et al., 2009). In ESEM, multiple sets of CFA and/or ESEM factors can be included in the loading matrix Λ. Specifically, the CFA factors are identified as in traditional SEM in which each factor is associated with a different set of indicators. ESEM factors can be divided into blocks of factors so that different sets of indicators can be used to estimate ESEM factors within different blocks. However, each indicator can be assigned to more than one set of CFA and/or ESEM factors (see Asparouhov & Muthén, 2009; Marsh et al., 2014 for more details). Assignments of items to CFA and/or ESEM factors should rely on a priori theoretical and practical considerations and preliminary tests conducted with the data (Marsh et al., 2009, 2014).

Given that the basic ICM-CFA model is nested under the corresponding ESEM, conventional approaches to model comparison can be used to compare the fit of the two models. CFA models typically do not provide an adequate fit to the data and tend to be misspecified due to the restrictive assumption that each indicator is allowed to load on only one factor. Such independent cluster models appear to be rare in populations of interest. Typically, when true positive cross-loadings are constrained to be zero in ICM-CFA models, the factor correlations are likely to be positively biased, which might undermine the discriminant and predictive validity of the factors that form instruments (Marsh et al., 2014). Indeed, based on simulated data, the ESEM solution consistently provided improved model fit and more accurate factor correlation estimates than ICM-CFA solution (Marsh et al., 2010).

Bayesian Structural Equation Modeling (BSEM)

Bayesian analysis is a broad topic and has been well established in mainstream statistics (Kaplan, 2014; Van de Schoot et al., 2017). In the Bayesian approach, a prior distribution $p(\theta)$ is specified for each of the CFA model parameters, i.e., $\theta = (\mu, \Lambda, \Psi, \Theta)$; this prior distribution reflects previous knowledge about the parameters (see Kaplan, 2014, for an introduction). Based on the observed data $Y = (y_1, \ldots, y_n)^T$ a posterior distribution $p(\theta|Y)$ is then determined, which is proportional to the likelihood function $p(\theta|Y)$ of the data given the model parameters multiplied by the prior distribution: $p(\theta|Y) = p(Y|\theta)p(\theta) / p(Y) \propto p(Y|\theta)p(\theta)$.

The likelihood for model (1) is

$$p(Y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) = \prod_{i=1}^{n}\left[(2\pi)^{-p/2}\left|\Lambda\Psi\Lambda^T + \Theta\right|^{-1/2} \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T(\Lambda\Psi\Lambda^T+\Theta)^{-1}(y_i-\mu)\right\}\right]. \quad (5)$$

Various prior distributions of θ may be used. In the SEM framework, it is conceptually convenient to specify the prior distributions of the model parameters as sets of common conjugate distributions (see Kaplan & Depaoli, 2012). For the CFA model, let $\theta_{norm} = \{\mu, \Lambda\}$ be the set of free model parameters which prior distributions are assumed to follow a normal distribution:

$$\theta_{norm} \sim N(\mu, \Omega) \quad (6)$$

where are the mean and variances hyperparameters of the normal prior, respectively. Different choices of μ and Ω will yield different degrees of informativeness for the prior distributions. For example, the variance of 0.01 for a cross-loading yields a prior where 95% of the loading variation is between -0.2 and 0.2 (see below for more discussion).

The prior distribution that is typically used for the covariance matrix of multivariate normally distributed variables, such as Ψ and Θ, is known as the inverse-Wishart distribution (Barnard, McCulloch, & Meng, 2000; Gelman et al., 2013; Kaplan, 2014). The inverse-Wishart distribution is a conjugate prior for multivariate normally distributed variables implying that when combining with the likelihood function, it will result in a posterior distribution that belongs to the same distributional family. Another important advantage of the inverse-Wishart distribution is that it ensures positive definiteness of the covariance matrix. Let $\theta_{IW} = \{\Psi, \Theta\}$ be the set of free model parameters that are assumed to follow an inverse-Wishart distribution:

$$\theta_{IW} \sim IW(\mathbf{R}, df) \quad (7)$$

where $\mathbf{R}$ is a positive definite scale matrix and df is the number of degrees of freedom with $df > p-1$, where p is the number of observed variables. The larger the df, the higher the certainty about the information in $\mathbf{R}$, and the more informative the distribution is (Gelman et al., 2013; see below for detailed discussion).

An important benefit of the BSEM approach is the flexible specification of models that would be unidentified in a likelihood-based approach (e.g., in CFA where all cross-loadings are given small-variance priors

or where all residual covariances are specified; Bollen, 1989; also see Scheines, Hoijtink, & Boomsma, 1999). By replacing fixed-zero parameters relating to cross-loadings and residual covariances in ICM-CFA with small variance priors, the BSEM approach provides a more realistic model specification (see below for more discussion).

For Bayesian estimation, the most common algorithm is based on Markov chain Monte Carlo (MCMC) sampling (Kaplan, 2014). The general idea of MCMC is to draw specially constructed samples from the posterior distribution $p(\theta|Y)$ of the model parameters rather than attempting to analytically solve for the moments and quantiles of the posterior distribution. In the present study we use the Gibbs sampler (Geman & Geman, 1984) as implemented in Mplus (Muthén & Muthén, 2018). The Gibbs sampler begins with an initial set of starting values for the CFA model parameters: $\theta_{norm}^{(0)}, \theta_{IW}^{(0)}$. The Gibbs sampler then produces $\theta_{norm}^{(s+1)}, \theta_{IW}^{(s+1)}$ from $\theta_{norm}^{(s)}, \theta_{IW}^{(s)}$ as follows:

$$1.\ \text{Sample } \theta_{norm}^{(s+1)} \text{ from } p(\theta_{norm}|\theta_{IW}^{(s)}, y) \tag{8}$$

$$2.\ \text{Sample } \theta_{IW}^{(s+1)} \text{ from } p(\theta_{IW}|\theta_{norm}^{(s)}, y) \tag{9}$$

where $s = 1,2,...,S$ are the Monte Carlo iterations. The computational details can be found in (Asparouhov & Muthén, 2010). Using Gibbs sampling, the empirical distribution of the N_1 MCMC samples after N_0 burn-in iterations[2], denoted as $\{\mu^{(s)}, \Lambda^{(s)}, \Psi^{(s)}, \Theta^{(s)} | N_0 < s \le N_0 + N_1\}$, approximates the posterior distribution $p(\theta|Y)$ on which Bayesian estimates and inference are based. For example, the mean or mode of $p(\theta|Y)$ is often used as the Bayesian point estimate, and the percentiles of $p(\theta|Y)$ are used to form credible intervals.

BSEM with Informative Cross-Loadings Priors [BSEM-CL]. As mentioned above, the ICM-CFA model is based on the highly restrictive assumption that all cross-loadings are fixed to zero in Λ. In practice, most indicators present both a certain level of random noise as well as construct-relevant association with other constructs (see Asparouhov et al., 2015). In BSEM-CL, the cross-loadings are allowed to be estimated by determining cross-loadings priors. The proposed approach operates as follows: λ_{jk} is the element in the *j*th row and *k*th column of independent Gaussian prior distributions $N(\mu_{jk}, \sigma_{jk}^2)$ are assigned to λ_{jk} as

[2] Once the Markov chain has stabilized, the iterations prior to the stabilization (referred to as the “burn-in” phase) are discarded.

$$p(\Lambda|\Sigma_0, \mathrm{K}_0) \propto \exp\left\{\sum_{j=1}^{p}\sum_{k=1}^{q} -\frac{(\lambda_{jk}-\mu_{jk})^2}{2\sigma_{jk}^2}\right\}, \tag{10}$$

where μ_{jk} and σ_{jk}^2 are hyperparameter assumed to be known from prior knowledge, and Σ_0 and K_0 are matrices containing all μ_{jk} and σ_{jk}^2. The λ_{jk} s are further divided into two groups. The first group consists of *main (hypothesized) loadings* generally implemented in standard CFA as supported by substantive knowledge. The main loadings are given diffuse (non-informative) priors (i.e., $\mu_{jk} = 0$ with large σ_{jk}^2) to allow λ_{jk} to take on values that deviate substantially from zero. The second group comprises the remaining elements of λ_{jk} s that are fixed to zero in the ICM-CFA model. In the original BSEM-CL model proposed by Muthén and Asparouhov (2012), these strict constraints were replaced by "soft" constraints characterized by prior distribution of λ_{jk} with small variance σ_{jk}^2 (i.e., $\sigma_{jk}^2 = 0.01$) and $\mu_{jk} = 0$, which reflect the prior beliefs that these λ_{jk} have large prior probability near 0. This informative prior structure concentrates the posterior distributions for λ_{jk} around zero. However, if prior knowledge indicates that a large number of cross-loadings are positive, it may be more appropriate to use $\mu_{jk} > 0$ (e.g., $\mu_{jk} = 0.1$) with small variance σ_{jk}^2.

Hence, in terms of the parameter specification, BSEM-CL is similar to ESEM with target rotation, which allows researchers to have more a priori control on the expected factor structure. However, BSEM-CL enables researchers to specify a prior distribution for cross-loadings by varying the prior mean and variance and thus make stronger assumptions about the strength of the cross-loadings. Such specification is not readily available in an ESEM approach even with target rotation. To some extent, target rotation can be adjusted by specifying the target value according to a researcher's judgement, normally using zero target value for cross-loadings. Target rotation, however, does not allow user-specified stringency of closeness to zero. Therefore, BSEM-CL can be viewed to lie on a continuum between CFA and ESEM with target rotation (Muthén & Asparouhov, 2012). Using a combination of real and simulated data, Muthén and Asparouhov (2012) demonstrated that BSEM-CL is superior to ICM-CFA in terms of model fit and the coverage of parameters; although the changes in prior variance for cross-loadings may

affect factor correlations (but not main loadings), the influence was small and of little substantive importance. No study to our knowledge, however, has systematically compared BSEM-CL to ESEM.

BSEM with Informative Residual Covariances Priors [BSEM-RC]. Another important feature of BSEM is that all residual covariances among observed indicators can be freely estimated using informative priors. Zyphur and Oswald (2013) claimed that it is impossible to assume exactly zero covariance among residuals because the content between items could covary to some small extent beyond the trait being measured, even in a unidimensional scale. Basically, BSEM-CL and BSEM-RC follow the same idea, which is to explicitly model some otherwise unmodeled source of influence on the indicators in a measurement model (Asparouhov et al., 2015). While cross-loadings model the relationships between indicators and nontarget factors, residual covariances model shared sources of influence on the indicators that are unrelated to the factors, such as method effects (e.g., negatively worded or parallel worded items). The failure to include a priori correlated uniquenesses (CUs; the specific residual covariance between two observed indicators) can result in inflated factor correlations, biased parameter estimates, and even improper solutions such as a nonpositive definite Ψ (Marsh et al., 2010, 2013).

Given that freeing all residuals covariances Θ would lead to an unidentified model (Bollen, 1989), it is difficult to discern which residuals should covary in the likelihood-based framework. BSEM-RC provides a possible approach to this problem by applying an informative inverse-Wishart prior $IW(\mathbf{R}, df)$ on Θ. The means and covariance matrix of the inverse-Wishart distribution are a function of the elements r_{mn} on row m and column n from $\mathbf{R}$ (e.g., $p \times p$ scale matrix), with degrees of freedom df and number of variables p. The density of the inverse-Wishart distribution is

$$\frac{\left|\mathbf{R}^{df/2}\right|}{2^{dfp/2}\Gamma_p(df/2)}\left|\mathbf{X}\right|^{-(df+p+1)/2} e^{-tr(\mathbf{R}\mathbf{X}^{-1})/2} \tag{11}$$

where Γ_p and $tr()$ are the multivariate Gamma function and the trace function, respectively. The mean of the inverse-Wishart distribution is

$$E[\mathbf{X}] = \frac{\mathbf{R}}{df - p - 1} \tag{12}$$

and the variance of each element of the inverse-Wishart distribution is

$$Var[x_{mn}] = \frac{(df - p + 1)r_{mn}^2 + (df - p - 1)r_{mm}r_{nn}}{(df - p)(df - p - 1)^2(df - p - 3)} \tag{13}$$

where the elements r_{mm} are on row m and column m from $\mathbf{R}$ and r_{nn} are on row n and column n from $\mathbf{R}$. The variances for the diagonal elements of the inverse-Wishart distribution simplify to

$$Var[x_{mm}] = \frac{2r_{mm}^2}{(df - p - 1)^2(df - p - 3)} \tag{14}$$

Equation (13) indicates that when df increases, the denominator will increase more rapidly than the numerator, and thus the variance will become smaller. It implies that the larger the value of df, the more informative the prior is. Equation (13) also indicates the size of variance is partially determined by $\mathbf{R}$:the smaller the elements of $\mathbf{R}$, the smaller the variance, and thus the more informative the prior is. Nevertheless, setting the scale to large values also impacts the position of the inverse-Wishart distribution in parameter space (see Equation 12). Hence, specifying an inverse-Wishart distribution requests balance the df and the size of $\mathbf{R}$. In practice, a typically used informative inverse-Wishart prior is an identity matrix $\mathbf{R} = \boldsymbol{I}$ with varying df. Note that to obtain a proper posterior where the marginal mean and variance are defined, df should be greater than $p+3$. For example, following the specification strategy recommended by Muthén and Asparouhov (2012), an identity matrix $\mathbf{R} = \boldsymbol{I}$ and $df = p+6$ for θ_{IW} gives prior means of zero and variance of roughly 0.01 for residual covariances (see p. 335 in Muthén & Asparouhov, 2012 for a detailed description). Note that the specification of the priors in BSEM depends on the scale of the observed variables and that the guidelines by Muthén and Asparouhov assume that the variables have a *SD* close to one (see Muthén & Asparouhov, 2012, p. 316, for a discussion).

Using an inverse-Wishart prior specification was shown to outperform other prior specification approaches for residual covariances in terms of good convergence and coverage for main loadings and correlations in the simulation study (see Asparouhov & Muthén, 2010, Muthén & Asparouhov, 2012 for more discussion). However, it is not feasible to specify priors to the specific residual covariance elements (e.g., freely estimated single correlated uniqueness) using an inverse-Wishart prior because the inverse-Wishart distribution assumes a prior for the whole covariance matrix and does not allow to modify single entries of the matrix. Specifically, the parameter df in the inverse-Wishart distribution (see Equations 7, 11-14) is equal for all parameters in the same inverse

Wishart prior block.

The present study uses the inverse-Wishart prior method to apply informative priors to residual covariances. While both BSEM-CL and BSEM-RC involve adding to the model a set of potentially misspecified parameters with small priors, BSEM-RC requires heavier computations because of larger numbers of estimated parameters and slow MCMC convergence (Muthén and Asparouhov, 2012; Muthén et al., 2015). Although very few studies have directly compared these two approaches, it is expected that BSEM-RC provides a better model fit given more free parameters (see below for more discussion). However, the bias and coverage of estimated parameters (e.g., main loadings, factor correlations) between these two approaches needs further study.

BSEM with Informative Cross-loadings and Residual Covariances Priors [BSEM-CLRC]. The BSEM technique also allows for simultaneous inclusion of informative, normal priors for all cross-loadings and inverse-Wishart priors for residual covariances (i.e., BSEM-CLRC; Muthén & Asparouhov, 2012). It takes into account the presence of trivial cross-loadings in the CFA model and many minor residual covariances among the observed indicators. Recent empirical studies have shown that, compared to BSEM-CL, BSEM-CLRC provides a better model fit given that large numbers of fixed parameters are converted to free parameters (e.g., Fong & Ho, 2013; Stromeyer et al., 2015). Again, little is known about which approach leads to more accurate parameter estimates in relation main loadings, cross-loadings, and factor correlations.

Despite improved model fit in BSEM-RC and BSEM-CLRC, the complexity in model specifications has received growing concerns. For example, Stromeyer et al. (2015) argued that the relaxing of restrictions on all residual covariances might result in perfect model fit even in the presence of fundamental model misspecifications. MacCallum et al. (2012) also expressed similar concerns in that the complexity of BSEM models with freely estimated residual covariances results in an increase in estimation error, which in turn might diminish stability and generalizability of the solution. In other words, improved fit is obtained at the expense of modeling idiosyncratic sample characteristics that are unlikely to generalize in subsequent samples (Myung, 2000; Zucchini, 2000). These concerns emphasize the importance of cross-validation in evaluating BSEM-RC and BESM-CLRC models. In a recent empirical study, Asparouhov et al., (2015) cross-validated the BSEM-CLRC solution between two independent samples and found strong support for measurement invariance where all parameters are held equal across samples. However, the measurement model was employed in Asparouhov et al.'s (2015) study is relatively

simple, containing only 17 observed indicators and 5 latent factors. In the present study, we expand this approach and compare and cross-validate different BSEM models (i.e., BSEM-CL, BSEM-RC, BSEM-CLRC) using a more complex factor structure (60 observed indicators and 5 latent factors) with longitudinal and *k*-fold cross-validation approaches using empirical data.

Applications of ESEM and BSEM Studies in the Literature

In the last decade, ESEM has been increasingly used in clinical and applied psychological research (see Marsh et al., 2014 for a review). It has been extended to evaluate longitudinal and multi-group measurement invariance tests, differential item functioning, and relations between latent variables with other constructs (Marsh et al., 2009, 2010, 2013, 2014). However, BSEM (Muthén & Asparouhov, 2012) has also recently garnered interest in psychological research (Van de Schoot et al., 2017). Nevertheless, researchers have used this technique in many alternative, potentially inconsistent ways because of the great flexibility in model specifications in BSEM. We reviewed recent studies utilizing BESM approaches for factor analyses (Table 1) based on different approaches to setting informative priors. However, only two of these are simulation studies in which BSEM estimates can be compared with known population values. Unfortunately, neither of them has directly compared BSEM models with different subsets of informative priors, based on which it is difficult to give practical guidelines for researchers to apply different BSEM optimal strategies and estimation procedures when developing a measurement model. Given that ESEM and BSEM (particularly BSEM-CL) adhere to similar logic (see above), the main purpose in this article is to systematically evaluate and compare ESEM and BSEM models with different subsets of informative priors based on simulated and real data and derive constructive and practical guidelines for applied researchers.

The Present Investigation

The purpose of the present investigation is to evaluate and compare ESEM and BSEM approaches designed to resolve the dilemmas between EFA and CFA. To achieve this goal, we conducted two studies: 1) in a simulation study, we evaluated the appropriateness of ESEM and BSEM (BSEM-CL, BSEM-RC, BSEM-CLRC) models in relation to known population parameters under a variety of different conditions including varying specifications of target rotations in ESEM and of informative priors in BSEM; 2) in an empirical study with real data, we compared and cross-validated different models based on the most widely used Big-Five personality instrument (12 items for each factor; Costa & McCrae, 1992).

Based on a limited amount of BSEM research, we test the following a priori hypotheses across two studies:

Hypothesis 1(H1): Model fit: We hypothesize that BSEM-RC and BSEM-CLRC fit the data better (e.g., having low model rejection rate) than BSEM-CL and ESEM given a large additional number of freely estimated parameters.

Hypothesis 2(H2): Close performance between ESEM and BSEM-CL. We anticipate that ESEM will perform more closely to BSEM-CL than BSEM-RC and BSEM-CLRC in terms of model fit, bias, coverage, and power in estimation of major loadings and factor correlations as they function on a similar logic.

Research question 1(Q1): Comparison between ESEM and different BSEM models. We leave as a research question which model (ESEM vs. different BSEM models) is superior in accurately estimating parameters, particularly when the model specifications were substantially manipulated, such as varying the number, location, and size of the targeted values in ESEM and the distribution of informative priors in different BESM models.

Research question 2(Q2): comparison between simulation and real data results. Given that factor structure is usually more complex in reality than that in simulation, we leave open the question as to the consistency of results between simulation and real data.

Study 1: Stimulation Study

In this simulation study, two factor loading structures were used. In order to enhance comparability, the first loading structure, based on the simulation design that Muthén and Asparouhov (2012) used to introduce BSEM, addresses several critical issues left unanswered by the Muthén and Asparouhov demonstration. Compared to the first loading structure, the second and the third were more complex (with multiple major cross-loadings for each factor instead of just one). Thus, the three simulation designs allow us to closely compare CFA, ESEM, and different BSEMs with different subsets of priors, and evaluate results in relation to a priori hypotheses.

Method

Data generation. On the basis of the Muthén and Asparouhov (2012) simulation, we generated data using three latent factor models with five indicator variables for each factor. The first structure of the loading pattern (Design 1) in Table 2 is considered where A denotes a main loading, B denotes major cross-loadings, and C denotes minor cross-loadings. In Design 1, one major cross-loading and nine minor cross-loadings were

incorporated for each factor—a total of three major cross-loadings and 27 minor cross-loading across the three factors. The simulation design factors manipulated for the first loading structure included: (a) the sizes of the three major cross-loadings (0.1, 0.2, and 0.3) that are considered to be of little importance, some importance, and importance respectively (Cudeck & O'Dell, 1994); (b) sample size (N = 200, 500, and 1000); and (c) approaches (CFA, ESEM, BSEM-RC, BSEM-RC, and BSEM-CLRC). In total, Design 1 resulted in 45 conditions. The other parameters were set such that: the main loadings were all 0.8, the minor cross-loadings were all 0.01, the correlations among the three factors were all 0.5, and the residual variances of indicator variables were all 0.5. The factor metric is determined by fixing the variances of each factor at 1.

In Design 2, four major cross-loadings (0.1, 0.2, 0.3, and 0.4) and six minor cross-loadings (i.e., 0.01) were incorporated for each factor (see Table 2)—a total of 12 major cross-loadings across the three factors. Note that all cross-loadings were positive (i.e., the sum/average of the sizes of the cross-loadings for each factor = 1.06/0.106), which results in an unbalanced (positively oriented) factor structure. A balanced factor structure (i.e., Design 3) was then investigated where four major cross-loadings for each factor were set to -0.1, 0.2, 0.3, -0.4, respectively and six minor cross-loadings were set to a combination of 0.01 and -0.01 (see Table 2). Hence, Design 3 led to a completely balanced factor structure (i.e., the sum of the sizes of the cross-loadings for each factor = 0). In Designs 2 and 3, the model specifications were substantially manipulated for both ESEM and BSEM models, resulting in 14 model designs (see below for more details) coupled with three sample sizes (N = 200, 500, and 1000). In total, 42 conditions were tested for each design (2 and 3). The other parameters were defined as same as those in Design 1. A total of 500 replications were used in both simulation designs.

ESEM specification. The ESEM models were estimated based on oblique target rotation (Asparouhov & Muthén, 2009; Browne, 2001). According to the most common specification of target rotation, all cross-loadings were "targeted" to be close to zero by setting the target value to be 0, while all of the main loadings were freely estimated in standard ESEM models used in simulation and real data studies. In Designs 2 and 3 of the simulation study, the specification of target rotation was varied in terms of the number and location of targets as well as the size of the targeted values in relation to matrices A and B in Figure 1. Specifically, the number of targets were manipulated by freely estimating 2 major or 2 minor cross-loadings; the location of targets were manipulated by only targeting on minor cross-loadings (i.e., freely estimating 4 major cross-loadings); and the size of target values

were manipulated by specifying to 0.1 (rather than zero; see Appendix 1 for the specified targeted pattern matrix). In addition, ESEM with a mechanical rotation criterion (i.e., geomin[3]) was added and compared with target rotation. In total, six ESEM models were symmetrically evaluated in Designs 2 and 3.

It should be noted that in ESEM "the order of the latent factors is interchangeable and each factor is interchangeable with its negative" (p. 436, Asparouhov & Muthén, 2009); these indeterminacies (i.e., the order and sign pattern) are particularly important in simulation studies (Asparouhov & Muthén, 2009). Without evaluating and correcting the order and sign pattern for each replication, the results would be biased in relation to parameter bias, mean square error, and coverage in simulation studies (Myers, Ahn, Lu, Celimli, & Zopluoglu, 2017). As such, we carefully reviewed all ESEM solutions and corrected (i.e., reordered or re-signed) parameter estimates for each replication so that all replications uniformly aligned with the pattern defined by the population values.

Choice of Priors in BSEM. The posterior distribution of Bayesian estimation was approximated by using an MCMC algorithm with the Gibbs sampler method. Note that choices of the prior variance are associated with the scale of the observed variables. For example, a prior variance of 0.01 corresponds to a small loading for an observed variable with unit variance, but it corresponds to an even smaller loading for an observed variable with variance larger than one (Muthén and Asparouhov, 2012). For convenience, observed variables were standardized to establish a common scale. In BSEM-CL normal priors with mean zero and variance 0.01 were used for the cross-loadings' priors; and standard non-informative prior distributions were used for other parameters: main loadings ~N(0, infinity), residual variances ~ Inverse Gramma Γ^{-1}(-1,0), and intercepts ~N(0, infinity). We used *Mplus default* improper prior $IW(0, -q-1)$ for the latent factor covariance matrix (where q is the number of latent factors). This is a widely used diffuse prior and allows the variance parameters to be any nonnegative value from 0 to infinity and the covariance parameters to be any value from –infinity to +infinity (Depaoli & van de Schoot, 2017). It should also be noted that the informative priors are applied not only to the major cross-loadings used to generate the data, but to all minor cross-loadings in the analysis model to reflect a real-data analysis situation. In relation to residual covariance priors in BSEM-RC and BSEM-CLRC, the inverse-Wishart prior $IW(\mathbf{R}, df)$ with $\mathbf{R}$

[3] In geomin rotation, the constant ε was set to .05 which has been widely used in empirical studies (Marsh et al., 2009; 2010, 2014). A recent simulation study (Celimli, Myers, & Ahn, 2018) found that the geomin rotation with ε = .05 provided more stable but less accurate factor solutions than the default geomin rotation (where ε = .001 in Mplus) with very small effect size; the accuracy of factor solutions in geomin rotation with ε = .05 increased when the factors are more correlated.

= ***I*** and *df* = *p* + 6 (*p* = number of indictor variables) was used, corresponding to prior mean and variance for residual covariances of zero and 0.01 respectively (Gucciardi & Zyphur, 2016; MacKinnon, 2008; Muthén and Asparouhov, 2012). Table 3 lists the specific priors used in each BSEM approach (also see Supplemental Materials, Appendix 7 for the annotated Mplus syntax).

Variation of prior specification. In the simulation designs 2 and 3, we have further considered various informative priors for factor loadings and residual covariances, in addition to the standard priors setup ($M = 0$, $Var = 0.01$). First, the variance of cross-loadings priors in BSEM-CL and of residual covariances priors in BSEM-RC were varied to 0.02 and 0.005. Second, informative priors with mean 0.1(rather than zero) for cross-loadings were implemented in BSEM-CL to reflect in situations where researchers have a priori information (based on theory or prior research) indicating that the cross-loadings are likely to be positive. In total, eight ESEM models were symmetrically evaluated in designs 2 and 3.

Model fit in BSEM. Convergence of BSEM models is evaluated by the potential scale reduction (PSR; Asparouhov & Muthén, 2010). PSR is the ratio of total variance across chains and pooled variance within a chain. A PSR value of 1.00 represents perfect convergence (Muthén & Muthén, 1998–2015; Kaplan & Depaoli, 2012). With a large number of parameters, a PSR < 1.10 for each parameter indicates that the convergence of the MCMC sequence is obtained (Muthén & Muthén, 1998-2015; Gelman, Carlin, Stern, & Rubin, 2004). In this study, we used PSR < 1.05 as an appropriate convergence criterion (Zyphur & Oswald, 2013). For each replication, BSEM models were estimated with 10,000 MCMC iterations with two Markov Chains in Mplus (Muthén & Muthén, 1998-2017), on which PSRs were assured to be < 1.05 (see Tables 4&5). We report the model rejection rate that is computed as the proportion of replications (in each condition of the simulation design) with a Bayesian posterior predictive *p* (PP *p*) value for BSEM models (or maximum likelihood [ML] *p* value for CFA and ESEM) of smaller than 0.05. Small PP *p* values (i.e., <.05) indicate poor model fit because this means that the observed data rarely fit better than generated data (e.g., < 5% of the time). We also reported another two indices for comparing Bayesian models: deviance information criterion (DIC) and Bayesian information criteria (BIC; Muthén, 2010). Smaller BIC and DIC values indicate better models, and models can be compared using the DIC even when they are not nested (Zyphur & Oswald, 2013). DIC is preferable to BIC when sample sizes are large, coupled with a large number of observed indicators (Asparouhov et al., 2015).

To provide a comprehensive evaluation of different ML and Bayesian approaches, we considered a variety of measures of accuracy and precision. We reported the mean and *SD* of relative bias (difference between the estimated and the true value divided by the true value) for main loadings and factor correlations across the 500 replicates. Generally, a relative bias less than 5% could be considered negligible, and less than 10% could be acceptable. In addition, we reported the 95% coverage that refers to the proportion of the replications for which the 95% Bayesian credibility interval covers the true parameter values used to generate data in BSEM models. We also reported the corresponding 95% coverage for ESEM models using a ML bootstrap confidence interval. Specifically, we drew 500 bootstrap samples from each replication to estimate a confidence interval. It is also interesting to study what corresponds to power in a frequentist setting for a Bayes setting, particularly with respect to major cross-loadings. For Bayes, power is computed as the proportion of the replications for which the 95% Bayesian credibility interval (or the ML bootstrap confidence interval) does not cover zero (see *Power* in Tables 6, 8, and 9, also see Appendices 2&3 in Supplemental Materials for the summary of cross-loadings).

Results: Design 1

The ML 5% rejection rate for ESEM was appropriately small (5%-9%), whereas the nominal 5% rejection rate of the Bayes PP p value was close to zero (BSEM-CL, < 1%), or actually zero (BSEM-RC and BSEM-CLRC). Tables 6 reports the average relative bias of parameters (main loadings and factor correlations) across 45 conditions. To provide a comprehensive evaluation of the impact of different conditions on simulation results (i.e., models, sample sizes, sizes of major cross-loading), an ANOVA with the conditions of the simulation design as factors was employed (see Table 7). Results showed that the variances of relative bias across conditions for main loadings were largely explained by different models (R-square 89.1%) rather than sizes of major cross-loadings (R-square 5%) and sample sizes (R-square 1%). The sizes of relative bias in relation to main loadings across different models were all acceptable (0.6% to 10.5%). Even though ESEM resulted in the smallest relative bias of main loadings followed by BSEM-CL, all models have similar and small *SD*s of relative bias.

For factor correlations, the sizes of bias varied for the different models (R-square 54.7%) and size of major cross-loadings (R-square 37.5%) but not sample sizes (R-square 0.4%). When sizes of major cross-loadings were small and moderate (i.e., 0.1 and 0.2), BSEM-RC and BSEM-CLRC showed slightly smaller relative bias than

ESEM and BSEM-CL. These differences disappeared when the major cross-loading was 0.3. Note that overall the differences in the *SD* of relative bias among different models were relatively small (< 0.8%).

Of particular relevance to the present investigation, we compared coverage and power results across different conditions (see Table 5). The differences in 95% coverage for main loadings and factor correlations were largely explained by different models (R-square 78.6% and 65.1%, respectively). ESEM, BSEM-CL, and BSEM-CLRC had similar and good coverage (> .900), whereas BSEM-RC resulted in relatively low coverage when major cross-loadings and sample sizes were large. In addition, all models showed excellent power to detect main loadings and factor correlations across different sample sizes.

Results: Design 2

In Design 2, a more complex and unbalanced (positively oriented) factor structure was utilized, in which multiple positive major cross-loadings instead of one were incorporated for each factor. We started with the standard comparison among the models (ESEM, BSEM-CL, BSEM-RC, and BSEM-CLRC) evaluated in Design 1, and then compared different variation of model specification in relation to BSEM and ESEM. In total, 14 models (6 ESEM models and 8 BSEM models) with three different sample sizes were evaluated in Design 2.

Similar to Design 1, BSEM-CL showed slightly lower rejection rate than ESEM[4] (.022 to .064 for BSEM-CL and .056 to.100 for ESEM). BSEM-RC and BSEM-CLRC showed a zero rejection rate in terms of Bayesian *p* (PP *p*) value. The relative bias and its *SD* and coverage for main loadings and factor correlations across conditions were largely explained by different models (R-square 58.8% to 99.0%, see Table 7). As seen in Tables 8 and 9, ESEM consistently showed the smaller relative bias in estimation of main loadings (4.3% to 5.7%) and factor correlations (19.3% to 22.9%), and better coverage than different BSEM models. Even though similar sizes of relative biases for main loadings (12.1% to 18.4%) were found across different BSEM models, BSEM-RC showed slightly larger *SD* of bias (15.7% to 17.4%) and lower coverage (.445 to .548) in main loadings than other BSEM models. Relative bias and coverage for factor correlations were substantially large across different BSEM models

[4] We also compared BIC between ESEM and BSEM-CL and found that ESEM had consistently smaller BIC than BSEM-CL to a small extent (diff= 53 to 64 to across different sample sizes). In addition, given that the DIC was developed as the Bayesian counterpart of AIC in frequentist analysis, we compared AIC in ESEM with DIC in BSEM-CL and found that the differences were tiny (diff= 7 to 12 to across different sample sizes). Even though these fit indices are fairly good approximations for model comparisons between ML and Bayesian estimation, there is no full simulation studies that have confirmed that. Hence, the small differences in these fit indices should be treated as inconclusive (see http://www.statmodel.com/cgi-bin/discus/show.cgi?9/6256 for further discussion in Mplus discussion forum). Also see below for model fit comparison among different BSEM models (Table 5).

(36.4% to 40.6% and .000 to .275, respectively), although all models showed 100% power to detect factor correlations.

ESEM with different specifications. Note that changing the number and location of targets, sizes of target values, and the rotation method in ESEM resulted in identical model fit. As seen in Tables 8 and 9, ESEM with geomin rotation showed small and negative relative bias in estimation of main loadings (-3.2% to -2.8%) and factor correlations (-16.7% to -17.9%) with small *SD* and good coverage and power. Overall, the size of these measures in geomin rotation were quite similar with that in target rotation. When the number of targets were manipulated, freely estimating minor cross-loadings resulted in larger relative bias and lower coverage, with the reverse being true for freely estimating major cross-loadings (see Figure 2). Particularly, when all major cross-loadings were freely estimated (i.e., only minor cross-loadings were targeted), ESEM showed the smallest relative bias and coverage. Finally, when the targeted values were changed to .1, the ESEM model resulted in substantively smaller relative bias (-.6% to 8.2%) than the typical ESEM model where targeted values were set to zero.

BSEM with different specifications of priors. To better evaluate the influence of prior specifications on BSEM model solutions, we also included BIC and DIC in addition to rejection rate for model fit comparisons (see Table 5). Whereas BSEM-RC and BSEM-CLRC had smaller (zero) rejection rate than BSEM-CL, BSEM-CL provided smaller BIC and DIC than BSEM-RC ($\Delta \approx 350$ and $\Delta \approx 50$, respectively) and BSEM-CLRC ($\Delta \approx 650$ and $\Delta \approx 50$, respectively). When the variances of cross-loadings priors in BSEM-CL and residual covariances priors in BSEM-RC were set to 0.05 and 0.02 (instead of 0.01), the BSEM models remained highly similar model fit and estimation solutions (see Tables 8&9 and Figure 3). However, when the mean of cross-loadings priors in BSEM-CL was set to 0.1, the BSEM solution improved substantially and resulted in very small relative bias and excellent coverage and power in estimation of main loadings and factor correlations, even though the model fit remained similar.

Results: Design 3

In simulation designs considered thus far, we have only considered positive cross-loadings, which resulted in an unbalanced and positive-oriented factor structure. In Design 3, we evaluated a balanced factor structure by incorporating both positive and negative cross-loadings (e.g., -0.1, 0.2, 0.3, -0.4, for major loadings). Again, we started with the standard comparison among the models (ESEM, BSEM-CL, BSEM-RC, and BSEM-CLRC), and

then compare different variation of model specifications in relation to BSEM and ESEM. In total,14 models (6 ESEM models and 8 BSEM models) with three different sample sizes were evaluated in Design 3.

Similar to Designs 2, BSEM-RC and BSEM-CLRC showed a zero rejection rate, followed by BSEM-CL (.006 to .010) and ESEM (.054 to .082). The relative bias and its *SD* and coverage for main loadings and factor correlations across conditions were largely explained by different models (R-square 64.5% to 98.7%, Table 7). As seen in Tables 8 and 9, small sizes of relative biases for main loadings were found across the ESEM (-2.9% to -2.3%) and BSEM-CL (1.9% to 2.2%) and BSEM-CLRC (1.8% to 3.2%), whereas the bias of factor correlations in ESEM (-18.6% to -16.2%) was much larger than BSEM-CL (3.9% to 4.0%) and BSEM-CLRC (2.6% to 4.7%). ESEM, BSEM-CL, and BSEM-RC showed excellent coverage and power for main loadings and factor correlations. Although most of BSEM-RC resulted in acceptable size of relative bias for main loadings (-4.5% to -6.6%) and factor correlations (-4.7% to 19.1%), particularly when sample sizes were small, the *SD* of relative bias for main loading (25.7% to 28.0%) and factor correlations (12.0% to 44.4%) in BSEM-RC was much larger than that in other models. It is also evident that BSEM-RC resulted in lower coverage and power in detecting main loadings and factor correlations.

ESEM with different specifications. ESEM with geomin rotation showed smaller relative bias (-6.4% to -6.0%) but lower coverage (.623 to .883) in estimation of main loadings than that with target rotation to a small extent. However, geomin rotation resulted in much larger relative bias for factor correlations (-47.7% to -47.0%) with smaller *SD* and lower coverage than target rotation (see Figure 2). Again, freely estimating minor cross-loadings resulted in larger relative bias and lower coverage, with the reverse being true for freely estimating major cross-loadings. When the targeted values were changed to .1, the ESEM model resulted in much larger relative bias -63.7% to -56.6%) than the typical ESEM model where targeted values were set to zero.

BSEM with different specifications of priors. Similar to Design 2, BSEM-CL provided smaller BIC and DIC than BSEM-RC (Δ≈400 and Δ≈50, respectively) and BSEM-CLRC (Δ≈600 and Δ≈50, respectively), although BSEM-RC and BSEM-CLRC showed a smaller rejection rate than BSEM-CL. Again, changing variances of informative priors on cross-loadings in BSEM-CL and residual covariances in BSEM-RC led to similar model fit and estimation solutions (See Tables 8&9 and Figure 3). However, when the mean of cross-loadings priors in

BSEM-CL was changed to 0.1, the relative biases for main loadings and factor correlations substantially increased (-47% to -46.6%) with larger *SD* and lower coverage and power, even though the model fit remained similar.

Summary of the Three Stimulation Designs.

Given that in factor analyses applied researchers usually start with CFA that are likely to be misspecified in reality, we also evaluated CFA in Design 1. Results revealed that CFA was ill fitting and resulted in the worst model fit and the largest bias in estimating factor correlations (see Appendix 2).

In relation to model fit, BSEM-RC and BSEM-CLRC consistently showed lower rejection rates than BSEM-CL given an enormous increase in the number of free parameters, whereas BSEM-CL showed lower DIC to a very small extent. Although the differences in BIC favored by BSEM-CL (low value is preferred) were much larger than those in DIC, this finding should be interpreted cautiously because BIC unnecessarily penalizes the BSEM model by counting small-variance prior parameters as actual parameters and thereby overshadows information provided by BSEM (Asparouhov et al, 2015). Overall, relaxing the restrictions on either cross-loadings or residual covariances (or both) in BSEM did not lead to large differences in model fit. Compared to BSEM models, ESEM resulted in lower rejection rate but is somewhat closer to BSEM-CL (also see more discussion in the footnote 4).

For the estimation of main loadings and factor correlations, the pattern of results was substantially varied by the factor structures. ESEM resulted in more accurate parameter estimates in main loadings and factor correlations than different BSEM models in most of cases in designs 1 and 2 where only positive cross-loadings were implemented. This advantage was stronger particularly when the stimulated factor structure was complex (Design 2) and the sample size was small. When both positive and negative cross-loading were introduced in a balanced factor structure (Design 3), the BSEM-CL and BSEM-CLRC provided more accurate estimation in factor correlations than ESEM, whereas these three models resulted in small bias, and its *SD* as well as good coverage and power for main loadings. In terms of the direction of bias, BSEM-CL and BSEM-CLRC tend to result in more positive bias in estimating main loadings and factor correlations than ESEM, particularly when the factor structure was unbalanced. BSEM-CL provided unstable estimation solutions in terms of larger *SD* bias, lower coverage, and less power than other models, although sometimes the sizes of relative bias in BSEM-CL were acceptable.

In relation to different model specification, specifying mean of cross-loadings to 0.1 in BSEM-CL and target value to 0.1 in ESEM performed much better than the ESEM and BSEM models (where mean and target values were zero, respectively) in the unbalanced factor structure, with the reverse being true in the balanced factor structure (see Figure 2). However, change variances of cross-loadings and residual covariances in BSEMs led to similar results. Target rotation was, superior to geomin rotation in estimating factor correlations when the factor structure was balanced; however, Geomin rotation produced more stable factor solutions. For ESEM with target rotation, changing the number of targets by freely estimating major cross-loadings improved the model solutions. In contrast, freely estimating minor cross-loadings led to more biased results. In a typical case, when only minor cross-loadings were targeted, ESEM provided almost accurate estimates parameters. This pattern of results consistently showed in both designs 2 and 3.

Study 2: Real data - NEO-FFI Big-Five Personality Example

An empirical example used data from a large German study (Transformation of the Secondary School System and Academic Careers [TOSCA]; Trautwein, Neumann, Nagy, Lüdtke, & Maaz, 2010; Marsh et al., 2010). The Big-Five personality factors (Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness) were measured by using the German version of the NEO-FFI (Borkenau & Ostendorf, 1993), where 12 items were used to measure each of the five factors. Using these data, Marsh et al. (2010) applied ESEM to demonstrate that the a priori scales showed a well-defined five-factor solution and that ESEM resulted in substantially more differentiated (less correlated) factors than did CFA. They also defined an a priori set of correlated uniquenesses (CUs, correlated errors between indicators) inherent to the design of the NEO-FFI (see below). Their results provided apparently the first acceptable fit to the Big-Five factor structure based on the 60 NEO-FFI items, and was used to counter suggestions that factor analysis might not be an appropriate tool in personality research. Their study was also one of the strongest demonstrations of the usefulness of ESEM in applied research. Hence, these data provide an ideal setting for comparing CFA, ESEM, and BSEM with different (cross-loadings or residual covariances) informative priors.

Method

Data. The 60-item NEO-FFI (Costa & McCrae, 1992) provides a short measure of the Big-Five personality factors. For each factor, 12 items from the 180 items of the longer NEO-PI (and the full 240-item NEO-PI-R;

McCrae & Costa, 1989) were selected. The NEO-FFI responses by late-adolescent Germans showed high reliability, validity, and comparability with responses of the original English-language version (Borkenau & Ostendorf, 1993; Trautwein et al., 2010). Two waves of data were used in this study. At Wave 1 , the students (N = 3,390; 45% men, 55% women) were in their final year of upper secondary schooling; T2 was assessed (N = 1,570, 39% men, 61% women) 2 years after graduation from high school. Marsh et al. (2010) revealed that sample attrition effects were statistically significant in some domains, but the effect sizes were small and indicative of only small selectivity effects. Coefficient alpha reliabilities of the five factors at Wave 1 and Wave 2 were acceptable (.72 - .87, also see Marsh et al., 2010).

A priori CUs. In the full NEO-PI-R (with 240 items), each of the Big-Five factors is represented by six facets, and each facet is represented by multiple items (see McCrae & Costa, 2004). However, in the 60-item NEO-FFI, all items were selected to best represent each of the Big-Five factors without reference to the facets. Marsh et al. (2010) posited that the correlations between items that came from the same facet of a specific Big-Five factor would be higher than those between items that came from different facets of the same Big-Five factor—beyond correlations that could be explained in terms of the common Big-Five factor that they represented. They found that test–retest factor correlations were substantially inflated and might result in improper solutions due to the failure to include CUs relating each pair of items from the same facet. In total, an a priori set of 57 CUs were included in this study.

Priors Choice. In line with the simulation study, normal priors with mean zero and variance 0.01 were used for cross-loadings priors. In BSEM-CL, the a priori CUs were freely estimated by using noninformative (diffuse) normal priors with mean zero and variance 1000 (hereafter BSEM-CL+CUs). In BSEM-RC and BSEM-CLRC, the inverse-Wishart prior $IW(\mathbf{R}, df)$ with $\mathbf{R} = \boldsymbol{I}$ and $df = 66$ (60[number of indictor variables] + 6) was used for residual covariances, corresponding to mean zero and SD roughly 0.1, respectively (Muthén & Asparouhov, 2012). Due to high auto-correlation among the MCMC iterations, only every 10th iteration was used with a total of 100,000 iterations to describe the posterior distribution (Muthén & Asparouhov, 2012).

Goodness of fit. We evaluated a number of traditional indices (Marsh, Hau, & Grayson, 2005): the comparative fit index (CFI), the root- mean-square error of approximation (RMSEA), and the Tucker- Lewis Index (TLI). Values greater than .95 and .90 for CFI and TLI typically indicate excellent and acceptable levels of fit to

the data. RMSEA values of less than .06 and .08 are considered to reflect good and acceptable levels of fit to the data. Apart from posterior predictive *p* values and ML likelihood-ratio chi-square values, we also reported another two indices for comparing Bayesian models: deviance information criterion (DIC) and Bayesian information criteria (BIC; Muthén, 2010). Smaller BIC and DIC values indicate better models, and models can be compared using the DIC even when they are not nested (Zyphur & Oswald, 2013). DIC is preferable to BIC when sample sizes are large, coupled with a large number of observed indicators (Asparouhov et al, 2015).

Cross-validation and External validity. The cross-validation is important for BSEM approaches, particularly for BSEM with residual covariance priors (BSEM-RC and BSEM-CLRC) because a large number of freely estimated parameters can lead to over-fitting. We compared model fit and parameter estimates (factor loadings, CUs, and factor correlations) across different approaches based on Wave 1 data and then cross-validated the parameter estimates using Wave 2 data. More specifically, we sampled the Wave 2 data with replacement 10,000 times and then computed Root Mean Square Residual (RMSR) by comparing the variance-covariance matrix of indicator variables at Wave 2 and the estimated (implied) variance-covariance matrix based on different models at Wave 1. In addition, we calculated RMSEA by employing all parameter estimates from the Wave 1 solution as fixed parameters to estimate the same model based on Wave 2 data. Based on the same logic, we also cross-validated the Wave 2 parameter estimates to Wave 1 data.

Each model (e.g., CFA, ESEM) was estimated five times to different partitions of the data (80% of the data each time), the results (i.e., parameter estimates as fixed values) were applied to the remaining 20% of the sample (Grimm, Mazza, & Davoudzadeh, 2016). Also, we provide another set of cross-validation analysis by estimating each model five times to different partitions of the data (20% of the data each time) and applying the results to the remaining 80% of the sample. RMSEA was reported for five-fold cross-validation results. Finally, we tested the construct validity of the big-five factors in relation to external criteria (e.g., life-satisfaction, positive/negative affect) across different estimation procedures.

Results

Model fit. Consistent with previous studies (e.g., Marsh et al., 2010), the CFA solution did not provide an acceptable fit to the data (e.g., CFI = .684, see Table 10). The next CFA model incorporated a priori CUs; results were still inadequate, albeit improved (e.g., CFI = .750). The corresponding ESEM solution fit the data much

better. Although the fit of the ESEM without a priori CUs was still not acceptable (e.g., CFI = .850), the inclusion of a priori CUs fitted the data reasonably well (e.g., CFI = .912). Aligned with CFA and ESEM, the BSEM with cross-loadings priors with a priori CUs (BSEM-CL +CUs) resulted in much better fit to the data, compared to that with no a priori CUs (ΔDIC = 3069; ΔBIC = 2720). However, the low PP *p* values ($p < .05$) indicated poor fit for both BSEM-CL models. When residual covariance priors were incorporated, the fit of the BSEM models (BSEM-RC and BSEM-CLRC) improved substantially but had many more freely estimated parameters compared to BSEM-CL+CUs (Δ number of parameters = 1476). Although more parameters were freely estimated in BSEM-CLRC than in BSEM-RC (Δparameters = 240), both models resulted in similar model fit. Given that the models with no a priori CUs provided relatively poor model fit, the subsequent analyses focused on the CFA, ESEM and BSEM models with the a priori CUs.

Factor loadings. To enhance interpretability, the items measuring Neuroticism were reversed coded to represent a measure of Emotional Stability. For main loadings, different approaches resulted in similar and substantial sizes of loadings except for BSEM-RC where the size of factor loadings was slightly smaller (see Figure 4, also see Appendix 4 for the Mean, *SD*, and Range of loadings for the Big-Five factors across models). Next, we compared cross-loadings across ESEM+CUs, BSEM-CL+CUs, and BSEM-CLRC. While sizes of cross-loadings were substantially smaller than those of main loadings in the three models, BSEM-CLRC resulted in slightly smaller cross-loadings than ESEM+CUs and BSEM-CL+CUs.

CUs. Consistent with previous studies (e.g., Marsh et al. 2010), the a priori CUs were significant across different estimation procedures (*M* = from .087 to .168; see Appendix 5). In the model where the full set of CUs was considered (i.e., BSEM-RC and BSEM-CLRC), the sizes of the a priori CUs were substantially larger than those of other CUs (Δ*M* =.116 and .087, respectively).

Correlations. All factor correlations were statistically significant in CFA and ESEM solutions, but the coefficients in CFA+CUs (Mean[*M*] = .184, *SD* = .191) were systematically higher than those in ESEM+CUs (*M* = .095, *SD* = .114, Table 11). Although BSEM-CL+CUs and BSEM-CLRC had a similar pattern of correlations with ESEM+CUs, the correlations in BSEM-CL+CUs and BSEM-CLRC were slightly higher (*M* = .112, *SD* = .160; *M* = .137, *SD* = .168, respectively). BSEM-RC resulted in the much lower correlations, and most of these were close to zero and less than 0.1 in absolute value (*M* = .009, *SD* = .093). The correlations in BSEM-CL+CUs

showed low degree of uncertainty with a small posterior *SD*. However, the posterior *SD*s were considerably larger in BSEM-CL+CUs and BSEM-CLRC. Thus, a majority of correlation coefficients were insignificant in the sense of their 95% posterior distribution credibility intervals covering zero.

Cross-validation. We cross-validated the findings across two waves of data and reported mean and 2.5% and 97.5% Quantiles of RMSR as well as RMSEA with 90% confidence interval (Table 12). Results showed that there was a stronger cross-validation support for the ESEM and BSEM models, compared to CFA. The support for ESEM+CUs and BSEM-CL+CUs was almost identical and slightly weaker than that for BSEM-RC and BSEM-CLRC when cross-validating the results from Wave 1 (N = 3,390) to Wave 2 (N = 1,750) data. However, these four models had similar cross-validation support when cross-validating from Wave 2 to Wave 1 data. Five-fold cross-validation analysis also showed that BSEM-RC and BSEM-CLRC cross-validated better than others when cross-validating the results from 80% (N = 2,711) to 20% (N = 679) data at Wave 1, whereas ESEM+CUs cross-validated best followed by BSEM-CL+CUs when cross-validating the results from 20% to 80% data (see Table 13). These findings indicated that the cross-validation results vary by sample sizes. When the sample size (of the training data) was large, BSEM-RC and BSEM-CLRC provided slightly more predictive accuracy than ESEM+CUs and BSEM-CL+CUs; the reverse was true when the sample size was small.

External validity. We evaluated the construct validity of the Big-Five constructs in relation to five external criteria (i.e., life-satisfaction, positive/negative affect, self-esteem, and Emotional Stability self-concept) across different estimation procedures (see Appendix 6). Specifically, consistent with prior research (e.g., Diener, Suh, Lucas, & Smith ,1999; McCrae & Costa, 1999), CFA+CUs, ESEM+CUs, BSEM-CL+CUs, and BSEM-CLRC showed that Emotional Stability was highly correlated with negative affect (rs = -.638, -.630, -.633, -.631, respectively) and Extraversion was substantially correlated with positive affect (rs = .570, .505, .524, .548, respectively). However, the sizes of corresponding correlations coefficients in BSEM-RC were significantly smaller (r = -.408 for Emotional Stability and negative affect, r =.296 for Extraversion and positive affect). Similarly, as expected (e.g., Asendorpf & vanAken, 2003; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006), high correlations of Emotional Stability with self-esteem and Emotional Stability self-concept were evident across models except for BSEM-RC (rs = .350 and .453, respectively). In total, CFA+CUs, ESEM+CUs, BSEM-CL+CUs, and BSEM-CLRC revealed similar and substantially higher correlation patterns between the Big-Five

factors and the five external criteria than BSEM-RC. This indicates that BSEM-RC results in much weaker support for the external validity of the big-five constructs than other models.

Comparisons between simulation and real data results. The pattern of results revealed in the simulation study in Study 1 was largely consistent with the findings based on real data. Firstly, CFA, even including a priori CUs, consistently had a poorer fit to the data than ESEM. Again, BSEM-RC and BSEM-CLRC again provided similar and better model fit than BSEM-CL. Secondly, BSEM-CL and BSEM-CLRC had slightly higher main loadings and factor correlations across the Big-Five factors than ESEM (consistent with the finding that they tended to result in more inflated estimated parameters than ESEM in the simulation study). However, the differences in estimation of factor loadings among ESEM, BSEM-CL, and BSEM-CLRC were smaller than the simulation results. BSEM-RC resulted in the lowest main loadings and factor correlations among different estimation procedures, which is not consistent with the simulation results where it tends to have positively biased estimated parameters. Importantly, BSEM-RC also provided weak support for the convergent validity of the big-five factors for external validity criteria.

Overall Discussion

This study evaluated CFA, ESEM and BSEM approaches based on two simulation designs and one real data example which covered different sample sizes and a variety of degrees of model misspecification (complexity) of the factor structure. Thus, the juxtaposition of simulation and real data studies provide insights into the performance of different estimation procedures. Table 14 summaries key findings of the present study and indicates whether these findings supported our expectations. The critical findings are discussed as follows.

Comparison ESEM with BSEM-CL

BSEM-CL and ESEM (with target rotation) work on a similar logic: taking into account unmodeled source of influence on the indicators through conversion from fixed-to-zero cross-loadings to approximately fixed-to-zero cross-loadings while having a priori control on the expected factor structure. In this regard, BSEM-CL performs more closely to ESEM than other BSEM models in terms of bias, *SD* of bias, coverage, and power, particularly in large sample sizes where the likelihood dominates the estimation of posteriors. Particularly, we found that changing targeted value to 0.1 in ESEM resulted in a similar pattern of results by changing mean of priors on

cross-loadings to 0.1 in BSEM-CL, indicating that targeted value in ESEM and mean of cross-loadings work in the similar way.

However, BSEM-CL differs from ESEM in two major ways. Firstly, BSEM-CL provides researchers with more control on cross-loadings by specifying different degrees of small variance priors (additional to small mean priors) and thus acts in a more confirmatory nature than ESEM (see below for further discussion). Secondly, in ESEM, the optimal rotation is determined only on the basis of the unrotated loadings as in EFA (Muthén & Asparouhov, 2012). This means that the effects of residual covariances are not considered in the optimal rotation. By contrast, the optimal rotation in BSEM-CL is determined by all parts of the model (Muthén & Asparouhov, 2012). In the empirical example, the inclusion of a priori CUs allows us to examine the influence of residual covariances on these two models. However, both models result in almost identical estimation of a priori CUs (most of them are statistically significant) and similar model solutions and cross-validation results.

BSEM with Different Subsets of Informative Priors

Muthén and Asparouhov (2012) proposed an alternative BSEM technique that designates subsets of parameters that are assigned informative priors. Small-variance priors can be assigned to different subsets (i.e., cross-loadings, residual covariances) or a combination of subsets of parameters in different models. One of the key aims of this article is to systematically evaluate BSEM models with different subsets of informative priors.

In the simulation study, BSEM-CL is well-fitting and further inclusion of residual covariance priors (i.e., BSEM-CLRC) only results in slightly better model fit. These residual covariance priors, however, appear to have small and negative effects on estimation of factor loadings – main loadings become more positively biased and cross-loadings become more negatively biased (see Appendices 2&3). However, these differences are quite small. Again, in the real data example both BSEM-CL+CUs and BSEM-CLRC perform very similarly, even though further inclusion of residual covariance priors (i.e., BSEM-CLRC) achieves a large improvement in model fit. This indicates that the BSEM-CL+CUs misfit is likely due to small and unimportant residual correlations and the main parameter estimates tend to remain unchanged.

BSEM-RC provide slightly more biased estimated parameters for main loadings and factor correlations than other BSEM models, the solution of BSEM-RC is much more unstable particularly when both positive and negative biases are included, evident by large *SD* of bias. Thus, it partially explains why the results in the

simulation study are substantially different from those in the real data example for BSEM-RC. Specifically, the factor loadings and correlations are considerably smaller in BSEM-RC than those in BSEM-CLRC (as well as in ESEM and BSEM-CL), which leads to weak support for the external validity. Another potential reason for these differences is that no residual covariances were proposed in the simulation study, whereas in the real data study 57 pairs of the a priori CUs that came from the same facet of a specific Big-Five factor were included and shown to be important in terms of goodness of fit (Marsh, et al., 2010). In this case, the variances of observed indicators can be largely explained by residual covariances, which leads to attenuated main loadings and factor correlations in BSEM-RC. Additionally, a possibility is that this study applies small variance on inverse-Wishart priors for the residual covariances matrix (Muthén & Asparouhov, 2012), in which the a priori CUs cannot be specified with their own priors (i.e., noninformative prior) in BSEM-RC. Although this method was found to perform better than others in the relatively simple simulation design with only two large residual covariances (Muthén & Asparouhov, 2012), further evaluation of influence of the underlying mechanism of residual covariances on factor structure is clearly warranted.

Another issue of BSEM-RC and BSEM-CLRC is that the estimation of large numbers of additional parameters (associated with residual covariances) brings with it an enormous increase in the posterior *SD* when the factor structure is complex (e.g., many factors and observed indictors). Although imposing small variance (e.g., .01) on the priors for these new parameters rather than freely estimating them may alleviate the negative impact of the increased estimation error, the stability and generalizability of model solutions is still be affected. Study 2 examines this issue by cross-validating our findings using longitudinal and 5-fold cross-validation techniques, which are often ignored in the model comparisons in relation to BSEM (e.g., Lu et al., 2016). Consistent with previous research (MacCallum & Tucker, 1991; Cudeck & Browne, 1983), cross-validation results indicate that more complex models (i.e., BSEM-RC and BSEM-CLRC) have a smaller likelihood of cross-validating than the simple model (i.e., BSEM-CL+CUs) when sample size is small, whereas the reverse is true when sample size is large. To further examine the impact of model specification with the choice of different priors on cross-validation, we used both more informative priors (*SD* = 0.05 and 0.01) and less informative priors (*SD* = 0.3) for residual covariances. All BSEM-RC and BSEM-CLRC resulted in good convergence (PSR < 1.05) and model fit (PP *p* value = from .474 to .525), and led to similar cross-validation results. Our findings suggest that

BSEM-RC and BSEM-CLRC should be used cautiously when the factor structure is complex and the sample size is small, given that they may capture idiosyncratic sample characteristics. However, further investigation for this important issue is still needed.

Model specification and factor structure

This study is the first to provide a comprehensive evaluation of the relations between model specifications and different simulated factor structures by varying the number, location, and size of targeted values in ESEM and the distribution of informative priors in different BESM models. Results suggest that the performance of different model specification is highly associated with factor structures: changing mean of cross-loadings priors to 0.1 in BSEM-CL and targeted value to 0.1 in ESEM performed worse in a balanced factor structure (average of the sizes of the cross-loadings for each factor = 0) but much better in an positive-oriented unbalanced factor structure (where only positive major cross-loadings are included). Typically, in our simulation study, the average of the sizes of the cross-loadings for each factor was 0.106 in the unbalanced factor structure. As such, changing the targeted value and mean of cross-loadings priors to 0.1 produced almost perfect estimated parameters (relative bias < 1.2%). Alternatively, in ESEM freeing substantive cross-loadings in the target loading matrix can also improve model estimation, whereas freeing trivial cross-loadings will deteriorate model estimation. It is expected that the logic behind changing the number and location of targeted value in ESEM also works in BSEM-CL, where using non-informative priors for substantive cross-loadings will improve model estimation, the reverse is true in using non-informative priors for trivial cross-loadings. Certainly, ESEM and BSEM-CL would perform even better when targeted values and informative priors for specific cross-loadings are set close to the population values. Nevertheless, these non-zero targeted value and informative priors should be applied very cautiously and should not be based on ex-post facto adjustments to models following preliminary analyses with the same data.

As mentioned above, a strength of BSEM-CL is the flexibility of specifying prior variance of cross-loadings. As prior variance of cross-loadings increased (from 0.005 to 0.02), factor correlations became less biased but main loadings became more biased, however, these differences were small. Similarly, specifying different prior variances of residual covariances (from .005 to 0.02) in BSEM-CR did not change the pattern of results. Thus, our study confirms previous findings (Muthén & Asparouhov, 2012), indicating that the prior variance choice did not have an important impact on the results in terms of model fit and biases of factor loadings and factor correlations.

Implications

The variance and flexibility in model specifications in BSEM make it challenging to guide researchers in deciding the most appropriate or optimal strategies and estimation procedures when developing a measurement model. However, based on our findings derived from both simulated and real data, we propose some constructive strategies for practices. Before providing these recommendations, however, we caution readers should not mistake these strategies as golden rules. Indeed, readers should be cautious of all golden rules presented in relation to SEM (Marsh, Hau, & Wen, 2004). Thus, we encourage readers to think about the unique situation of their own data and modelling needs; considering the strategies below as useful guides only.

First, EFA with mechanical rotation should be used in early pilot studies of a measurement instrument. Researchers can move to ESEM or BSEM approaches once they gain knowledge about the factor indicators and the factors (see below for discussion about informative priors for main loadings). Researchers should start with ESEM where targeted values are set to 0 for cross-loadings and BSEM-CL where only weakly informative priors on variance (i.e., variance = 0.01 with Mean = 0) are incorporated for cross-loadings. These two models can be used as benchmarks against which choices of targeted value and other priors can be compared. And then more substantive (previous) knowledge can be incorporated into the estimation process (via targeted value and informative priors on mean). Given that model specification on targeted values and informative priors have a significant impact on model estimation, the elicitation of targeted values and informative priors should be based on evidence-based approach rather than personal opinion that in principle bias toward a specific outcome (Kaplan, 2014). Specifically, the elicitation can be based on the results of a meta-analysis and previous publications. Compared to ESEM, an advantage of BSEM-CL is having more control on the variance of cross-loadings. In such cases, the informative priors can be chosen with smaller and smaller variances, reflecting a switch from an exploratory to confirmatory nature (also see Depaoli & van de Schoot, 2017).

We recommend not using BSEM-CLRC when model fit for ESEM and BSEM-CL is reasonable and particularly when sample size is small. The reason is that BSEM-CLRC require heavy computational burden and do not provide more accurate parameters estimates. Particularly, handling a multivariate variance prior (e.g., the covariances matrix) has technical complexities where some severe issues can arise; it is quite difficult to encode prior knowledge into probability distributions and requires detailed consideration during implementation in

practice (Depaoli & van de Schoot, 2017). However, when ESEM and BSEM-CL result in ill fit and sample size is large, BSEM-CLRC may be preferred. In such cases, researchers should start with small *df* values (see Equation 7) of inverse Wishart prior on residual covariances (i.e., relatively large variance priors) and increase *d* values by checking the rate of convergence in the Bayesian iterations and PP *p* value. We recommend that BSEM-RC should be used cautiously given the instability of model estimation and poor cross-validation and external validity.

Overall, we recommend that researchers experiment with a variety of priors, verify frequency coverage of key parameters estimates, assess sensitivity of results, and report all available findings (see Hamra, MacLehose, & Cole, 2013 for more discussion). More recently, Depaoli and van de Schoot (2017) developed a succinct checklist: the WAMBS-checklist (When to worry and how to Avoid the Misuse of Bayesian Statistics) which provides a guideline for Bayesian users to evaluate the influence of different priors and to interpret Bayesian results.

Limitations and Directions for Further Research

There are several limitations of the current study that motivate future research. First, an advantage of the BSEM technique proposed by Muthen & Asparouhov (2012) is that it produces posterior distributions for cross-loadings and residual covariances that can be used in line with modification indices (MIs). Researchers can free parameters, where the credibility interval does not cover zero, using noninformative priors and re-estimate the model. This technique benefits researchers to examine the degree of deviation of all parameters from zero in a single step, rather than relying on one-at-a-time MIs under conventional ML-based SEM (MacCallum et al., 2012). The one-at-a-time nature is associated with an inflated risk of capitalizing on chance (MacCallum et al., 1992; Marsh et al., 2017). However, model modifications under BSEM depend on which subsets of parameters have been specified to have small-variance priors. Furthermore, the estimation of the residual covariances parameters are not independent. In other words, the BSEM model may show more than one statistically significant residual covariances to compensate the misfit, although only one covariance is misfitted in the CFA model (see Asparouhov et al., 2015 for more discussion). Thus, it is beneficial to further compare BSEM with MIs-like respecification to ESEM and other BSEM models.

Second, like previous BSEM studies, all main loadings are specified by noninformative prior distribution (a diffuse prior) in the present study, which allows the data to dominate the estimation of posteriors through the likelihood (Zyphur & Oswald, 2013). However, in practice applied researchers might have ‘better available’

information/knowledge about main loadings that can be incorporated into a prior distribution, compared to cross-loadings. Very few studies have put informative priors on main loadings that are expected to be large according to *evidence-based* knowledge. For example, Rindskopf (2012) suggests one can have a normal prior with a mean of 0.6/0.7 with a *SD* of .15 or have a normal prior with an unknown mean. In both cases, the *SD* should be large enough to allow reasonable variation in main loadings. Therefore, the specification of main loadings with informative priors needs further research.

Another avenue for further investigation is to examine how different heterogeneous errors for the indicator variables (all were set to 0.5 in this study) influence the estimation, given that heterogeneity is not uncommon in real data sets and can cause problems.

Reference

Asendorpf, J. B., & Van Aken, M. A. G. (2003). Personality-relationship transaction in adolescence: Core versus surface personality characteristics. *Journal of Personality*, 71, 629–666.

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

Asparouhov, T., & Muthén , B. (2010). *Bayesian analysis using Mplus: Technical implementation.* Manuscript submitted for publication. Retrieved from www.statmodel.com

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*(6), 1561–1577. https://doi.org/10.1177/0149206315591075

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281-1311.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605–34.

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, 11, 515–524

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

Celimli.S, Myers, N. D., & Ahn, S. (2018). Geomin Versus Target Rotation in Exploratory Factor Analysis: Correlated Factors and Large and Complex Pattern Matrices. *Proceedings Paper presented at the American Educational Research Association Annual Meeting*, NYC, USA.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12, 381–398.

Costa, P. T., Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment*, 4, 5–13.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 147–167.

Cudeck, R. & MacCallum, R. (Eds.) (2007) *Factor analysis at 100: Historical developments and future directions.* Mahwah, NJ: Lawrence Erlbaum

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, 115, 475-487.

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*, 6:1963.

Depaoli, S., & van de Schoot, R. (2017). Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist. *Psychological Methods*, *22*(2), 240–261.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*, 276–302.

Fong, T. C. T., & Ho, R. T. H. (2013). Factor analyses of the Hospital Anxiety and Depression Scale: a Bayesian structural equation modeling approach. *Quality of Life Research*, *22*(10), 2857–2863.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling. *Psychological Assessment*, *25*(2), 496–508.

Grimm, K. J., Mazza, G. L., & Davoudzadeh, P. (2016). Model Selection in Finite Mixture Models: A k-Fold Cross-Validation Approach. *Structural Equation Modeling*, *24*(2), 1–11.

Gucciardi, D. F., & Zyphur, M. J. (2016). Exploratory structural equation modelling and Bayesian estimation. In N. Ntoumanis & N. D. Myers (Eds.), *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists* (pp.172-194). Chichester, England: Wiley

Hamra G. B., MacLehose R. F., & Cole S. R (2013). Sensitivity analyses for sparse-data problems—using weakly informative Bayesian priors. *Epidemiology*, 24, 233–239.

Jennrich, R. I., & Sampson, P.F. (1966). Rotation for simple loadings. *Psychometrika*, 31, 313–323.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 650–673). New York, NY: Guilford.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lu, Z.-H., Chow, S.-M., & Loken, E. (2016). Bayesian Factor Analysis as a Variable-Selection Problem: Alternative Priors and Consequences. *Multivariate Behavioral Research*, *51*(4), 519-539.

Lu, Z., Chow, S., & Loken, E. (2017). A comparison of Bayesian and frequentist model selection methods for factor analysis models. *Psychological Methods*, *22*(2), 361–381.

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*(3), 340–345.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490–504.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Quantitative Methods in Psychology*, *109*(3), 502–511.

McCrae, R. R., & Costa, P. T., Jr. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research*, 24, 107–124

McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York: Guilford Press.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2017). What to do When Scalar Invariance Fails: The Extended Alignment Method for Multi-Group Factor Analysis Comparison of Latent Means Across Many Groups. *Psychological Methods*, *21*(4), 1–22.

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

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–91.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194–1218.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, *74*(2), 403–456.

Muthén, B. O. (2010). *Bayesian Analysis In Mplus: A Brief Introduction*. Manuscript submitted for publication. Retrieved from www.statmodel.com

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–35.

Muthén, B. O., & Muthén, L. (1998–2017). *Mplus user's guide* (8ed.). Los Angeles, CA: Authors.

Myers, N. D., Ahn, S., & Jin, Y. (2013). Rotation to a partially specified target matrix in exploratory factor analysis: How many targets? *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(1), 131–147.

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*(2), 494–505.

Myers, N. D., Ahn, S., Lu, M., Celimli, S., & Zopluoglu, C. (2017). Reordering and Reflecting Factors for Simulation Studies With Exploratory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(1), 112–128.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190–204.

Press, S. J. (2003). *Subjective and objective Bayesian statistics: Principles, models, and applications* (2nd ed.). New York, NY: Wiley

Rindskopf, D. (2012). Next steps in Bayesian structural equation models: Comments on, variations of, and extensions to Muthén and Asparouhov (2012). *Psychological Methods*, *17*(3), 336–339.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64, 37–52.

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*(2), 491–520.

Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., & Maaz, K. (Eds.). (2010). Schulleistungen von Abiturienten: *Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217–239.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, 44, 41-61.

Zyphur, M. J., & Oswald, F. L. (2013). Bayesian Estimation and Inference: A User's Guide. *Journal of Management*, *41*(2), 390-420.

Table 1.

Overview of Previous Studies Using BSEM Approaches

Studies	Type	BSEM approaches
Muthén &Asparouhov (2012)	Simulation and empirical	BSEM-CL, BSEM-CLRC
Golay, Reverte, Rossier, Favez, & Lecerf (2013)	Empirical	BSEM-CL
Zyphur & Oswald (2013)	Empirical	BSEM-RC
Fong & Ho (2013)	Empirical	BSEM-CL, BSEM-CLRC
De Bondt, & Van Petegem (2015)	Empirical	BSEM-CL
Stromeyer, Miller, Sriramachandramurthy, & DeMartino (2015)	Empirical	BSEM-CL, BSEM-CLRC
Asparouhov, Muthén, & Morin (2015)	Empirical	BSEM-CL, BSEM-CLRC
Lu, Chow, & Loken, 2016	Simulation and empirical	BSEM-CL
Gucciardi & Zyphur (2016)	Empirical	BSEM-CL

Table 2.

Simulation Designs

	Design 1			Design 2			Design 3		
Variable	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3
y1	A	C	B	A	C	.4	A	D	-.4
y2	A	C	C	A	.3	C	A	.3	C
y3	A	C	C	A	.2	C	A	.2	C
y4	A	C	C	A	C	C	A	D	C
y5	A	C	C	A	C	.1	A	D	-.1
y6	B	A	C	.4	A	C	-.4	A	D
y7	C	A	C	C	A	.3	C	A	.3
y8	C	A	C	C	A	.2	C	A	.2
y9	C	A	C	C	A	C	C	A	D
y10	C	A	C	.1	A	C	-.1	A	D
y11	C	B	A	C	.4	A	D	-.4	A
y12	C	C	A	.3	C	A	.3	C	A
y13	C	C	A	C	C	A	D	C	A
y14	C	C	A	C	.1	A	D	-.1	A
y15	C	C	A	.2	C	A	.2	C	A

Note. A (Major factor loadings) = 0.8; B (Major cross-loadings) = 0.1, 0.2, and 0.3(three conditions); C (Minor cross-loadings) = .01; D (Minor cross-loadings) = -.01; Factor correlations = 0.5 (see Appendix 7 for the Mplus syntax)

Table 3.

Choice of Priors in BSEM for Simulation Study

BSEM approaches	Informative priors	Non-informative priors
BSEM-CL	Cross-loadings ~ $N(0, .01)$	main loadings ~N(0, infinity), residual variances ~ Γ^{-1}(-1,0), latent factor covariances ~ $IW(0, -q-1)$ intercepts ~N(0, infinity).
BSEM-RC	Residual covariance matrix Θ: diagonal elements: $IW(1, p+6)$ off-diagonal elements: $IW(0, p+6)$	main loadings ~N(0, infinity), latent factor covariances ~ $IW(0, -q-1)$ intercepts ~N(0, infinity)
BSEM-CLRC	Cross-loadings ~ $N(0, .01)$ Residual covariance matrix Θ: diagonal elements ~ $IW(\mathbf{I}, p+6)$ off-diagonal elements ~ $IW(0, p+6)$	main loadings ~N(0, infinity), latent factor covariances ~ $IW(0, -q-1)$ intercepts ~N(0, infinity)

Note. p is the number of observed variables; q is the number of latent variables; BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors.Also see p. 775 in *Mplus Uers' Guide* (8th) for detailed description.

Table 4.

Model Rejection Rate (5%) for ML and Bayesian Models in the Simulation Study.

		Rejection rate (5%)			
		ML *p*	Bayes PP *p* (PSR)		
Sizes of Major cross-loading	Sample size	ESEM	BSEM-CL	BSEM-RC	BSEM-CLRC
		Design 1			
.1	200	.086	.000(1.009)	.000(1.005)	.000(1.010)
.1	500	.048	.000(1.004)	.000(1.035)	.000(1.025)
.1	1000	.066	.006(1.015)	.000(1.042)	.000(1.044)
.2	200	.088	.002(1.010)	.000(1.006)	.000(1.011)
.2	500	.052	.000(1.005)	.000(1.035)	.000(1.023)
.2	1000	.068	.006(1.015)	.000(1.040)	.000(1.042)
.3	200	.092	.006(1.010)	.000(1.007)	.000(1.012)
.3	500	.054	.000(1.005)	.000(1.034)	.000(1.023)
.3	1000	.068	.006(1.015)	.000(1.039)	.000(1.042)
		Design 2			
-	200	.100	.064(1.011)	.000(1.018)	.000(1.008)
-	500	.056	.034(1.006)	.000(1.024)	.000(1.033)
-	1000	.062	.022(1.003)	.000(1.040)	.000(1.040)
		Design 3			
-	200	.082	.008(1.008)	.000(1.017)	.000(1.013)
-	500	.056	.010(1.006)	.000(1.027)	.000(1.034)
-	1000	.054	.006(1.005)	.000(1.036)	.000(1.046)

Table 5.
Model fit for BSEM models in Simulation Designs 2 & 3.

Design 2	Model description	Para meters	PSR			Rejection rate			BIC			DIC		
		-	200	500	1000	200	500	1000	200	500	1000	200	500	1000
BSEM-CL	Cross-loadings (M = 0, Var = .01)	78	1.011	1.006	1.003	.064	.034	.022	7940	19352	38319	7650	19002	37920
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	78	1.010	1.003	1.002	.412	.530	.288	7961	19380	38343	7663	19022	37939
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	78	1.011	1.011	1.005	.006	.004	.004	7925	19339	38311	7642	18993	37913
BSEM-RC	Residual covariances (M = 0, Var = .01)	153	1.018	1.024	1.040	.000	.000	.000	8250	19738	38763	7697	19051	37973
BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	153	1.016	1.031	1.038	.000	.000	.000	8250	19738	38763	7698	19052	37973
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	153	1.023	1.038	1.042	.000	.000	.000	8249	19738	38763	7696	19051	37973
BSEM-CLRC	Residual covariances+cross-loadings (M=0, Var=.01)	183	1.008	1.033	1.040	.000	.000	.000	8406	19924	38972	7697	19051	37971
BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	78	1.012	1.010	1.006	.016	.004	.008	7929	19342	38313	7644	18994	37914
Design 3														
BSEM-CL	Cross-loadings (M = 0, Var = .01)	78	1.008	1.006	1.005	.008	.010	.006	7938	19365	38359	7653	19017	37961
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	78	1.009	1.003	1.003	.196	.078	.028	7961	19382	38371	7668	19030	37969
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	78	1.007	1.010	1.010	.002	.008	.004	7927	19359	38356	7648	19014	37959
BSEM-RC	Residual covariances (M = 0, Var = .01)	153	1.017	1.027	1.036	.000	.000	.000	8265	19776	38822	7702	19063	38011
BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	153	1.019	1.036	1.040	.000	.000	.000	8266	19777	38820	7703	19064	38013
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	153	1.029	1.039	1.043	.000	.000	.000	8265	19776	38822	7700	19063	38010
BSEM-CLRC	Residual covariances+cross-loadings (M=0, Var=.01)	183	1.013	1.034	1.046	.000	.000	.000	8417	19949	39022	7706	19073	38015
BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	78	1.010	1.007	1.006	.008	.010	.004	7935	19363	38358	7653	19016	37960

Note. PSR = the potential scale reduction

Table 6.
Relative bias, Coverage, and Power across Models based on Simulation Design 1.

Model	MajorCL	Size	Relative Bias		SD of Relative Bias		95% coverage		Power	
			MFL	FC	MFL	FC	MFL	FC	MFL	FC
ESEM	0.1	200	0.6%	3.5%	10.5%	12.3%	.959	.997	1.000	1.000
BSEM-CL	0.1	200	4.7%	7.7%	10.0%	12.5%	.944	.986	1.000	1.000
BSEM-RC	0.1	200	8.9%	-3.9%	9.3%	11.7%	.887	.946	1.000	1.000
BSEM-CLRC	0.1	200	8.8%	-3.3%	9.3%	11.4%	.976	.997	1.000	1.000
ESEM	0.1	500	1.5%	6.5%	6.6%	7.7%	.953	.993	1.000	1.000
BSEM-CL	0.1	500	3.5%	8.3%	6.4%	7.8%	.961	.999	1.000	1.000
BSEM-RC	0.1	500	8.1%	-4.8%	6.4%	7.1%	.877	.913	1.000	1.000
BSEM-CLRC	0.1	500	7.0%	-2.8%	5.9%	7.0%	.997	1.000	1.000	1.000
ESEM	0.1	1000	1.8%	7.5%	4.7%	5.5%	.947	.953	1.000	1.000
BSEM-CL	0.1	1000	3.1%	8.6%	4.7%	5.5%	.977	1.000	1.000	1.000
BSEM-RC	0.1	1000	7.9%	-5.0%	5.2%	5.0%	.868	.853	1.000	1.000
BSEM-CLRC	0.1	1000	6.4%	-2.1%	4.4%	5.0%	1.000	1.000	1.000	1.000
ESEM	0.2	200	1.2%	5.8%	10.6%	11.9%	.959	.994	1.000	1.000
BSEM-CL	0.2	200	5.8%	12.4%	10.2%	12.1%	.932	.969	1.000	1.000
BSEM-RC	0.2	200	9.3%	2.9%	11.4%	11.4%	.839	.949	1.000	1.000
BSEM-CLRC	0.2	200	9.8%	2.1%	9.9%	11.1%	.965	.998	1.000	1.000
ESEM	0.2	500	2.1%	8.8%	6.6%	7.5%	.947	.984	1.000	1.000
BSEM-CL	0.2	500	4.7%	12.8%	6.6%	7.5%	.946	.989	1.000	1.000
BSEM-RC	0.2	500	8.4%	2.2%	9.4%	7.0%	.812	.957	1.000	1.000
BSEM-CLRC	0.2	500	8.1%	2.5%	6.6%	6.8%	.991	1.000	1.000	1.000
ESEM	0.2	1000	2.5%	9.8%	4.7%	5.3%	.937	.904	1.000	1.000
BSEM-CL	0.2	1000	4.3%	13.1%	4.7%	5.3%	.962	.997	1.000	1.000
BSEM-RC	0.2	1000	8.2%	2.2%	8.7%	4.9%	.800	.952	1.000	1.000
BSEM-CLRC	0.2	1000	7.5%	3.2%	5.3%	4.8%	.996	1.000	1.000	1.000
ESEM	0.3	200	1.4%	6.7%	10.7%	11.6%	.958	.995	1.000	1.000
BSEM-CL	0.3	200	6.8%	17.0%	10.5%	11.8%	.918	.921	1.000	1.000
BSEM-RC	0.3	200	9.4%	11.1%	14.7%	11.1%	.803	.840	1.000	1.000
BSEM-CLRC	0.3	200	10.5%	8.0%	11.2%	10.8%	.941	.990	1.000	1.000
ESEM	0.3	500	2.4%	9.6%	6.7%	7.4%	.944	.975	1.000	1.000
BSEM-CL	0.3	500	5.7%	17.1%	6.7%	7.3%	.929	.943	1.000	1.000
BSEM-RC	0.3	500	8.4%	10.9%	13.3%	6.7%	.797	.721	1.000	1.000
BSEM-CLRC	0.3	500	9.0%	8.4%	8.1%	6.6%	.967	1.000	1.000	1.000
ESEM	0.3	1000	2.7%	10.6%	4.7%	5.2%	.933	.877	1.000	1.000
BSEM-CL	0.3	1000	5.4%	17.1%	4.8%	5.2%	.940	.965	1.000	1.000
BSEM-RC	0.3	1000	8.1%	11.1%	13.0%	4.7%	.799	.531	1.000	1.000
BSEM-CLRC	0.3	1000	8.5%	9.0%	6.9%	4.7%	.974	1.000	1.000	1.000

Note. MFL = Main factor loading; FC = Factor correlation; BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors. Power refers to proportion of 95% credibility interval not covering 0 in a Bayes setting and proportion of 95% confidence interval not covering 0 in a frequentist setting, respectively.

Table 7.

ANOVA Testing Variance Explained by Different Design Conditions

	Relative Bias		SD of Relative Bias		95% coverage		Power	
	MFL	FC	MFL	FC	MFL	FC	MFL	FC
				Design 1				
Model	89.1%	54.7%	17.5%	1.5%	78.6%	65.1%	-	-
Size of major cross-loading	5.0%	37.5%	17.5%	0.4%	8.9%	7.3%	-	-
Sample sizes	1.0%	0.4%	46.5%	97.8%	0.6%	5.3%	-	-
				Design 2				
Model	99.0%	99.5%	74.0%	37.6%	90.1%	88.8%	-	-
Sample sizes	0.1%	0.1%	22.4%	58.8%	4.1%	4.5%	-	-
				Design 3				
Model	98.7%	96.8%	93.7%	64.5%	81.9%	84.7%	-	-
Sample sizes	0.3%	0.8%	4.0%	21.8%	6.6%	7.19%	-	-

Note. MFL = Main factor loading; FC = Factor correlation; cov = 95% coverage; Power refers to proportion of 95% credibility interval not covering 0 in a Bayes setting and proportion of 95% confidence interval not covering 0 in a frequentist setting, respectively.

Table 8.

Relative bias, Coverage, and Power in relation to Main Loadings across Models based on Simulation Designs 2 & 3.

Model	Description	Relative Bias			SD of Relative Bias			95% coverage			Power		
Design 2 (unbalanced factor structure)		200	500	1000	200	500	1000	200	500	1000	200	500	1000
ESEM	Target rotation	4.3%	5.3%	5.7%	11.4%	7.2%	5.2%	.939	.905	.840	1	1	1
ESEM(Geomin)	Geomin rotation	-3.2%	-2.8%	-2.6%	9.8%	6.1%	4.3%	.947	.947	.925	1	1	1
ESEM(Free:2MinCL)	Free 2 minor cross-loadings (.01)	6.7%	8.0%	8.4%	12.0%	7.7%	5.6%	.916	.840	.701	1	1	1
ESEM(Free:2MajCL)	Free 2 major cross-loadings (.2, .4)	2.1%	2.9%	3.2%	11.3%	7.1%	5.1%	.949	.939	.908	1	1	1
ESEM(Free:4MajCL)	Free 4 major cross-loadings (.1,.2,.3, .4)	-0.1%	0.3%	0.6%	11.1%	6.9%	4.9%	.949	.955	.956	1	1	1
ESEM(~.1)	Target value = .1	-0.7%	0.4%	1.1%	13.8%	11.7%	11.2%	.895	.785	.566	1	1	1
BSEM-CL	Cross-loadings (M = 0, Var = .01)	14.5%	12.7%	12.1%	11.2%	7.3%	5.4%	.804	.748	.675	1	1	1
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	14.9%	13.0%	12.3%	11.3%	7.1%	5.1%	.732	.642	.504	1	1	1
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	14.3%	12.6%	12.1%	11.6%	7.6%	5.6%	.859	.854	.850	1	1	1
BSEM-RC	Residual covariances (M = 0, Var = .01)	17.9%	16.7%	16.5%	17.4%	16.1%	15.7%	.548	.484	.455	1	1	1
BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	18.1%	16.8%	16.6%	17.4%	16.1%	15.7%	.538	.473	.448	1	1	1
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	17.8%	16.6%	16.4%	17.3%	16.1%	15.7%	.559	.493	.465	1	1	1
BSEM-CLRC	Residual covariances+cross-loadings (M=0, Var=.01)	18.4%	16.9%	16.4%	13.5%	10.6%	9.7%	.855	.893	.904	1	1	1
BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	1.1%	0.1%	0.0%	10.2%	6.4%	4.6%	.956	.977	.991	1	1	1
Design 3 (balanced factor structure)													
ESEM	Target rotation	-2.9%	-2.4%	-2.3%	9.9%	6.4%	4.8%	.949	.919	.887	1	1	1
ESEM(Geomin)	Geomin rotation	-6.4%	-6.1%	-6.0%	9.7%	6.7%	5.3%	.883	.774	.623	1	1	1
ESEM(Free:2MinCL)	Free 2 minor cross-loadings (.01)	-2.6%	-2.2%	-2.1%	10.2%	6.7%	5.1%	.938	.912	.873	1	1	1
ESEM(Free:2MajCL)	Free 2 major cross-loadings (.2, -.4)	0.1%	0.9%	1.1%	10.4%	6.5%	4.7%	.955	.952	.949	1	1	1
ESEM(Free:4MajCL)	Free 4 major cross-loadings (-.1,.2,.3, -.4)	-1.4%	-0.6%	-0.4%	10.6%	6.6%	4.7%	.947	.947	.949	1	1	1
ESEM(~.1)	Targeted value = .1	-6.5%	-6.5%	-6.5%	10.7%	7.7%	6.3%	.872	.691	.527	1	1	1
BSEM-CL	Cross-loadings (M = 0, Var = .01)	1.9%	1.9%	2.2%	11.1%	6.8%	5.0%	.932	.961	.974	1	1	1
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	0.8%	0.9%	1.4%	12.5%	7.3%	5.1%	.878	.927	.949	1	1	1
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	3.3%	2.8%	2.8%	10.6%	6.8%	5.0%	.954	.980	.992	1	1	1
BSEM-RC	Residual covariances (M = 0, Var = .01)	**-5.6%**	**-6.2%**	**-4.7%**	**26.2%**	**26.0%**	**27.8%**	**.701**	**.576**	**.450**	.992	1	.999
BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	-5.0%	-5.8%	-4.5%	26.5%	26.2%	28.0%	.676	.556	.416	.993	.999	.999
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	-6.1%	-6.6%	-5.0%	25.8%	25.7%	27.6%	.737	.601	.479	.991	.999	.998
BSEM-CLRC	Residual covariances+cross-loadings (M=0, Var=.01)	3.2%	2.3%	1.8%	17.3%	15.1%	14.7%	.936	.964	.976	1	1	1

BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	-5.9%	-5.9%	-5.7%	11.1%	7.4%	5.8%	.853	.784	.697	1	1	1

Note. *Note*. Power refers to proportion of 95% credibility interval not covering 0 in a Bayes setting and proportion of 95% confidence interval not covering 0 in a frequentist setting, respectively.

Table 9.
Relative bias, Coverage, and Power in relation to Factor Correlations across Models based on Simulation Designs 2 & 3.

Model	Description	Relative Bias			SD of Relative Bias			95% coverage			% sig coeff		
Design 2 (unbalanced factor structure)		200	500	1000	200	500	1000	200	500	1000	200	500	1000
ESEM	Target rotation	19.3%	22.0%	23.0%	10.0%	6.3%	4.7%	.923	.531	.129	1	1	1
ESEM(Geomin)	Geomin rotation	-17.9%	-17.0%	-16.7%	7.9%	5.0%	3.5%	.998	.971	.777	1	1	1
ESEM(Free:2MinCL)	Free 2 minor cross-loadings (.01)	26.4%	29.3%	30.2%	9.8%	6.1%	4.4%	.737	.106	.001	1	1	1
ESEM(Free:2MajCL)	Free 2 major cross-loadings (.2, .4)	11.7%	14.2%	15.1%	11.6%	7.3%	5.3%	.981	.895	.593	1	1	1
ESEM(Free:4MajCL)	Free 4 major cross-loadings (.1,.2,.3, .4)	0.1%	2.5%	3.3%	14.7%	9.3%	6.7%	.991	.991	.985	1	1	1
ESEM(~.1)	Target value = .1	-0.6%	5.1%	8.2%	16.2%	11.5%	7.3%	.987	.972	.919	1	1	1
BSEM-CL	Cross-loadings (M = 0, Var = .01)	40.6%	39.1%	38.1%	9.5%	5.9%	4.2%	.200	.035	.003	1	1	1
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	42.4%	40.9%	39.5%	9.5%	5.9%	4.2%	.084	.003	.000	1	1	1
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	38.7%	37.5%	36.9%	9.5%	5.9%	4.3%	.491	.339	.225	1	1	1
BSEM-RC	Residual covariances (M = 0, Var = .01)	40.4%	40.2%	40.2%	8.1%	5.0%	3.4%	.025	.000	.000	1	1	1
BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	40.4%	40.2%	40.2%	8.1%	4.9%	3.4%	.022	.000	.000	1	1	1
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	40.3%	40.1%	40.2%	8.2%	5.0%	3.4%	.027	.000	.000	1	1	1
BSEM-CLRC	Residual covariances+cross-loadings (M=0, Var=.01)	36.4%	36.7%	37.0%	8.5%	5.2%	3.7%	.275	.071	.009	1	1	1
BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	1.1%	-0.5%	-1.2%	14.5%	8.9%	6.2%	.989	.999	1	.999	1	1
Design 3 (balanced factor structure)													
ESEM	Target rotation	-18.6%	-16.7%	-16.2%	11.9%	7.6%	5.3%	.981	.908	.708	1	1	1
ESEM(Geomin)	Geomin rotation	-47.7%	-47.1%	-47.0%	8.6%	5.5%	3.8%	.541	.003	.000	1	1	1
ESEM(Free:2MinCL)	Free 2 minor cross-loadings (.01)	-16.1%	-14.4%	-14.1%	12.6%	8.1%	5.7%	.985	.947	.799	1	1	1
ESEM(Free:2MajCL)	Free 2 major cross-loadings (.2, -.4)	-1.8%	0.9%	1.8%	12.5%	8.1%	5.8%	.999	.998	.995	1	1	1
ESEM(Free:4MajCL)	Free 4 major cross-loadings (-.1,.2,.3, -.4)	-3.9%	-1.0%	-0.1%	14.4%	9.3%	6.8%	.991	.993	.989	1	1	1
ESEM(~.1)	Target value = .1	-56.6%	-61.2%	-63.7%	20.1%	14.8%	10.0%	.356	.058	.005	.796	.981	1
BSEM-CL	Cross-loadings (M = 0, Var = .01)	4.3%	3.9%	4.0%	13.4%	8.2%	5.7%	.989	1	1	1	1	1
BSEM-CL(Var=.005)	Cross-loadings (M = 0, Var = .005)	6.1%	4.8%	4.4%	13.7%	8.3%	5.8%	.959	.991	.998	1	1	1
BSEM-CL(Var=.02)	Cross-loadings (M = 0, Var = .02)	3.1%	3.1%	3.6%	13.0%	8.1%	5.8%	.999	1	1	1	1	1
BSEM-RC	Residual covariances (M = 0, Var = .01)	-2.5%	4.9%	18.5%	43.3%	34.0%	13.0%	.767	.497	.221	.670	.700	.893

BSEM-RC(Var=.005)	Residual covariances (M = 0, Var = .005)	0.4%	7.0%	19.1%	40.9%	31.4%	12.0%	.743	.466	.173	.704	.719	.917
BSEM-RC(Var=.02)	Residual covariances (M = 0, Var = .02)	-4.7%	2.8%	17.7%	44.4%	35.9%	14.7%	.795	.539	.277	.624	.682	.873
BSEM-CLRC	R4esidual covariances+cross-loadings (M=0, Var=.01)	2.6%	3.9%	4.7%	13.0%	8.0%	6.0%	.998	1	1	.998	.999	.998
BSEM-CL(M=1)	Mean = .1, Var = .01 for cross-loadings	-47.0%	-47.0%	-46.6%	17.6%	10.7%	7.5%	.401	.134	.016	.765	.956	.999

Note. Power refers to proportion of 95% credibility interval not covering 0 in a Bayes setting and proportion of 95% confidence interval not covering 0 in a frequentist setting, respectively.

Table 10.
Model fit for Empirical Data Study.

Maximum likelihood analyses						
Model	Parameters	df	p-value	CFI	TLI	RMSEA
CFA	190	1700	0	.684	.671	.053
CFA+CUs	244	1646	0	.750	.731	.048
ESEM	410	1480	0	.850	.820	.039
ESEM+CUs	464	1426	0	.912	.891	.030
Bayesian analysis						
Model	Parameters	2.5% PP limit	97.5% PP limit	PP p-value	DIC	BIC
BSEM-CL	430	7422	7688	0	424152	426837
BSEM-CL+CUs	484	4304	4571	0	421083	424117
BSEM-RC	1960	-176	171	.518	418026	430265
BSEM-CLRC	2200	-176	169	.526	418027	432215

Note. BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors. PP = posterior predictive; CUs = a priori correlated uniquenesses.

Table 11.

Correlation among the Big-Five Factors (Emotional Stability = reversed Neuroticism[RN])

Model		C	E	RN	O	M(SD)
CFA+ CUs	Agreeableness (A)	-				
	Conscientiousness (C)	.243(.022)*	-			
	Extraversion (E)	.253(.021)*	.395(.023)*	-		
	Emotional Stability (RN)	.305(.019)*	.142(.023)*	.502(.019)*	-	
	Openness (O)	-.092(.022)*	.061(.024)*	.081(.023)*	-.054(.022)*	.184(.191)
ESEM+CUs	Agreeableness (A)	-				
	Conscientiousness (C)	.078(.016)*	-			
	Extraversion (E)	.189(.017)*	.181(.015)*	-		
	Emotional Stability (RN)	.239(.015)*	.071(.015)*	.227(.015)*	-	
	Openness (O)	.006(.018)	-.051(.017)*	.090(.018)*	-.085(.017)*	.095(.114)
BSEM-CL+ CUs[1]	Agreeableness (A)					
	Conscientiousness (C)	.099(.103)				
	Extraversion (E)	.237(.101)*	.232(.105)*			
	Emotional Stability (RN)	.313(.087)*	.107(.100)	.318(.095)*		
	Openness (O)	-.050(.099)	-.069(.099)	.057(.105)	-.124(.095)	.112(.160)
BSEM-RC	Agreeableness (A)					
	Conscientiousness (C)	.076(.045)				
	Extraversion (E)	.018(.036)	.007(.097)			
	Emotional Stability (RN)	.054(.047)	-.044(.050)	.192(.065)*		
	Openness (O)	-.146(.031)*	.042(.036)	-.019(.039)	-.087(.050)	.009(.093)
BSEM-CLRC	Agreeableness (A)					
	Conscientiousness (C)	.160(.123)				
	Extraversion (E)	.208(.104)*	.235(.126)*			
	Emotional Stability (RN)	.312(.096)*	.131(.126)	.432(.104)*		
	Openness (O)	-.060(.095)	.020(.115)	.030(.110)	-.100(.100)	.137(.168)

Note. [1] A priori correlated uniquenesses were freely estimated by using noninformative priors; BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors; CUs = correlated uniquenesses;* indicates $p < .05$ for CFA and ESEM but it indicates significance in the sense of their 95% posterior distribution credibility intervals not including zero for BSEM models. We also report standard errors of correlation coefficients for CFA and ESEM and posterior standard deviation for BSEM.

Table 12

Root Mean Square Residual (RMSR) and RMSEA for Cross-validation Analysis (between Wave 1 and 2 Data)

	Cross-validation from Wave 1 to Wave 2				
	RMSR			RMSEA	
	Mean	Quantile (2.5%)	Quantile (97.5%)	RMSEA	90 Percent C.I.
CFA+CUs	.049	.046	.052	.054	[.053, 0.55]
ESEM+CUs	.036	.033	.039	.043	[.042, 0.44]
BSEM+CL priors +CUs[1]	.036	.033	.039	.043	[.042, 0.44]
BSEM+RC priors	.033	.031	.036	.036	[.035, 0.37]
BSEM+CL priors + RC priors	.033	.031	.036	.036	[.035, 0.37]
	Cross-validation from Wave 2 to Wave1				
CFA+CUs	.047	.046	.049	.048	[.047, .048]
ESEM+CUs	.033	.032	.035	.044	[.043, .045]
BSEM-CL+CUs	.033	.032	.035	.044	[.043, .044]
BSEM-RC	.032	.030	.033	.045	[.044, .045]
BSEM-CLRC	.033	.031	.035	.045	[.044, .046]

Table 13

RMSEA for 5-Fold Cross-validation Analysis

	From 80% to 20%	
	RMSEA	90 Percent C.I.
CFA+CUs	.046	[.044, 0.47]
ESEM+CUs	.031	[.029, 0.33]
BSEM+CL priors +CUs[1]	.039	[.038, 0.41]
BSEM+RC priors	.026	[.024, 0.28]
BSEM+CL priors + RC priors	.026	[.024, 0.28]
	From 20% to 80%	
CFA+CUs	.047	[.046, .048]
ESEM+CUs	.034	[.033, .035]
BSEM-CL+CUs	.040	[.039, .041]
BSEM-RC	.045	[.044, .045]
BSEM-CLRC	.045	[.044, .045]

Table 14
Summary for the Key Findings

	Hypothesis	Support for predictions	Inconsistent with predictions
H1	Model fit	BSEM-RC and BSEM-CLRC fit the data better (e.g., having low model rejection rate) than BSEM-CL and ESEM	BSEM-CL showed lower DIC to a very small extent
H2	Close performance between ESEM and BSEM-CL.	ESEM will perform more closely to BSEM-CL than BSEM-RC and BSEM-CLRC in terms of model fit, bias, coverage, and power in estimation of major loadings and factor correlations	-
		Research question	
Q1	Comparison between ESEM and different BSEM models in the simulation study	• The pattern of results in main loadings and factor correlations was substantially varied by the factor structures: ESEM resulted in more accurate parameters estimates than BSEM-CL and BSEM-CLRC in unbalanced factor structures (all cross-loadings were positive), the reverse being true in a balanced factor structure (i.e., the sum of the sizes of the cross-loadings for each factor = 0). • BSEM-CL and BSEM-CLRC tended to result in more inflated estimated parameters than ESEM. • BESM-CL provided unstable estimation solutions in terms of the larger bias SD, lower coverage, and less power. • Specifying mean of cross-loadings to 0.1 in BSEM-CL and target value to 0.1 in ESEM changed the pattern substantially. However, change variances of cross-loadings and residual covariances in BSEMs lead to similar results. • The prior variance choice did not have an important impact on the results	
Q2	comparison between simulation and real data results	• The pattern of results revealed in the simulation study was largely consistent with the findings based on real data. • The differences between different model solutions were smaller than those in simulation study.	

Note. DIC = deviance information criterion; BSEM-CL = BSEM + cross-loading priors; BSEM-RC=BSEM + residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors;

$$\Lambda = \begin{pmatrix} .8 & .01 & .2 \\ .8 & .01 & .01 \\ .8 & .01 & .01 \\ .2 & .8 & .01 \\ .01 & .8 & .01 \\ .01 & .8 & .01 \\ .01 & .2 & .8 \\ .01 & .01 & .8 \\ .01 & .01 & .8 \end{pmatrix} A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} B = \begin{pmatrix} ? & 0 & 0 \\ ? & 0 & 0 \\ ? & 0 & 0 \\ 0 & ? & 0 \\ 0 & ? & 0 \\ 0 & ? & 0 \\ 0 & 0 & ? \\ 0 & 0 & ? \\ 0 & 0 & ? \end{pmatrix}$$

Figure 1. An example of loadings matrix and rotation matrices in EFA with target rotation.
Note. Matrix A designated whether each pattern coefficient was (1) or was not (0) a target. Matrix B provided values that targeted elements would be rotated toward and denoted nontargeted elements with a ? sign. Matrix provided population values.

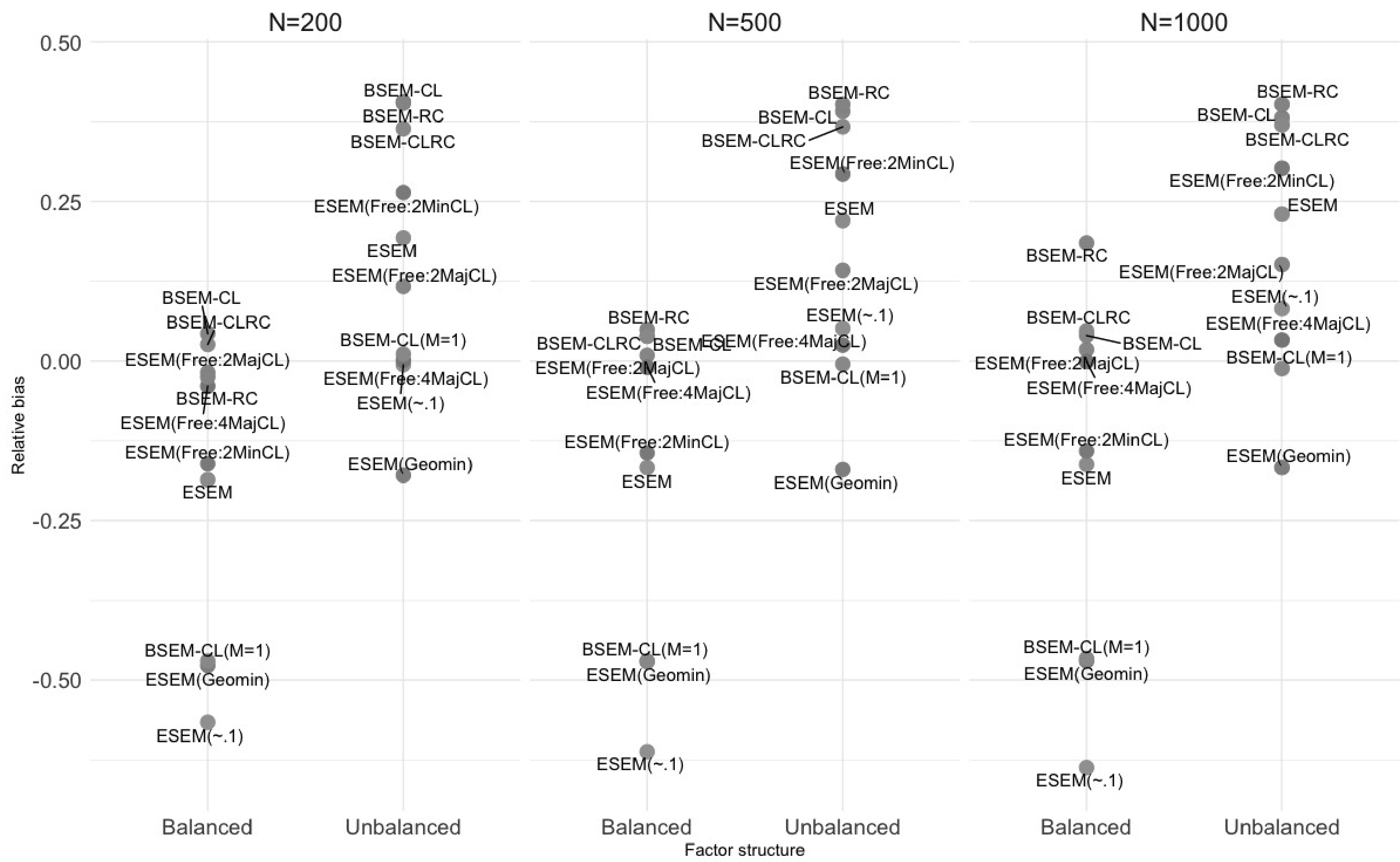

Figure 2. Relative bias of factor correlations across models based on unbalanced (Design 2) and balanced factor structure (Design 3).
Note. BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CLRC=+cross-loading priors + residual covariance priors; ESEM(Free:2MinCL) = ESEM with Free 2 minor cross-loadings (.01); ESEM(Free:2MajCL) = ESEM with Free 2 major cross-loadings; ESEM(Free:4MajCL) = Free 4 major cross-loadings; ESEM(~.1) = ESEM with targeted value = 0.1.

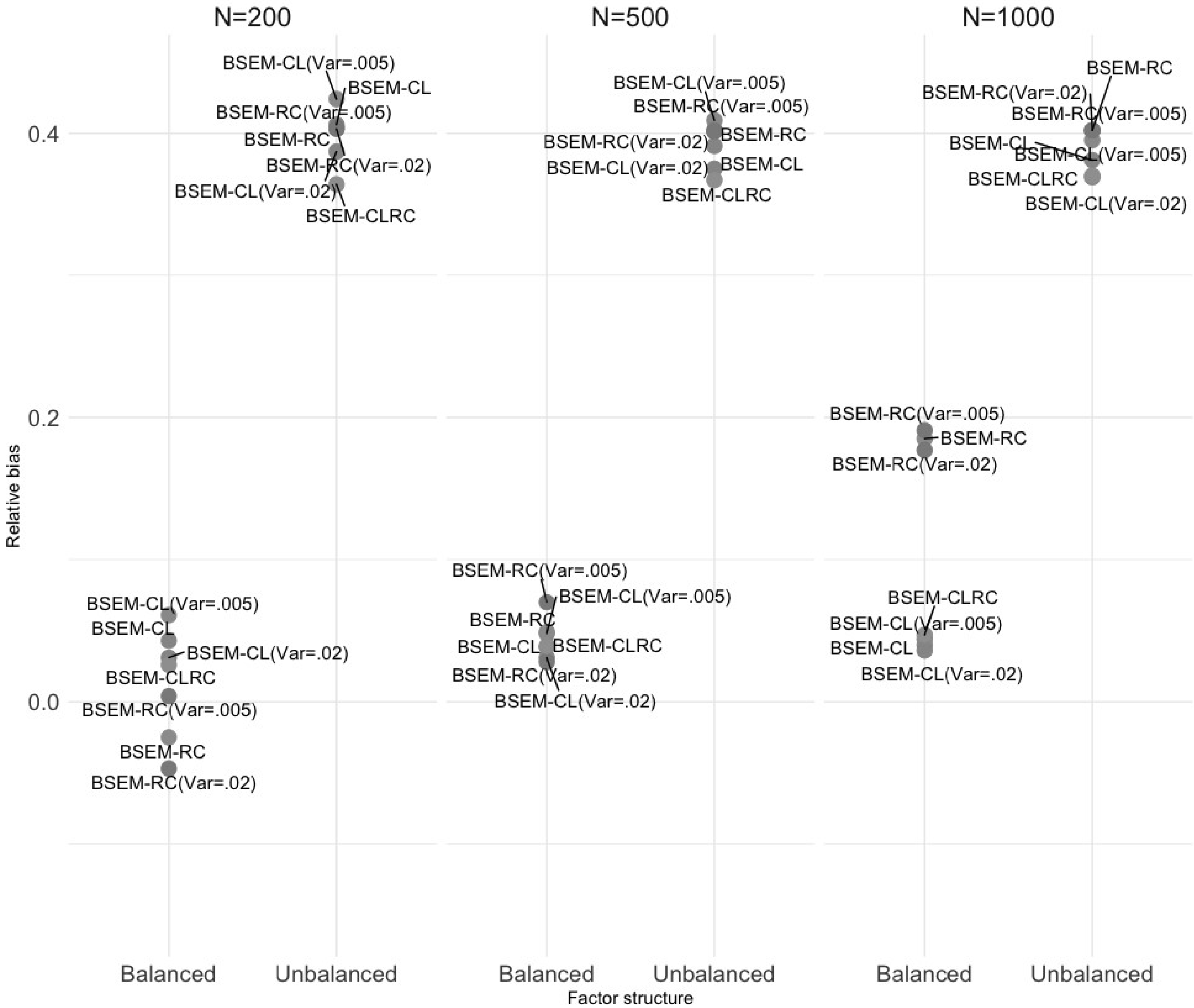

Figure 3. Relative bias of factor correlations across BESM models based on unbalanced (Design 2) and balanced factor structure (Design 3).

Note. BSEM-CL = BSEM+cross-loading priors (M = 0, Var = .01); BSEM-CL(Var=.005) = BSEM+cross-loading priors (M = 0, Var = .005); BSEM-CL(Var=.02) = BSEM+cross-loading priors (M = 0, Var = .02); BSEM-RC = BSEM+Residual covariances (M = 0, Var = .01); BSEM-RC(Var=.005) = BSEM+Residual covariances (M = 0, Var = .005); BSEM-RC(Var=.02) = BSEM+Residual covariances (M = 0, Var = .02); BSEM-CLRC= BSEM+Residual covariances and cross-loadings (M=0, Var=.01); BSEM-CL(M=1) = BSEM+cross-loading priors (M = 1, Var = .01).

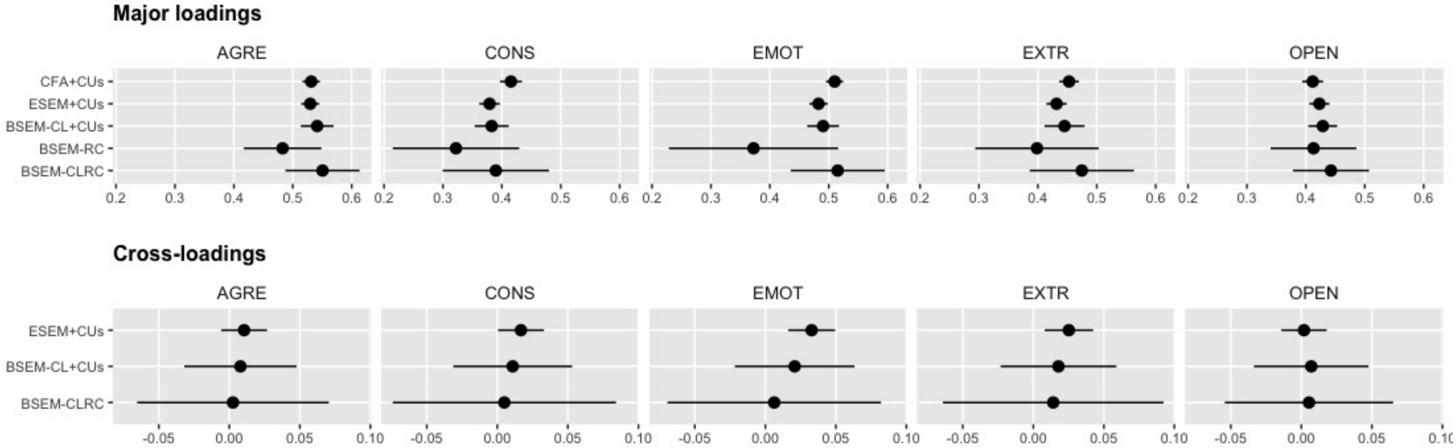

Figure 4. Factor loadings across models based on the Big-five Data.

Note. Dot points present average major loadings or cross-loadings for each factor; error bars present +/- SE of correlation coefficients for CFA and ESEM and +/- posterior SD for BSEM models. BSEM-CL = BSEM+cross-loading priors; BSEM-RC=BSEM+residual covariance priors; BSEM-CL RC=+cross-loading priors + residual covariance priors