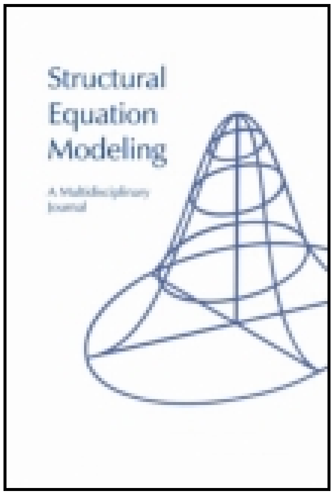

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

A Bayesian Approach to Multilevel Structural Equation Modeling With Continuous and Dichotomous Outcomes

Sarah Depaoli[a] & James P. Clifton[a]
[a] University of California, Merced
Published online: 03 Mar 2015.

Structural Equation Modeling: A Multidisciplinary Journal, 22: 327–351, 2015

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2014.937849

Routledge
Taylor & Francis Group

A Bayesian Approach to Multilevel Structural Equation Modeling With Continuous and Dichotomous Outcomes

Sarah Depaoli and James P. Clifton
University of California, Merced

Multilevel Structural equation models are most often estimated from a frequentist framework via maximum likelihood. However, as shown in this article, frequentist results are not always accurate. Alternatively, one can apply a Bayesian approach using Markov chain Monte Carlo estimation methods. This simulation study compared estimation quality using Bayesian and frequentist approaches in the context of a multilevel latent covariate model. Continuous and dichotomous variables were examined because it is not yet known how different types of outcomes—most notably categorical—affect parameter recovery in this modeling context. Within the Bayesian estimation framework, the impact of diffuse, weakly informative, and informative prior distributions were compared. Findings indicated that Bayesian estimation may be used to overcome convergence problems and improve parameter estimate bias. Results highlight the differences in estimation quality between dichotomous and continuous variable models and the importance of prior distribution choice for cluster-level random effects.

Keywords: Bayesian estimation, multilevel modeling, structural equation modeling

Multilevel modeling (MLM) and structural equation modeling (SEM) are gaining popularity in the social sciences because these statistical techniques offer the flexibility to answer a variety of complex questions (see, e.g., Hox & Roberts, 2011; Matsueda, 2012). MLM provides a framework for analyzing hierarchical data by accounting for variation due to the sampling of units at more than one level (Raudenbush & Bryk, 2002), whereas SEM provides a framework for testing causal relationships and accounting for error in the measurement of constructs in multivariate data by modeling means and covariances (Bollen, 1989). Although these approaches were developed for different purposes, the types of problems that motivate the use of MLM and SEM often cooccur.

For instance, educational studies that use MLM to examine student variation in academic ability across higher level units such as classrooms or schools frequently use data from tests that are not perfectly reliable measures of ability. Similarly, psychometric studies that use SEM to measure complex psychological phenomena are often conducted with hierarchical data such as those that arise in institutions (e.g., families, hospitals, organizations). Neither modeling framework is solely appropriate for these examples because each approach ignores an important component of the data, and their use could compromise the quality of model estimates. In the former example, biased estimates might be obtained if constructs measured by multiple indicators are represented with composite scores instead of latent variables (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011); in the latter, parameter estimates could be inaccurate if the hierarchical structure of the data is ignored (Julian, 2001; Muthén & Satorra, 1995). To deal with the shortcomings of each approach, methodologists have combined features of MLM and SEM into a flexible framework that extends conventional latent variable modeling to accommodate hierarchical data, called multilevel SEM (Gottfredson, Panter, Daye, Allen, & Wightman, 2009; Johnson, Burlingame, Olsen, Davies, & Gleave, 2005; Marsh et al., 2012).

Correspondence should be addressed to Sarah Depaoli, Assistant Professor of Quantitative Psychology, School of Social Sciences, Humanities, and Arts, University of California, Merced, Merced, CA 95343. E-mail: sdepaoli@ucmerced.edu

An important consideration in the application of multilevel SEM is determining the appropriate method of estimation. Multilevel SEMs are most often estimated in a frequentist framework using maximum likelihood (ML), which chooses a set of model parameters that maximize the likelihood of observing a set of outcomes in the data. However, due to the complexity of multilevel SEM, ML frequently leads to problems such as negative estimates for cluster-level variance components (Hox & Maas, 2001; Lüdtke et al., 2011; Meuleman & Billiet, 2009; Muthén & Satorra, 1995) and biased parameter estimates (Li & Beretvas, 2013; Lüdtke et al., 2008; Preacher, Zhang, & Zyphur, 2011). Therefore, ML might not always be a suitable estimation method in this modeling context. Alternatively, one can apply a Bayesian approach to multilevel SEM using Markov chain Monte Carlo (MCMC) estimation algorithms (Gelfand & Smith, 1990; Gilks, Richardson, & Spiegelhalter, 1996). MCMC methods do not require multidimensional numerical integration and can be used to estimate models that are intractable or computationally inefficient with ML (e.g., two-level models with random slopes for observed categorical variables; Asparouhov & Muthén, 2012). Aside from this estimation advantage, Bayesian methodology has unique benefits (discussed later) in the context of hierarchical and latent variable models.

AIMS OF THIS STUDY

At present, little research is available to aid researchers in the selection of priors for use with multilevel SEM. A small simulation study by Hox, van de Schoot, and Matthijsse (2012) examined sample size limits for multilevel SEM from a Bayesian framework in the context of cross-cultural research where countries represent cluster-level units. They found that Bayesian estimation with noninformative priors requires fewer countries to obtain accurate parameter estimates when compared with ML. However, the average group size in this setting was quite large ($>$ 1,000 observations per country), and the type of prior was not varied in this study to determine its impact on parameter recovery. It is currently unknown whether a Bayesian approach will provide unbiased estimates for group sizes that are typical in other social science settings (e.g., educational or psychological research), or the extent to which a fully Bayesian approach with informative priors will improve estimation accuracy.

Additionally, investigations into the accuracy and efficiency of multilevel SEM have focused exclusively on models with continuous outcomes (e.g., Julian, 2001; Lüdtke et al., 2011; Muthén & Satorra, 1995), even though multilevel data with dichotomous and ordered categorical outcomes are common in the social sciences. For instance, multilevel measurement modeling with categorical indicators represents the basis of multilevel item response theory (IRT; Fox, 2010; Fox & Glas, 2001; Muthén & Asparouhov, 2012b). There are a number of applications of multilevel SEM with categorical outcomes in the research literature using both frequentist (e.g., Gottfredson et al., 2009; Little, 2013; Mitchell & Bradshaw, 2013) and Bayesian (e.g., Diya, Li, Heede, Sermeus, & Lesaffre, 2013; Goldstein, Bonnet, & Rocher, 2007) approaches, yet simulation research has not been conducted to determine whether the sample size recommendations that have been put forth for continuous outcomes also hold for categorical outcomes. Moreover, a Bayesian approach is uniquely beneficial in the context of categorical data because it can be used to estimate categorical variable models that cannot be analyzed with currently available frequentist approaches (see, e.g., Ansari & Jedidi, 2000; Dunson, 2000; Muthén & Asparouhov, 2012b; Steele & Goldstein, 2006).

Accordingly, this article has two major aims that are explored via simulation. The first aim is to investigate the impact of prior distributions in the context of multilevel SEM using a number of prior distribution specifications. The second aim is to compare multilevel SEM with continuous versus dichotomous factor indicators in terms of accuracy, efficiency, convergence, and 95% confidence or credible interval coverage. This article is organized as follows. In the next section we briefly expand on the types of problems for which multilevel SEM is used. This is followed by a discussion of estimation issues, where we review research on ML estimation of multilevel SEM, challenges that might arise with a frequentist approach, and how Bayesian estimation may be used to circumvent these issues. Thereafter, we define the model used in the simulation, including the specification of prior distributions. We then present the design of the simulation, followed by the results of the study. Finally, in the Discussion section we compare the relative merits of Bayesian and frequentist approaches to multilevel SEM in the context of continuous and dichotomous models. The article concludes with recommendations on the specification of priors in practical applications of multilevel SEM.

MULTILEVEL SEM

In this section we provide a brief introduction to some common two-level SEMs with a single measurement occasion, but note that multilevel SEM is a general modeling approach that can also be applied to longitudinal data (e.g., Kaplan, Kim, & Kim, 2009) and data with more than two levels of nesting (e.g., Preacher, 2011). For a more in-depth description of multilevel SEMs we refer the reader to Rabe-Hesketh, Skrondal, and Zheng (2012).

One of the largest areas of application of multilevel SEM is multilevel measurement modeling (e.g., Dyer, Hanges, & Hall, 2005; F. Li, Duncan, Harmer, Acock, & Stoolmiller, 1998; Little, 2013; Toland & De Ayala, 2005), which is an extension of conventional factor analysis to multilevel data. An interesting feature of multilevel factor analysis is the ability to allow the factor structure to explain between-group variation, whereby different factor loading patterns or a different number of factors can be specified at each level. Multilevel SEM has also been developed to handle multilevel mediation (Muthén, 1989), including models with upper level mediator or outcome variables (e.g., Li & Beretvas, 2013; Preacher et al., 2011; Preacher, Zyphur, & Zhang, 2010). It also provides a framework for combining path analysis and factor analysis with multilevel data (for applications, see Cheung & Au, 2005; Everson & Millsap, 2004; Johnson et al., 2005). Some examples include the multilevel multiple indicator multiple cause (MIMIC) model (Finch & French, 2011; Morselli, Spini, & Devos, 2012), and the multilevel latent covariate model (Lüdtke et al., 2008; Lüdtke et al., 2011; Marsh et al., 2009; Rabe-Hesketh, Skrondal, & Pickles, 2004), which represents the main focus of this article.

A primary motivation for using MLM is the ability to study the substantive effects of group-level constructs, known as contextual effects (Bosker & Snijders, 1999). Different types of MLMs can be specified in which intercepts or slopes are allowed to vary randomly across contexts. Randomly varying intercepts or slopes can also be specified in a multivariate setting where latent variables are measured by multiple observed indicators (Asparouhov & Muthén, 2012). However, the interpretation of contextual effects becomes complicated if the measurement model is allowed to differ within and between groups because factors at each level of nesting could have a different meaning. If one assumes that the same constructs are being measured at each level, factor loadings can be constrained to be equal across levels to equate the metric of the latent variables. Simplifying the measurement model in such a manner is practically useful because it allows for a straightforward interpretation of contextual effects as a simple decomposition of the between-group and within-group effects (Kim & Yoon, 2011; Mehta & Neale, 2005). Forcing metric equivalence can also provide more accurate and efficient estimates of contextual effects than when composite scores are used in place of latent variables (Lüdtke et al., 2008; Lüdtke et al., 2011).

ESTIMATION

As with conventional latent variable modeling, determining the appropriate estimation method is an important step in the application of multilevel SEM. The choice of an estimator largely depends on the type of outcomes being modeled (e.g., continuous, categorical, censored, count), model complexity, and whether assumptions (e.g., normality) have been met.

For multilevel analysis with categorical outcomes, diagonally weighted least squares (DWLS) estimators have been developed to bypass the need for multidimensional numerical integration (e.g., Asparouhov & Muthén, 2007). DWLS is an approach that estimates model parameters based on the covariance structure instead of the individual data. Two popular DWLS estimators implemented in M*plus* include the weighted least squares with mean-adjusted (WLSM) and weighted least squares with mean and variance-adjusted (WLSMV; Muthén & Muthén, 1998–2013). Both estimators provide identical parameter estimates and are considered robust to violations of normality, but WLSMV differs in the calculation of the chi-square test statistic and associated degrees of freedom (Asparouhov & Muthén, 2007).

Asymptotically, ML and its variants are unbiased and efficient. However, a frequentist approach to multilevel SEM can lead to biased and variable estimates as well as problems with convergence when the number of groups and group sizes are small (Hox & Maas, 2001; Hox, Maas, & Brinkhaus, 2010; Lüdtke et al., 2011; Meuleman & Billiet, 2009; Preacher et al., 2011). One factor that interacts with sample size to affect these outcomes is strength of the intraclass correlation (ICC), which is an index of the proportion of variability explained at the cluster level. The ICC (ρ) is defined as:

$$\rho = \frac{Var(B)}{Var(B) + Var(W)}, \tag{1}$$

where $Var(B)$ denotes between-group variability and $Var(W)$ denotes variability within groups. It can be seen that as $Var(B)$ decreases, the ratio of between-group to total variability also decreases and the value of the ICC becomes smaller. A value of zero for ρ indicates that the data are independent. Research on the accuracy of multilevel SEM shows that estimates of within-level (e.g., individual-level) parameters are generally unbiased with a small number of groups, small group sizes, and low ICCs, but between-level (e.g., group-level) parameter estimates are not recovered accurately under these conditions (e.g., Finch & French, 2011; Hox & Maas, 2001; Lüdtke et al., 2011; Muthén & Satorra, 1995). Because it is not always feasible to collect data on a large number of groups (e.g., due to budget constraints), a number of Monte Carlo studies have been conducted to develop sample size recommendations for multilevel SEM.

Hox et al. (2010) looked at the properties of a number of frequentist estimators in the context of a multilevel confirmatory factor analysis (CFA) model, finding that robust maximum likelihood and the DWLS estimators implemented in *Mplus* produced accurate factor loading estimates with as few as 50 groups and an average group size of 5, although the DWLS approaches outperformed MLR in terms of 95% confidence interval coverage. Other studies have found that factor loading estimates are generally unbiased for multilevel CFA with a small number of groups (Hox & Maas, 2001; Muthén & Satorra, 1995; Wu & Kwok, 2012). For more complex models that combine measurement and structural effects, a larger number of groups are required for accurate parameter recovery (Finch & French, 2011; Lüdtke et al., 2008); Lüdtke et al., 2011). However, large group sizes (Cheung & Au, 2005; Meuleman & Billiet, 2009) and large ICCs (Lüdtke et al., 2011; Muthén & Satorra, 1995; Preacher et al., 2011; Wu & Kwok, 2012) have been shown to partially compensate for a small number of groups. By the same token, the factors that affect accuracy of multilevel SEM also hold for efficiency, whereby less variability is observed as the number of clusters, cluster sizes, and ICCs increase (Lüdtke et al., 2008; Lüdtke et al., 2011; Preacher et al., 2011).

Convergence is a major concern in the estimation of multilevel SEM (see, e.g., Johnson et al., 2005; Toland & De Ayala, 2005). Inadmissible estimates (e.g., negative variances) and non-convergence are frequently encountered during estimation, and these problems become worse with small group sizes, a small number of groups, and low ICCs (Hox & Maas, 2001; Li & Beretvas, 2013; Lüdtke et al., 2011; Meuleman & Büliet, 2009; Muthén & Satorra, 1995; Ryu, 2011).

Despite the growing number of applications of multilevel SEM with categorical variables (e.g., Diya et al., 2013; Goldstein et al., 2007; Gottfredson et al., 2009; Little, 2013; Mitchell & Bradshaw, 2013; Steele & Goldstein, 2006), the simulation studies that have been conducted to develop lower bounds on sample size in multilevel SEM have been based solely on models with continuous variables. It is not yet known how different types of outcomes—most notably, categorical—affect parameter recovery in multilevel SEM. In theory, robust DWLS estimators should perform well for multilevel SEM with categorical outcomes when there is a relatively small number of clusters (e.g., 50–100 groups; Asparouhov & Muthén, 2007). DWLS has been found to be almost as accurate as ML in the context of a two-level CFA model with continuous indicators (Hox et al., 2010), but the accuracy and efficiency of DWLS for categorical indicators has not been studied under various conditions of group size, number of groups, and ICC. At the same time, convergence rates have not been investigated in this context. For SEM with categorical factor indicators, convergence rates appear to become worse with a decreasing number of categories per indicator (Yang-Wallentin, Jöreskog, & Luo, 2010), so it follows that problems with convergence might be more pronounced for multilevel SEMs with categorical versus continuous outcomes. In short, frequentist approaches could lead to biased and variable estimates and problems with convergence when little information is available in the data (i.e., a small number of clusters, small cluster sizes, and low ICCs). In the next section, we briefly explain how a Bayesian approach could be used to avoid some of the issues that arise during estimation of multilevel SEM, borrowing from the literature on the use of Bayesian procedures in MLM and SEM.

BENEFITS OF A BAYESIAN APPROACH FOR MULTILEVEL SEM

Bayesian estimation of SEMs (Kaplan & Depaoli, 2012; Lee, 2007; Muthén & Asparouhov, 2012a; Scheines, Hoijtink, & Boomsma, 1999) and MLMs (Congdon, 2010; Gelman & Hill, 2007; Lynch, 2007) is a prevalent methodological topic because Bayesian methods are useful for solving problems that are unique to these modeling frameworks. Compared with ML, Bayesian estimation of hierarchical models can produce more accurate and efficient parameter estimates when the data consist of a small number of groups (Asparouhov & Muthén, 2010; Baldwin & Fellingham, 2013). Likewise, bias and variability can be improved in a Bayesian framework because the data and the prior borrow strength from one another, shrinking estimates toward the prior mean (Gelman & Hill, 2007). Another problem common to MLM is that small ICCs can lead to estimates of zero for cluster-level variances. In these cases, ICC estimates will suggest that the data are independent, and researchers who interpret these estimates might consequently—and misleadingly—believe that the hierarchical structure of the data can be ignored. Disregarding the multilevel nature of data with small but nonzero ICCs is problematic because it can cause an inflation of Type I errors (Baldwin et al., 2011; Murray & Hannan, 1990; Siddiqui, Hedeker, Flay, & Hu, 1996). In a Bayesian analysis, boundary estimates are not an issue because variance components can be given priors that shrink estimates away from zero (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). Obtaining proper estimates for variances can also be an issue in the context of SEM. Bayesian estimation can be used to prevent problems with inadmissible solutions that are caused by negative variance estimates (Lee, 1981; Martin & McDonald, 1975), which frequently occur during estimation with small samples (Boomsma, 1987).

Taken together, these findings suggest that a Bayesian approach to multilevel SEM should produce more accurate and efficient estimates because of shrinkage. It should also eliminate problems with convergence due to negative variance estimates because priors can be used to bound estimates to positive values. For more information on Bayesian analysis, we refer the reader to Kaplan and Depaoli (2013), Rupp, Dey, and Zumbo (2004), and van de Schoot et al. (2014) for a simple introduction; and Gelman, Carlin, Stern, & Rubin, (2004) for a more technical treatment of the topic.

THE MODEL CONSIDERED IN THIS INVESTIGATION

The multilevel SEM used in the simulations reported here consists of a latent covariate and a latent response variable, each of which is measured by three indicators with randomly

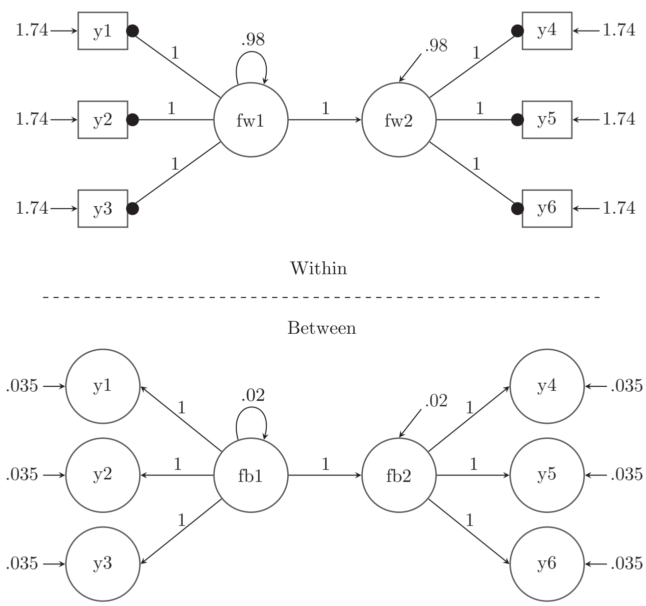

FIGURE 1 Data generating model for continuous indicators with an intraclass correlation coefficient of .02.

varying intercepts. Similar models have been presented by Lüdtke et al. (2011), Marsh et al. (2009), and Rabe-Hesketh et al. (2004). Figure 1 depicts the model for one of the conditions defined in the Design section. In the within–between framework (Muthén, 1989, 1994), the covariance structure is partitioned into within-group and between-group components and separate models are estimated for each component. The within-group component (denoted by a W subscript) represents unit-level variation and the between-group component (denoted by a W subscript) represents variation between clusters. The response vector $\mathbf{y}_{ij}$ for observation i in cluster j is decomposed as:

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \mathbf{y}_{Wij} + \mathbf{y}_{Bj}, \tag{2}$$

where the within-cluster $\mathbf{y}_{Wij}$ and between-cluster $\mathbf{y}_{Bj}$ components are orthogonal and independent, and $\boldsymbol{\mu}$ represents the grand means. The item responses $\mathbf{y}_{ij}$ are normally distributed with cluster means $\boldsymbol{\mu}_j$ as the expected value (i.e., random intercepts) and covariance matrix $\boldsymbol{\Sigma}_W$. The random effects $\boldsymbol{\mu}_j$ are normally distributed with expected value $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_B$.

For observed dichotomous indicators, normally distributed continuous latent variables denoted by $\mathbf{y}^*_{ij}$ are assumed to underlie the category choice (Asparouhov & Muthén, 2007). For $k = 1, \ldots, K$ dichotomous indicators, a threshold specification is used to link the underlying latent response variable y^*_{ijk} to the observed indicators, whereby $y^*_{ijk} = 1$ if $y_{ijk} > 0$ and $y^*_{ijk} = 0$ otherwise.

Following conventional SEM, the between-group and within-group components each consist of separate measurement and structural models. The measurement models relate the observed indicators to the latent constructs and the structural models relate the latent constructs to one an other. For the within-group component, the measurement model depicted in Figure 1 is given by:

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \boldsymbol{\Lambda}_W \eta_{Wij} + \boldsymbol{\varepsilon}_{Wij} \tag{3}$$

where $\boldsymbol{\Lambda}_W$ is a 6×2 factor loading matrix that is postmultiplied by a 2×1 vector of unit-level latent variables $\boldsymbol{\eta}_{Wij}$ that contains the latent response variable and the latent covariate. The $\boldsymbol{\eta}_{Wij}$ vector is distributed multivariate normal with an expectation of zero and a 2×2 covariance matrix $\boldsymbol{\Psi}_W$. Because the latent variables are uncorrelated, the $\boldsymbol{\Psi}_W$ matrix contains the factor variances along the diagonal and the off-diagonal elements consist of zeros. The $\boldsymbol{\varepsilon}_{Wij}$ term denotes a 6×1 vector of residuals that are distributed multivariate normal with an expectation of zero and a diagonal covariance matrix $\boldsymbol{\Theta}_W$, with error terms along the diagonal. The structural model for the within-group component has the form:

$$\boldsymbol{\eta}_{Wij} = \mathbf{B}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\zeta}_{Wij}, \tag{4}$$

where $\mathbf{B}_W$ is a 2×2 matrix of structural coefficients defined as:

$$\begin{bmatrix} 0 & b_W \\ 0 & 0 \end{bmatrix},$$

where b_W represents the regression of the outcome on the covariate. The $\mathbf{B}_W$ matrix is postmultiplied by the vector of unit-level latent variables to obtain the estimate of the covariate effect. Finally, $\boldsymbol{\zeta}_{Wij}$ is a 2×1 vector of errors distributed multivariate normal with an expected value of zero and a diagonal covariance matrix $\boldsymbol{\Omega}_W$, with error terms along the diagonal.

The measurement model and the structural model at the cluster level are respectively given by:

$$\boldsymbol{\mu}_j = \boldsymbol{\mu} + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\varepsilon}_{Bj}, \tag{5}$$

$$\boldsymbol{\eta}_B = \mathbf{B}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\zeta}_{Bj}. \tag{6}$$

Here, $\boldsymbol{\Lambda}_B$, $\boldsymbol{\eta}_{Bj}$, $\boldsymbol{\varepsilon}_{Bj}$, $\mathbf{B}_B$, and $\boldsymbol{\zeta}_{Bj}$ are the between-group terms corresponding to the within-group terms $\boldsymbol{\Lambda}_w$, $\boldsymbol{\eta}_{wij}$, $\boldsymbol{\varepsilon}_{wij}$, $\mathbf{B}_{\mathrm{w}}$, and $\boldsymbol{\zeta}_{wij}$. Additionally, the covariance matrices $\boldsymbol{\Psi}_B$, $\boldsymbol{\Theta}_B$, and $\boldsymbol{\Omega}_B$ are the cluster-level counterparts to the unit-level covariance matrices $\boldsymbol{\Psi}_W$, $\boldsymbol{\Theta}_W$, and $\boldsymbol{\Omega}_W$.

Specification of Priors

The extent to which parameters are recovered accurately in a Bayesian analysis depends in large part on the quality and amount of information modeled in the prior. There are three main categories of priors that are typically discussed in terms of level of informativeness: noninformative (used interchangeably here with *diffuse*), weakly informative, and informative. Diffuse priors contribute vague information to a model to integrate full uncertainty into the estimation process. Weakly informative priors contain more information compared to diffuse, but use less information than is available as to exhibit some degree of uncertainty (Gelman, 2006). Finally, an informative prior incorporates a great deal of certainty about the value of the model parameter. Weakly informative priors might be preferred to informative priors when it is unclear how to construct a fully informative model and it is therefore appropriate to allow the likelihood to contribute more information than the prior (Chung et al., 2013). Additionally, weakly informative priors could be used as a procedure for regularization or to restrict estimates to a reasonable range of values (Gelman, Jakulin, Pittau, & Su, 2008).

A successful Bayesian analysis requires a careful selection of priors, even if priors are meant to be noninformative. Priors are quantified with the specification of hyperparameters that typically control the location and scale of the distribution. For example, a prior for a normally distributed variable has a mean hyperparameter and a variance hyperparameter that correspond to the distribution's location and scale, respectively. The variance hyperparameter for a noninformative prior is likely to be much larger than that of an informative prior to reflect greater uncertainty about the range of values the parameter can take on. With large sample sizes, noninformative priors have little impact on posterior inference because the prior becomes overpowered by the likelihood. On the other hand, model estimates are more sensitive to prior specification with small samples because fewer data are available for estimation.

For the current model, parameters of interest include the regression and variance parameters at each level of the model. Although the following priors are described as being multivariate, the M*plus* software program used here implements univariate priors on the individual elements of a vector. We use multivariate notation here simply to reduce notation in the article, but it is important to note that all priors implemented were univariate as described in Muthén and Muthén (1998–2013).

The prior for the vector of regression parameters is the multivariate normal distribution. For the within-group regression parameters $\boldsymbol{\delta}_w = (\boldsymbol{\Lambda}_w, b_w)$, the prior is denoted as:

$$\boldsymbol{\delta}_w \sim MVN\left(\boldsymbol{\nu}_w, \boldsymbol{\sigma}_w^2\right),$$

where $\boldsymbol{\nu}_w$ is a vector containing the mean hyperparameters and $\boldsymbol{\sigma}_w^2$ is a vector containing the variance hyperparameters. Correspondingly, the prior for the regression parameters $\boldsymbol{\delta}_B = (\boldsymbol{\Lambda}_B, b_B)$ at the cluster level is given by:

$$\boldsymbol{\delta}_B \sim MVN\left(\boldsymbol{\nu}_B, \boldsymbol{\sigma}_{B^2}\right),$$

where $\boldsymbol{\nu}_B$ and $\boldsymbol{\sigma}_B^2$ represent vectors with the mean and variance hyperparameters, respectively. The prior for the variances is the inverse-gamma distribution, which has shape (α) and scale (β) parameters that control the density. Inverse-gamma priors are specified for the diagonal elements of the covariance matrices associated with the variances at each level of the model. At the within-group level, the priors for each of the variance parameters are given by:

$$\boldsymbol{\Theta}_W \sim \Gamma^{-1}\left(\alpha_{w_1}, \beta_{w_1}\right) \text{ and } \boldsymbol{\Psi}_w \sim \Gamma^{-1}\left(\alpha_{w_2}, \beta_{w_2}\right),$$

where α_w is the shape hyperparameter and $\boldsymbol{\beta}_w$ is the scale hyperparameter (for ease of notation, $\boldsymbol{\tau}_w = \left(\left[diag\right]\boldsymbol{\Theta}_w, \left[diag\right]\boldsymbol{\Psi}_w\right)\right)$. The prior for each between-group variance component is similarly denoted as:

$$\boldsymbol{\Theta}_B \sim \Gamma^{-1}\left(\alpha_{B_1}, \beta_{B_1}\right) \text{ and } \boldsymbol{\Psi}_B \sim \Gamma^{-1}\left(\alpha_{B_2}, \beta_{B_2}\right),$$

where α_B and β_B are the shape and scale hyperparameters, respectively (for ease of notation, $\boldsymbol{\tau}_B = \left(\left[diag\right]\boldsymbol{\Theta}_B, \left[diag\right]\boldsymbol{\Psi}_B\right)\right)$.

In the context of MLM, parameter estimates are particularly sensitive to prior specifications for cluster-level variance components (see, e.g., Asparouhov & Muthén, 2010; Baldwin & Fellingham, 2013; Browne & Draper, 2006; Gelman, 2006). As a result, a wide range of Γ^{-1} specifications have been considered in the research literature. One such specification is the Γ^{-1} (.001, .001) prior, which has been implemented in many examples using the BUGS software as a noninformative prior for between-level variances (Spiegelhalter, Thomas, Best, Gilks, & Lunn, 2003). However, findings suggest that setting the α and β hyperparameters to such small values might actually have a greater impact on parameter estimates—and therefore contribute more information—than originally thought (Asparouhov & Muthén, 2010; Gelman, 2006). Specifically, the Γ^{-1} (.001, .001) specification gives more weight to smaller values and can lead to an underestimation of large hierarchical variance parameters. Uniform priors have also been considered as a noninformative prior distribution for cluster-level variance components in multilevel models (e.g., Asparouhov & Muthén, 2010). However, uniform priors are also problematic in that they tend to overestimate small hierarchical variance parameters (Gelman, 2006). In light of these findings, a major focus of this study was to investigate the impact that specifying uniform and Γ^{-1} (.001, .001) priors (as well as other forms of Γ^{-1}) for cluster-level variance components have on parameter estimates in the context of multilevel SEM at varying levels of ICC.

DESIGN

Data were generated using M*plus* Version 7.11 (Muthén & Muthén, 1998–2013) with 1,000 repetitions per cell; code is available upon request. Bayesian analysis was carried out using a single MCMC chain comprising 50,000 iterations with the first 25,000 discarded as burn-in and the remaining 25,000 used to estimate the posterior distribution, although convergence might have been obtained sooner via the convergence criterion implemented. Chain convergence was monitored through the Brooks and Gelman (1998) convergence diagnostic as implemented in the M*plus* software program, with a default convergence criterion of 0.05. The data-generating model consisted of a latent explanatory variable and a latent response variable at the individual and group levels, with each latent variable measured by three continuous or dichotomous factor indicators. For the purpose of model identification, the first indicator of each latent variable was fixed to 1.0 in all conditions and the within-group residual variances ($\boldsymbol{\varepsilon}_{wij}$) were fixed to 1.0 for the dichotomous probit model. Conditions varied in this study were the number of clusters (4 levels), cluster size (3 levels), ICC (5 levels), type of factor indicator (2 levels), invariance of factor loadings (2 levels), and estimator/prior specification (9 levels).[1]

Number of clusters. Monte Carlo investigations have reported different findings regarding the number of groups that are necessary for obtaining unbiased estimates in multilevel SEM (see e.g., Hox et al., 2010; Li & Beretvas, 2013; Lüdtke et al., 2011; Preacher et al., 2011). As a result, we looked at a range of clusters; the number of clusters were specified as 40, 50, 100, and 200.

Cluster size. For each cluster, sample sizes of 5, 10, and 20 were generated to reflect group sizes typically seen in educational and psychological studies (e.g., Johnson et al., 2005; Marsh et al., 2012; Mathisen, Torsheim, & Einarsen, 2006).

ICC. Values of .02, .05, .10, .20, and .40 were chosen for the ICC of the observed indicators. Previous simulation studies on multilevel SEM have typically focused on ICCs between .05 and .20 (e.g., Hox et al., 2010; Lüdtke et al., 2011; Muthén & Satorra, 1995; Preacher et al., 2011). However, smaller ICCs are regularly seen in educational and behavioral research (e.g., Muthén, 1991), and ignoring the hierarchical structure of data with ICCs as small as .02 can lead to inflated Type I errors (Baldwin et al., 2011; Murray & Hannan, 1990; Siddiqui et al., 1996).

In multilevel SEM, the ICC for each observed indicator (ρ_k) can be expressed as a function of the factor loadings, variances, and residual variances at each level of the model using the following formula (Muthén, 1991):

$$\rho_k = \left(\lambda_{Bk}^2 \times \boldsymbol{\Psi}_B + \boldsymbol{\Theta}_B\right) / \left[\left(\lambda_{Bk}^2 \times \boldsymbol{\Psi}_B + \boldsymbol{\Theta}_B\right) + \left(\lambda_{Wk,}^2 \times \boldsymbol{\Psi}_W + \boldsymbol{\Theta}_W\right)\right], \tag{7}$$

where λ_k refers to the factor loading for item *k*. The ICC levels for this study were created using Equation 7 by changing the variances and residual variances at each level of the model while setting factor loadings at a constant value of 1.0. Thus, factor loadings had the same value of ρ_k in each ICC condition. Note that the within-group residual variances were fixed to 1.0 for the dichotomous model because estimation was carried out with probit regression. As a reference, Figure 1 displays parameter values for the continuous outcome model with ICC = .02.

Indicator type. The type of factor indicator is a condition that has not been systematically varied in previous simulation studies examining multilevel SEM. In this study, we compared models with continuous or dichotomous outcomes; other types of outcomes were not considered to keep the number of simulations manageable.

Invariance of loadings. Two conditions were specified with respect to the factor structure at each level of the model. In one condition, factor loadings were freely estimated on the within and between levels. In the other condition, factor loadings were held invariant such that $\boldsymbol{\Lambda}_W = \boldsymbol{\Lambda}_B = \boldsymbol{\Lambda}$ to force metric equivalence of the latent variables (see Lüdtke et al., 2008).

Estimator. Frequentist estimation was accomplished using ML with robust standard errors and chi-squares. Estimation with continuous indicators was performed using the robust full information ML (FIML) estimator (robust ML, or MLR) in M*plus,* whereas estimation with dichotomous indicators was performed using a robust DWLS estimator (WLSM).[2]

[1]The reader should be aware that the terminology *level* is used in two different contexts for the remainder of the article. First, we continue to use level to denote the different nested components in the hierarchical model, in which the lower level represents within-group variation and the highest level represents variation between groups. Second, we also use the term level to denote the different categories of a factor (e.g., factor: average cluster size, levels: $N_j = 5, 10,$ and 20) to keep with conventional terminology used to discuss simulation factors. The context is always stated to distinguish between the different uses of this term.

[2]We chose to use WLSM instead of WLSMV in this study because both estimators result in identical parameter estimates, but WLSM requires one fewer step and was therefore more efficient for the purpose of the simulation. In addition, we used a different ML estimator for each type of indicator for two reasons. First, we selected a DWLS approach for dichotomous indicators because FIML estimation would have required eight dimensions of integration and could not be performed in M*plus.* Second, we chose MLR

Bayesian estimation was carried out using MCMC estimation with a Gibbs sampling algorithm (Geman & Geman, 1984). Three types of prior distributions were specified for Bayesian estimation: noninformative (diffuse), informative, and weakly informative, with the latter consisting of six levels that were defined by varying the precision of the priors. There are two points worth mentioning with respect to consistencies in prior distribution specifications across Bayesian estimation levels. First, regression parameters were specified to have the same prior across the within-group and between-group components of the model. Second, variance parameters for the within-group component of the model were always specified with the elements in $\boldsymbol{\tau}_W \sim \Gamma^{-1}(-1, 0)$, which is the default in M*plus* and corresponds to a uniform prior of $U[0, \infty)$ (Asparouhov & Muthén, 2010). Diffuse priors were defined using the default specification in M*plus,* whereby $\boldsymbol{\delta}_W$, $\boldsymbol{\delta}_B \sim N(0, 10^{10})$ for continuous indicators and $\boldsymbol{\delta}_W$, $\boldsymbol{\delta}_B \sim N(0, 5)$ for categorical indicators, and the elements in $\boldsymbol{\tau}_B \sim \Gamma^{-1}(-1, 0)$ for both types of indicator.

For Bayesian estimation with informative and weakly informative priors, the mean hyperparameter of the regression priors was specified at the true population value of 1.0 to give the regression priors the greatest mass in the general neighborhood of the generating values. The weakly informative prior distribution levels were varied in two ways. First, three levels of informativeness were specified for the regression priors via the variance hyperparameter such that $\boldsymbol{\delta}_W$ and $\boldsymbol{\delta}_B \sim N(1, 1)$, $N(1, .5)$, or $N(1, .25)$. Second, the variance components at the cluster level were varied to have either $\Gamma^{-1}(-1, 0)$ or $\Gamma^{-1}(.001, .001)$ priors. To study the interaction of different forms of priors, both of these conditions were crossed, resulting in six levels of weakly informative prior distributions. With respect to the informative prior distribution level, the regression priors were specified to have a tighter variance hyperparameter than the weakly informative prior distribution levels such that $\boldsymbol{\delta}_W$, $\boldsymbol{\delta}_B \sim N(1, .1)$ and the cluster-level variances were specified with $\Gamma^{-1}(.001, .001)$ priors.

Priors for continuous regression parameters can be thought of in terms of their 95% limits to construct plausible bounds. The informative and weakly informative prior distribution levels considered in this study were constructed in this manner. For instance, the weakly informative prior distribution level where $\boldsymbol{\delta}_W$, $\boldsymbol{\delta}_B \sim N(1, .25)$ has a 95% bound of $1 \pm 1.96 \times \sqrt{.25} = .02$ to 1.98, which we deemed a reasonable range for applied settings. Figure 2 depicts the 95% limits of the informative and weakly informative prior distribution levels for the continuous model.

for continuous indicators because it is more commonly used than asymptotic ML or DWLS approaches in applications of multilevel SEM with continuous outcomes (e.g., Marsh et al., 2012; Roesch et al., 2010; Thoonen, Sleegers, Oort, Peetsma, & Geijsel, 2011).

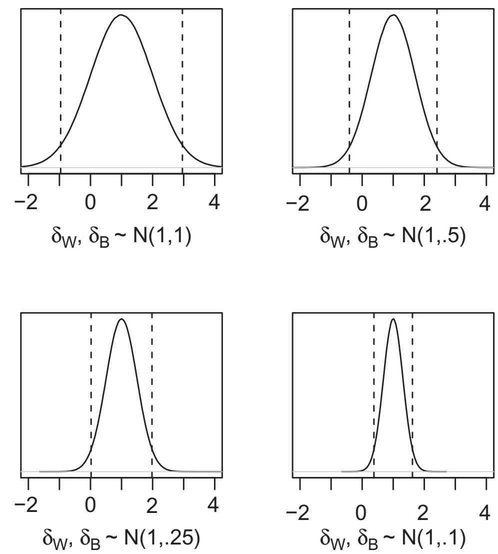

FIGURE 2 95% limits of the informative and weakly informative prior distributions for the continuous model.

Data Analysis

We present results for the structural path linking the latent covariate to the latent response variable at each level of the model together with select results on the accuracy of factor loading estimates. We assessed accuracy of the covariate effect by computing relative percentage bias, and we measured accuracy of the factor loadings by computing the mean absolute bias of the freely estimated factor loadings.[3] In addition, we present the root mean square error (RMSE) for the covariate effect as a composite measure of accuracy and efficiency. In the presence of parameter estimate bias, RMSE combines bias and variability of the estimated covariate effect around the population covariate effect. We also present findings on Type I error coverage for the structural path, calculated as the proportion of replications in which the true covariate effect was captured by the confidence or credible interval. Finally, we calculated convergence rates by computing the percentage of replications with admissible estimates in each condition.[4]

[3]Relative percentage bias was computed using the formula $\left[(\hat{b} - b)\big/b\right] \times 100$, where $\hat{b}$ denotes the average covariate effect estimate across simulated data sets and b denotes the population covariate effect. Absolute bias was calculated as $\left[\frac{1}{n}\sum_{k=1}^{n}\left(\hat{\lambda}_k - \lambda_k\right)\big/\lambda_k\right] \times 100$, where $\hat{\lambda}_k$ denotes the factor loading estimate for item k.

[4]The most common reason for nonconvergence was the occurrence of negative estimates for cluster-level variances. Solutions with negative variance estimates are useless in practice, and in the context of Monte Carlo studies, including replications with negative variance estimates can bias simulation results. In M*plus,* repetitions with negative variance estimates were not automatically discarded when averaging results over simulated data sets. To deal with this issue, we saved the repetitions for each of the study conditions and manually removed solutions with negative variance estimates. The external Monte Carlo feature of M*plus* was subsequently used to rerun the simulations with only the admissible solutions.

TABLE 1
Percentage of Admissible Solutions for the Frequentist Estimation Conditions

		Noninvariant Factor Loadings							
		Continuous Indicators				*Categorical Indicators*			
		$J = 40$	$J = 50$	$J = 100$	$J = 200$	$J = 40$	$J = 50$	$J = 100$	$J = 200$
$N_j = 5$	ICC = .02	11.2	13.4	11.6	14.1	0.1	0.0	0.1	0.1
	ICC = .05	6.7	7.3	10.1	21.9	0.2	0.1	0.7	2.6
	ICC = .10	8.6	11.2	61.7	61.5	1.0	1.5	5.5	20.1
	ICC = .20	24.6	36.1	72.8	94.6	7.0	13.3	40.0	78.1
	ICC = .40	63.3	74.4	96.8	100.0	38.4	52.7	86.6	98.5
$N_j = 10$	ICC = .02	9.6	8.2	10.9	19.1	0.1	0.3	0.8	1.5
	ICC = .05	8.7	37.0	29.3	56.9	0.4	1.4	4.7	18.9
	ICC = .10	24.3	33.3	67.4	93.6	7.3	12.4	37.0	74.7
	ICC = .20	52.9	65.4	93.6	99.4	36.1	49.1	83.0	98.5
	ICC = .40	78.4	93.7	98.7	99.9	71.2	82.5	98.0	100.0
$N_j = 20$	ICC = .02	6.6	7.4	19.6	45.6	0.7	0.6	4.3	15.9
	ICC = .05	19.8	28.5	65.4	93.7	5.7	10.6	36.7	75.7
	ICC = .10	49.5	58.8	92.9	99.9	33.7	47.1	85.8	97.9
	ICC = .20	87.5	83.1	98.6	99.9	67.5	79.8	98.5	100.0
	ICC = .40	83.6	91.6	99.7	99.9	86.4	93.8	99.8	100.0
		Invariant Factor Loadings							
		Continuous Indicators				*Categorical Indicators*			
		$J = 40$	$J = 50$	$J = 100$	$J = 200$	$J = 40$	$J = 50$	$J = 100$	$J = 200$
$N_j = 5$	ICC = .02	25.2	29.4	39.7	45.7	0.0	0.0	0.0	0.6
	ICC = .05	23.6	27.6	39.3	54.2	0.5	0.6	1.8	5.7
	ICC = .10	29.1	37.7	61.7	82.4	2.7	4.2	13.2	34.3
	ICC = .20	56.8	70.9	88.8	97.1	14.5	19.1	51.1	83.8
	ICC = .40	85.5	90.2	98.0	99.9	44.7	57.2	87.0	98.6
$N_j = 10$	ICC = .02	24.3	27.2	38.7	54.2	0.1	0.5	1.4	2.9
	ICC = .05	32.3	37.0	59.7	80.7	1.9	2.4	12.5	29.4
	ICC = .10	58.4	64.5	87.0	97.1	14.6	18.8	53.4	84.7
	ICC = .20	81.2	87.3	97.1	99.4	49.7	59.3	91.7	98.4
	ICC = .40	89.3	93.7	98.8	99.9	78.0	85.4	98.1	99.9
$N_j = 20$	ICC = .02	29.1	35.3	57.5	72.5	1.5	2.0	7.9	25.0
	ICC = .05	54.1	64.6	87.3	97.5	14.1	21.2	51.1	84.8
	ICC = .10	78.8	84.5	96.5	99.9	52.4	62.0	90.6	99.3
	ICC = .20	87.5	91.9	98.8	99.9	80.1	86.8	98.2	99.9
	ICC = .40	91.8	94.7	99.6	99.9	91.1	94.7	99.7	100.0

Note. ICC = intraclass correlation coefficient; J = number of groups; N_j = group size.

RESULTS

Convergence

All of the requested replications converged for the Bayesian estimation conditions for the continuous model, and convergence rates were nearly perfect for Bayesian estimation of the categorical model (> 98%, on average). Bayesian estimation resulted in poor convergence rates only in the categorical model for the diffuse prior distribution level and the weakly informative prior distribution levels with a uniform prior specified for cluster-level variances (i.e., $\Gamma^{-1}(-1, 0)$). However, these low convergence rates for Bayesian estimation occurred only when there were a small number of clusters (e.g., $J = 40$), small cluster sizes (e.g., $N_j = 5$), and extremely low ICCs (e.g., .02), with convergence improving substantially at larger ICCs. Inadmissible solutions were quite problematic for frequentist estimation conditions, with the convergence rate being noticeably lower for the categorical model compared with the continuous model (see Table 1 for frequentist convergence rates). The percentage of converged solutions in the frequentist estimation conditions was positively related to the number of clusters, cluster size, and strength of the ICC. For the continuous model, convergence rates improved considerably when factor loadings were held invariant. Note that results for the cells of frequentist estimation with low convergence rates should be interpreted with caution due to the possibility of sample selection bias.

Within-Group Component

Results for the within level of the model are presented only in text because the majority of estimation problems occurred on the between level. However, tables of the within-level findings are available from the authors on request.

Continuous indicator model. Consistent with previous research on multilevel SEM (e.g., Finch & French, 2011; Hox & Maas, 2001; Lüdtke et al., 2011; Muthén & Satorra, 1995), frequentist estimation of the continuous model resulted in unbiased estimates of the within-level factor loadings and covariate effect. Likewise, the within-level parameter estimates were recovered accurately for each level of Bayesian estimation. In the continuous model, Bayesian estimation with informative priors resulted in nominal 95% coverage of the covariate effect across all conditions. For the diffuse and weakly informative prior distribution levels, coverage of the structural path was negatively related to the average cluster size and ICC. The average operating alpha level was somewhat high for Bayesian estimation with diffuse priors (.06), whereas coverage was good for each of the weakly informative prior distribution levels except when the group size was 10 or larger and the ICC $\geq$ .20. For frequentist estimation, coverage of the covariate effect was negatively related to average group size, whereby good coverage was observed for $N_j = 5$, but coverage was slightly low (.94) for larger group sizes.

Categorical indicator model. For the categorical model, Bayesian estimation with informative priors resulted in accurate estimates of the within-level factor loadings and covariate effect across study conditions. Similarly, Bayesian estimation with weakly informative priors produced accurate parameter estimates across the majority of study conditions. Regarding frequentist estimation, estimates of the within-level factor loadings and covariate effect displayed problematic bias for frequentist estimation when $J \leq 50$ and $N_j = 5$. For Bayesian estimation with diffuse priors, the within-level factor loadings were generally estimated with greater accuracy than the within-level covariate effect. When the cluster size was small $\left(N_j = 5\right)$, Bayesian estimation with diffuse priors required 100 clusters to obtain accurate estimates of the within-level factor loadings and 200 clusters to obtain an accurate estimate of the within-level covariate effect. When $N_j \geq 10$, Bayesian estimation with diffuse priors often resulted in problematic bias for the structural path at $J = 40$ and $J = 50$, whereas bias for the factor loadings at $J = 40$ and $J = 50$ was generally acceptable. With respect to coverage of the within-level structural path, Bayesian estimation with informative priors was at the nominal 95% level across conditions, and the average level of coverage for frequentist estimation was close to the nominal level. In contrast, coverage was poor for Bayesian estimation with diffuse and weakly informative priors, and became worse as the sample size per group and ICC increased. In addition, coverage was better for the weakly informative prior distribution levels in which the regression priors were more informative and cluster-level variances were specified with Γ^{-1} (.001, .001) priors.

Between-Group Component

Overall, Bayesian estimation with informative priors considerably outperformed the other estimation conditions at the between level of the model. Across all conditions, the informative prior distribution level resulted in negligible bias of the between-group factor loadings and covariate effect, as well as small values of RMSE and good credible interval coverage of the cluster-level structural path. When diffuse priors were implemented, some of the parameter estimates were quite biased, as reported later. On inspecting trace plots, it was clear that spikes (i.e., extreme or out-of-bound estimates at a given MCMC iteration) were present. Spikes are likely responsible for the extreme levels of bias that are reported for some of the conditions and we discuss the handling of spikes in applied research contexts in the section, “Concluding Remarks and Recommendations for Applied Researchers.”

Bias of the covariate effect. Relative percentage bias of the between-level structural effect is presented in Tables 2 through 5 for the noninvariant factor loading level and Tables 6 through 9 for the invariant factor loading level. Cells with bold values refer to conditions in which relative bias was more extreme than ±10%. Parameters were recovered most accurately under Bayesian estimation with informative priors, followed by Bayesian estimation with weakly informative priors, frequentist estimation, and Bayesian estimation with diffuse priors. In general, each estimator recovered the cluster-level covariate effect more accurately as the amount of information provided by the data increased (i.e., as the number of clusters, average cluster size, and ICC increased).

For $J \leq 100$, frequentist estimation with noninvariant factor loadings generally resulted in overestimation of the between-level covariate effect. For instance, Tables 2 through 4 show that the direction of bias was positive for the majority of the frequentist conditions in which problematic bias was observed. Holding factor loadings invariant led to a substantial improvement in bias, whereby unbiased estimates were obtained with as few as 40 clusters as long as the within-group sample size was 10 or more and the ICC was moderate to large ($\geq$ .10; see Table 6). At lower values of ICC, frequentist estimation with invariant factor

TABLE 2
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Noninvariant Factor Loadings (Number of Clusters = 40)

		Continuous Indicators																	
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*MLR*	
$N_j = 5$	ICC = .02	**58337.21**	(*.**)	**−12.11**	(0.45)	**−24.21**	(0.46)	−4.40	(0.36)	**−16.62**	(0.36)	−2.08	(0.26)	−9.07	(0.27)	−0.72	(0.19)	**178.30**	(10.22)[1]
	ICC = .05	**82331.26**	(0.09)	−6.89	(0.48)	**−22.19**	(0.46)	−5.66	(0.36)	**−15.55**	(0.36)	−2.59	(0.29)	**−11.13**	(0.29)	−1.19	(0.20)	**187.99**	(13.04)
	ICC = .10	**−77028.29**	(*.**)	−5.39	(0.50)	**−20.93**	(0.45)	−2.31	(0.38)	**−13.95**	(0.36)	−1.27	(0.29)	−8.13	(0.28)	−0.83	(0.20)	**41.30**	(2.88)
	ICC = .20	**−131992.30**	(0.31)	−1.06	(0.51)	**−13.63**	(0.41)	−0.08	(0.39)	**−11.99**	(0.36)	−1.69	(0.30)	−7.87	(0.29)	−0.51	(0.21)	**51.33**	(2.40)
	ICC = .40	**57515.48**	(0.11)	**10.39**	(0.50)	−4.10	(0.39)	4.88	(0.40)	−6.77	(0.35)	0.72	(0.31)	−4.62	(0.28)	1.49	(0.22)	**18.97**	(1.04)
$N_j = 10$	ICC = .02	**−33844.70**	(0.11)	**−12.47**	(0.47)	**−23.75**	(0.45)	−6.56	(0.35)	**−13.71**	(0.33)	−3.26	(0.25)	**−11.31**	(0.26)	−1.28	(0.17)	**54.42**	(2.81)
	ICC = .05	**7045.21**	(*.**)	−6.86	(0.50)	**−18.65**	(0.43)	−2.87	(0.36)	**−13.57**	(0.34)	−2.42	(0.27)	−8.51	(0.26)	−1.83	(0.19)	**66.84**	(2.66)
	ICC = .10	**−5289.37**	(0.12)	−3.17	(0.51)	**−15.26**	(0.44)	−2.43	(0.38)	**−10.81**	(0.34)	−1.67	(0.28)	−8.86	(0.27)	−0.94	(0.19)	**27.71**	(1.04)
	ICC = .20	**163503.67**	(*.**)	6.35	(0.50)	−7.82	(0.41)	4.12	(0.39)	−8.64	(0.31)	0.60	(0.31)	−5.73	(0.26)	−0.26	(0.22)	**17.34**	(0.99)
	ICC = .40	**216985.55**	(2.04)	**10.91**	(0.48)	−3.09	(0.38)	5.62	(0.40)	−4.05	(0.33)	3.75	(0.32)	−4.70	(0.27)	1.03	(0.22)	**15.19**	(0.83)
$N_j = 20$	ICC = .02	**−78322.39**	(*.**)	**−12.53**	(0.49)	**−23.00**	(0.44)	−7.04	(0.36)	**−16.84**	(0.35)	−3.83	(0.25)	**−11.12**	(0.26)	−1.69	(0.17)	**99.16**	(6.78)
	ICC = .05	**100164.77**	(2.49)	−6.13	(0.47)	**−15.89**	(0.43)	−4.90	(0.39)	**−13.35**	(0.35)	−2.29	(0.28)	−9.63	(0.27)	−2.11	(0.19)	**35.05**	(1.49)
	ICC = .10	**−22553.40**	(0.10)	3.31	(0.47)	−9.33	(0.40)	−0.26	(0.38)	−9.83	(0.34)	−0.91	(0.29)	−7.15	(0.26)	−0.71	(0.20)	**31.94**	(2.12)
	ICC = .20	**54679.32**	(*.**)	**11.32**	(0.47)	−5.36	(0.36)	4.11	(0.37)	−7.13	(0.31)	1.81	(0.31)	−4.07	(0.25)	0.34	(0.20)	0.39	(0.45)
	ICC = .40	**212157.01**	(0.23)	**14.62**	(0.49)	−0.70	(0.36)	7.97	(0.40)	−3.20	(0.31)	4.61	(0.32)	−2.42	(0.27)	2.16	(0.22)	**10.14**	(0.63)
		Categorical Indicators																	
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*WLSM*	
$N_j = 5$	ICC = .02	**−47.46**	(6.49)	−7.73	(0.67)	**−15.79**	(0.77)	−5.00	(0.53)	−9.96	(0.57)	−1.46	(0.39)	−5.58	(0.45)	−0.64	(0.27)	**−27.08**	(0.27)
	ICC = .05	**−56.39**	(6.92)	−6.29	(0.51)	**−18.31**	(0.54)	−2.18	(0.38)	**−10.18**	(0.38)	−2.90	(0.28)	−8.57	(0.31)	−0.85	(0.20)	**−21.64**	(0.47)
	ICC = .10	**−17.44**	(8.24)	−0.41	(0.51)	**−13.49**	(0.57)	0.60	(0.40)	−9.16	(0.41)	−0.83	(0.29)	−7.80	(0.31)	−0.56	(0.20)	**52.25**	(1.84)
	ICC = .20	**160.85**	(9.80)	3.12	(0.53)	−3.65	(0.59)	3.39	(0.43)	−2.73	(0.44)	1.97	(0.33)	−3.92	(0.34)	1.59	(0.23)	**33.62**	(1.16)
	ICC = .40	**295.73**	(8.99)	**16.03**	(0.58)	**16.71**	(0.63)	9.92	(0.48)	6.75	(0.49)	5.21	(0.37)	0.87	(0.39)	1.60	(0.26)	**36.17**	(1.56)
$N_j = 10$	ICC = .02	**−59.37**	(2.46)	−8.99	(0.54)	**−16.60**	(0.55)	−1.21	(0.43)	**−12.18**	(0.45)	−0.63	(0.33)	−2.24	(0.34)	0.29	(0.23)	**149.90**	(1.50)
	ICC = .05	**−55.13**	(2.39)	−3.43	(0.48)	**−16.38**	(0.46)	−1.97	(0.36)	**−11.84**	(0.35)	−0.61	(0.28)	−6.51	(0.26)	−0.35	(0.18)	**−10.99**	(0.56)
	ICC = .10	**−16.15**	(2.14)	2.57	(0.50)	**−10.03**	(0.44)	1.10	(0.40)	−7.81	(0.35)	0.08	(0.30)	−5.13	(0.27)	0.06	(0.20)	**76.63**	(2.02)
	ICC = .20	**58.59**	(4.58)	9.34	(0.50)	1.11	(0.43)	4.17	(0.41)	−0.37	(0.35)	2.24	(0.30)	−0.63	(0.29)	0.60	(0.22)	**35.44**	(1.16)
	ICC = .40	**114.50**	(3.21)	**15.16**	(0.51)	**10.82**	(0.48)	9.85	(0.42)	6.06	(0.38)	4.57	(0.33)	1.57	(0.32)	1.80	(0.24)	**17.03**	(0.75)
$N_j = 20$	ICC = .02	**−49.78**	(4.93)	−8.24	(0.52)	**−19.30**	(0.49)	−6.06	(0.39)	**−12.27**	(0.38)	−3.08	(0.30)	−7.93	(0.29)	−2.37	(0.20)	**155.16**	(2.96)
	ICC = .05	**−41.60**	(3.81)	1.01	(0.49)	**−13.12**	(0.42)	−1.63	(0.36)	**−10.36**	(0.33)	−0.15	(0.28)	−7.18	(0.26)	−2.39	(0.18)	**38.83**	(1.17)
	ICC = .10	1.17	(5.21)	7.46	(0.46)	−5.81	(0.38)	0.63	(0.37)	−5.75	(0.31)	0.24	(0.28)	−4.87	(0.25)	0.82	(0.19)	**24.11**	(0.95)
	ICC = .20	**15.08**	(0.71)	9.64	(0.44)	1.40	(0.35)	3.28	(0.36)	−0.45	(0.31)	0.49	(0.28)	−2.16	(0.26)	−0.37	(0.21)	**18.10**	(0.72)
	ICC = .40	**21.43**	(0.84)	**10.99**	(0.46)	6.77	(0.40)	7.33	(0.39)	2.10	(0.33)	2.55	(0.31)	−0.04	(0.27)	0.62	(0.23)	**10.80**	(0.54)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 3

Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Noninvariant Factor Loadings (Number of Clusters = 50)

Continuous Indicators

		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*MLR*
$N_j = 5$	ICC = .02	**−7093.85** (0.05)	−9.22 (0.46)	**−24.41** (0.45)	−4.54 (0.35)	**−15.57** (0.36)	−3.16 (0.26)	**−11.23** (0.27)	−1.71 (0.18)	**33.32** (1.56)
	ICC = .05	**−16264.29** (*.**)	−9.19 (0.49)	**−22.31** (0.45)	−4.65 (0.36)	**−14.33** (0.36)	−2.29 (0.27)	−9.62 (0.27)	−1.15 (0.18)	**36.81** (1.60)
	ICC = .10	**−30444.74** (0.19)	−7.16 (0.52)	**−19.80** (0.45)	−1.83 (0.37)	**−15.27** (0.37)	−3.46 (0.28)	−9.85 (0.28)	−1.05 (0.19)	**39.28** (1.57)
	ICC = .20	**175536.88** (*.**)	2.10 (0.52)	**−13.24** (0.43)	−1.75 (0.41)	**−10.56** (0.35)	0.65 (0.30)	−7.73 (0.28)	−0.64 (0.20)	**33.50** (2.29)
	ICC = .40	**21835.15** (2.95)	**10.06** (0.51)	−4.99 (0.38)	5.83 (0.40)	−5.80 (0.32)	3.32 (0.32)	−4.59 (0.28)	1.63 (0.23)	**16.72** (0.81)
$N_j = 10$	ICC = .02	**4594.27** (2.95)	−9.01 (0.46)	**−22.37** (0.43)	−6.63 (0.35)	**−15.72** (0.35)	−3.67 (0.26)	**−10.09** (0.25)	−1.12 (0.17)	**47.01** (2.27)
	ICC = .05	**−5300.00** (0.06)	−6.65 (0.51)	**−19.43** (0.43)	−4.66 (0.36)	**−13.71** (0.34)	−2.23 (0.28)	−9.62 (0.27)	−1.76 (0.18)	**−22.74** (1.11)
	ICC = .10	**42911.12** (*.**)	0.04 (0.51)	**−12.84** (0.42)	−0.87 (0.39)	**−10.04** (0.34)	−1.77 (0.29)	−8.67 (0.28)	−0.97 (0.20)	**26.25** (1.37)
	ICC = .20	**−49010.64** (0.27)	8.07 (0.51)	−5.84 (0.40)	3.28 (0.39)	−5.70 (0.33)	2.08 (0.31)	−5.78 (0.27)	0.54 (0.22)	**25.61** (2.19)
	ICC = .40	**57129.94** (*.**)	**13.25** (0.51)	0.17 (0.37)	8.77 (0.41)	−2.06 (0.34)	4.61 (0.31)	−2.80 (0.28)	3.03 (0.23)	4.83 (0.39)
$N_j = 20$	ICC = .02	**150965.11** (0.08)	−8.44 (0.47)	**−20.54** (0.43)	−5.36 (0.35)	**−14.94** (0.33)	−4.45 (0.26)	**−10.72** (0.25)	−2.43 (0.16)	**22.86** (1.37)
	ICC = .05	**−101164.35** (*.**)	−0.54 (0.50)	**−14.49** (0.42)	−3.43 (0.38)	**−11.60** (0.34)	−3.21 (0.27)	−9.09 (0.27)	−1.30 (0.19)	**25.10** (1.42)
	ICC = .10	**−47579.34** (0.55)	6.64 (0.48)	−7.64 (0.38)	0.93 (0.36)	−9.17 (0.32)	0.75 (0.29)	−7.35 (0.26)	−0.68 (0.19)	**19.26** (1.13)
	ICC = .20	**34323.69** (0.40)	**10.27** (0.46)	−4.08 (0.35)	3.78 (0.36)	−5.25 (0.30)	2.75 (0.29)	−6.22 (0.26)	0.63 (0.20)	**10.92** (0.66)
	ICC = .40	**171816.24** (*.**)	**11.70** (0.46)	−0.38 (0.34)	9.55 (0.38)	−2.24 (0.30)	4.51 (0.30)	−2.17 (0.25)	1.84 (0.20)	9.56 (0.62)

Categorical Indicators

		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*WLSM*
$N_j = 5$	ICC = .02	**−44.79** (6.61)	0.67 (0.62)	**−18.91** (0.72)	−3.31 (0.49)	−9.65 (0.53)	0.83 (0.37)	−5.23 (0.42)	−1.18 (0.27)	−.- (−.-)
	ICC = .05	**−44.54** (5.62)	−5.80 (0.51)	**−16.58** (0.50)	−4.22 (0.38)	**−12.14** (0.37)	−0.95 (0.27)	−7.72 (0.29)	−0.03 (0.20)	**17.66** (0.18)
	ICC = .10	**39.09** (6.40)	0.74 (0.51)	**−12.58** (0.51)	−0.07 (0.38)	−9.81 (0.38)	0.15 (0.29)	−5.18 (0.30)	−0.42 (0.19)	**32.42** (0.91)
	ICC = .20	**104.88** (5.36)	7.67 (0.56)	−3.05 (0.50)	3.77 (0.41)	−2.08 (0.40)	1.64 (0.31)	−2.88 (0.31)	0.20 (0.22)	**32.82** (0.95)
	ICC = .40	**281.08** (6.01)	**14.82** (0.56)	**14.84** (0.59)	9.60 (0.46)	8.59 (0.46)	3.91 (0.35)	3.12 (0.34)	0.00 (0.25)	**32.66** (1.26)
$N_j = 10$	ICC = .02	**−61.62** (2.14)	−6.27 (0.53)	**−19.07** (0.53)	−2.41 (0.40)	**−10.61** (0.41)	−2.67 (0.32)	−6.95 (0.33)	−1.04 (0.22)	**17.12** (0.57)
	ICC = .05	**−34.58** (2.45)	−4.32 (0.49)	**−14.57** (0.44)	−0.47 (0.37)	**−11.71** (0.35)	−0.91 (0.27)	−6.19 (0.26)	0.29 (0.18)	7.89 (0.65)
	ICC = .10	**−13.16** (146)	4.98 (0.49)	−6.31 (0.42)	0.87 (0.38)	−7.66 (0.35)	0.54 (0.29)	−6.20 (0.28)	0.05 (0.19)	**47.45** (1.67)
	ICC = .20	**47.17** (2.02)	9.82 (0.48)	4.19 (0.41)	5.38 (0.38)	0.40 (0.34)	2.85 (0.31)	−1.05 (0.28)	0.89 (0.22)	**25.38** (0.88)
	ICC = .40	**75.71** (2.18)	**14.83** (0.49)	**10.84** (0.45)	7.59 (0.41)	4.55 (0.35)	5.12 (0.32)	2.68 (0.31)	1.03 (0.24)	**13.98** (0.60)
$N_j = 20$	ICC = .02	**−64.28** (1.35)	−8.34 (0.49)	**−17.04** (0.46)	−3.68 (0.36)	**−11.26** (0.36)	−1.47 (0.29)	−7.22 (0.27)	−0.54 (0.19)	**73.99** (1.62)
	ICC = .05	**−21.74** (1.11)	1.17 (0.46)	−9.61 (0.40)	1.50 (0.36)	−8.29 (0.33)	−0.24 (0.28)	−6.11 (0.26)	−0.49 (0.18)	**29.52** (0.92)
	ICC = .10	3.54 (1.05)	5.93 (0.43)	−4.36 (0.36)	3.93 (0.36)	−3.66 (0.32)	0.26 (0.28)	−3.79 (0.25)	−0.22 (0.19)	**23.80** (0.87)
	ICC = .20	**10.09** (0.62)	8.44 (0.42)	1.68 (0.35)	3.94 (0.34)	−0.81 (0.29)	2.51 (0.29)	−1.45 (0.26)	0.59 (0.21)	**11.22** (0.52)
	ICC = .40	**15.04** (0.53)	**10.23** (0.43)	6.61 (0.37)	6.72 (0.35)	2.59 (0.32)	3.53 (0.29)	−0.09 (0.27)	0.04 (0.22)	8.82 (0.49)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 4
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Noninvariant Factor Loadings (Number of Clusters = 100)

Continuous Indicators

		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*MLR*
$N_j = 5$	ICC = .02	**53781.37** (*.**)	**−11.08** (0.45)	**−24.13** (0.46)	−6.39 (0.33)	**−16.09** (0.33)	−2.65 (0.24)	−9.33 (0.25)	−1.98 (0.17)	**47.67** (2.73)
	ICC = .05	**15143.93** (1.88)	−8.86 (0.47)	**−21.23** (0.45)	−3.66 (0.35)	**−13.18** (0.34)	−2.97 (0.26)	−8.33 (0.27)	−0.60 (0.18)	**45.20** (2.18)
	ICC = .10	**24196.57** (0.12)	−2.20 (0.51)	**−14.74** (0.42)	−0.89 (0.38)	**−11.73** (0.34)	−0.15 (0.29)	−7.82 (0.27)	−1.23 (0.19)	−5.21 (1.37)
	ICC = .20	**−44658.14** (*.**)	8.32 (0.50)	−4.85 (0.37)	4.65 (0.38)	−6.24 (0.32)	2.12 (0.31)	−5.53 (0.26)	0.94 (0.21)	**12.89** (0.83)
	ICC = .40	**116983.46** (0.13)	**14.22** (0.46)	−0.12 (0.33)	9.26 (0.38)	−1.09 (0.29)	4.10 (0.28)	−1.98 (0.25)	2.83 (0.21)	4.74 (0.45)
$N_j = 10$	ICC = .02	**−63587.20** (0.13)	−7.67 (0.47)	**−20.38** (0.42)	−6.15 (0.33)	**−13.91** (0.33)	−2.87 (0.25)	**−10.19** (0.26)	−2.02 (0.16)	**49.30** (2.30)
	ICC = .05	**124599.04** (*.**)	−2.65 (0.50)	**−14.01** (0.40)	−1.83 (0.37)	**−10.72** (0.34)	−2.09 (0.28)	**−10.21** (0.27)	−1.98 (0.18)	**64.51** (3.75)
	ICC = .10	**113801.57** (1.54)	6.80 (0.21)	−6.14 (0.37)	2.11 (0.37)	−7.14 (0.32)	0.66 (0.30)	−6.63 (0.26)	0.15 (0.19)	**20.16** (1.76)
	ICC = .20	**86099.33** (0.19)	**11.68** (0.46)	−0.85 (0.34)	6.31 (0.36)	−3.12 (0.29)	3.51 (0.29)	−3.21 (0.24)	1.61 (0.20)	**10.47** (0.78)
	ICC = .40	**15924.46** (*.**)	**12.74** (0.40)	1.99 (0.30)	8.97 (0.34)	0.65 (0.29)	4.59 (0.28)	−0.93 (0.24)	2.51 (0.20)	5.61 (0.45)
$N_j = 20$	ICC = .02	**−16408.81** (0.09)	−9.63 (0.49)	**−19.01** (0.43)	−5.41 (0.36)	**−14.69** (0.35)	−4.30 (0.26)	−9.28 (0.25)	−2.17 (0.17)	**29.52** (1.37)
	ICC = .05	**−124079.06** (*.**)	5.55 (0.46)	−8.39 (0.38)	−0.70 (0.34)	−8.75 (0.31)	−1.62 (0.27)	−6.91 (0.25)	−1.08 (0.19)	**20.10** (1.59)
	ICC = .10	**205265.23** (0.14)	**10.42** (0.45)	−2.04 (0.33)	4.80 (0.34)	−3.01 (0.28)	2.20 (0.27)	−4.06 (0.23)	0.43 (0.19)	**10.70** (0.90)
	ICC = .20	**161902.16** (*.**)	**12.65** (0.44)	1.09 (0.30)	6.53 (0.34)	−1.23 (0.27)	4.02 (0.28)	−1.97 (0.23)	2.20 (0.19)	5.38 (0.40)
	ICC = .40	**60805.00** (1.28)	**13.34** (0.41)	3.15 (0.30)	9.56 (0.34)	1.45 (0.26)	5.07 (0.26)	−0.06 (0.22)	2.84 (0.19)	4.25 (0.34)

Categorical Indicators

		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*WLSM*
$N_j = 5$	ICC = .02	**−53.26** (2.60)	−5.36 (0.54)	**−19.31** (0.59)	−6.68 (0.42)	−9.30 (0.45)	−2.19 (0.33)	−6.50 (0.34)	−0.27 (0.23)	**11.62** (0.12)
	ICC = .05	**−70.56** (2.86)	−3.88 (0.48)	**−17.48** (0.45)	−2.83 (0.36)	**−10.82** (0.34)	−1.62 (0.27)	−7.97 (0.27)	−0.26 (0.18)	**25.31** (0.99)
	ICC = .10	**−14.25** (2.91)	3.58 (0.50)	−9.46 (0.43)	2.74 (0.37)	−8.97 (0.35)	1.10 (0.28)	−6.21 (0.28)	1.48 (0.19)	**64.11** (2.00)
	ICC = .20	**42.00** (2.28)	**10.58** (0.50)	1.89 (0.41)	3.95 (0.39)	−0.87 (0.33)	1.93 (0.31)	−0.50 (0.27)	1.53 (0.22)	**19.97** (0.74)
	ICC = .40	**86.89** (2.50)	**13.63** (0.46)	**11.90** (0.43)	8.09 (0.38)	6.68 (0.36)	3.91 (0.30)	3.15 (0.30)	1.32 (0.23)	**10.53** (0.49)
$N_j = 10$	ICC = .02	**−68.41** (1.58)	−8.11 (0.49)	**−19.72** (0.48)	−4.80 (0.37)	**−11.41** (0.37)	−2.63 (0.28)	−8.17 (0.28)	−0.71 (0.20)	**24.24** (0.64)
	ICC = .05	**−27.28** (3.03)	−0.85 (0.48)	−9.60 (0.40)	−1.48 (0.36)	−9.84 (0.34)	−0.76 (0.27)	−5.82 (0.25)	0.34 (0.18)	**22.11** (0.83)
	ICC = .10	**281.11** (90.92)	6.03 (0.48)	−2.41 (0.38)	2.34 (0.36)	−3.11 (0.31)	0.98 (0.28)	−3.84 (0.26)	0.50 (0.20)	**30.29** (1.04)
	ICC = .20	**12.20** (0.64)	**10.07** (0.41)	3.20 (0.33)	5.68 (0.35)	1.39 (0.29)	1.42 (0.27)	−0.51 (0.24)	0.65 (0.20)	9.93 (0.45)
	ICC = .40	**16.62** (0.49)	**10.63** (0.37)	9.34 (0.34)	6.22 (0.32)	5.00 (0.29)	3.72 (0.26)	2.37 (0.25)	1.29 (0.21)	4.55 (0.34)
$N_j = 20$	ICC = .02	**−67.60** (1.81)	**−11.15** (0.50)	**−17.35** (0.45)	−4.79 (0.38)	**−12.39** (0.35)	−2.92 (0.28)	−7.86 (0.28)	−1.47 (0.19)	**20.55** (0.83)
	ICC = .05	0.09 (1.47)	5.97 (0.45)	−4.33 (0.37)	1.26 (0.35)	−5.24 (0.30)	−0.31 (0.27)	−4.99 (0.25)	−1.11 (0.19)	**17.30** (0.69)
	ICC = .10	0.39 (1.81)	9.02 (0.41)	0.12 (0.32)	3.80 (0.33)	−0.93 (0.29)	1.70 (0.27)	−1.92 (0.24)	−0.01 (0.19)	9.26 (0.49)
	ICC = .20	6.16 (0.34)	8.05 (0.34)	2.68 (0.29)	4.01 (0.29)	1.53 (0.26)	2.94 (0.24)	0.02 (0.22)	0.39 (0.19)	4.60 (0.32)
	ICC = .40	6.78 (0.33)	7.27 (0.32)	4.76 (0.28)	5.09 (0.28)	2.09 (0.25)	2.65 (0.24)	0.78 (0.22)	1.20 (0.19)	3.21 (0.27)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 5

Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Noninvariant Factor Loadings (Number of Clusters = 200)

		Continuous Indicators																	
		Diffuse		*Weak la*		*Weak lb*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*MLR*	
$N_j = 5$	ICC = .02	**90564.38**	(0.23)	**−11.70**	(0.49)	**−22.53**	(0.45)	−4.77	(0.35)	**−15.60**	(0.34)	−3.77	(0.25)	**−11.29**	(0.25)	−1.20	(0.16)	**108.44**	(5.54)
	ICC = .05	**96067.36**	(*.**)	−6.37	(0.50)	**−18.91**	(0.45)	−3.91	(0.36)	**−13.52**	(0.34)	−3.84	(0.27)	−9.47	(0.26)	−1.71	(0.18)	**46.56**	(2.29)
	ICC = .10	**−24481.65**	(2.71)	2.28	(0.49)	−8.08	(0.38)	1.14	(0.39)	−7.32	(0.32)	−0.38	(0.29)	−7.01	(0.26)	−0.45	(0.19)	**24.03**	(1.20)
	ICC = .20	**58502.32**	(0.23)	**12.28**	(0.46)	−0.34	(0.32)	7.33	(0.38)	−2.39	(0.30)	3.38	(0.29)	−3.57	(0.25)	1.73	(0.20)	6.47	(0.49)
	ICC = .40	**13696.01**	(*.**)	9.82	(0.34)	2.53	(0.28)	7.76	(0.31)	0.56	(0.25)	3.75	(0.24)	−0.23	(0.21)	2.19	(0.19)	3.12	(0.28)
$N_j = 10$	ICC = .02	**52190.69**	(*.**)	−9.36	(0.49)	**−16.93**	(0.45)	−3.23	(0.37)	**−12.88**	(0.33)	−2.77	(0.27)	−9.71	(0.27)	−1.68	(0.16)	**98.66**	(5.00)
	ICC = .05	**73941.79**	(0.12)	3.87	(0.49)	−6.66	(0.40)	1.66	(0.37)	−7.90	(0.32)	0.91	(0.29)	−6.40	(0.27)	−0.09	(0.19)	**17.06**	(1.03)
	ICC = .10	**13818.79**	(*.**)	**12.09**	(0.47)	1.35	(0.35)	7.39	(0.36)	−1.81	(0.29)	4.21	(0.28)	−2.59	(0.25)	2.51	(0.21)	8.27	(0.53)
	ICC = .20	**19944.87**	(2.07)	**12.58**	(0.39)	2.83	(0.30)	8.69	(0.32)	0.37	(0.26)	5.09	(0.26)	−0.07	(0.22)	2.95	(0.19)	4.16	(0.33)
	ICC = .40	6.48	(0.09)	9.24	(0.31)	3.58	(0.25)	6.33	(0.27)	2.72	(0.23)	5.21	(0.23)	1.20	(0.21)	2.46	(0.18)	2.75	(0.26)
$N_j = 20$	ICC = .02	**118172.14**	(2.04)	−5.72	(0.48)	**−11.57**	(0.42)	−5.20	(0.38)	**−10.65**	(0.33)	−2.24	(0.28)	−7.68	(0.26)	−1.74	(0.18)	**19.82**	(0.90)
	ICC = .05	**19856.66**	(0.09)	8.28	(0.43)	−1.80	(0.34)	4.75	(0.34)	−3.07	(0.29)	1.47	(0.27)	−3.80	(0.24)	−0.16	(0.18)	4.49	(0.46)
	ICC = .10	**21823.43**	(*.**)	9.88	(0.38)	2.16	(0.29)	5.38	(0.30)	−0.60	(0.26)	3.43	(0.25)	−1.72	(0.21)	1.25	(0.18)	3.08	(0.31)
	ICC = .20	7.01	(0.17)	7.87	(0.31)	1.36	(0.24)	5.63	(0.27)	0.92	(0.23)	3.58	(0.22)	−0.43	(0.20)	1.18	(0.17)	2.01	(0.25)
	ICC = .40	3.85	*.**	6.78	(0.27)	2.18	(0.23)	5.44	(0.24)	1.12	(0.21)	4.11	(0.22)	0.79	(0.19)	1.30	(0.17)	1.58	(0.22)
		Categorical Indicators																	
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*WLSM*	
$N_j = 5$	ICC = .02	**−65.68**	(2.30)	−8.25	(0.52)	**−17.05**	(0.50)	−5.03	(0.39)	**−13.43**	(0.40)	−0.91	(0.27)	−8.76	(0.30)	−0.64	(0.20)	**309.40**	(3.09)
	ICC = .05	**−69.96**	(4.29)	−1.82	(0.50)	**−13.41**	(0.42)	−0.65	(0.35)	−8.66	(0.33)	−0.31	(0.26)	−6.87	(0.25)	−0.47	(0.17)	**47.93**	(1.18)
	ICC = .10	**−11.04**	(3.52)	6.41	(0.48)	−4.83	(0.39)	2.13	(0.36)	−4.68	(0.32)	1.19	(0.28)	−4.88	(0.26)	−0.12	(0.19)	**19.51**	(0.92)
	ICC = .20	**−239.53**	(81.32)	8.13	(0.42)	3.46	(0.35)	6.20	(0.36)	1.73	(0.30)	2.89	(0.28)	−0.40	(0.24)	−0.40	(0.20)	9.18	(0.46)
	ICC = .40	**17.19**	(0.59)	9.48	(0.34)	7.25	(0.32)	6.52	(0.30)	4.80	(0.28)	3.99	(0.24)	3.13	(0.24)	0.76	(0.20)	4.49	(0.31)
$N_j = .20$	ICC = .02	**−76.79**	(3.42)	−6.13	(0.51)	**−16.97**	(0.45)	−3.47	(0.36)	**−10.94**	(0.34)	−1.93	(0.27)	−6.69	(0.27)	−0.56	(0.19)	7.08	(0.51)
	ICC = .05	−8.16	(11.24)	4.83	(0.47)	−3.97	(0.38)	1.88	(0.36)	−4.65	(0.32)	0.97	(0.26)	−4.16	(0.25)	0.16	(0.19)	**15.61**	(0.69)
	ICC = .10	9.79	(0.67)	8.69	(0.42)	3.28	(0.35)	5.30	(0.34)	0.00	(0.29)	2.14	(0.28)	−1.98	(0.25)	0.87	(0.19)	**11.75**	(0.51)
	ICC = .20	6.04	(0.33)	7.49	(0.33)	4.43	(0.28)	6.10	(0.28)	2.34	(0.25)	3.15	(0.24)	1.26	(0.22)	1.75	(0.18)	4.19	(0.30)
	ICC = .40	6.69	(0.27)	7.69	(0.26)	5.62	(0.25)	6.38	(0.24)	4.41	(0.23)	3.85	(0.21)	3.34	(0.20)	2.24	(0.17)	2.27	(0.23)
$N_j = 20$	ICC = .02	**−31.68**	(7.27)	−7.46	(0.51)	**−14.36**	(0.43)	−2.81	(0.37)	**−10.50**	(0.36)	−3.42	(0.29)	−7.26	(0.27)	−2.20	(0.19)	**22.73**	(0.84)
	ICC = .05	**25.75**	(6.52)	5.98	(0.39)	−0.83	(0.32)	2.92	(0.32)	−1.72	(0.27)	0.80	(0.26)	−3.10	(0.24)	−1.48	(0.18)	**10.70**	(0.52)
	ICC = .10	2.73	(0.33)	5.87	(0.32)	1.39	(0.27)	2.54	(0.27)	−0.22	(0.24)	1.40	(0.23)	−1.19	(0.21)	−0.05	(0.18)	3.94	(0.29)
	ICC = .20	2.09	(0.24)	4.00	(0.24)	1.30	(0.22)	2.71	(0.23)	1.27	(0.20)	2.02	(0.20)	0.65	(0.19)	1.04	(0.16)	1.75	(0.21)
	ICC = .40	3.04	(0.22)	3.90	(0.22)	2.63	(0.21)	3.43	(0.21)	2.58	(0.20)	2.82	(0.19)	1.70	(0.18)	1.50	(0.16)	1.29	(0.18)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 6
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Invariant Factor Loadings (Number of Clusters = 40)

		Continuous Indicators								
		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*MLR*
$N_j = 5$	ICC = .02	**−33.34** (1.51)	−7.08 (0.43)	**−18.80** (0.41)	−5.90 (0.32)	**−14.33** (0.33)	−2.91 (0.24)	−8.66 (0.25)	−1.65 (0.17)	**−34.07** (1.22)
	ICC = .05	**−20.49** (2.30)	−6.94 (0.45)	**−16.51** (0.41)	−3.91 (0.33)	**−13.56** (0.33)	−2.98 (0.25)	−8.31 (0.26)	−1.14 (0.17)	**−29.19** (1.22)
	ICC = .10	**−43.25** (8.65)	−4.99 (0.51)	**−14.29** (0.41)	−3.64 (0.38)	**−11.91** (0.34)	−2.62 (0.27)	−8.94 (0.28)	−1.51 (0.19)	**−14.45** (1.30)
	ICC = .20	−8.41 (1.93)	2.95 (0.51)	**−10.98** (0.40)	1.59 (0.40)	−9.58 (0.34)	−0.96 (0.30)	−7.62 (0.28)	−1.15 (0.21)	**−10.82** (0.84)
	ICC = .40	2.05 (0.78)	**14.11** (0.52)	−4.65 (0.37)	8.48 (0.42)	−4.60 (0.33)	5.75 (0.31)	−4.99 (0.28)	1.91 (0.23)	0.74 (0.53)
$N_j = 10$	ICC = .02	**−34.73** (4.27)	−6.00 (0.43)	**−17.06** (0.39)	−6.28 (0.31)	**−13.87** (0.32)	−4.40 (0.24)	−8.98 (0.24)	−1.11 (0.16)	**−40.37** (1.34)
	ICC = .05	**−26.42** (3.03)	−4.58 (0.47)	**−14.29** (0.39)	−4.51 (0.34)	**−11.69** (0.32)	−3.12 (0.25)	−9.35 (0.25)	−0.40 (0.17)	**−27.99** (1.45)
	ICC = .10	**−17.20** (1.31)	2.01 (0.48)	**−11.29** (0.38)	−1.46 (0.36)	−9.09 (0.32)	−2.18 (0.27)	−8.61 (0.26)	−1.47 (0.19)	−4.28 (0.74)
	ICC = .20	3.79 (2.23)	**12.53** (0.51)	−6.52 (0.37)	6.37 (0.38)	−5.53 (0.32)	3.52 (0.30)	−5.74 (0.26)	1.34 (0.20)	−0.96 (0.49)
	ICC = .40	1.69 (0.46)	**16.61** (0.49)	−2.32 (0.35)	**10.72** (0.39)	−2.33 (0.30)	5.47 (0.31)	−3.01 (0.27)	3.20 (0.22)	3.44 (0.43)
$N_j = 20$	ICC = .02	**−16.99** (1.67)	−8.39 (0.50)	**−17.32** (0.41)	−7.20 (0.35)	**−13.36** (0.32)	−4.76 (0.25)	−9.67 (0.24)	−2.93 (0.16)	**−35.60** (1.25)
	ICC = .05	−8.50 (1.21)	−1.74 (0.47)	**−13.72** (0.38)	−4.16 (0.37)	**−11.77** (0.32)	−3.27 (0.28)	−9.57 (0.25)	−2.72 (0.18)	**−17.31** (0.67)
	ICC = .10	**−10.01** (1.64)	6.48 (0.47)	−8.16 (0.36)	4.21 (0.38)	−8.16 (0.31)	1.41 (0.28)	−6.84 (0.26)	−0.16 (0.19)	−2.74 (0.59)
	ICC = .20	−1.05 (1.84)	**13.11** (0.47)	−3.52 (0.34)	9.48 (0.38)	−4.10 (0.29)	5.23 (0.31)	−3.68 (0.26)	3.30 (0.20)	0.39 (0.45)
	ICC = .40	1.60 (0.42)	**17.45** (0.47)	0.08 (0.34)	**12.08** (0.38)	−1.23 (0.30)	7.88 (0.31)	−2.49 (0.27)	3.76 (0.22)	2.79 (0.41)
		Categorical Indicators								
		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak .25a*	*Weak .25b*	*Informative*	*WLSM*
$N_j = 5$	ICC = .02	**326.34** (9.93)	−6.00 (0.67)	2.13 (0.77)	−1.11 (0.50)	−5.45 (0.59)	−1.01 (0.41)	−0.90 (0.42)	−1.68 (0.27)	−.– (-.–)
	ICC = .05	**498.96** (15.49)	−1.30 (0.43)	−7.39 (0.56)	−2.98 (0.33)	−8.74 (0.40)	−0.81 (0.24)	−4.19 (0.30)	−1.13 (0.18)	3.25 (0.94)
	ICC = .10	**546.60** (15.06)	0.11 (0.49)	−4.79 (0.59)	−0.11 (0.35)	−7.21 (0.44)	−1.55 (0.27)	−3.80 (0.31)	−0.71 (0.18)	4.42 (1.21)
	ICC = .20	**617.07** (12.86)	**12.50** (0.54)	6.14 (0.60)	6.17 (0.41)	2.83 (0.46)	2.40 (0.31)	−0.70 (0.35)	−0.13 (0.22)	5.05 (0.94)
	ICC = .40	**622.79** (11.23)	**23.39** (0.61)	**22.52** (0.66)	**13.53** (0.48)	8.48 (0.50)	6.03 (0.38)	5.19 (0.39)	3.16 (0.26)	**20.62** (0.93)
$N_j = 10$	ICC = .02	**143.44** (4.44)	−2.11 (0.49)	−1.61 (0.57)	−1.23 (0.38)	−2.22 (0.44)	0.12 (0.31)	−3.07 (0.34)	−0.80 (0.23)	**29.03** (0.29)
	ICC = .05	**93.49** (4.45)	−0.05 (0.42)	−7.61 (0.43)	−0.24 (0.34)	−8.68 (0.33)	−1.81 (0.24)	−6.79 (0.25)	−1.25 (0.16)	**−27.89** (0.74)
	ICC = .10	**59.99** (16.30)	7.97 (0.51)	−2.69 (0.42)	0.97 (0.37)	−4.28 (0.34)	1.91 (0.27)	−3.08 (0.26)	−0.07 (0.19)	0.44 (0.74)
	ICC = .20	**147.91** (4.42)	**13.95** (0.49)	9.70 (0.49)	7.66 (0.40)	3.58 (0.36)	3.60 (0.31)	0.51 (0.28)	1.37 (0.21)	5.38 (0.66)
	ICC = .40	**241.25** (5.47)	**19.71** (0.52)	**17.40** (0.54)	**11.72** (0.42)	**10.09** (0.42)	7.42 (0.32)	2.97 (0.33)	3.35 (0.24)	5.85 (0.58)
$N_j = 20$	ICC = .02	1.34 (1.09)	−1.71 (0.44)	−7.03 (0.43)	−2.10 (0.33)	−7.65 (0.35)	−2.47 (0.27)	−5.45 (0.27)	−1.08 (0.19)	5.05 (0.59)
	ICC = .05	−3.14 (0.72)	−0.57 (0.45)	−8.50 (0.38)	−0.63 (0.35)	−7.38 (0.31)	−1.27 (0.26)	−6.32 (0.24)	−0.74 (0.17)	**−12.04** (0.67)
	ICC = .10	5.89 (0.70)	6.86 (0.45)	−3.33 (0.37)	3.61 (0.36)	−3.56 (0.30)	1.99 (0.28)	−3.88 (0.25)	−0.38 (0.18)	−4.17 (0.55)
	ICC = .20	**13.22** (0.68)	**12.56** (0.46)	3.80 (0.35)	6.39 (0.35)	1.70 (0.31)	3.07 (0.29)	0.12 (0.25)	1.73 (0.20)	0.61 (0.44)
	ICC = .40	**27.76** (1.09)	**16.05** (0.46)	8.03 (0.40)	**11.12** (0.38)	4.86 (0.34)	5.40 (0.31)	2.28 (0.28)	2.49 (0.23)	3.00 (0.43)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 7
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Invariant Factor Loadings (Number of Clusters = 50)

		Continuous Indicators																	
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*MLR*	
$N_j = 5$	ICC = .02	**−95.66**	(30.33)	−8.65	(0.43)	**−19.28**	(0.40)	−7.09	(0.31)	**−13.87**	(0.32)	−2.83	(0.24)	**−10.13**	(0.25)	−1.39	(0.17)	**−28.78**	(1.46)
	ICC = .05	**−16.43**	(2.52)	−8.72	(0.46)	**−16.48**	(0.40)	−4.98	(0.35)	**−13.30**	(0.32)	−3.31	(0.25)	−9.03	(0.25)	−0.94	(0.18)	**−15.62**	(1.16)
	ICC = .10	**−33.48**	(4.01)	−6.52	(0.49)	**−15.13**	(0.40)	−3.10	(0.38)	**−11.80**	(0.33)	−3.36	(0.27)	−9.33	(0.27)	−1.91	(0.18)	**−16.55**	(1.11)
	ICC = .20	−7.44	(1.26)	4.18	(0.48)	−9.11	(0.39)	1.24	(0.39)	−9.37	(0.34)	−0.71	(0.29)	−7.03	(0.28)	−0.20	(0.21)	−3.61	(0.66)
	ICC = .40	2.81	(0.68)	**14.60**	(0.50)	−1.26	(0.37)	9.22	(0.39)	−3.30	(0.32)	5.34	(0.33)	−3.45	(0.27)	2.14	(0.22)	3.51	(0.47)
$N_j = 10$	ICC = .02	−2.13	(6.22)	−8.07	(0.44)	**−17.73**	(0.39)	−6.88	(0.32)	**−13.40**	(0.32)	−3.08	(0.22)	−9.62	(0.24)	−1.67	(0.15)	**−34.99**	(1.17)
	ICC = .05	0.60	(2.46)	−5.27	(0.47)	**−13.49**	(0.40)	−3.54	(0.37)	**−11.38**	(0.33)	−1.66	(0.27)	−8.50	(0.25)	−1.70	(0.17)	**−22.74**	(1.11)
	ICC = .10	**20.18**	(9.45)	2.13	(0.51)	**−10.98**	(0.38)	1.36	(0.39)	−8.79	(0.33)	−0.38	(0.29)	−7.56	(0.27)	−0.14	(0.19)	−4.26	(0.65)
	ICC = .20	−0.10	(0.57)	**12.27**	(0.49)	−2.95	(0.36)	7.35	(0.40)	−5.75	(0.32)	5.06	(0.30)	−3.73	(0.26)	1.51	(0.21)	2.68	(0.55)
	ICC = .40	1.79	(0.40)	**16.15**	(0.47)	−0.03	(0.34)	**12.16**	(0.40)	−1.48	(0.30)	8.83	(0.32)	−1.00	(0.26)	4.03	(0.23)	4.83	(0.39)
$N_j = 20$	ICC = .02	**−29.37**	(1.46)	−9.01	(0.47)	**−16.45**	(0.39)	−6.21	(0.33)	**−12.81**	(0.32)	−4.66	(0.24)	−9.63	(0.24)	−2.42	(0.15)	**−38.49**	(1.33)
	ICC = .05	−8.96	(1.74)	−0.70	(0.49)	**−11.36**	(0.37)	−0.41	(0.37)	−9.85	(0.32)	−1.70	(0.28)	−8.59	(0.25)	−1.89	(0.18)	−9.62	(0.70)
	ICC = .10	−4.33	(0.63)	9.45	(0.46)	−5.55	(0.34)	3.84	(0.36)	−6.36	(0.30)	2.58	(0.28)	−5.51	(0.25)	0.19	(0.19)	0.10	(0.50)
	ICC = .20	−3.97	(1.09)	**14.45**	(0.46)	−2.32	(0.31)	8.93	(0.36)	−3.66	(0.28)	5.90	(0.29)	−3.89	(0.24)	2.58	(0.19)	3.65	(0.43)
	ICC = .40	3.48	(0.39)	**14.87**	(0.44)	−0.23	(0.31)	**12.32**	(0.37)	−0.82	(0.28)	8.25	(0.30)	−1.67	(0.25)	3.75	(0.21)	4.07	(0.38)
		Categorical Indicators																	
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak .25a*		*Weak .25b*		*Informative*		*WLSM*	
$N_j = 5$	ICC = .02	**251.58**	(7.93)	−6.01	(0.60)	−1.28	(0.71)	−3.81	(0.49)	−2.11	(0.53)	0.00	(0.37)	−1.83	(0.42)	0.12	(0.26)	**-.–**	(-.–)
	ICC = .05	**342.12**	(11.03)	−2.72	(0.43)	−9.59	(0.50)	−3.86	(0.31)	−7.69	(0.38)	−1.26	(0.25)	−6.31	(0.28)	−0.63	(0.17)	**−69.75**	(0.92)
	ICC = .10	**392.71**	(11.48)	0.13	(0.48)	−3.70	(0.55)	−1.13	(0.35)	−5.55	(0.38)	−1.38	(0.27)	−5.58	(0.29)	−1.25	(0.17)	**39.19**	(2.50)
	ICC = .20	**505.89**	(12.92)	**11.69**	(0.53)	8.06	(0.56)	4.81	(0.40)	2.04	(0.43)	2.36	(0.30)	−3.01	(0.31)	−0.05	(0.22)	**12.54**	(1.03)
	ICC = .40	**504.97**	(9.45)	**23.01**	(0.58)	**22.53**	(0.65)	**12.81**	(0.46)	**12.23**	(0.46)	6.21	(0.34)	5.27	(0.35)	2.91	(0.26)	**17.70**	(0.99)
$N_j = 10$	ICC = .02	**83.02**	(3.74)	1.35	(0.48)	−4.22	(0.52)	−0.66	(0.37)	−3.51	(0.42)	−0.02	(0.29)	−2.28	(0.34)	−0.07	(0.21)	6.53	(1.12)
	ICC = .05	**28.45**	(2.62)	0.98	(0.44)	−7.75	(0.40)	−1.09	(0.34)	−7.90	(0.32)	−0.54	(0.24)	−6.58	(0.25)	−0.31	(0.16)	**−13.68**	(0.57)
	ICC = .10	**40.36**	(3.22)	5.60	(0.48)	−1.87	(0.41)	3.33	(0.39)	−4.54	(0.34)	0.40	(0.28)	−2.97	(0.25)	0.14	(0.19)	−5.29	(0.64)
	ICC = .20	**81.65**	(2.84)	**13.63**	(0.50)	**10.47**	(0.45)	8.76	(0.39)	4.55	(0.36)	4.94	(0.30)	1.74	(0.29)	2.40	(0.21)	4.05	(0.55)
	ICC = .40	**113.39**	(3.19)	**20.21**	(0.51)	**15.38**	(0.48)	**12.58**	(0.40)	9.49	(0.38)	6.78	(0.32)	3.80	(0.30)	3.58	(0.23)	3.94	(0.47)
$N_j = 20$	ICC = .02	−2.01	(0.84)	−2.64	(0.42)	−7.14	(0.39)	−1.47	(0.34)	−6.69	(0.33)	−1.58	(0.25)	−6.24	(0.27)	−1.06	(0.18)	**−28.12**	(0.86)
	ICC = .05	**−12.97**	(7.11)	2.34	(0.44)	−6.58	(0.35)	0.07	(0.35)	−5.95	(0.31)	−0.44	(0.27)	−4.69	(0.24)	−0.93	(0.16)	**−10.36**	(0.70)
	ICC = .10	4.80	(0.76)	6.49	(0.42)	−1.52	(0.34)	3.53	(0.34)	−2.61	(0.30)	1.99	(0.27)	−2.67	(0.24)	0.15	(0.19)	−1.05	(0.50)
	ICC = .20	7.88	(0.45)	**10.30**	(0.41)	3.11	(0.34)	8.25	(0.35)	2.96	(0.30)	4.47	(0.28)	1.22	(0.25)	1.47	(0.20)	2.58	(0.40)
	ICC = .40	**14.26**	(0.59)	**11.79**	(0.41)	6.16	(0.36)	8.34	(0.35)	4.91	(0.32)	5.99	(0.29)	1.83	(0.28)	2.30	(0.22)	2.96	(0.36)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 8
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Invariant Factor Loadings (Number of Clusters = 100)

		Continuous Indicators								
		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak 25a*	*Weak .25b*	*Informative*	*MLR*
$N_j = 5$	ICC = .02	**−39.97** (5.31)	−8.69 (0.44)	**−16.68** (0.37)	−5.09 (0.31)	**−13.79** (0.32)	−2.94 (0.23)	**−10.39** (0.24)	−2.10 (0.14)	**−30.55** (2.09)
	ICC = .05	**−37.02** (4.65)	−7.35 (0.46)	**−14.65** (0.39)	−4.68 (0.35)	**−13.14** (0.32)	−4.96 (0.25)	−9.16 (0.25)	−1.12 (0.16)	**−13.59** (1.92)
	ICC = .10	**−24.53** (4.87)	−0.44 (0.48)	−9.99 (0.37)	−2.53 (0.37)	−9.13 (0.32)	−1.21 (0.27)	−7.88 (0.26)	−0.86 (0.17)	−5.21 (1.37)
	ICC = .20	0.61 (0.72)	**12.28** (0.47)	−3.95 (0.33)	7.01 (0.38)	−4.69 (0.29)	4.57 (0.30)	−4.79 (0.26)	1.17 (0.21)	2.04 (0.43)
	ICC = .40	3.42 (0.41)	**14.18** (0.42)	−0.62 (0.28)	**10.94** (0.36)	0.54 (0.27)	6.69 (0.29)	−1.32 (0.23)	3.51 (0.21)	4.38 (0.33)
$N_j = 10$	ICC = .02	−3.22 (4.06)	−7.18 (0.45)	**−14.82** (0.39)	−6.33 (0.34)	**−13.03** (0.31)	−3.46 (0.24)	−8.28 (0.23)	−2.17 (0.14)	−5.62 (1.78)
	ICC = .05	−8.19 (1.30)	−0.79 (0.46)	**−11.24** (0.38)	−1.24 (0.36)	−9.55 (0.32)	−2.14 (0.28)	−8.00 (0.25)	−2.10 (0.18)	**−13.46** (0.70)
	ICC = .10	−6.98 (1.73)	9.99 (0.47)	−3.97 (0.33)	4.28 (0.37)	−5.27 (0.29)	2.41 (0.28)	−4.54 (0.24)	1.01 (0.19)	−0.09 (0.47)
	ICC = .20	0.54 (0.43)	**12.16** (0.41)	−0.73 (0.29)	9.19 (0.34)	−0.86 (0.26)	6.27 (0.29)	−2.24 (0.23)	2.98 (0.20)	4.05 (0.35)
	ICC = .40	3.05 (0.29)	**11.23** (0.37)	1.50 (0.27)	9.20 (0.33)	1.27 (0.24)	6.15 (0.27)	−0.23 (0.22)	3.91 (0.20)	3.09 (0.27)
$N_j = 20$	ICC = .02	**−23.84** (2.00)	−8.11 (0.48)	**−14.97** (0.37)	−5.42 (0.37)	**−11.75** (0.32)	−4.54 (0.26)	−9.50 (0.25)	−2.93 (0.16)	**−20.15** (1.55)
	ICC = .05	3.78 (1.34)	6.35 (0.44)	−5.75 (0.34)	3.63 (0.36)	−5.23 (0.29)	1.12 (0.28)	−5.70 (0.24)	−0.11 (0.18)	1.49 (0.52)
	ICC = .10	0.71 (0.44)	**11.40** (0.40)	−0.97 (0.30)	7.67 (0.34)	−2.47 (0.26)	5.16 (0.28)	−2.77 (0.23)	2.47 (0.19)	4.65 (0.37)
	ICC = .20	1.58 (0.30)	**10.88** (0.36)	0.09 (0.26)	7.66 (0.31)	−0.25 (0.24)	6.05 (0.27)	−0.11 (0.22)	2.77 (0.19)	3.63 (0.28)
	ICC = .40	2.13 (0.27)	**10.33** (0.34)	1.32 (0.25)	7.59 (0.29)	0.91 (0.23)	5.81 (0.26)	0.77 (0.22)	3.48 (0.19)	3.23 (0.24)
		Categorical Indicators								
		Diffuse	*Weak 1a*	*Weak 1b*	*Weak .5a*	*Weak .5b*	*Weak 25a*	*Weak .25b*	*Informative*	*WLSM*
$N_j = 5$	ICC = .02	**78.85** (3.41)	−3.56 (0.49)	−1.60 (0.56)	−4.65 (0.39)	−4.91 (0.44)	−2.63 (0.29)	−3.35 (0.36)	−2.56 (0.21)	−.– (-.–)
	ICC = .05	**124.75** (15.77)	−3.92 (0.45)	−9.70 (0.39)	−3.02 (0.32)	−8.39 (0.32)	−1.67 (0.24)	−7.24 (0.24)	−1.87 (0.16)	**−18.12** (0.75)
	ICC = .10	**79.14** (4.13)	4.36 (0.47)	−2.61 (0.41)	2.35 (0.36)	−4.09 (0.32)	−0.44 (0.26)	−4.19 (0.26)	−0.27 (0.17)	−2.15 (0.66)
	ICC = .20	**124.57** (3.93)	**12.01** (0.47)	8.11 (0.43)	8.02 (0.36)	3.06 (0.35)	4.38 (0.29)	1.98 (0.27)	1.15 (0.21)	5.84 (0.53)
	ICC = .40	**154.20** (3.87)	**16.16** (0.46)	**18.72** (0.49)	**11.46** (0.39)	**11.03** (0.37)	7.18 (0.31)	5.90 (0.30)	1.84 (0.23)	7.05 (0.42)
$N_j = 10$	ICC = .02	0.03 (1.03)	−2.98 (0.44)	−4.14 (0.41)	−2.18 (0.33)	−5.18 (0.33)	−1.21 (0.25)	−4.71 (0.27)	−1.01 (0.18)	**−23.68** (0.59)
	ICC = .05	**−25.25** (6.17)	2.21 (0.45)	−7.56 (0.36)	0.89 (0.35)	−6.34 (0.30)	−0.10 (0.25)	−4.84 (0.24)	−0.21 (0.16)	−2.87 (0.61)
	ICC = .10	**−36.32** (11.83)	8.12 (0.44)	−0.76 (0.35)	4.83 (0.36)	−1.79 (0.30)	2.25 (0.28)	−2.83 (0.24)	0.17 (0.19)	1.50 (0.55)
	ICC = .20	**11.05** (0.48)	**11.50** (0.39)	5.99 (0.32)	7.24 (0.33)	2.85 (0.29)	4.71 (0.27)	2.36 (0.24)	2.02 (0.20)	2.52 (0.36)
	ICC = .40	**17.38** (0.53)	**12.65** (0.37)	9.42 (0.34)	9.41 (0.32)	7.85 (0.30)	7.47 (0.27)	5.11 (0.26)	3.22 (0.20)	1.29 (0.29)
$N_j = 20$	ICC = .02	**23.29** (10.98)	−3.25 (0.46)	−8.51 (0.38)	−1.47 (0.32)	−6.64 (0.32)	−1.38 (0.24)	−5.25 (0.24)	−0.91 (0.16)	0.49 (0.70)
	ICC = .05	−0.72 (0.90)	5.69 (0.42)	−3.09 (0.33)	1.82 (0.35)	−3.61 (0.28)	1.49 (0.27)	−3.47 (0.24)	0.03 (0.18)	0.24 (0.51)
	ICC = .10	3.55 (0.42)	9.13 (0.39)	2.01 (0.30)	6.02 (0.32)	−0.28 (0.26)	2.98 (0.26)	−0.63 (0.22)	1.34 (0.19)	2.44 (0.34)
	ICC = .20	4.41 (0.29)	6.85 (0.30)	3.11 (0.27)	6.85 (0.28)	1.84 (0.24)	4.32 (0.24)	1.24 (0.21)	2.03 (0.19)	1.65 (0.25)
	ICC = .40	5.16 (0.29)	6.39 (0.27)	4.20 (0.26)	4.82 (0.26)	3.11 (0.24)	4.13 (0.23)	2.32 (0.22)	1.86 (0.18)	1.43 (0.23)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

TABLE 9
Relative Percentage Bias and Root Mean Square Error (RMSE) for the Cluster-Level Structural Coefficient With Invariant Factor Loadings (Number of Clusters = 200)

		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak 25a*		*Weak .25b*		*Informative*		*MLR*	
Continuous Indicators																			
$N_j = 5$	ICC = .02	**−15.09**	(1.99)	−8.90	(0.45)	**−16.09**	(0.39)	−6.03	(0.33)	**−13.00**	(0.31)	−4.49	(0.24)	**−10.20**	(0.25)	−1.62	(0.14)	**−25.62**	(1.20)
	ICC = .05	0.88	(3.99)	−1.94	(0.47)	**−11.22**	(0.38)	−4.31	(0.36)	**−10.65**	(0.32)	−3.37	(0.27)	−9.03	(0.26)	−2.30	(0.16)	**−16.46**	(0.86)
	ICC = .10	0.00	(1.05)	8.07	(0.48)	−5.71	(0.35)	4.73	(0.38)	−5.05	(0.29)	2.18	(0.29)	−6.00	(0.25)	0.01	(0.19)	−0.78	(0.53)
	ICC = .20	2.23	(0.37)	**12.94**	(0.43)	−0.35	(0.28)	9.58	(0.36)	−0.72	(0.27)	5.79	(0.29)	−1.54	(0.23)	3.09	(0.21)	3.41	(0.31)
	ICC = .40	1.52	(0.24)	8.14	(0.29)	0.91	(0.23)	6.50	(0.27)	0.64	(0.22)	4.17	(0.23)	0.02	(0.20)	1.79	(0.18)	2.37	(0.21)
$N_j = 10$	ICC = .02	**−34.79**	(4.23)	−5.27	(0.48)	**−12.83**	(0.39)	−4.39	(0.35)	**−10.33**	(0.32)	−3.09	(0.26)	−8.10	(0.24)	−1.62	(0.15)	**−17.90**	(1.71)
	ICC = .05	−6.05	(1.68)	8.27	(0.50)	−4.75	(0.35)	5.68	(0.38)	−4.58	(0.31)	1.70	(0.29)	−4.12	(0.25)	−0.18	(0.19)	−1.79	(0.51)
	ICC = .10	3.09	(0.47)	**15.22**	(0.43)	0.48	(0.29)	9.89	(0.35)	0.37	(0.27)	6.63	(0.28)	−0.77	(0.23)	3.43	(0.20)	3.26	(0.33)
	ICC = .20	3.40	(0.29)	8.77	(0.32)	2.20	(0.25)	8.20	(0.29)	1.35	(0.23)	5.92	(0.25)	1.10	(0.21)	3.97	(0.19)	2.35	(0.23)
	ICC = .40	2.39	(0.22)	5.53	(0.23)	1.78	(0.20)	5.33	(0.23)	2.19	(0.20)	4.04	(0.20)	1.06	(0.18)	2.74	(0.17)	1.70	(0.19)
$N_j = 20$	ICC = .02	0.30	(2.80)	−1.59	(0.45)	−8.31	(0.36)	−2.86	(0.36)	−9.04	(0.31)	−2.58	(0.26)	−7.33	(0.25)	−1.67	(0.17)	−7.10	(0.54)
	ICC = .05	−0.36	(0.43)	**10.25**	(0.41)	−1.82	(0.28)	6.11	(0.34)	−2.52	(0.27)	3.61	(0.27)	−2.93	(0.22)	1.29	(0.19)	3.15	(0.32)
	ICC = .10	0.50	(0.26)	6.72	(0.31)	−0.43	(0.23)	4.60	(0.26)	−0.40	(0.21)	3.69	(0.23)	−1.44	(0.20)	2.17	(0.18)	2.30	(0.23)
	ICC = .20	0.49	(0.21)	4.39	(0.23)	0.45	(0.20)	3.64	(0.22)	0.11	(0.18)	2.65	(0.20)	−0.02	(0.18)	1.74	(0.16)	1.39	(0.18)
	ICC = .40	0.02	(0.18)	3.59	(0.21)	0.60	(0.18)	3.12	(0.20)	0.18	(0.17)	2.75	(0.18)	0.31	(0.17)	1.73	(0.15)	1.04	(0.16)
Categorical Indicators																			
		Diffuse		*Weak 1a*		*Weak 1b*		*Weak .5a*		*Weak .5b*		*Weak 25a*		*Weak .25b*		*Informative*		*WLSM*	
$N_j = 5$	ICC = .02	**10.74**	(2.31)	−3.66	(0.43)	−3.65	(0.42)	−3.80	(0.32)	−4.93	(0.35)	−1.61	(0.25)	−4.61	(0.28)	−0.10	(0.19)	**25.20**	(0.69)
	ICC = .05	−9.26	(1.73)	0.94	(0.46)	−8.42	(0.38)	0.01	(0.33)	−7.52	(0.31)	−1.04	(0.24)	−6.46	(0.24)	−0.17	(0.15)	3.26	(0.62)
	ICC = .10	6.68	(1.70)	8.20	(0.46)	−1.93	(0.35)	2.60	(0.35)	−2.39	(0.30)	1.80	(0.26)	−2.82	(0.24)	0.11	(0.18)	−5.62	(0.49)
	ICC = .20	**12.82**	(0.54)	**11.88**	(0.40)	6.26	(0.32)	8.47	(0.34)	4.71	(0.28)	4.89	(0.27)	1.98	(0.24)	1.58	(0.19)	3.39	(0.33)
	ICC = .40	**21.75**	(0.76)	9.50	(0.33)	9.92	(0.33)	7.48	(0.30)	6.35	(0.28)	5.42	(0.25)	4.22	(0.24)	2.17	(0.19)	2.57	(0.27)
$N_j = 10$	ICC = .02	−7.63	(1.62)	−3.20	(0.44)	−5.45	(0.39)	−2.25	(0.32)	−5.36	(0.30)	−1.15	(0.24)	−3.57	(0.24)	−0.92	(0.16)	0.94	(0.82)
	ICC = .05	2.28	(0.94)	6.66	(0.44)	−3.39	(0.34)	3.38	(0.35)	−3.07	(0.29)	1.91	(0.28)	−3.28	(0.24)	0.35	(0.17)	−1.37	(0.54)
	ICC = .10	3.43	(0.37)	**10.32**	(0.39)	1.60	(0.29)	7.34	(0.35)	1.14	(0.27)	4.02	(0.27)	0.16	(0.23)	1.36	(0.19)	2.47	(0.35)
	ICC = .20	6.21	(0.27)	7.57	(0.28)	4.62	(0.24)	6.58	(0.26)	3.74	(0.22)	5.19	(0.23)	3.32	(0.21)	2.68	(0.18)	2.00	(0.25)
	ICC = .40	6.39	(0.25)	6.32	(0.24)	5.70	(0.24)	5.77	(0.22)	5.44	(0.23)	4.73	(0.21)	4.10	(0.20)	2.77	(0.17)	0.78	(0.20)
$N_j = 20$	ICC = .02	**−24.87**	(4.39)	−1.84	(0.48)	−5.27	(0.36)	−1.69	(0.36)	−5.58	(0.31)	−1.28	(0.26)	−5.08	(0.24)	−0.44	(0.16)	1.89	(0.69)
	ICC = .05	2.05	(0.41)	5.54	(0.34)	−0.06	(0.27)	3.12	(0.30)	−1.24	(0.25)	1.89	(0.25)	−1.25	(0.22)	0.50	(0.18)	3.11	(0.32)
	ICC = .10	1.00	(0.24)	4.85	(0.27)	0.24	(0.22)	3.92	(0.24)	0.16	(0.22)	2.57	(0.22)	−0.34	(0.19)	0.84	(0.17)	1.86	(0.22)
	ICC = .20	1.31	(0.19)	2.95	(0.20)	1.34	(0.19)	2.37	(0.19)	1.43	(0.18)	2.02	(0.18)	0.72	(0.17)	1.34	(0.15)	0.71	(0.17)
	ICC = .40	1.85	(0.19)	2.90	(0.19)	1.72	(0.18)	2.94	(0.18)	1.40	(0.17)	1.59	(0.17)	1.68	(0.17)	1.19	(0.15)	0.55	(0.16)

Note. MLR = robust maximum likelihood; WLSM = weighted least squares with mean-adjusted; ICC = intraclass correlation coefficient; J = number of groups; N_j = group size. For the weak conditions, numbers and letters (a = Γ^{-1} (.001, .001) and b = Γ^{-1} (−1,0)) refer to the variance and scale hyperparameter components of the normal and inverse-gamma priors, respectively. For instance, Weak 1a refers to the condition in which regression parameters $\sim N(1, 1)$ and cluster-level variance parameters $\sim \Gamma^{-1}$ (.001, .001), and Weak .25b refers to the condition in which regression parameters $\sim N(1, .25)$ and cluster-level variance parameters $\sim \Gamma^{-1}(-1, 0)$. Bold values indicate problematic bias levels greater than 10.00%. RMSE for each condition is presented in parentheses. For cells with asterisks (*.**), RMSE could not be computed because the estimated standard error was too large. A dashed line (-.–) refers to cells in which values could not be computed because none of the requested replications converged.

loadings often led to an underestimation of the between-level covariate effect.

Bayesian estimation with diffuse priors resulted in the most problematic relative bias for the between-level covariate estimate, whereby parameter recovery was noticeably worse for the continuous model compared with the categorical model. For Bayesian estimation with weakly informative priors, estimates were slightly more accurate for the categorical model compared with the continuous model and when factor loadings were held invariant versus freely estimated. Regarding the different estimation levels, parameter recovery improved as the prior for the regression parameters became more informative. When bias was problematic, the covariate effect was consistently underestimated for low to moderate values of ICC and overestimated for high values of ICC. Furthermore, placing Γ^{-1} (.001, .001) priors on the cluster-level variances led to more accurate estimation of the between-level covariate effect for low to moderate values of ICC than when $\Gamma^{-1}(-1,0)$ priors were placed on cluster-level variances. Conversely, using the $\Gamma^{-1}(-1,0)$ cluster-level variance prior led to more accurate estimates than the Γ^{-1} (.001, .001) cluster-level variance prior for high values of ICC. This pattern is visible in Table 2 for the weakly informative prior distribution levels where $\boldsymbol{\delta}_W$ and $\boldsymbol{\delta}_B \sim N(1, 1)$. Specifically, it can be seen that the weakly informative prior distribution level with an Γ^{-1} $(-1,0)$ prior placed on between-group variance components (labeled Weak lb in Table 2) often displayed problematic bias for low to moderate values of ICC, whereas the weakly informative prior distribution level with an Γ^{-1} (.001, .001) prior placed on between-group variance components (labeled Weak la in Table 2) did not. On the other hand, Table 2 shows that a reversal in bias occurred for these estimation levels when the ICC was large. To clarify, the weakly informative prior distribution level with an Γ^{-1} (.001, .001) prior specification for cluster-level variances exhibited problematic bias at high values of ICC, whereas the weakly informative prior distribution level with an Γ^{-1} $(-1,0)$ prior specification for cluster-level variances led to an accurate estimate under these conditions.

Bias of factor loadings. Figure 3 displays patterns of the mean absolute bias of the factor loadings as a function of the ICC, number of clusters, and cluster size. Each of the weakly informative prior distribution levels produced similar results regarding the pattern of absolute bias of the factor loadings and were collapsed in Figure 3 to reduce cluttering in the graph. Compared with bias of the structural path (displayed in Tables 2–5), it can be seen in Figure 3 that bias of the factor loadings followed the same general patterns. Specifically, bias was highest for Bayesian estimation with diffuse priors, followed by frequentist estimation, Bayesian estimation with weakly informative priors, and Bayesian estimation with informative priors. Moreover, bias was negatively related to the strength of the ICC, number of clusters, and sample size per cluster. Patterns of problematic bias were largely the same for the cluster-level factor loadings and the cluster-level covariate effect, with the following exceptions. First, in the categorical indicator model, slightly less bias was observed for the factor loadings versus the covariate effect for frequentist estimation and Bayesian estimation with diffuse priors. Second, in the continuous model, factor loadings were slightly less biased than the covariate effect for frequentist estimation, and factor loadings were substantially less biased than the covariate effect for Bayesian estimation with diffuse priors.

RMSE. RMSE of the between-level covariate effect is displayed in parentheses in Tables 2 through 5 for the noninvariant factor loading level and Tables 6 through 9 for the invariant factor loading level. RMSE findings mimic relative bias findings with respect to the different estimation levels. Across study conditions, the weakly informative and informative Bayesian estimation levels showed consistently small values of RMSE. For frequentist estimation and Bayesian estimation with diffuse priors, RMSE was negatively related to the number of groups, group size, and ICC. For instance, Tables 8 and 9 show that RMSE decreased for frequentist estimation and Bayesian estimation with diffuse priors as the amount of information provided by the data increased. In addition, frequentist estimation and Bayesian estimation with diffuse priors often produced smaller values of RMSE (a) when factor loadings were held invariant versus freely estimated, and (b) for the categorical model compared with the continuous model.

Coverage. For frequentist estimation, 95% coverage of the between-level covariate effect improved as the number of clusters and cluster size increased. In contrast, coverage became worse for each of the Bayesian estimation levels as J and N_j increased. Frequentist estimation often produced poor coverage at low values of ICC, with coverage improving at moderate values of ICC and becoming worse at high values of ICC. For each of the Bayesian estimation levels, coverage was inversely related to strength of the ICC. Figure 4 displays this pattern of results for the continuous model with noninvariant factor loadings. Patterns of coverage were similar for each of the weakly informative prior distribution levels regardless of the informativeness of the regression priors. However, coverage results diverged slightly depending on whether cluster-level variance priors were specified as Γ^{-1} (.001, .001) or Γ^{-1} $(-1,0)$. Consequently, the results presented in Figure 4 for Bayesian estimation with weakly informative priors are averaged across the different estimation levels, but separated by the type of prior that was specified for cluster-level variances. In Figure 4, it can be seen that coverage was good for all of the weakly informative prior distribution levels at low values of ICC. However, coverage became worse as the ICC increased, with this effect being more pronounced for the weakly informative

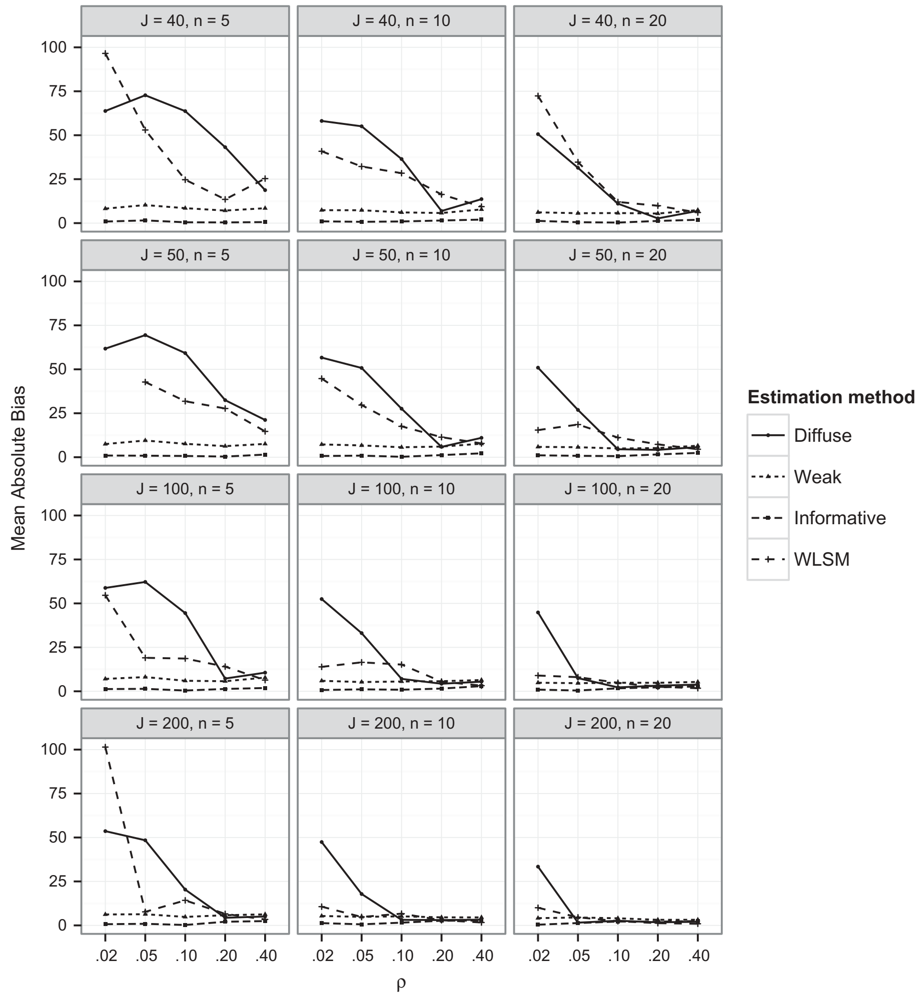

FIGURE 3 Mean absolute bias of the freely estimated cluster-level factor loadings for the categorical model with noninvariant factor loadings. J = number of groups; n = sample size per group; ρ = intraclass correlation; Weak = an average of all weakly informative prior distribution levels; WLSM = weighted least squares with mean-adjusted. Due to model nonconvergence, there is a missing datum for the frequentist estimation condition in which $J = 50$, $n = 5$, and $\rho = .02$.

prior distribution levels in which the between-level variances were specified as $\Gamma^{-1}(.001, .001)$ compared with the weakly informative prior distribution levels in which the variance components were specified as $\Gamma^{-1}(-1, 0)$.

Overall, frequentist confidence interval coverage was better for the categorical model ($M = .95$, $SD = .04$) compared with the continuous model ($M = .93$, $SD = .03$). For Bayesian estimation with diffuse priors, there was an interaction between the type of factor indicator and whether factor loadings were held invariant. Specifically, coverage in the categorical model was better for noninvariant factor loadings ($M = .97$, $SD = .05$) versus invariant factor loadings ($M = .93$, $SD = .09$); for the continuous model, coverage was better for invariant factor loadings ($M = .96$, $SD = .03$) versus noninvariant factor loadings ($M = .94$, $SD = .05$). Across the weakly informative prior distribution levels, coverage was better for the categorical model ($M = .98$, $SD = .02$) compared with the continuous model ($M = .97$, $SD = .03$) and when factor loadings were noninvariant ($M = .98$, $SD = .02$) versus invariant ($M = .97$, $SD = .03$). The best coverage for the weakly informative prior distribution levels occurred in the categorical model with noninvariant loadings, in which nominal coverage of Type I errors was observed for all estimation levels except with $J = 200$ and ICC $\geq .20$. In contrast, coverage for the continuous indicator model with invariant factor loadings was too low when ICC $\geq .20$, regardless of the number of groups.

DISCUSSION

Consistent with previous research (Hox & Maas, 2001; Li & Beretvas, 2013; Lüdtke et al., 2011; Meuleman & Billiet, 2009; Muthén & Satorra, 1995; Ryu, 2011; Ryu & West, 2009), frequentist estimation of multilevel SEM led to poor convergence rates when there was a small number of clusters,

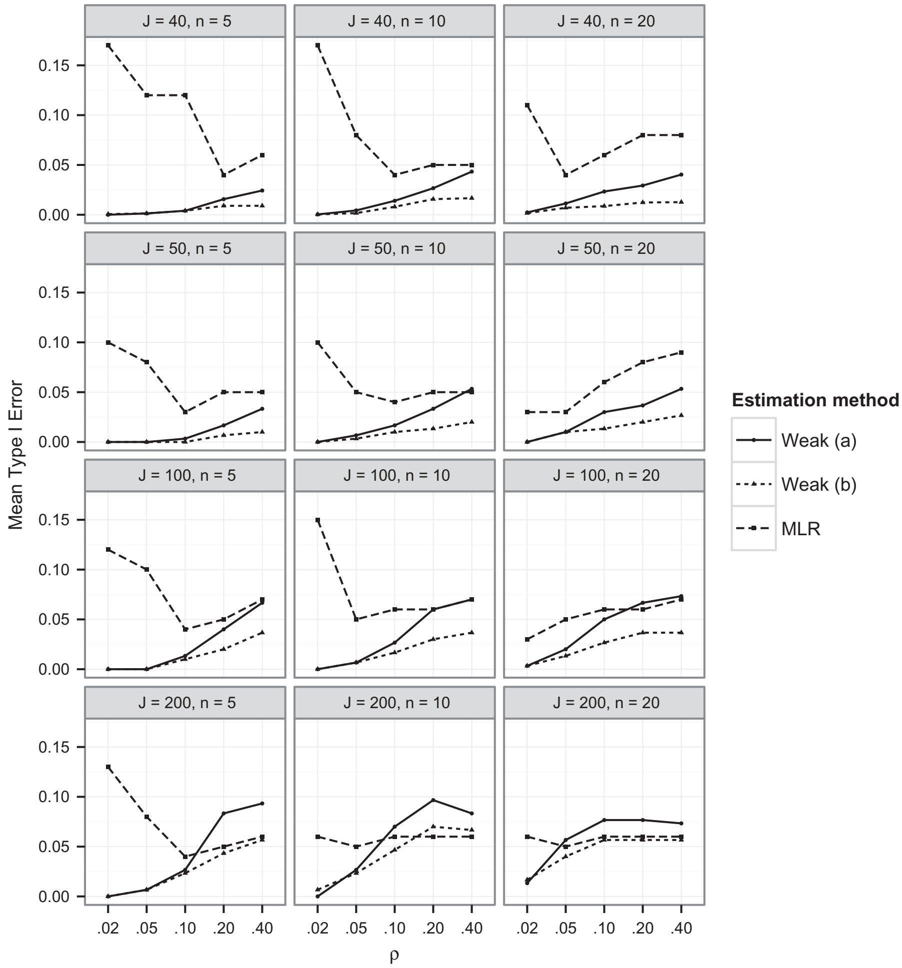

FIGURE 4 Type I error rates of the cluster-level covariate effect for the continuous model with noninvariant factor loadings. J = number of groups; n = sample size per group; ρ = intraclass correlation; Weak (a) = mean Type I error rate of the weakly informative prior distribution levels with cluster-level variance priors $\sim \Gamma^{-1}$ (.001, .001); Weak (b) = mean Type I error rate of the weakly informative prior distribution levels with cluster-level variance priors $\sim \Gamma^{-1}(-1, 0)$; MLR = robust maximum likelihood.

smaller cluster sizes, and low ICCs. Out of these three design factors, the ICC had the largest impact on convergence rates, likely because a lack of between-group variation often led to singularity of the between-level covariance matrix. Nonconverged solutions are useless in practice, and convergence problems are therefore a considerable hurdle in the application of multilevel SEM using frequentist techniques. It should be noted that applied researchers have several options for increasing convergence when using a frequentist approach (e.g., by using more informative starting values, adapting the model, using constraints, or increasing the number of iterations, although this latter option will not fix an inherent problem with convergence). Alternatively, the results of this simulation suggested that Bayesian estimation can be used to obtain admissible estimates in situations where a frequentist approach is likely to fail. The ability for Bayesian estimation to improve convergence is a noted advantage of using the framework in the context of multilevel latent variable modeling (e.g., Asparouhov & Muthén, 2012).

For the within level, each of the estimation methods led to negligible bias for the continuous model (in the range of ±5%). Previous research on ML estimation of multilevel SEM with continuous variables has found that within-level estimates were generally accurate with as few as 50 groups and five observations per group (Hox et al., 2010), and, in part, our findings indicated that within-level estimates were accurate with an even smaller number of groups ($J = 40$).

To examine accuracy in the estimation of the factor loadings, the mean absolute bias of the freely estimated factor loadings for each condition was calculated and compared to the relative bias of the covariate effect. Importantly, the factor loadings were generally recovered with slightly more accuracy than the covariate effect at each level of the model. Thus, researchers who are interested primarily in factor loading estimates when using multilevel SEM might be able to obtain accurate results with somewhat smaller sample sizes than when structural relationships are also of interest.

For the between level, Bayesian estimation with diffuse priors did a poor job of recovering parameter estimates in almost every modeling condition. Findings for the diffuse prior distribution level are contrary to the results of Hox et al. (2012), who found that Bayesian estimation of multilevel SEM with diffuse priors produced more accurate estimates than frequentist estimation. However, the simulation by Hox et al. (2012) considered estimation of multilevel SEM in the context of cross-cultural research where the average group size was considerably larger than that considered here. Thus, further research is needed to determine how the number of groups and group size interact to impact parameter recovery for Bayesian estimation of multilevel SEM with diffuse priors.

Categorical Items

Holding factor loadings invariant across levels led to a substantial improvement in recovery of the between-level covariate effect for the categorical model. This finding is consistent with that of Lüdtke et al. (2011), who found that simplifying the measurement model by placing an invariance constraint on the factor loadings can lead to more accurate parameter recovery in a multilevel latent covariate model with continuous variables. However, frequentist convergence problems were so severe for the categorical model that it would not be practically useful to use categorical items with 40 to 50 clusters unless the item ICCs were extremely large ($\geq$ .40). Notably, convergence rates were acceptable (> 80%) in only 23% of the study conditions for the categorical model, compared with 34% of the study conditions for the continuous model. Nonconvergence was therefore much more problematic in the categorical model versus the continuous model. It is important to note that only dichotomous items were considered in this investigation; convergence rates might be higher for items with a larger number of categories (see, e.g., Yang-Wallentin et al., 2010).

Recommendations on the Elicitation and Specification of Priors

Priors for models with categorical versus continuous variables. For each of the Bayesian estimation levels, parameter recovery was better in the categorical model compared with the continuous model. The tendency for priors to have a large impact on estimates for categorical variables is a recognized phenomenon in the SEM literature that has been referred to as prior assumption dependence (Asparouhov & Muthén, 2010). Although our findings indicated that Bayesian estimation might be particularly advantageous for multilevel SEM with categorical variables, the occurrence of prior assumption dependence implies that the elicitation of accurate priors might be more important for models with categorical variables than it is for models with continuous variables. In this study, regression priors for the informative and weakly informative prior distribution levels were specified to have greatest mass in the general neighborhood of the generating values. In practice, researchers might not have a good guess about the actual location of parameters, and specifying priors with inaccurate locations could have a detrimental impact on parameter estimates due to the prior assumption dependence (Depaoli, 2014). Further, it can be challenging to define priors for use with probit regression because of the difficulty in defining a realistic bound on model parameters. Priors can be defined more easily for continuous variables because they are on a probability scale. For instance, the specification of regression priors used in this study can be thought of in terms of their 95% limits for the continuous model (see Figure 2). In the case of logistic or probit models, it is more difficult to construct regression priors because boundaries for the prior cannot be defined in such a straightforward manner. However, we recommend the default prior for categorical items implemented in M*plus* as a good starting point for estimation via Bayes. In particular, Gelman et al. (2008) indicated that an $N(0, 5)$ prior is a weakly informative prior in the context of logistic and probit models. This statement corroborates our findings in that estimates using default priors were much more accurate for categorical items compared to continuous.

Priors for between-level random effects. The research literature is mixed on whether the use of $\Gamma^{-1}(.001, .001)$ priors for random effects in two-level models are completely noninformative (Gelman, 2006). Early examples using the BUGS software routinely implemented $\Gamma^{-1}(.001, .001)$ priors as a so-called noninformative prior for cluster-level variances (Spiegel-halter et al., 2003), yet it appears to be quite informative in the context of multilevel SEM when the number of clusters is small. In this investigation, estimation accuracy was importantly related to the specification of priors for cluster-level variance components. As a result, we recommend that applied researchers interested in using a Bayesian approach to multilevel SEM with small samples consider the magnitude of between-group variation present in their data to inform the choice of priors. One way to assess the amount of variance located between groups is to first estimate a model under ML to obtain ICC estimates for each observed indicator. We recommend specifying a uniform prior for indicators with large ICCs such as the $\Gamma^{-1}(-1, 0)$ specification and a more informative prior for indicators with small to moderate ICCs. It is important to note that the $\Gamma^{-1}(.001, .001)$ prior used in this investigation is only one possible specification of an informative prior for cluster-level variance components and that setting the shape and scale hyperparameters to slightly larger values such as 0.1 or 1.0 might be useful as well.

Concluding Remarks and Recommendations for Applied Researchers

Bayesian estimation of multilevel SEM is an attractive alternative to frequentist estimation when priors are selected carefully. However, if diffuse priors are going to be used by applied researchers in the context of multilevel SEM, additional research is needed to determine the conditions under which their use leads to accurate parameter recovery. Some of the conditions in this study, particularly for continuous items with diffuse priors, showed very high bias levels. These extreme bias levels were likely a result of spikes in the trace plots, where the MCMC estimate for a given iteration jumps to an extreme and out-of-bound value. We recommend that researchers always look at the trace plots for every single parameter in the model to diagnose whether spikes were present. If spikes are present, then these out-of-bound values will impact parameter estimates, sometimes rather drastically. In this case, we recommend the researcher to use a more informative prior for parameters to avoid the presence of spikes.

In the case where informative or weakly informative priors are desired, there are additional issues to be aware of. Perhaps one of the more common forms of specifying informative or weakly informative priors is to use data-driven priors. Within this context of data-driven priors, it is common to use ML estimates (or akin) to inform the prior distribution (see, e.g., Berger, 2006; Brown, 2008; Candel & Winkens, 2003; van der Linden, 2008). One criticism of using a data-driven prior is that the sample data have been utilized twice in the estimation—once when constructing the prior and another when the posterior distribution was estimated. This "double-dipping" into the sample data could potentially distort parameter estimates, as well as artificially decrease the uncertainty in those estimates (Darnieder, 2011).[5] As a result, we recommend against using data-driven priors that are derived through a preestimation process.

The use of Bayesian estimation in the context of latent variable modeling is burgeoning due to recent methodological advances (see e.g., Muthén & Asparouhov, 2012a), and the need to study the impact of prior distributions in SEM is commensurate with its growing use. Until more research has been conducted to establish suitable methods for eliciting priors in the social sciences, we recommend that researchers interested in a Bayesian approach to multilevel SEM always conduct a sensitivity analysis to determine the extent to which model estimates depend on prior assumptions.

[5] An earlier point was made about using ML to derive ICC estimates. It is important to clarify here that we do not consider deriving ICC estimates in this manner as double-dipping in the sample data. Double-dipping refers specifically to the instance where a particular parameter is estimated, that estimate is then used to form a prior, and then a second estimation process implementing that prior takes place with the same data.

REFERENCES

Ansari, A., & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, *65*, 475–496. doi:10.1007/BF02296339

Asparouhov, T., & Muthén, B. (2007, August). *Computationally efficient estimation of multilevel high-dimensional latent variable models*. Paper presented at the Joint Statistical Meeting, Salt Lake City, UT.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus* (*M*plus Technical Report). Retrieved from statmodel.com/download/BayesAdvantages18.pdf

Asparouhov, T., & Muthén, B. (2012). *General random effect latent variable modeling: Random subjects, items, contexts, and parameters*. Retrieved from statmodel.com/download/NCME_revision2.pdf

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151–164. doi:10.1037/a0030642

Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowitz, J. S., . . . Watson, J. (2011). Intraclass correlation associated with therapists: Estimates and applications in planning psychotherapy research. *Cognitive Behaviour Therapy*, *40*, 15–33. doi:10.1080/16506073.2010.520731

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *3*, 385–402.

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

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), *Structural equation modeling in example: Applications in educational, sociological and behavioral research* (pp. 160–188). New York, NY: Cambridge University Press.

Bosker, R., & Snijders, T. A. B. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Brown, L. D. (2008). In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies. *The Annals of Applied Statistics*, *2*, 113–152.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*, 473–514.

Candel, J. J. M., & Winkens, B. (2003). Performance of empirical Bayes estimators of level-2 random parameters in multilevel analysis: A Monte Carlo study for longitudinal designs. *Journal of Educational and Behavioral Statistics*, *28*, 169–194.

Cheung, M. W.-L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, *12*, 598–619.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*, 685–709. doi:10.1007/S11336-013-9328-2

Congdon, P. (2010). *Applied Bayesian hierarchical methods*. Boca Raton, FL: CRC.

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Depaoli, S. (2014). The impact of inaccurate "informative" priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling*, *21*, 239–252. doi:10.1080/10705511.2014.882686

Diya, L., Li, B., Heede, K., Sermeus, W., & Lesaffre, E. (2013). Multilevel factor analytic models for assessing the relationship between nurse-reported adverse events and patient safety. *Journal of the Royal

Statistical Society: Series A (Statistics in Society), *177*, 237–257. doi:10.1111/rssa.12012

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*, 355–366. doi:10.1111/1467-9868.00236

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149–167. doi:10.1016/j.leaqua.2004.09.009

Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist*, *39*, 157–172. doi:10.1207/sl5326985ep3903_2

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling*, *18*, 229–252. doi:10.1080/10705511.2011.557338

Fox, J. P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288. doi:10.1007/BF02294839

Fox, J. P. (2010). *Bayesian item response modeling*. New York, NY: Springer.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 515–533.

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

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*, 1360–1383. doi:10.1214/08-AOAS191

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In W R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 1–19). New York, NY: Chapman & Hall.

Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioral Statistics*, *32*, 252–286. doi:10.3102/1076998606298042

Gottfredson, N. C, Panter, A., Daye, C. E., Allen, W. F., & Wightman, L. F. (2009). The effects of educational diversity in a national sample of law students: Fitting multilevel latent variable models in data with categorical indicators. *Multivariate Behavioral Research*, *44*, 305–331. doi:10.1080/00273170902949719

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, *8*, 157–174. doi:10.1207/S15328007SEM0802_l

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*, 157–170. doi:10.1111/j.1467-9574.2009.00445.x

Hox, J., & Roberts, J. K. (Eds.). (2011). *Handbook of advanced multilevel analysis*. New York, NY: Taylor & Francis.

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93.

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology*, *52*, 3l0–32l. doi:10.1037/0022-0167.52.3.310

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*, 352–352. doi:10.1207/S15328007SEM0803_1

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.

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

Kaplan, D., Kim, J. S., & Kim, S. Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 592–612). Thousand Oaks, CA: Sage.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, *18*, 212–228. doi:10.1080/10705511.2012.659623

Lee, S. Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, *46*, 153–160. doi:10.1007/BF02293896

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. West Sussex, UK: Wiley.

Li, F., Duncan, T. E., Harmer, P., Acock, A., & Stoolmiller, M. (1998). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling*, *5*, 294–306. doi:10.1080/10705519809540106

Li, X., & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling*, *20*, 241–264. doi:10.1080/10705511.2013.769391

Little, J. (2013). Multilevel confirmatory ordinal factor analysis of the Life Skills Profile–16. *Psychological Assessment*, *25*, 810–825. doi:10.1037/a0032574

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*, 444–467. doi:10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. doi:10.1037/a0012869

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer. doi:10.1007/978-0-387-71265-9

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*, 106–124. doi:10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764–802. doi:10.1080/00273170903333665

Martin, J. K, & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, *40*, 505–517. doi:10.1007/BF02291552

Mathisen, G. E., Torsheim, T., & Einarsen, S. (2006). The team-level model of climate for innovation: A two-level confirmatory factor analysis. *Journal of Occupational and Organizational Psychology*, *79*, 23–35. doi:10.1348/096317905X52869

Matsueda, R. L. (2012). Key advances in the history of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 17–42). New York, NY: Guilford.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*, 259–284. doi:10.1037/1082-989X.10.3.259

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, *3*, 45–58.

Mitchell, M. M., & Bradshaw, C. P. (2013). Examining classroom influences on student perceptions of school climate: The role of classroom management and exclusionary discipline strategies. *Journal of School Psychology*, *51*, 599–610. doi:10.1016/j.jsp.2013.05.005

Morselli, D., Spini, D., & Devos, T. (2012). Human values and trust in institutions across countries: A multilevel test of Schwartz's hypothesis of structural equivalence. *Survey Research Methods*, *6*, 49–60.

Murray, D. M., & Hannan, P. J. (1990). Planning for the appropriate analysis in school-based drug use prevention studies. *Journal of Consulting and Clinical Psychology*, *58*, 458–468. doi:10.1037/0022-006X.58.4.458

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585. doi:10.1007/BF02296397

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338–354. doi:10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398. doi:10.1177/0049124194022003006

Muthén, B. O., & Asparouhov, T. (2012a). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. doi:10.1037/a0026802

Muthén, B. O., & Asparouhov, T. (2012b). Item response modeling in M*plus*: A multi-dimensional, multi-level, and multi-timepoint example. Forthcoming in W. J. van der Linden & R. K. Hambleton (2015), *Handbook of item response theory: Models, statistical tools, and applications*. Boca Raton, FL: Chapman & Hall/CRC Press. Retrieved from statmodel.com/download/IRT1Version2.pdf

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316. doi:10.2307/271070

Muthén, L. K., & Muthén, B. (1998–2013). M*plus* (version 7). Los Angeles, CA: Muthén & Muthén.

Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research*, *46*, 691–731. doi:10.1080/00273171.2011.589280

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*, 161–182. doi:10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. doi:10.1037/a0020141

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. doi:10.1007/BF02295939

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2012). Multilevel structural equation modeling. In R. H. Hoyle (Ed.), *Handhook of structural equation modeling* (pp. 512–531). New York, NY: Guilford.

Raudenbush, S. W., & Bryk. A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Roesch, S. C., Aldridge, A. A., Stocking, S. N., Villodas, F., Leung, Q., Bartley, C. E., & Black, L. J. (2010). Multilevel factor analysis and structural equation modeling of daily diary coping data: Modeling trait and state variation. *Multivariate Behavioral Research*, *45*, 767–789. doi:10.1080/00273171.2010.519276

Rupp, A. A., Dey, D. K., & Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling*, *11*, 424–451. doi:10.1207/sl5328007sem1103_7

Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods*, *43*, 1066–1074. doi:10.3758/sl3428-011-0115-7

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, *16*, 583–601. doi:10.1080/10705510903203466

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37–52. doi:10.1007/BF02294318

Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study: Outcome and mediating variables, by sex and ethnicity. *American Journal of Epidemiology*, *144*, 425–433. doi:10.1093/oxfordjournals.aje.a008945

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., & Lunn, D. (2003). *BUGS: Bayesian inference using Gibbs sampling*. Cambridge, UK: Medical Research Council Biostatistics Unit.

Steele, F., & Goldstein, H. (2006). A multilevel factor model for mixed binary and ordinal indicators of women's status. *Sociological Methods & Research*, *35*, 137–153. doi:10.1177/0049124106289112

Thoonen, E. E., Sleegers, P. J., Oort, F. J., Peetsma, T. T., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly*, *47*, 496–536. doi:10.1177/0013161X11400185

Toland, M. D., & De Ayala, R. J. (2005). A multilevel factor analysis of students' evaluations of teaching. *Educational and Psychological Measurement*, *65*, 272–296. doi:10.1177/0013164404268667

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842–860. doi:10.1111/cdev.12169

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavorial Staistics*, *33*, 5–20.

Wu, J. Y. & Kwok, O. (2012). Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling*, *19*, 16–35. doi:10.1080/10705511.2012.634703

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*, 392–423. doi:10.1080/10705511.2010.489003