

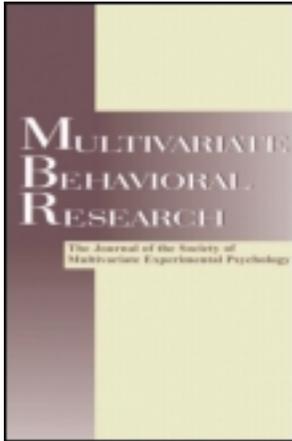
This article was downloaded by: [University of California, Los Angeles (UCLA)]

On: 05 April 2012, At: 12:08

Publisher: Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Multivariate Behavioral Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/hmbr20>

Mixture Factor Analysis for Approximating a Nonnormally Distributed Continuous Latent Factor With Continuous and Dichotomous Observed Variables

Melanie M. Wall ^a, Jia Guo ^b & Yasuo Amemiya ^c

^a Departments of Biostatistics and Psychiatry, Columbia University

^b Sanofi-Aventis US

^c IBM Thomas J. Watson Research Center

Available online: 03 Apr 2012

To cite this article: Melanie M. Wall, Jia Guo & Yasuo Amemiya (2012): Mixture Factor Analysis for Approximating a Nonnormally Distributed Continuous Latent Factor With Continuous and Dichotomous Observed Variables, *Multivariate Behavioral Research*, 47:2, 276-313

To link to this article: <http://dx.doi.org/10.1080/00273171.2012.658339>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Mixture Factor Analysis for Approximating a Nonnormally Distributed Continuous Latent Factor With Continuous and Dichotomous Observed Variables

Melanie M. Wall

Departments of Biostatistics and Psychiatry, Columbia University

Jia Guo

Sanofi-Aventis US

Yasuo Amemiya

IBM Thomas J. Watson Research Center

Mixture factor analysis is examined as a means of flexibly estimating nonnormally distributed continuous latent factors in the presence of both continuous and dichotomous observed variables. A simulation study compares mixture factor analysis with normal maximum likelihood (ML) latent factor modeling. Different results emerge for continuous versus dichotomous outcomes. For dichotomous outcomes, normal ML path estimates have bias that worsens as latent factor skew/kurtosis increases and does not diminish as sample size increases, whereas the mixture factor analysis model produces nearly unbiased estimators as sample sizes increase (500 and greater) and offers near nominal coverage probability. For continuous outcome variables, both methods produce factor loading estimates with minimal bias regardless of latent factor skew, but the mixture factor analysis is more efficient. The method is demonstrated using data motivated by a study on youth

Correspondence concerning this article should be addressed to Melanie M. Wall, Departments of Biostatistics and Psychiatry, Columbia University, 1051 Riverside Drive, Unit 48, New York, NY 10032. E-mail: mmwall@columbia.edu

with cystic fibrosis examining predictors of treatment adherence. In summary, mixture factor analysis provides improvements over normal ML estimation in the presence of skewed/kurtotic latent factors, but due to variability in the estimator relating the latent factor to dichotomous outcomes and computational issues, the improvements were only fully realized, in this study, at larger sample sizes (500 and greater).

Typically, for estimation purposes, latent factor measurement models and structural equation models in general assume the underlying latent factors are normally distributed. This normality assumption is most certainly not always true for real problems (Micceri, 1989) and assuming it is may lead to incorrect inference for model parameters. The purpose of this article is to examine the performance of using mixture models, specifically the “mixture factor analysis model” (B. Muthén, 2007, p. 3), as a means of weakening the distributional assumptions for the underlying latent factors and thus potentially robustifying the parameter estimation. The current study considers both continuous and dichotomous observed outcome variables that are manifest from an underlying continuous latent factor. The inclusion of both is of interest because nonnormality of the latent factors may differentially impact estimation of the linear relationship between a latent factor and a continuous outcome versus the nonlinear relationship between a dichotomous observed outcome variable and the latent factor.

Mixture models have become increasingly popular in the social sciences and public health research as an analytic approach for describing underlying latent subgroups in populations that may exhibit specifically distinct patterns across variables and relationships among them. In particular, growth mixture models (B. Muthén et al., 2002; B. Muthén & Shedden, 1999; Nagin, 1999; Verbeke & Lesaffre, 1996) have received a lot of attention as have factor and structural equation modeling mixture models (Arminger & Stein, 1997; Dolan & Van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997; Lee & Song, 2003; Lubke & Muthén, 2005; Yung, 1997; Zhu & Lee, 2001). In his overview of hybrid latent variable models (i.e., models including both categorical and continuous latent variables), B. Muthén (2007) provides a taxonomy of the different types of mixture models. In the current article we examine the so-called mixture factor analysis model, which assumes measurement invariance for the continuous latent factor but incorporates a mixture model (i.e., an underlying categorical latent class variable) for the continuous latent factor.

Commonly, the goal of mixture modeling is to identify distinct latent subgroups that make up a supposed heterogeneous population. In contrast, this article uses mixture models solely as a means to flexibly model distinctly nonnormal latent factor distributions. Finite mixtures (McLaclan & Peel, 2000; Titterton, Smith, & Makov, 1985) are useful to approximate intractable or complex distributions with a small number of simpler component distributions.

In particular, finite mixtures of normal distributions have been used extensively, especially in the statistical literature, as a mathematical tool to approximate non-normal, generally complex continuous distributions (Carroll, Roeder, & Wasserman, 1999; Escobar & West, 1995; Richardson & Green, 2002; Sorenson & Alspach, 1971; for a review see McLachlan & Peel, 2000). Bauer and Curran (2003, 2004) and Bauer (2007) provide excellent expositions describing the difficulties of discerning from data whether the mixtures represent true distinct underlying groups or just an approximation to a homogeneous but nonnormally distributed underlying group. Indeed, Bauer (2007) is particularly critical of the almost solitary use of mixture models for the purpose of identifying underlying subgroups when it is likely there are often no true subgroups in the population but just an arbitrary nonnormally distributed continuum. Our goal is to use a mixture factor analysis model that can approximate the unknown latent factor distribution well enough to ensure correct estimation and inference for the other parameters in the model.

The basic idea of using mixtures of normals to approximate a nonnormal continuous distribution is demonstrated in Figure 1 where a right skewed distribution, a standardized chi-square distribution with 1 degree of freedom, is overlaid with the best fitting normal (upper left) and the best fitting mixture of 2, 3, and 4 normals, respectively. Details of the method are described later, but as can be seen from the figure, as more components are added, the aggregation of the mixtures appears more similar to the target chi-square distribution. Indeed with enough components, the mixture distribution can be made arbitrarily close to the target distribution (Sorenson & Alspach, 1971).

As a motivating example, we consider a study of adolescents with cystic fibrosis that examined the effects that stressors and supports (social, familial, and personal) have on adherence to treatment regimes. Self-report questionnaire data were collected with multiple items intended to measure several different dimensions of stressors and supports as well as adherence behavior (Patterson, Wall, Berge, & Milla, 2009). Here we focus on the conceptual variable “illness-strains” and how it affects the dichotomous indicator of whether or not the patient adheres to treatment. The variable illness strain is modeled as a continuous latent variable underlying three measured variables labeled “emotional strains,” “appearance worries,” and “physical strains.” Figure 2A and 2B presents a graphical representation of the model where covariates, gender and age, are either included or not.

The potential limitation of the commonly used estimation method (normal maximum likelihood) for the structural equation model in Figure 2A and 2B is that it assumes the latent factor f representing illness strains is normally distributed. This assumption is likely unrealistic because the researchers expect a priori a right skewed distribution for illness strains among the sample where most adolescents will be in the low to moderate range but also a substantial minority

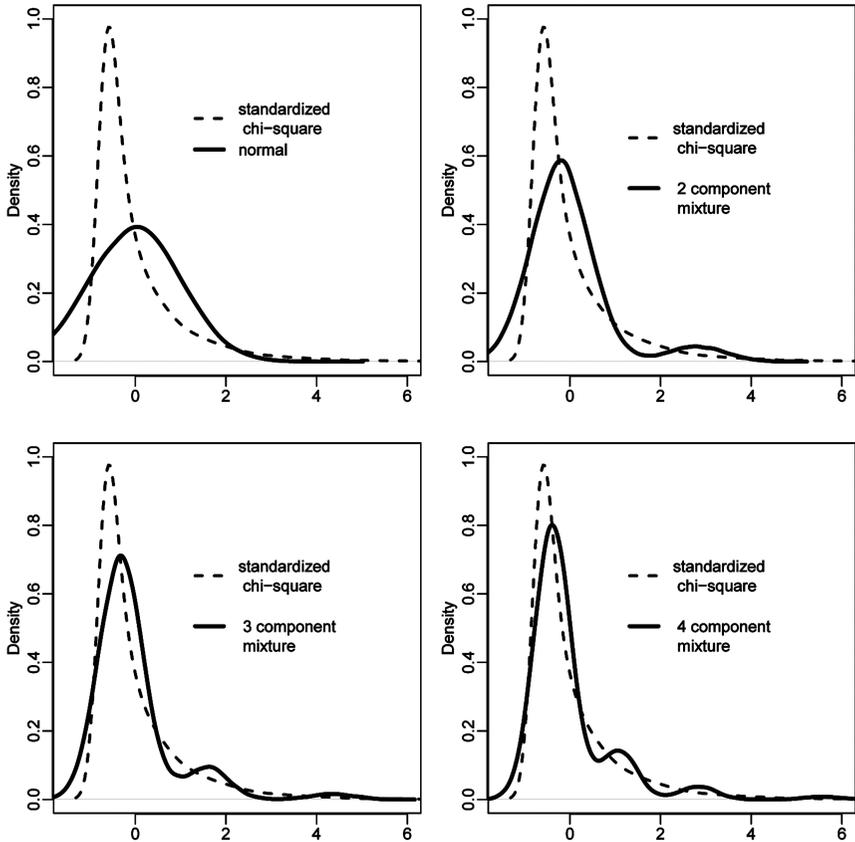
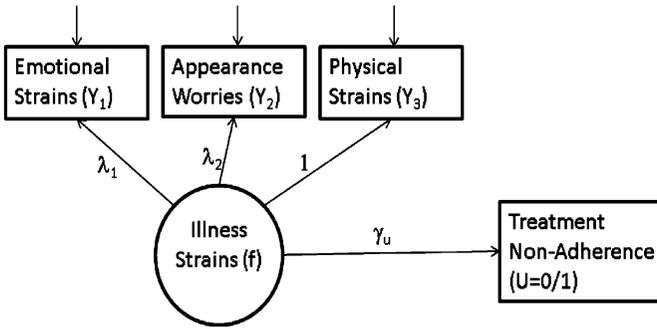


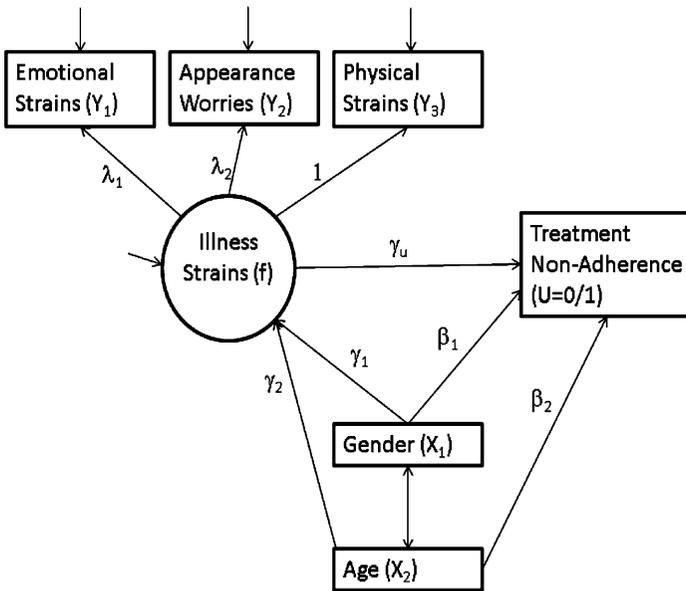
FIGURE 1 Comparison of a standardized chi-square distribution with 1 degree of freedom to the best fitting normal, 2-, 3-, and 4-component mixtures of normal models. Best fitting models are based on respective maximum likelihood estimates to a random sample of $n = 50,000$ observations from the standardized chi-square distribution. For $df = 1$, the mean = 1 and standard deviation = $\sqrt{2}$ are used to standardize the chi-square distribution.

will exhibit extreme levels of the latent factor. Simulated data mimicking that found in the original study is shown in Figure 3 where the three measurements for illness-related strains and worries are skewed and kurtotic. If we take Y_1 , Y_2 , and Y_3 to be linearly related to the latent factor, then it is probably not reasonable to assume the latent factor is itself normally distributed.

In this current article we weaken the normality assumption for the latent factor by instead assuming that it follows a mixture of normal distributions through the use of a mixture factor analysis model. A simulation study is used to

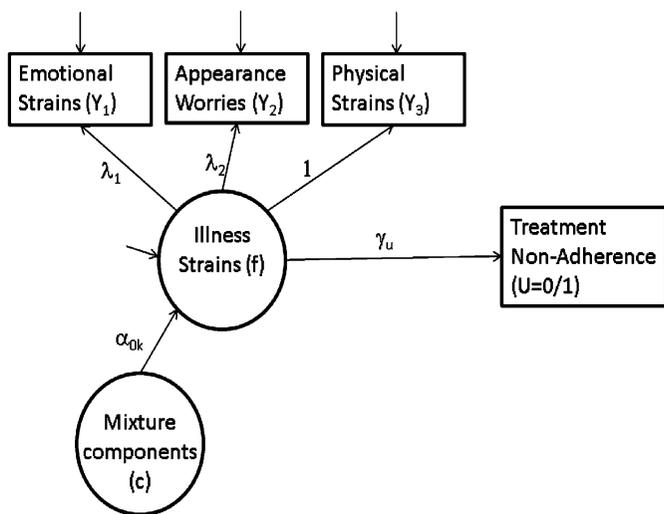


(a)

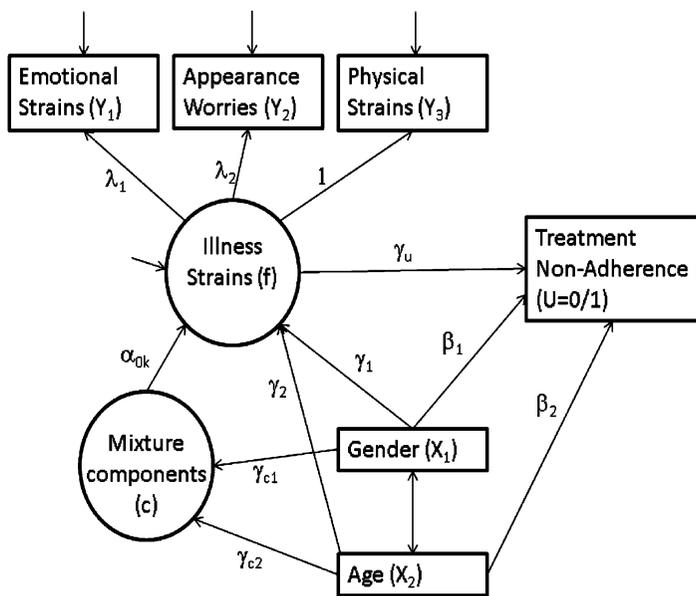


(b)

FIGURE 2 Models relating illness strains (a continuous latent factor f) in children with cystic fibrosis to nonadherence to treatment (a dichotomous observed variable U). A. Basic latent factor model, B. Adjustment for covariates, C. Inclusion of mixtures (latent classes c) to flexibly model the distribution of illness strains, D. Inclusion of mixtures and covariates. Additional single-headed arrows pointing to continuous variables, Y_1, Y_2, Y_3 , and latent factor f , represent measurement and equation errors. Following Mplus graphical notation, no separate error arrows are drawn for categorical variables U or latent c due to inseparability between mean and variance for categorical outcomes. (continued)



(c)



(d)

FIGURE 2 (Continued).

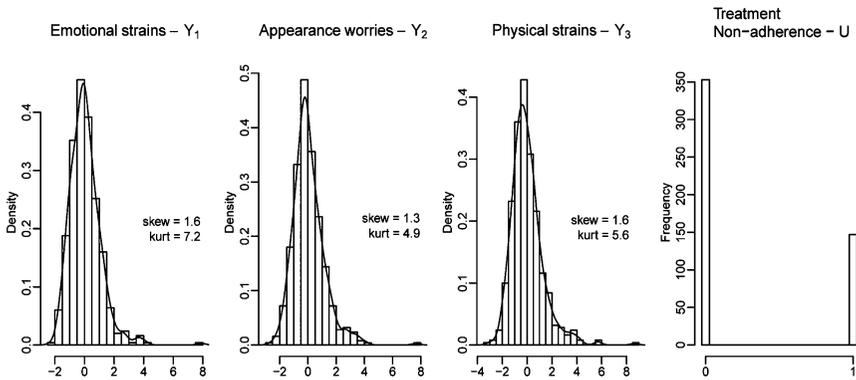


FIGURE 3 Histograms of data ($n = 500$) for simulated cystic fibrosis illness strains example.

investigate the performance of the mixture factor analysis under varying amounts of skew and kurtosis of the underlying factor. Of practical interest is whether it is necessary in the first place to worry about the lack of normality and so we also extensively compare the mixture factor analysis results with the traditional maximum likelihood (ML) solution, which assumes the latent factor is normally distributed. In particular we examine and contrast estimates and inference for the paths relating the latent factor to continuous and dichotomous observed indicators. Practical recommendations are made based on simulation results including the choice of the number of mixture components. Implementation of the mixture factor analysis approach is demonstrated using Mplus.

THE MIXTURE FACTOR ANALYSIS MODEL

For concrete illustration, we consider a model for the the cystic fibrosis illness strains example, but the modeling framework can be extended straightforwardly to problems with more observed indicators, covariates, and outcomes. Our working assumption is that the relationship between the latent factor “illness strains” and its indicators, covariates, and main outcome nonadherence is homogeneous across the population. The mixture factor analysis model proposed here solely targets the distribution of the latent factor in order to provide a more flexible distributional form than normality. The model is made up of two parts: (a) a typical measurement model relating observed indicators and outcomes to an underlying continuous latent factor and covariates that is invariant across all components of the latent factor mixture and (b) a K -component normal mixture

model for the underlying latent factor incorporating covariates into the probabilities and the means of each component. The K-component normal mixture is operationalized using a K-category latent class variable (latent categorical variable) directly influencing the continuous latent factor.

Measurement Model

We begin by defining the measurement model, which remains fixed regardless of how the latent factor distribution is specified. In other words, the measurement model parameters, which include the factor loadings, intercepts, error variances, and coefficients relating covariates to observed indicators, are held invariant across the multiple mixture components that are specified for the latent factor f_i .

Let $\mathbf{Y}_i = (Y_{1i}, Y_{2i}, Y_{3i})$ be the three observed scales shown in Figure 3 for each individual i . It is assumed that these continuous observed measures are linearly related to a single underlying latent factor f plus error through the usual linear factor model, that is,

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \end{pmatrix} = \begin{pmatrix} v_{01} \\ v_{02} \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 1 \end{pmatrix} f_i + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \end{pmatrix}, \quad (1)$$

where the measurement errors have mean 0 and are uncorrelated with one another, that is, $E(\epsilon_{ji}) = 0$, $Cov(\epsilon_{ji}, \epsilon_{j'i}) = 0$ for $j \neq j'$ with variance $Var(\epsilon_{ji}) = \theta_j$, $j = 1, 2, 3$. The measurement errors ϵ_{ji} are assumed to be normally distributed. Note that the intercept and factor loading for Y_3 are fixed to 0 and 1, respectively, for identifiability of the mean and scale of the latent factor f . The choice of fixing the parameters for Y_3 rather than Y_1 or Y_2 is arbitrary.

The main outcome of interest, nonadherence, which is being predicted by the latent variable can also be considered part of the measurement model because it is directly linked to the latent factor f . Specifically, the dichotomous indicator of nonadherence to treatment, U_i , is assumed to have probability π_i that is related to the illness strains latent factor, f_i , and observed covariates gender (X_{1i}) and age (X_{2i}) through a log odds or "logit" relationship, that is,

$$U_i \sim \text{Bernoulli}(\pi_i) \quad (2)$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \tau_0 + \gamma_u f_i \quad (3)$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \tau_0 + \gamma_u f_i + \beta_1 X_{1i} + \beta_2 X_{2i}, \quad (4)$$

where Equation (2) is combined with either Equation (3) or Equation (4) depending on whether covariates are included (Figure 2). The Bernoulli-logit measurement relationship between the dichotomous observed variable and the latent factor specified earlier follows the same form as the commonly used two-parameter logistic item response theory model (Embretson & Reise, 2000). A directly related formulation for dichotomous outcomes (Kamata & Bauer, 2008; Takane & de Leeuw, 1987) can be considered using the probit link and assuming an intermediate underlying continuous variable U^* exists that relates U to f (B. Muthén, 1984; B. Muthén, du Toit, & Spisic, 1997) and would be expected to yield similar results. We note that there is not a separate measurement error included for the dichotomous outcome U_i (i.e., following graphical notation similar to that used by Mplus for dichotomous outcomes, in Figure 2 there is no error term pointing to U). The measurement error in U_i is not modeled separately from the mean structure due to the fact that the variance of U_i (i.e., $\pi_i * (1 - \pi_i)$) is a direct function of its mean π_i . Finally, similar to the linear factor model in Equation (1), the model parameters, $\tau_0, \gamma_u, \beta_1, \beta_2$ for the relationships with dichotomous U are held invariant across the multiple mixture components that are specified for the latent factor f_i in the next section.

Normal Mixtures for the Latent Factor

Now we describe the specification of the latent factor distribution f . Figures 2A and 2B are drawn to represent traditional factor models in which the distribution of f_i or the conditional distribution of f_i given covariates is assumed to be normal, that is,

$$f_i \sim N(\alpha_0, \psi) \quad (\text{Figure 2A})$$

$$f_i = \alpha_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \delta_i \quad \text{and} \quad \delta_i \sim N(0, \psi) \quad (\text{Figure 2B})$$

$$\Rightarrow f_i | X_{1i}, X_{2i} \sim N(\alpha_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}, \psi).$$

Next, the distribution of f is made more flexible by assuming it follows a mixture of K normal distributions. The introduction of mixture components is illustrated in Figures 2C and 2D by the latent categorical variable c , which is pointing to the latent factor f . This latent class variable can take on any one of K different values for each individual i with $P(c_i = k) = \eta_k$, such that $\sum_{k=1}^K \eta_k = 1$. In other words, the distribution of f is taken to be a mixture of K components (recall Figure 1) where η_k is the marginal probability of each component (i.e., probability of each latent class). In the case with covariates (Figure 2D), besides their effect on f_i , we also allow the possibility that the covariates influence the probability of being in any one of the latent categories,

such that $P(c_i = k | X_{1i}, X_{2i}) = \eta_{ik}$. The mixture model is then formed by assuming for each category of c_i that the latent factor follows a normal distribution with different mean (and possibly different variance), that is,

$$f_i | c_i = k \sim N(\alpha_{0k}, \psi_k), \quad k = 1 \dots K \tag{Figure 2C}$$

$$\Rightarrow f_i \sim \sum_{k=1}^K \eta_k N(\alpha_{0k}, \psi_k)$$

$$f_i | c_i = k, X_{1i}, X_{2i} \sim N(\alpha_{0k} + \gamma_1 X_{1i} + \gamma_2 X_{2i}, \psi_k), \quad k = 1 \dots K \tag{Figure 2D}$$

$$\Rightarrow f_i \sim \sum_{k=1}^K \eta_{ik} N(\alpha_{0k} + \gamma_1 X_{1i} + \gamma_2 X_{2i}, \psi_k)$$

where $\log \left(\frac{\eta_{ik}}{1 - \eta_{ik}} \right) = \alpha_{ck} + \gamma_{c1} X_{1i} + \gamma_{c2} X_{2i}$.

Previously, we have written the mixture model distributions for f_i in conditional form, that is, $f_i | c_i$ conditioned on the component c_i , and in marginal form, that is, summing or aggregating across all categories of c_i . This distinction is useful as it points to the difference in focus between using the mixture model for purpose of identifying underlying clusters (conditional model) versus using it for the purpose of simply providing a more flexible distribution for f_i (marginal model). Consistent with our goal of providing a flexible distribution for f_i without focusing on the components per se, we have made the assumption that X_{1i} and X_{2i} relate to the latent factor f_i in the same way regardless of the latent component c_i , that is, γ_1 and γ_2 do not have k subscripts. Furthermore, although the model allows for the covariates to predict the probability of latent class membership, those paths from the covariates to the mixture component (i.e., γ_{c1} and γ_{c2}) will not be of primary interest because in this article we are not interested in any one component but instead the aggregation across the components.

Obtaining the Marginal Mean and Variance of a Normal Mixture

Consider again Figure 1. The skewed distribution shown there with the mixture of normals models overlayed represents an example of the true and fitted models for a latent factor f . The true population distribution for the latent factor in this illustration is a standardized chi-square distribution with 1 degree of freedom. This distribution has skew = 2.8 and kurtosis = 12 (Johnson & Kotz, 1982). Because it has been standardized (by subtracting 1 and dividing by sqrt(2)—recall the mean and variance of a chi-square distribution equal the df and $2*df$,

respectively), it has mean = 0 and standard deviation = 1. For demonstration $n = 50,000$ random samples from the standardized chi-square distribution were generated. In the first case (upper left plot in Figure 1) where a normal model is assumed for f , the estimated mean and variance of the best fitting normal distribution are both very close to 0 and 1 but we see that the symmetric normal distribution is poorly capturing the skewed nature of the chi-square. In the case where a 2-component normal mixture is assumed, that is, $f_i \sim \eta_1 N(\alpha_{01}, \psi_1) + \eta_2 N(\alpha_{02}, \psi_2)$, the best fit for the 2-component mixture finds the first normal component to have mean $\alpha_{01} = -.21$ with probability $\eta_1 = 0.93$ and the second component to have mean $\alpha_{02} = 2.80$ with probability $\eta_2 = .07$. A simplifying assumption is made when fitting the mixture models that the variances within each component are equal, that is, $\psi_1 = \psi_2 = \psi$. Both normal components are estimated to have variance $\psi = 0.41$. The resulting parameter estimates, that is, mixing proportions $\eta = (\eta_1, \dots, \eta_K)$, component means $\alpha_0 = (\alpha_{01}, \dots, \alpha_{0K})$, and component variance ψ , for the 3- and 4-component mixture models are as follows:

3 components $\eta = (.86, .12, .02)$, $\alpha_0 = (-.32, 1.58, 4.34)$, $\psi = .23$;

4 components $\eta = (.81, .14, .04, .01)$, $\alpha_0 = (-.38, 1.07, 2.83, 5.55)$, $\psi = .15$.

So we see that the right skewed nature of the true f is captured using the mixture models by a dominant (large probability) normal component centered near the mode and other smaller components with larger means to capture the right skew.

Furthermore we can obtain the estimated mean and variance for the marginal (aggregated) distribution for f from the mixture model component estimates. For example, for the 2-component model, $E(f) = \eta_1 \alpha_{01} + \eta_2 \alpha_{02}$ and $Var(f) = \eta_1(\psi_1 + \alpha_{01}^2) + \eta_2(\psi_2 + \alpha_{02}^2) - E(f)^2$, so from the fitted model we have an estimated mean of $0.93 * (-.21) + 0.07 * (2.80) \approx 0$ and estimated variance of $0.93 * (.41 + (-.21)^2) + .07 * (.41 + 2.80^2) \approx 1$. Hence, the estimated mean and variance of the aggregated 2-component mixture distribution are close to 0 and 1, which appropriately matches the mean and variance of true standardized chi-square population distribution for f . Similar calculations can be done for the 3- and 4-component mixture models and find their means and variance close to 0 and 1 as well.

ESTIMATION FOR MIXTURE FACTOR ANALYSIS

Given a parametric distribution for the latent factor f_i whether it be normal or a mixture of normals, estimation for all parameters in the models in Figure 2 can be obtained via maximum likelihood. Because of its usefulness as a method for

obtaining maximum likelihood estimates in the presence of missing data, and because the mixture factor analysis model can be operationalized straightforwardly as a missing data problem where latent class membership, c_i , is the “missing data,” the expectation maximization (EM) algorithm is commonly used for estimation in mixture models. The EM algorithm is an iterative method such that in the E-step of the t -th iteration, the conditional expectation of the complete data log-likelihood is evaluated as an integral over latent f_i and c_i , and in the M-step this expectation is maximized to obtain the $t + 1^{st}$ parameter estimates. The steps are iterated until there is negligible change in the t^{th} compared with $t + 1^{st}$ parameter estimates.

A substantially difficult computational problem is introduced into the EM algorithm when the form of the model is such that there is not a closed form solution to the integral in the E-step (Pinhero & Bates, 1995). Indeed, because of the nonlinear (logistic) relationship between the dichotomous observed variable and the latent factor f , a numerical integration method must be used to approximate the integral because there is not a closed form. The basic idea of numerical integration of a function is to use a weighted sum of the function evaluated over a set of integration points representing values of the latent factor. Gauss-Hermite quadrature approximation, which uses the normal distribution to calculate the weights, can be used for the integration and is a method implemented by default in Mplus 6.11 (L. K. Muthén & Muthén, 1998–2010) for the E-step. The weighted sum from the E-step is then maximized in the M-step and in Mplus this maximization is done using a combination of Quasi-Newton and Fisher scoring methods. A comprehensive exposition of details for computational methods used in maximum likelihood estimation is found in Skrondal and Rabe-Hesketh (2004, Chapter 6).

Robust standard errors can be calculated based on the “sandwich formula” (Freedman, 2006; Huber, 1967), which is a general method for estimating the covariance matrix of parameter estimates that does not rely on distributional assumptions or independence of the observations. The standard errors are called robust because they are consistent even when distributional assumptions underlying the original parameter estimates are incorrect. These standard errors can be directly obtained from Mplus.

The Mplus programs for fitting a mixture factor analysis with four latent class components with and without covariates are shown in the authors’ personal online appendix available at www.columbia.edu/~mmw2177.

Computational Issues

Implementation of the EM algorithm described earlier that incorporates numerical integration to handle the nonlinear link due to the dichotomous outcome can lead to several types of computational difficulties. First, Gauss-Hermite

quadrature is limiting in terms of the number of dimensions it can practically and accurately integrate. In our working example throughout this article, there is only one underlying latent factor (one dimension) and by default Mplus takes 15 integration points to approximate the integral. Increasing to two latent factors means that $15 \times 15 = 225$ integration points are taken and thus computation time increases exponentially as the number of factors increases. The Mplus 6 manual (L. K. Muthén & Muthén) notes repeatedly that “numerical integration becomes increasingly more computationally demanding as the number of factors and the sample size increase” (e.g., Ex 5.3, p. 59).

Second, mixture models are notorious for exhibiting multimodal and relatively flat likelihoods (McLachan & Peel, 2000). As such, it is possible for the EM algorithm to converge to a local rather than global maximum. Hence it is commonly recommended (McLachan & Peel, 2000) and, in fact, is the default in most softwares that perform mixture models (including Mplus) to use multiple starting values for the EM algorithm and then choose the estimates that lead to the largest likelihood. A summary of the number of times out of all the different starting values that the same maximum likelihood solution was reached can be taken as an indication of the complexity of the likelihood surface. Indeed if the same solution is obtained no matter what the starting values were then the likelihood was not difficult to maximize and vice versa. As detailed earlier, Mplus provides options for increasing the number of different random starting values and it refits the model for each set. It also identifies the maximum likelihood value associated with those fits and indicates if the best log-likelihood value was or was not replicated.

Finally, degenerate cases with zero probability estimated for one of the components or cases where 2 components have the same mean can occur. This boundary value phenomenon is not necessarily a problem because it can simply be an indication that too many mixture components are being fit to the data.

Increasing the Number of Components

As was seen in Figure 1, in general by adding more components to the mixture model, it is possible to better approximate the true distribution. But there are at least three problems with thinking “more is better” when it comes to data analysis with mixture models. The first is to remember that we do not observe the latent factor distribution. Indeed Figure 1 is a bit misleading because it may give the impression that we are fitting a mixture model to a distribution we can observe, but in real applications we do not observe the latent factor f , so it is not possible to directly compare the true f with its fitted distribution to verify if it “fits” well. The second reason that more components may not be better is lack of information in the finite sample. As was mentioned earlier, the likelihood surface for mixture models is commonly multimodal implying several solutions

(i.e., sets of parameter estimates) may fit the data equally well or at least very similarly. As the number of components increases, the multimodality of the likelihood increases making the estimation that much more difficult. The third reason is it may simply not be necessary to approximate the distribution any better if the focus of the research question is on other model parameters. This is the most practical of all the reasons and gets to the heart of the robustness issue and the motivation for our simulation study. Considering the distribution of the latent factor f a nuisance, the target parameters of interest in the model (e.g., γ_u) may be well estimated as long as the mixture model approximates f “close enough.” This idea of “close enough” is not made exact but is investigated via the simulation study.

A number of statistical criteria have been proposed to facilitate selecting the number of components in a mixture model. Information criteria are commonly used: Akaike Information Criterion (AIC; $= -2\log L + 2p$), Bayesian Information Criterion (BIC; $= -2\log L + p\log N$), and the sample-size adjusted BIC (ABIC) where the sample size N in the BIC is replaced by $n^* = (N + 2)/24$ (Sclove, 1987). These indices are based on the value of likelihood function (L), so they reward models that more accurately reproduce the observed data and extract a penalty for the number of parameters (p) in the model, favoring models with fewer mixture components. They are commonly used in choosing the number of components in mixture models, and simulation studies focused on latent class and growth mixture models have found both BIC and ABIC outperformed other tests for choosing the number of components (B. Muthén, 2007; Nylund, Asparouhov, & Muthén, 2007). In the following simulation study, in addition to comparing across different fixed number of components, we compare with a model where K is chosen using BIC.

SIMULATION STUDY

We set up a simulation study to investigate the performance of mixture factor analysis as a means of estimating target parameters in the presence of non-normality of the latent factor. An important related issue is the robustness (or nonrobustness) to nonnormality exhibited by the traditional maximum likelihood method, which assumes normality for the latent factors. Thus, we compare the performance of the mixture factor analysis with 2-, 3-, or 4-component mixtures of normals with the traditional maximum likelihood assuming the underlying factor is normally distributed. The model that assumes the underlying factor is normally distributed can be considered a degenerate mixture, that is, a mixture factor analysis with only 1 component.

The simulation study focuses on the basic 1-factor model with 3 continuous indicator variables and 1 dichotomous outcome variable as in Figure 2A albeit

with varying true distributions for f . It is of interest to see how the performance of the mixture factor analysis varies with respect to the degree of nonnormality of the latent factor (and consequently the observed variables) and the sample size. Varying measurement errors relating the continuous indicators to the underlying factor were also examined as a simulation factor. Patterns of results across varying levels of measurement error were similar to those presented later and are not included for simplicity of presentation (additional results available from the authors).

Three latent factor distributions were considered: High skew/kurtosis (standardized Beta (.1,30)), Mild skew/kurtosis (standardized chi-square(1 df)), and No skew/kurtosis (standardized normal). For each latent factor distribution we considered four different sample sizes: 100, 200, 500, and 1,000. Under each of these 12 (3 latent distributions * 4 sample sizes) scenarios, we generated 200 data sets from models (1), (2), and (3). The specific model parameters were fixed as $\tau_0 = -0.5$ and $\gamma_u = 1$ for the logistic relationship between the dichotomous outcome and the latent factor and $\nu_{01} = \nu_{02} = 0$ and $\lambda_1 = .7$ and $\lambda_2 = .8$ for the linear factor model (1). The latent factor under every scenario was generated to have mean equal to zero and variance equal to 1 albeit with different distributions (Beta, chi-square, or normal). Finally, the three measurement errors in model (1) were generated from independent normal distributions with mean equal to 0 and equal variances $\theta = 0.5$ corresponding to a reliability for the sum of the continuous observed variables of approximately 0.80 (i.e., $(0.7 + 0.8 + 1)^2 / [(0.7 + 0.8 + 1)^2 + 3 * 0.5] \approx 0.8$).

Generating Nonnormal Data

The choice of Beta (.1,30) and chi-square with 1 df distributions for the true latent factor distributions represent two different levels of right skew with positive kurtosis (high peakedness). Mattson (1997) summarized two methods for generating nonnormal data: Method A where multivariate nonnormal data are generated directly to have the desired level of skew/kurtosis and given covariance matrix and Method B where the multivariate nonnormal data are generated from a latent variable model where the components of the model (i.e., the latent factors and errors) come from specific distributions. It was argued (and developed) by Mattson that Method B is preferred for simulations because it provides a “truth” that can be compared with in terms of modeling. Indeed other authors (Reinartz, Echambadi, & Chin 2002) refer to Mattson’s Method B simply as “Mattson’s method” and advocate for it.

In this study we use Mattson’s method (Method B). But as Mattson (1997) pointed out, “an apparent limitation . . . is that univariate distributions with more extreme values of skewness and kurtosis must be used to match certain values of skewness and kurtosis of the observed variables” (p. 371). In this

study, this is exactly what leads us to use the highly skewed and kurtotic Beta (.1,30) and chi-square (1 *df*) distribution for the underlying latent factor. Indeed the population skew/kurtosis for a Beta (.1,30) is 6/52 and for a chi-square with 1 *df* is 2.8/12 (Johnson & Kotz, 1982). Nevertheless, the finite sample generated latent factor and the associated observed Y variables are not nearly as skewed/kurtotic. Table 1 shows the actual sampling distribution of the estimated skew and kurtosis across the 200 simulated data sets in each of the 12 scenarios considered. We observe that as the sample size decreases, the observed skew and kurtosis for the generated latent factors are much smaller than the population values (Reinartz et al., 2002). Moreover, because random normal measurement error is added to the latent factors to obtain the Y_j , the skew/kurtosis for the Y_j is even less. For example at a sample of size 100, the observed Y_3 variable under Beta (.1,30) has a median skew of 2.0 with a 5th to 95th percentile ranging from 0.5–4.4, and Y_3 using the chi-square (1 *df*) distribution for the latent factor has a median skew of 1.2 ranging from 0.5–2.4 across the 200 simulated data sets. Similarly the median kurtosis for Y_3 generated from the model with latent factor distribution Beta (.1,30) and $n = 100$ is 7.1 ranging from 0.9–28.1 and with chi-square (1 *df*) latent factor leads to a median kurtosis for Y_3 of 2.2 with 5th to 95th percentile 0.3–10.8. Thus, although the population distribution for the latent factor is the same across the different sample sizes, the skew and kurtosis of the sample latent factor and observed variables are smaller and are quite variable.

Estimated Parameters and Inference

In the illness strains example, the γ_u is the main target parameter of interest to the researchers, that is, it represents how the latent illness strains relate to nonadherence to the treatment. Furthermore, it is a parameter that is part of a nonlinear relationship (the logit link); hence it may be affected by nonnormality differently than the other linear factor loadings, for example, λ_1 , in model (1). Thus our presentation of results compares and contrasts the parameter estimation for γ_u and also one of the linear factor loadings λ_1 .

Four models—3 mixture factor analysis models with 4, 3, or 2 latent classes (labeled 4, 3, and 2) and the normal factor model (labeled 1)—were applied for parameter estimation and inference to each of the 200 simulated data sets under each of the 12 scenarios. One hundred different random starting values were used and the model with the largest likelihood value was chosen. Robust standard errors were used and 95% Wald-type confidence intervals were obtained using estimate ± 1.96 *(standard error). It is of interest to see how much improvement or deterioration in estimation comes from adding more components in the mixture factor analysis approach. A maximum of 4 components was taken due to a combination of computational difficulties and diminishing returns seen for adding more.

TABLE 1
 Sample estimates of skew and kurtosis for the latent factor f and one of the observed variables Y_3 for all 12 scenarios used in the simulation study

Latent Factor Distribution	Generated Variables	$n = 100$			$n = 200$			$n = 500$			$n = 1,000$		
		Skew	Kurtosis	Skew	Kurtosis	Skew	Kurtosis	Skew	Kurtosis	Skew	Kurtosis	Skew	Kurtosis
High skew/kurtosis Beta (1,1,30)	f	4.3 (3.0-7.0)	21.1 (8.9-55.9)	4.7 (3.4-7.2)	26.5 (12.3-64.1)	5.1 (3.9-7.4)	32.7 (17.9-76.9)	5.4 (4.3-7.8)	37.8 (22.2-90.6)	5.4 (4.3-7.8)	37.8 (22.2-90.6)	5.4 (4.3-7.8)	37.8 (22.2-90.6)
	Y_3	2.0 (0.5-4.4)	7.1 (0.9-28.1)	2.3 (1.1-4.6)	10.4 (2.9-35.1)	2.7 (1.6-4.5)	13.5 (5.4-39.0)	2.9 (2.0-4.6)	16.2 (7.9-43.9)	2.9 (2.0-4.6)	16.2 (7.9-43.9)	2.9 (2.0-4.6)	16.2 (7.9-43.9)
Mild skew/kurtosis χ^2_1	f	2.3 (1.5-3.7)	5.9 (1.7-19.1)	2.5 (1.8-3.6)	7.4 (3.0-18.8)	2.6 (2.1-3.6)	9.2 (5.2-20.3)	2.7 (2.2-3.5)	9.9 (6.2-19.2)	2.7 (2.2-3.5)	9.9 (6.2-19.2)	2.7 (2.2-3.5)	9.9 (6.2-19.2)
	Y_3	1.2 (0.5-2.4)	2.2 (0.3-10.8)	1.3 (0.7-2.3)	3.0 (0.7-10.4)	1.4 (1.0-2.1)	4.0 (1.8-10.4)	1.4 (1.1-2.0)	4.3 (2.3-9.3)	1.4 (1.1-2.0)	4.3 (2.3-9.3)	1.4 (1.1-2.0)	4.3 (2.3-9.3)
No skew/kurtosis Normal	f	0.0 (-0.4-0.4)	-0.1 (-0.6-0.8)	0.0 (-.3-0.3)	-0.1 (-0.5-0.6)	0.0 (-.2-0.2)	0.0 (-0.3-0.4)	0.0 (-0.1-0.1)	0.0 (-0.3-0.3)	0.0 (-.2-0.2)	0.0 (-0.3-0.4)	0.0 (-0.1-0.1)	0.0 (-0.3-0.3)
	Y_3	0.0 (-0.4-0.4)	-0.1 (-0.6-0.9)	0.0 (-.3-0.3)	0.0 (-0.5-0.6)	0.0 (-.2-0.2)	0.0 (-0.3-0.4)	0.0 (-0.1-0.1)	0.0 (-0.3-0.3)	0.0 (-.2-0.2)	0.0 (-0.3-0.4)	0.0 (-0.1-0.1)	0.0 (-0.3-0.3)

Note. Values shown are the median (5th-95th percentile) of sample skew and kurtosis across 200 simulated data sets.

Bias of the estimators. Summaries of the estimated path, $\hat{\gamma}_u$ for the dichotomous outcome, and loading $\hat{\lambda}_1$ for a continuous indicator under the 12 scenarios across the 200 simulated data sets are shown in Table 2. Focusing first on the results for $\hat{\gamma}_u$ representing the nonlinear relationship between the dichotomous outcome and the latent factor, we compare the side-by-side box-plots of the different estimators in Figure 4. We find that in both cases of high and mild skew/kurtosis, the normal factor model exhibits downward bias (i.e., mean estimated values less than the true value of 1) that persists even as the sample size increases. Because this bias does not decrease as the sample size increases, this indicates statistical inconsistency for the normal factor analysis estimator in the presence of nonnormality. On the other hand, the mixture

TABLE 2
Simulation results. Empirical bias, standard deviation (SD), root mean square error (RMSE), and 95% coverage probability for estimators of γ_u (top) and λ_1 (bottom)

	K	High Skew/Kurtosis Beta (.1, .30)				Mild Skew/Kurtosis χ^2_1				No Skew/Kurtosis Normal			
		n = 100	200	500	1,000	n = 100	200	500	1,000	n = 100	200	500	1,000
		True $\gamma_u = 1$											
Bias	1	-0.23	-0.23	-0.22	-0.24	0.00	-0.06	-0.07	-0.07	0.11	0.02	0.02	0.01
	2	0.04	-0.01	-0.05	-0.09	0.14	0.02	0.00	-0.02	0.10	0.02	0.02	0.01
	3	0.31	0.13	0.04	-0.02	0.17	0.04	0.01	-0.01	0.12	0.02	0.02	0.01
	4	0.39	0.20	0.09	0.02	0.17	0.04	0.00	0.00	0.11	0.02	0.02	0.01
SD	1	0.42	0.26	0.16	0.12	0.39	0.24	0.14	0.09	0.44	0.24	0.15	0.10
	2	0.64	0.37	0.25	0.18	0.53	0.32	0.17	0.11	0.44	0.25	0.15	0.10
	3	0.96	0.51	0.30	0.19	0.56	0.34	0.17	0.11	0.46	0.25	0.15	0.11
	4	1.12	0.70	0.33	0.20	0.55	0.33	0.16	0.11	0.48	0.25	0.16	0.11
RMSE	1	0.48	0.34	0.27	0.27	0.39	0.25	0.16	0.12	0.45	0.24	0.15	0.10
	2	0.64	0.37	0.25	0.21	0.55	0.32	0.17	0.11	0.45	0.25	0.15	0.10
	3	1.01	0.52	0.30	0.19	0.59	0.34	0.17	0.11	0.47	0.25	0.15	0.11
	4	1.18	0.72	0.34	0.20	0.58	0.33	0.16	0.11	0.49	0.25	0.16	0.11
Coverage Prob	1	0.79	0.77	0.64	0.39	0.96	0.91	0.93	0.92	0.93	0.96	0.97	0.96
	2	0.90	0.91	0.88	0.81	0.97	0.94	0.96	0.96	0.93	0.96	0.95	0.95
	3	0.89	0.94	0.95	0.90	0.96	0.95	0.97	0.97	0.95	0.96	0.95	0.95
	4	0.88	0.95	0.96	0.94	0.96	0.95	0.97	0.97	0.91	0.96	0.97	0.96
True $\lambda_1 = 0.7$													
Bias	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
SD	1	0.17	0.10	0.06	0.04	0.12	0.08	0.05	0.04	0.11	0.08	0.05	0.04
	2	0.13	0.08	0.05	0.03	0.11	0.08	0.05	0.03	0.11	0.08	0.05	0.04
	3	0.14	0.08	0.05	0.03	0.11	0.08	0.04	0.03	0.12	0.08	0.05	0.04
	4	0.14	0.08	0.05	0.03	0.11	0.07	0.04	0.03	0.13	0.08	0.05	0.04
RMSE	1	0.17	0.10	0.06	0.04	0.12	0.08	0.05	0.04	0.11	0.08	0.05	0.04
	2	0.13	0.08	0.05	0.03	0.11	0.08	0.05	0.03	0.11	0.08	0.05	0.04
	3	0.14	0.08	0.05	0.03	0.11	0.08	0.04	0.03	0.12	0.08	0.05	0.04
	4	0.14	0.08	0.05	0.03	0.11	0.07	0.04	0.03	0.13	0.08	0.05	0.04
Coverage Prob	1	0.94	0.94	0.95	0.94	0.93	0.94	0.97	0.96	0.97	0.96	0.95	0.98
	2	0.83	0.90	0.92	0.94	0.92	0.91	0.96	0.96	0.97	0.96	0.95	0.98
	3	0.83	0.90	0.92	0.92	0.92	0.91	0.96	0.94	0.97	0.96	0.97	0.98
	4	0.82	0.89	0.91	0.93	0.92	0.92	0.94	0.96	0.94	0.97	0.97	0.97

Note. Estimates are summarized across 200 simulated datasets under 12 scenarios (4 sample sizes times 3 latent distributions) based on mixture factor analysis with K = 1, 2, 3, or 4 components where K = 1 refers to the normal distributed latent factor model.

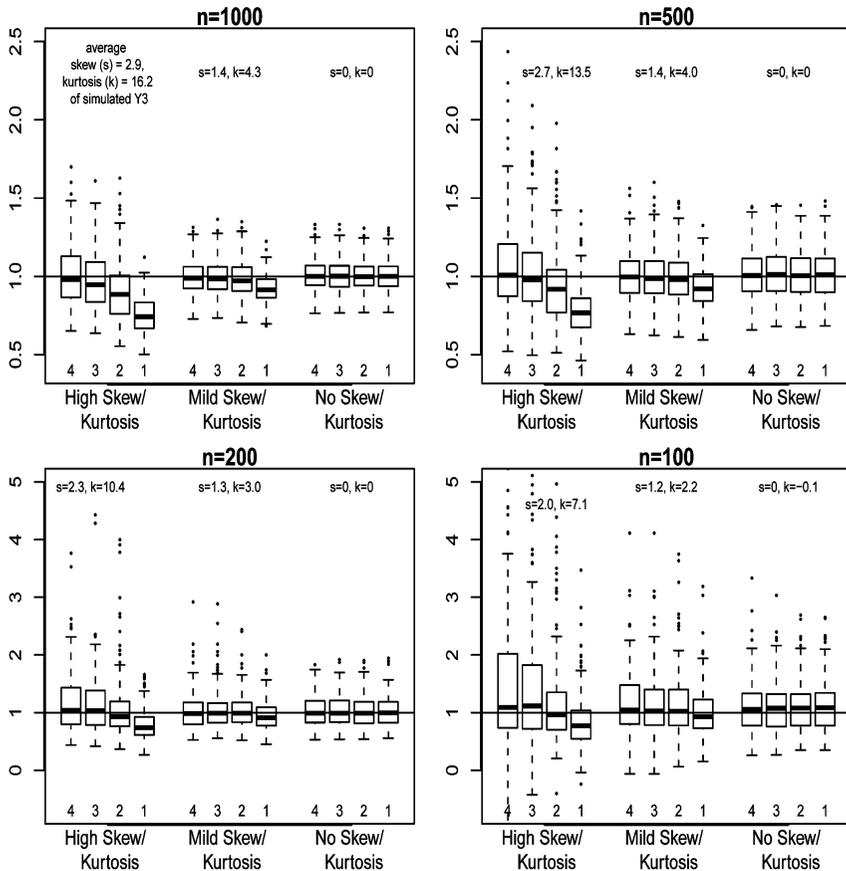


FIGURE 4 Boxplots of estimated γ_U (log odds ratio relating U to f) across 200 simulated data sets within each of 12 scenarios (4 sample sizes times 3 latent distributions) using mixture factor analysis with 4, 3, 2, or 1 components (i.e., 1 = the normal factor model). True $\gamma_U = 1$. Horizontal bar within each boxplot represents the median and the box itself denotes the 25th and 75th percentiles. Estimates deviating above or below the 25th and 75th percentiles by more than 1.5 times the interquartile range are shown with points. Note y-axis scale change in top graphs compared with bottom graphs. Average (i.e., median) skew (s) and kurtosis (k) values (taken from Table 1) associated with the simulated Y_3 variable are included to help facilitate interpretation due to varying values across different scenarios.

factor analysis models are correcting the bias to varying degrees depending on the sample size, degree of skewness, and number of components. For example, with high skew/kurtosis and a sample of size 1,000, the normal factor model has substantial empirical bias (i.e., $\sum(\hat{\gamma}_U - 1)/200$) equal to -0.24 ,

whereas the mixture factor analysis with 4 latent classes has empirical bias equal to 0.02, much closer to zero. In the case of $n = 100$ and 200 in the presence of high skew/kurtosis, as more components are added in the mixture factor analysis we find that median bias improves (horizontal lines shown within boxplots in Figure 4 are closer to 1 with more components), but mean bias (i.e., the usual definition of bias) increases. This is a direct result of the right skew found in the mixture factor analysis estimator that is accentuated at smaller sample sizes. In the case where the normal latent factor model is correct, it is important to observe that the mixture factor analysis models produce little bias for $\hat{\gamma}_u$ similar to that of the normal factor model despite being more complicated than necessary in this case where the normal model is correct.

In contrast to the bias found using the normal factor model to estimate γ_u in the presence of a latent factor with skew and kurtosis, there was no substantial bias found for the linear relationship parameter λ_1 using the normal factor model, Table 2, even in the case of high skew/kurtosis for the underlying latent factor. This result is not surprising given the theoretical work (Anderson & Amemiya, 1988; Browne & Shapiro, 1988) showing that the estimators for the linear factor analysis models using the maximum likelihood method assuming normality for the underlying latent variables are asymptotically consistent even if the normality assumption for the latent factors is violated. The mixture factor analysis estimators, regardless of the number of components, perform equally well in terms of bias for λ_1 .

Variability and root mean square error of the estimators. Inspection of the boxplots in Figure 4 finds that there is substantially higher variability (i.e., lower efficiency) in the estimator from the mixture factor models, particularly with 3 and 4 components and sample sizes 100 and 200. For example, in the high skew/kurtosis factor case when the sample size is 100, the standard deviation of the empirical sampling distribution of $\hat{\gamma}_u$ using the normal factor model is 0.42 whereas for the mixture factor analysis model with 4 latent classes the standard deviation is more than double that at 1.12. In the same scenario when the sample size increases to 1,000, the variability of $\hat{\gamma}_u$ obtained from the 4 component mixture factor analysis decreases to 0.20 and is closer to that of the normal factor model estimator, which equals 0.12.

In contrast to the larger variability found for $\hat{\gamma}_u$ using the mixture factor model compared with the normal factor model, the linear factor loading estimator $\hat{\lambda}_1$ based on the mixture factor model shows better efficiency than the normal factor model estimator when there is skew/kurtosis in the underlying factor, Table 2. For example, in the high skew/kurtosis factor scenario when the sample size is 100, the standard deviation of the empirical sampling distribution of $\hat{\lambda}_1$ using the normal factor model is larger, 0.17, compared with the mixture factor analysis

model with 4 latent classes, which is 0.14. This efficiency gain for $\hat{\lambda}_1$ from the mixture factor model compared with the normal factor model remains as sample size increases, although the gain decreases in size.

In the scenarios where no skew/kurtosis exists in the underlying factor, we find the mixture factor analysis models performing equally well in terms of variability of both the estimators $\hat{\gamma}_u$ and $\hat{\lambda}_1$ to the normal factor model (i.e., the correct model in these scenarios). This is important because it might have been expected that because the mixture model is more complex than need be in this case, it would lose efficiency, but this does not appear to happen in the different cases considered.

Given the trade-off between bias and variance for $\hat{\gamma}_u$ (i.e., estimation using the normal model is biased but less variable, whereas estimation using the mixture models is less biased but more variable), we also examined the empirical root mean square error (RMSE), which takes both into account. The empirical RMSE is calculated as the square root of the average of $(\hat{\gamma}_u - 1)^2$ or $(\hat{\lambda}_1 - 0.7)^2$ over 200 simulated data sets where 1 and 0.7 correspond to the true values for γ_u and λ_1 , respectively. Figure 5 and Table 2 display the empirical RMSE for all the estimators, $\hat{\gamma}_u$ and $\hat{\lambda}_1$, across all the scenarios and smaller values indicate better performance. In the high skew/kurtosis case, when we consider $\hat{\gamma}_u$, we find that for small sample sizes ($n = 100$ and 200), as the number of components increases, the performance decreases with the normal model being the best and the 4-component model being the worst. As the sample size increases to $n = 1,000$, the performance of the mixture factor analysis models for $\hat{\gamma}_u$ has improved and each has lower RMSE than the normal model. A similar phenomenon occurs for mild skew/kurtosis case where the normal model performs better at small sample sizes and by $n = 500$ and above the mixture factor model estimators, $\hat{\gamma}_u$, perform similarly. We find in contrast to $\hat{\gamma}_u$ that the linear loading $\hat{\lambda}_1$ using a mixture factor model performs better at all sample sizes when there is mild or high skew/kurtosis in the data. Finally, comparing across levels of skew/kurtosis in Figure 5, we find the performance of $\hat{\gamma}_u$ is overall worse regardless of which estimator is used when there is more skew/kurtosis in the latent factor (i.e., average RMSE in high skew/kurtosis case is higher than mild or no skew/kurtosis), yet the performance for $\hat{\lambda}_1$ is not nearly as affected overall by skew/kurtosis.

Coverage probability. Having summarized previously the performance of the parameter estimators in terms of bias, variability, and RMSE, we now turn to the examining their respective confidence intervals. The empirical 95% coverage probabilities for $\hat{\gamma}_u$ and $\hat{\lambda}_1$ using the different methods are shown in Table 2. Note, with 200 simulated data sets, the simulation error for estimating coverage probability is approximately $\pm 3\%$. The large downward bias found for $\hat{\gamma}_u$ when

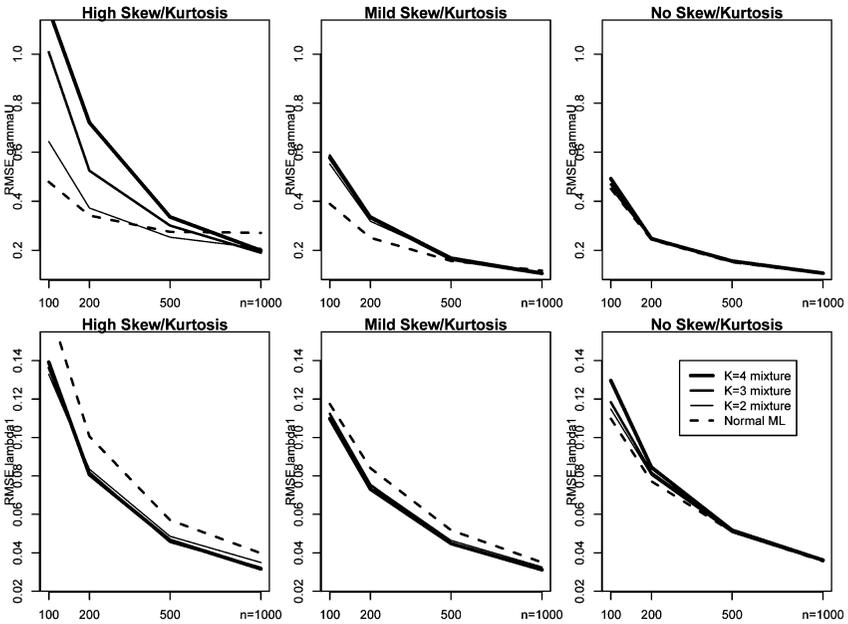


FIGURE 5 Empirical root mean square error (RMSE) for the different estimators, $\hat{\gamma}_u$ (top three plots), and $\hat{\lambda}_1$ (bottom three plots). RMSE is calculated across 200 simulated data sets within each of 12 scenarios. *Note.* ML = maximum likelihood; y-axis scale change in top graphs compared with bottom graphs.

the latent factor was assumed to be normal in the case of high skew/kurtosis leads to coverage probabilities far from nominal 95% that only get worse as sample size increased (39% coverage when $n = 1,000$ for $\hat{\gamma}_u$ using the normal factor model). Indeed, in the high skew/kurtosis scenario, the normal ML confidence interval underestimated the true value for $\hat{\gamma}_u$ in all cases where it did not cover the true value. On the other hand, the coverage probability for the linear loading $\hat{\lambda}_1$ using the normal ML confidence interval was good at all sample sizes and skew/kurtosis levels. The mixture factor analysis model in the high skew/kurtosis case had improving coverage for $\hat{\gamma}_u$ as the number of components increased and sample size was 200 and greater. Coverage probability for the mixture factor model for $\hat{\lambda}_1$ was near nominal levels at sample size 500 and greater. In the mild skew/kurtosis case, the coverage using the mixture factor models performed similarly across the number of components with good coverage at all sample sizes for $\hat{\gamma}_u$ and nominal coverage for $n = 500$ and larger sample sizes for $\hat{\lambda}_1$. In the no skew/kurtosis scenario for the latent factor, all methods showed good coverage probability.

Model Comparison and Empirical Choice of K

Which model does the data choose as best fitting? The traditional normal ML and mixture factor analysis approaches are likelihood based, hence an information criterion can be used to compare model fit. Here we present comparisons using BIC but note the general trend of the results is very similar whether AIC or ABIC were used. Table 3 presents the percentage of times out of the 200 simulated data sets that the particular model was chosen based on the smallest BIC value. Note that $K = 1$ corresponds to the normal ML case. Recall that in the case of high and mild skew/kurtosis with Beta and chi-square distributions for the latent factors, none of the models is exactly correct, whereas in the case of No skew/kurtosis (i.e., normally distributed factor), the $K = 1$ model is the correct model. First examining the high skew/kurtosis case we see that the normal $K = 1$ model is practically never chosen. Furthermore, as the sample size increases, the chosen model has increasing K indicating the data can support a more complex model. Similar results are found for the mild skew/kurtosis case. On the other hand, in the case where the true model is $K = 1$ we find that the BIC always chooses the correct model for sample size 200 or greater and nearly so in the smallest $n = 100$ case. Results using ABIC and AIC were similar to Table 3 but with a tendency to choose larger K across all scenarios.

Does the best fitting model according to BIC result in “best performance” in terms of estimators for γ_u and λ_1 ? The short answer is “yes” for λ_1 and “it depends on the sample size” for γ_u . The last four rows of Table 3 show the empirical RMSE associated with the estimator for λ_1 and γ_u chosen based on

TABLE 3

Top part of table: Percentage of times (out of 200 simulated data sets) K was chosen^a based on BIC criterion under each scenario. Bottom part of table: Empirical root mean square error (RMSE) for $\hat{\gamma}_u$ and $\hat{\lambda}_1$ based on the “Best”^b estimator in each of the 200 simulated data sets.

K	High Skew/Kurtosis Beta (.1, .30)				Mild Skew/Kurtosis χ^2_1				No Skew/Kurtosis Normal			
	n = 100	200	500	1,000	n = 100	200	500	1,000	n = 100	200	500	1,000
1	1	0	0	0	2	0	0	0	97	100	100	100
2	35	9	0	0	53	22	0	0	2	0	0	0
3	52	53	14	1	39	61	42	8	1	0	0	0
4	11	38	86	99	5	16	58	92	0	0	0	0
RMSE												
$\hat{\gamma}_u$ “Best”	1.07	0.61	0.34	0.20	0.60	0.33	0.17	0.11	0.45	0.24	0.15	0.10
$\hat{\gamma}_u$ Normal	0.48	0.34	0.27	0.27	0.39	0.25	0.16	0.12	0.45	0.24	0.15	0.10
$\hat{\lambda}_1$ “Best”	0.13	0.08	0.05	0.03	0.11	0.07	0.04	0.03	0.11	0.08	0.05	0.04
$\hat{\lambda}_1$ Normal	0.17	0.10	0.06	0.04	0.12	0.08	0.05	0.04	0.11	0.08	0.05	0.04

^aMixture factor models with K = 1 (normal), 2, 3, and 4 components were compared and the model with smallest Bayesian Information Criterion (BIC) was chosen for K.

^bThe “Best” estimator corresponds to the model with the smallest BIC.

whichever of the models with 1, 2, 3, or 4 components had the smallest BIC, called “Best,” compared with the normal factor model ($K = 1$). Note that we allow the “Best” estimator to be the $K = 1$ normal factor model estimator when it has smallest BIC. Similar to what we found before and in Figure 5 when considering the simulation results separately by the number of components, the RMSE for the estimator of λ_1 from the “Best” model is uniformly better than the normal model (at all sample sizes) in the presence of skew and kurtosis. On the other hand, the performance of the estimator for γ_u from the “Best” model is losing out to the simple normal model at sample sizes $n = 100$ and 200 and is similar or better in terms of RMSE than the normal model when $n = 500$ or greater.

Additional Simulation Scenarios That Vary the True Value of γ_u

Substantial bias and poor coverage probability were found when using the normal factor model estimator for the path between the latent factor and the dichotomous outcome, γ_u , in the case of high skew/kurtosis, and it did not improve as the sample size increased. Specifically, the true value of γ_u in the previous simulation study was 1 and the normal factor model estimated it (on average in the case when $n = 1,000$ and High skew/kurtosis) to be 0.76. Nevertheless, a pragmatic researcher may care less about biased estimation and more about testing the parameter to be different from zero. Thus, it is important to know what the Type I error and power are associated with the normal factor model in the case with high skew/kurtosis. Following the same latent factor model as before with 3 continuous and 1 dichotomous observed variables, using the high skew/kurtosis latent factor distribution and a sample size $n = 1,000$, five additional simulation scenarios were generated that included fixing the γ_u at 0.0, 0.25, 0.5, 1.5, and 2 (in addition to 1, which it was fixed at originally). Two hundred simulated data sets were used in each scenario and $\hat{\gamma}_u$ were obtained using the normal factor model and the mixture factor analysis model with 4 components.

Interesting results emerge regarding the performance of the normal factor model, Table 4. Despite the high skew/kurtosis, the normal factor model estimator of γ_u was unbiased when $\gamma_u = 0$ and had nominal (i.e., 0.05) Type I error. As the true γ_u increased, the downward bias in the normal factor model estimator increased (bias of -0.85 when $\gamma_u = 2$), but, despite the bias, it had near perfect power to reject zero in all cases when $\gamma_u \geq 0.5$. The mixture factor analysis model estimator had nominal Type I error; slightly better power; and clearly better bias, RMSE, and 95% coverage probability. Hence, whereas the mixture factor model provides an overall better estimator of γ_u in these cases of high skew/kurtosis with $n = 1,000$, the normal factor model performs adequately for the purpose of testing (i.e., it has correct Type I error and decent power).

TABLE 4
 Comparison of estimators for γ_u from the normal factor model and mixture factor model
 with 4 components in the high skew/kurtosis case with $n = 1,000$
 at varying true values for γ_u .

	$\gamma_u = 0.0$	$\gamma_u = 0.25$	$\gamma_u = 0.5$	$\gamma_u = 1$	$\gamma_u = 1.5$	$\gamma_u = 2$
Normal (K = 1)						
Bias	0.00	0.00	-0.05	-0.24	-0.52	-0.85
RMSE	0.08	0.08	0.10	0.27	0.54	0.87
95% Coverage	0.95	0.96	0.88	0.39	0.05	0.01
Power to reject 0 ^a	0.047	0.89	1.00	0.99	1.00	1.00
Mixture (K = 4)						
Bias	0.00	0.01	0.00	0.02	0.03	0.05
RMSE	0.07	0.08	0.11	0.20	0.31	0.48
95% Coverage	0.96	0.96	0.95	0.94	0.94	0.93
Power to reject 0	0.042	0.93	1.00	1.00	1.00	0.99

Note. Empirical summaries are across 200 simulated data sets. All other parameters are fixed to be the same as in the previous simulation.

^aThe "Power to reject 0" is the percentage of times out of 200 that the absolute value of the estimate divided by the standard error was greater than 1.96. When $\gamma_u = 0$, the values for "Power to reject 0" are equivalent to the empirical Type I error, which is nominally fixed at 0.05.

RMSE = empirical root mean square error.

Computational Issues

Generally, the more complicated the model, the more complicated the computation. In this simulation study, we find that under every scenario considered, maximum likelihood using the normal factor model converged using Mplus and in each case the maximum likelihood value was replicated at least once with different starting values. To varying degrees, this was not the case for the mixture factor models. Table 5 shows the percentage of times (out of the first 200 data sets generated for the first simulation study) that the model did not converge. Two types of errors appear in Mplus output file indicating the model did not converge correctly. They are, "The model estimation has reached a saddle point or a point where the observed and the expected information matrices do not match," and "The model estimation did not terminate normally due to a nonpositive definite Fisher information matrix." In both cases Mplus adds the following explanation/suggestion: "This is often due to the starting values but may also be an indication of model nonidentification. Change your model and/or starting values." These problems of convergence occur more as the number of components increases and when the sample size is smaller. Specifically, when $n = 100$, the 4-component mixture factor model did not converge in 10%–16% of the simulated data sets. But, when the sample size was $n = 1,000$ the multicomponent mixture models only had convergence problems in the case with

TABLE 5
 Summary of percentage of time computational issues arose in the simulation study using the $K = 2$ -, 3-, or 4-component mixture factor models.

<i>K</i>	<i>High Skew/Kurtosis</i> <i>Beta (.1, .30)</i>				<i>Mild Skew/Kurtosis</i> χ^2_1				<i>No Skew/Kurtosis</i> <i>Normal</i>			
	<i>n</i> = 100	200	500	1,000	<i>n</i> = 100	200	500	1,000	<i>n</i> = 100	200	500	1,000
Nonconvergence ^a												
2	2	0	0	0	1	0	0	0	1	4	3	6
3	6	1	0	0	2	0	0	0	4	3	4	4
4	14	4	0	0	10	4	2	0	16	9	7	6
Max not replicated ^b												
2	1	1	4	9	0	0	0	0	2	1	1	1
3	12	5	8	4	5	4	2	3	17	9	18	12
4	24	20	6	2	29	32	14	4	28	37	24	21
Collapse components ^c												
2	0	0	0	0	1	0	0	0	1	1	2	1
3	9	1	0	0	3	0	0	0	4	6	3	3
4	29	9	2	0	13	9	1	0	19	12	3	3

^a“Nonconvergence” indicates the maximum likelihood estimation in Mplus did not converge (i.e., Mplus indicated a warning and did not produce parameter estimates).

^bGiven that the model converged, “Max not replicated” indicates that the maximum likelihood estimators reported are based on a maximum likelihood value that was not replicated across the 100 random starting values.

^c“Collapse components” indicates that at least one of the parameters was held fixed by Mplus to avoid singularity of the information matrix.

normally distributed data, not converging 4%–6% of the time. Our simulation results presented earlier in Table 2 are based on 200 simulated data sets that yielded converged model estimation across all four models. That is, if one of the models did not converge for a particular data set, that data set was thrown out for all the models and an additional data set was simulated. This was done to ensure that we were comparing estimators from the different models based on the same data sets.

Considering the data sets where all the models did converge, it was possible for the maximum likelihood value to not be replicated across the 100 different random starting values. This occurred more frequently as more components were added and at smaller sample sizes, and like the convergence problem, occurred more when the true underlying factor had no skew/kurtosis (i.e., the scenario where the mixture factor model is more complicated than need be). In the presence of skew/kurtosis with $n \leq 200$, the maximum likelihood value associated with the 4-component mixture factor model was not replicated 20%–32% of the time. In the scenario where the latent factor had no skew or kurtosis, lack of replication for the 4-component model ranged from 21% to 37% and did not diminish as the sample size increased. The warning given by Mplus when this occurs is, “The best log-likelihood value was not replicated. The solution may not be trustworthy due to local maxima. Increase the number of random starts.”

One thing worth noting is that just because the best log-likelihood value was not replicated across different starting values, this does not necessarily mean that it is not the global maximum. And, on the other hand, just because the best log-likelihood value was replicated, this does not confirm that the value is indeed the global maximum. Hence based on this warning or lack thereof, one cannot conclude definitively that the maximum likelihood value has been found. This uncertainty is almost always present in numerical procedures for finding maximums but is particularly pervasive in mixture modeling.

An additional warning is output by Mplus for some cases, which indicates that “one or more parameters were fixed to avoid singularity of the information matrix. The singularity is most likely because the model is not identified or because of empty cells in the joint distribution of the categorical variables in the model.” This warning happens predominately due to 2 components in the mixture factor model having the same mean value, that is, α_{0k} is the same for two different k , indicating the components are on top of each other and hence cannot be distinguished. Essentially, this warning indicates the data do not support an additional component in the mixture factor model. As expected, this warning occurs more with more components in the model and smaller sample size, occurring 13%–29% of the time with 4 components and $n = 100$.

There was no evidence of an association between the data sets with non-replicated log-likelihoods and those that had one or more parameters fixed to avoid singularity. This indicates that these warnings are associated with different features of the observed data.

Further simulation investigation of warnings at smaller sample sizes.

The pervasiveness of warnings that occur in the mixture models with sample sizes, $n = 100$ and 200 , in the scenario with high skew/kurtosis begs the question of whether the large RMSE found for the estimator of γ_u using the 3- and 4-component models for these sample sizes is perhaps due to the numerical procedure not appropriately finding the maximum likelihood value. To examine this we employed two additional strategies targeting the high skew/kurtosis scenario with sample sizes $n = 100$ and 200 : (a) increased the number of random starting values from 100 to 200 and (b) considered the resulting estimators only for the cases where no warning was given.

When we reran the simulation with doubled random starting values, for $n = 100$ for the 4, 3, and 2 component models, respectively, the likelihood was not replicated 14%, 4%, and 0% and a warning about parameters being fixed to avoid singularity occurred 24%, 3%, and 0%. For $n = 200$, the nonreplicated likelihood occurred for the 4-, 3-, and 2-component models, respectively, 11%, 2%, and 1% and the fixed parameters to avoid singularity 5%, 0%, and 0%. So there was some decrease in the warnings for nonreplicated likelihoods and parameters being fixed to avoid singularity. But, the important question is whether

this decrease in warnings (attributable to increasing the number of random starting values) actually improved estimation of γ_u . Somewhat surprisingly, the answer was no. That is, the RMSE for 2, 3, and 4 components for $n = 100$ and $n = 200$ were nearly exactly the same (within simulation error) of the values presented in Table 2 when the random starting values were doubled.

We further examined the RMSE for γ_u in the high skew/kurtosis case when $n = 100$ and $n = 200$ just for those data sets where the maximum likelihood value was replicated across all of the models. Again, these restricted RMSE values were very similar to the ones previously reported and were not significantly different from the estimates in the data sets that had a warning that the maximum value had not been replicated. However, a distinction in RMSE was found in the $n = 100$ case with high skew/kurtosis using the 4-component model when we compared the estimators from the data sets with and without the warning indicating that parameters had been fixed (i.e., collapsed components) compared to those with that warning. Specifically, the RMSE was lower without this warning, 0.93 compared to 1.67 with the warning (which occurred 24% of the time). Thus, it appears that at least some part of the large variability in the mixture factor analysis estimator at small sample sizes is due to parameters being computationally fixed to avoid singularity of the information matrix.

ILLUSTRATIVE NUMERICAL EXAMPLE

The data presented here (Figure 3) are simulated data (available from the first author), which closely mimic the correlations and skew/kurtosis observed in the real data (Patterson et al., 2009). Cystic fibrosis (CF) is the most common inherited life-shortening, multisystemic disease among Caucasians. A typical treatment regimen includes daily chest physiotherapy, regular aerobic physical activity, a high-calorie/high-fat diet in addition to the medications. It requires regular, lifelong adherence to the treatment routine to obtain effective results. A self-report questionnaire was administered to CF patients and three scales were derived measuring emotional strains, appearance worries, and physical worries such that higher values on the scales indicated more frequent strains or worries. The Cronbach's alpha for the 3 indicators of illness strains was 0.78. Researchers hypothesized that the latent construct illness strains measured by the three scales would increase the odds of nonadherence to treatment. It was also of interest to consider potential covariates of gender and age. The data ($n = 500$) were simulated from a model with chi-square 1 *df* underlying distribution (the mild skew/kurtosis case in the simulation study section) so that the underlying factor was related to nonadherence with a log odds ratio (logit) of 1 and covariates of gender and uniformly distributed mean-centered age were generated in such

a way that gender was significantly related to adherence but age was not after controlling for the underlying factor.

Parameter Estimates and Inference

Based on results of the simulation study, these data, which exhibit skew in the continuous observed variables ranging from 1.3 to 1.6 and kurtosis from 4.9 to 7.2, might be suspected to lead to some bias in the γ_u parameter when a latent factor model assuming normality is used. Furthermore, given the moderately large sample size of $n = 500$, we expect a mixture factor model to be able to improve the bias without dramatically increased variability in the estimator. Table 6 presents the resulting estimates obtained from Mplus using 500 random starting values. The first column in Table 6 provides benchmark values for what each estimated relationship would be if the underlying factor were observed (which it is in this case because it is simulated data).

We see that the normal factor model estimates of γ_u (0.83 and 0.69 without/with covariates, respectively) are smaller than the benchmark values (0.94 and 0.83, respectively) indicating downward bias as was seen in the simulation. We note that a 95% confidence interval formed by taking ± 1.96 times the

TABLE 6
Parameter estimates (standard errors) from illustrative numerical example with $n = 500$
without covariates (top) and with covariates (bottom).

	<i>Regression on f observed</i>	<i>Normal Factor Model</i>	<i>Mixture-2 Factor Model</i>	<i>Mixture-3 Factor Model</i>	<i>Mixture-4 Factor Model</i>
Without covariates					
γ_u	0.94 (0.13)	0.83 (0.14)	0.89 (0.16)	0.94 (0.17)	0.91 (0.16)
λ_1	0.72 (0.03)	0.68 (0.06)	0.70 (0.06)	0.72 (0.06)	0.74 (0.06)
BIC	—	4,798	4,652	4,613	4,597
With covariates					
γ_u	0.83 (0.14)	0.69 (0.14)	0.77 (0.16)	0.84 (0.17)	0.81 (0.17)
λ_1	0.72 (0.03)	0.67 (0.06)	0.70 (0.06)	0.71 (0.06)	0.73 (0.06)
β_1 (female)	0.55 (0.22)	0.58 (0.23)	0.56 (0.23)	0.53 (0.23)	0.54 (0.23)
β_2 (age)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.01 (0.04)	0.00 (0.04)
BIC	—	4,702	4,571	4,541	4,535

Note. First column represents the estimates obtained from regressions on the actual simulated values for the underlying factor and provides benchmark for comparison. Columns labeled Normal-1 and Mixture-2, 3, 4 present results corresponding to fitting the normal factor model and the mixture factor models with respective number of components. For this example, the number of random starting values was set at 500 for the initial iteration stage followed by iteration to convergence of the best 50.

BIC = Bayesian Information Criterion.

standard error of the normal factor model estimator of γ_u would cover the benchmark value but not the true population value of 1 for the case with covariates (i.e., $0.69 \pm 1.96*0.14$ does not cover 1). Based on the simulation study we expected the estimate for the linear factor loading λ_1 from the normal model to perform well despite skew or kurtosis in the data and find that it is the case. We also find the the covariate parameter estimates ($\hat{\beta}_1$ and $\hat{\beta}_2$) based on the normal factor model are very close to the benchmark estimates and in both cases have 95% confidence intervals that surround the population values (0.30 and 0.00, respectively).

Both with and without covariates, the BIC indicates better fit (i.e., is smaller) with increasing numbers of components up to 4. The BIC for the 5-component model (estimates not shown) was larger (indicating worse fit) than the 4-component model: BIC = 4,601 for 5 components without covariates and BIC = 4,562 with covariates. The mixture factor models correct the downward bias seen for $\hat{\gamma}_u$ in the normal factor model. That is, a 95% confidence interval for the mixture factor model with 4 components (actually also for the 2- and 3-component models) surrounds the benchmark and population value for γ_u without and with covariates. Regarding variability, the standard errors of $\hat{\gamma}_u$ for the mixture factor models are slightly larger than those for the normal factor model, but they are not more variable for the linear loading λ_1 or the covariate parameter estimates.

Factor Score Estimates

The main focus of this article has been on estimating model parameters (e.g., γ_u and λ_1) using the mixture factor model compared with the normal factor model in the presence or absence of skew/kurtosis in the underlying latent factor. Although the distribution of the underlying factor has so far been considered a nuisance, it is instructive to examine how the normal factor model compared with the mixture factor model predicts the underlying factor values for individuals. Figure 6 graphs the factor score estimates for this example from the normal factor model and the mixture factor model with 4 components (Mixture-4 model) without covariates and compares them with the true generated underlying factor values. We first notice that the histogram of the factor scores from the normal factor model do not look normal and are right skewed (albeit not as much) like the true underlying factor. Despite the fact that the normal factor model assumes normality for the latent factor, the posterior distribution of the latent factor given the data, that is, the factor scores, which are calculated as the expected value of f given the data and the model are not expected to be normal when the data itself is not normal. However, as can be seen in the scatterplot of the true underlying factor versus the factor scores from the normal factor model

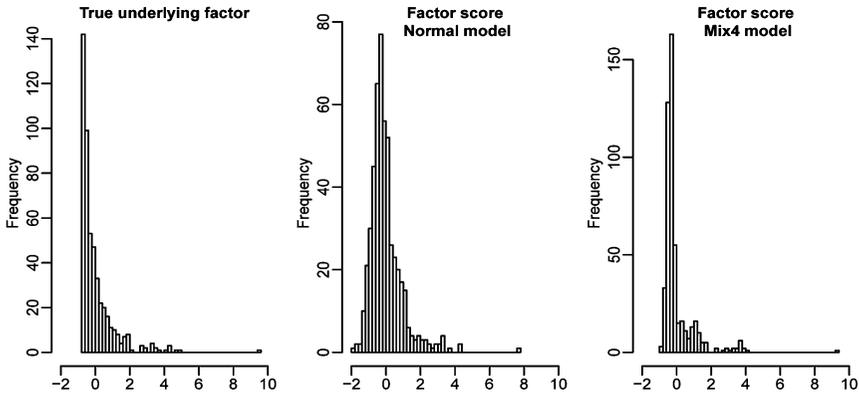


FIGURE 6 Histograms of simulated underlying factor values and factor score estimates obtained from the normal factor model and the mixture factor model with 4 components (Mix4) for the illustrative numerical example.

(Figure 7), the normal factor model has shrunk the large values down more than it should (i.e., on the high end true values are larger than predicted by the normal model) and it spreads out values on the low end more than it should (i.e., more variability on the low end in the normal factor model predictions than there is in the true latent factor). Both of these “misses” by the normal factor model are corrected to some extent by the factor mixture model with 4 components. That is, the factor score estimates from the mixture factor model line up more closely with the true underlying factor values (bottom left of Figure 7). We further note that the estimated probabilities of class membership found in the mixture factor model with 4 components were 82.8%, 13.5%, 3.4%, and 0.2%. The component with probability 0.2% included just one observation corresponding to the single large value of the true underlying factor near 10. Thus, despite whether this latent value might be described as an outlier or just a typical observation from a skewed distribution (as it is here), the mixture factor model provides an accurate predicted value for it.

DISCUSSION

In a latent factor model with both continuous and dichotomous observed variables, it was found that misspecifying the latent variable as normal and using normal maximum likelihood leads to downward bias in the estimated path relating the factor to the dichotomous outcome that worsens as the true latent factor distribution deviates further from normality (e.g., becomes more

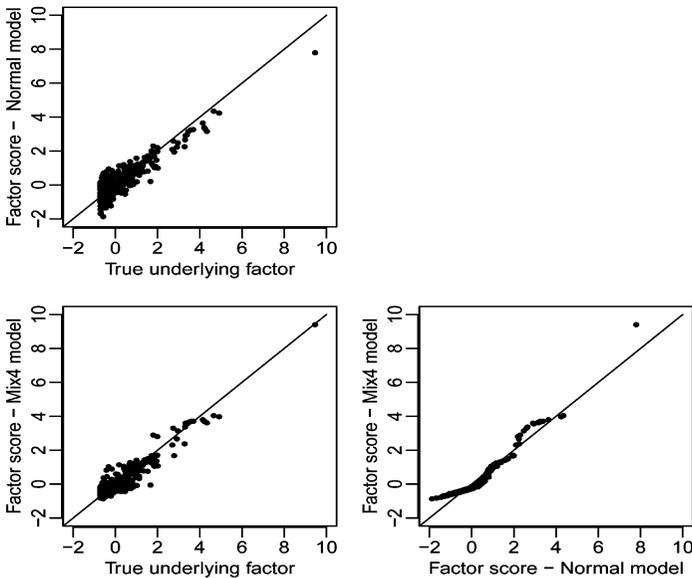


FIGURE 7 Q-Q plot comparisons of simulated underlying factor values with factor score estimates obtained from the normal factor model and the mixture factor model with 4 components (Mix4) for the illustrative numerical example.

skewed and kurtotic) and also worsens as the true magnitude of the relationship becomes stronger. As a result of the bias, the coverage probability for confidence intervals based on the normal factor model was very poor and worsened as sample size increased. Interestingly, in contrast, the Type I error and power associated with testing the path for the dichotomous outcome variable using maximum likelihood assuming normality were not negatively affected by the presence of skew and kurtosis in the latent factor. As expected based on previous work (Anderson & Amemiya, 1988; Browne & Shapiro, 1988), the linear factor loadings relating continuous observed variables to the underlying latent factors estimated using normal ML were robust to misspecification of normality of the latent factors. This study identified these properties through a simulation study and proposed to alleviate the bias of the normal factor model and potentially improve efficiency by weakening the distributional assumption for the underlying latent variable using a mixture factor analysis model.

The current application of the mixture factor analysis model considers the distribution of the latent factor f a nuisance that is modeled solely in order to obtain good estimation and inference for the target parameters of interest in the

model (e.g., γ_u and λ_1). The mixture factor analysis model provides an attractive solution to the problem of misspecification of the latent factor because it is a very flexible means for modeling nonnormal continuous distributions and can be directly implemented in Mplus. But, with flexibility comes variability and based on the current simulation study, the variability in the estimators from the mixture factor analysis model can be extreme for the path relating the factor to the dichotomous outcome variable when sample sizes are small ($n = 100$ or 200) in the case of high skew/kurtosis. Indeed for these smaller sample sizes, flexibly approximating the nonnormal latent factor with a mixture factor analysis model produced substantially worse RMSE for the path to the dichotomous outcome variable than what was obtained by simply assuming normality. But, at sample size 500 or greater, the mixture factor model performed better than the normal factor model in the presence of skew/kurtosis for the path relating the latent factor to the dichotomous outcome. Of note, even at smaller sample sizes, the continuous variable linear factor loading estimator was less variable (i.e., more efficient) using the mixture factor analysis model than the normal factor model. The coverage probability for both estimators based on confidence intervals from the mixture factor analysis models were near nominal for sample size 500 or greater. Also, the mixture factor analysis model performed equally well to the normal factor model in the case where there was no skew or kurtosis (i.e., where the normal factor model was correct). Hence, the mixture factor analysis model is a potentially useful estimation method for sample size 500 and greater given its improvements in bias, coverage probability, and RMSE for paths relating the factor to dichotomous and continuous observed variables compared with the normal factor model.

In this study, results were examined separately by the number of mixture components and also the BIC was used to choose the number of components. Because the sampling distributions of the estimated paths relating the factor to continuous and dichotomous observed variable behaved differently with respect to increases in the number of underlying components, it is hard to conclude which number of components was “optimal” in the different scenarios with skew/kurtosis because it depends on what criteria is used. The BIC performed well for identifying that only 1 component was needed in the scenario where the true latent factor was normal. Previous simulation studies have suggested that BIC performs well for identifying the “true” number of components (B. Muthén, 2007; Nylund et al., 2007), but recall in our high and mild skew/kurtosis scenarios there is not a “true” mixture model underlying the factor; it is simply a nonnormal homogeneous distribution. Further study is needed to identify a method for choosing the number of components that approximate the true latent factor “well enough” in order to obtain parameter estimates with good properties. At present we recommend the use of BIC to guide the choice of the number of components but suggest limiting the upper value for K regardless of

BIC. In this study, limiting K to be 4 appeared to provide enough flexibility so that the parameters of interest were well estimated with sample size 500 and greater.

Computational problems were regularly encountered in Mplus with the mixture factor analysis model when more components were fit and sample sizes were small. A recent paper by An and Bentler (2011) introduces a variant of the EM algorithm, namely, a Monte Carlo EM utilizing the Gibbs sampler. An and Bentler (2011) fit mixture factor analysis models with both continuous and dichotomous indicators and find in the data examples they considered that their algorithm “outperforms estimation algorithms implemented in Mplus in terms of estimation accuracy when the latent factor structure becomes more complex” (p. 2643). At present their algorithm is not implemented in existing software and it remains to be investigated whether it also has computational problems in small sample sizes with increasing components.

We note that there are parallels between the results found here regarding parameter estimation in the presence of misspecification of the latent factor distribution and results from the mixed effects model literature. In particular, paralleling the robustness property we found for the linear factor loading, several studies have shown that maximum likelihood estimation for linear fixed effects parameters based on Gaussian assumptions for the random effects is robust to non-Gaussian random effects distributions (Butler & Louis, 1992; Verbeke & Lesaffre, 1997; Zhang & Davidian, 2001). On the other hand, in the context of *generalized* linear mixed effects models (with nonlinear links such as logistic as for our dichotomous variable factor loading), there is evidence that random effects parameter estimation may be severely compromised (Agresti, Caffo, & Ohman-Strickland, 2004; Heagerty & Kurland, 2001; Litiere, Alonso, & Molenberghs, 2007; Neuhaus, Hauck, & Kalbfleisch, 1992; Richardson & Green, 2002).

Of course, our findings and recommendations are limited to the extent of our simulation experiment. There are a number of limitations that may encourage future lines of research. First, we considered only one latent factor. It may be that when using the mixture model approach to approximate a multivariate vector of latent factors, its performance may not be as good because of the added complexity of multiple dimensions. Second, the influence of the covariates within the model needs to be explored further. We did not include covariates in our simulation. This exclusion allowed us to pinpoint the effect of nonnormality of the latent factor on the parameter estimates without confounding the issue of covariate adjustment. It may be in some cases that the nonnormality of the latent factor can be explained by certain nonnormally distributed observed predictor variables and that conditional on the observed predictors, the latent factor is indeed normal. Related to this, a careful examination of the interpretation of the relationship between the predictors in the mixture model and the latent

factors should be explored taking into account the predictors' effect on the factor potentially through their effect on the mixture components. It is also of interest to explore the performance of the mixture model approach in models with more indicators and in particular a larger proportion of noncontinuous observed variables. Although not investigated, it seems that if there were a larger proportion of dichotomous items, it might be more difficult to identify the mixtures.

Finally, we note that other approaches instead of the mixture factor analysis model may be considered for dealing with nonnormality of the latent factors. A natural method to consider in comparison to the mixture of normals approach used herein is a mixture of mass points approach. This method has been referred to as "nonparametric maximum likelihood estimation" (Aitkin, 1999; Laird, 1978; Schafer, 2001; for a review see Skrondal & Rabe-Hesketh, 2004) and is implementable in Mplus (see example 7.26 in the Mplus 6 manual). Within the item response theory (IRT) setting (i.e., all dichotomous or ordered categorical observed variables measuring a single latent factor), Mislevy (1984) presents, in a mathematical exposition, several parametric methods for estimating the latent distribution in an IRT model including a method similar to the mixture of normals approach presented herein. Also within the IRT setting, nontrivial bias has been found in parameter estimates when normal IRT models are fit to data where the latent distribution is nonnormal (van den Oord, 2005; Zwiderman & Wollenberg, 1990). A method called Ramsay-curve IRT (Woods, 2006) was developed specifically to assess and correct for potential nonnormality in the latent factor within an IRT model. This spline-based method decreases bias similar to the mixture factor model approach and would be of interest to compare in future work. Software to implement the Ramsay-curve IRT, called RCLOG, is available (Woods & Thissen, 2004) but is not yet included in standard latent variable modeling packages such as Mplus.

In summary, this study contributes to understanding the influence of nonnormality in the estimation of model parameters in a latent factor model including both continuous and dichotomous outcome variables, and further we have introduced a potentially viable (readily implementable) mixture factor analysis approach for improved estimation in such cases.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant U01-HD061940 from the National Institute of Child Health and Development. We also give special thanks to the careful attention and constructive suggestions made by the associate editor, Stephen West.

REFERENCES

- Agresti, A., Caffo, B., & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics and Data Analysis*, 47, 639–653.
- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 117–128.
- An, X., & Bentler, P. (2011). Extended mixture factor analysis model with covariates for mixed binary and continuous responses. *Statistics in Medicine*, 30, 2634–2647.
- Anderson, T. W., & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, 16, 759–771.
- Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research*, 26, 148–182.
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42, 757–786.
- Bauer, D. J., & Curran, P. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3), 338–363.
- Bauer, D. J., & Curran, P. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9(1), 3–29.
- Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, 41, 193–208.
- Butler, S. M., & Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine*, 11, 1981–2000.
- Carroll, R. J., Roeder, K., & Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics*, 55, 44–54.
- Dolan, C. V., & Van der Maas, H. J. L. (1998). Fitting multivariate normal mixtures subject to structural equation modeling. *Psychometrika*, 63, 227–253.
- Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577–588.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Freedman, D. A. (2006). On the so-called “Huber sandwich estimator” and “Robust standard errors.” *The American Statistician*, 60(4), 299–302.
- Heagerty, P. J., & Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88, 973–985.
- Huber, P. (1967). The behavior of the maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 221–233). Berkeley: University of California Press.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39–59.
- Johnson, N. L., & Kotz S. (1982). *Encyclopedia of statistical sciences* (Vols. 1–9). New York, NY: Wiley.
- Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling*, 15(1), 136–153.
- Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73, 805–811.
- Lee, S. Y., & Song, X. Y. (2003). Bayesian model selection for mixtures of structural equation models with an unknown number of components. *British Journal of Mathematical and Statistical Psychology*, 56, 145–165.
- Litire, S., Alonso, A., & Molenberghs, G. (2007). Type I and Type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63, 1038–1044.

- Lubke, G., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21–39.
- Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research, 32*, 355–373.
- McLachlan, G. J., & Peel, G. (2000). *Finite mixture models*. New York, NY: Wiley.
- Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin, 105*(1), 156–166.
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359–381.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49*, 115–132.
- Muthén, B. (2007). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age.
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., . . . Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*, 459–475.
- Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
- Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4*, 139–157.
- Neuhaus, J. M., Hauck, W. W., & Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika, 79*, 755–762.
- Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.
- Patterson, J. M., Wall, M. M., Berge, J., & Milla, C. (2009). Associations of psychosocial factors with health outcomes among youth with cystic fibrosis. *Pediatric Pulmonology, 44*(1), 46–53.
- Pinhero, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*, 12–35.
- Reinartz, W. J., Echambadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research, 37*, 227–244.
- Richardson, S., & Green, P. J. (2002). Mixture models in measurement error problems with reference to epidemiological studies. *Journal of the Royal Statistical Society, Series A, 165*, 549–566.
- Schafer, D. W. (2001). Semiparametric maximum likelihood for measurement error model regression. *Biometrics, 57*, 53–61.
- Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, structural equation models*. Boca Raton, FL: CRC Press.
- Sorenson, H. W., & Alspach, D. L. (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica, 7*, 465–479.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408.
- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, UK: Wiley.

- van den Oord, E. J. C. G. (2005). Estimating Johnson curve population distributions in MULTILOG. *Applied Psychological Measurement, 29*, 45–64.
- Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91*, 217–221.
- Verbeke, G., & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis, 53*, 541–556.
- Woods, C. M. (2006). Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. *Psychological Methods, 11*(3), 253–270.
- Woods, C. M., & Thissen, D. (2004). *RCLOG v.1: Software for IRT parameter estimation with the latent population distribution represented using spline-based densities* (Tech Rep No. 1). Chapel Hill: University of North Carolina, L. L. Thurstone Psychometric Laboratory.
- Yung, Y. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297–330.
- Zhang, D., & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics, 57*, 795–802.
- Zhu, H. T., & Lee, S. Y. (2001). A Bayesian analysis of finite mixtures in the LISREL model. *Psychometrika, 66*, 133–152.
- Zwinderman, A. H., & Van den Wollenberg, A. L. (1999). Robustness of marginal maximum likelihood estimation in the Rasch model. *Applied Psychological Measurement, 14*(1), 73–81.