Extracting Spurious Latent Classes in Growth Mixture Modeling With Nonnormal Errors

Kiero Guerra-Peña[1] and Douglas Steinley[2]

Abstract

Growth mixture modeling is generally used for two purposes: (1) to identify mixtures of normal subgroups and (2) to approximate oddly shaped distributions by a mixture of normal components. Often in applied research this methodology is applied to both of these situations indistinctly: using the same fit statistics and likelihood ratio tests. This can lead to the overextraction of latent classes and the attribution of substantive meaning to these spurious classes. The goals of this study are (1) to explore the performance of the Bayesian information criterion, sample-adjusted BIC, and bootstrap likelihood ratio test in growth mixture modeling analysis with nonnormal distributed outcome variables and (2) to examine the effects of nonnormal time invariant covariates in the estimation of the number of latent classes when outcome variables are normally distributed. For both of these goals, we will include nonnormal conditions not considered previously in the literature. Two simulation studies were conducted. Results show that spurious classes may be selected and optimal solutions obtained in the data analysis when the population departs from normality even when the non-normality is only present in time invariant covariates.

[1]Pontificia Universidad Católica Madre y Maestra, Santiago, Dominican Republic
[2]University of Missouri–Columbia, MO, USA

Corresponding Author:
Kiero Guerra-Peña, Department of Psychology, Pontificia Universidad Católica Madre y Maestra, Autopista Duarte Km 1½, Santiago 51000, Dominican Republic.
Email: kpena@pucmm.edu.do

Growth mixture modeling (GMM) has gained great popularity in the social sciences over the past two decades. This methodology provides a more flexible modeling of longitudinal data analysis compared to traditional methods such as hierarchical linear modeling and repeated-measures ANOVA (Bauer & Curran, 2003a). Particularly interesting is the identification of latent trajectories classes within a population. These latent classes, obtained through GMM, could be interpreted as groups of individuals with substantively different developmental paths (Moffitt, 1993) or with different patterns of substance abuse (Sher, Jackson, & Steinley, 2011).

Nevertheless, few applied researchers are aware that finite mixture modeling, and by extension GMM, can be used for two main applied purposes: (1) to identify mixtures of normal subgroups within a larger population and (2) to approximate oddly shaped distributions by a mixture of normal components (Bauer & Curran, 2003a). In the first situation, the mixture of normal subpopulations can be correctly identified using substantive information, such as identifying two normal populations that represent hip bone density for male and female humans. On the other hand, when the second situation occurs, the mixture components are used by the finite mixture procedure to approximate the data better, and thus substantive interpretation is often inappropriate. For example, in the case of a population with an exponential distribution, the finite normal mixture model will most likely need more than one normal population to approximate the heavy right tail of this distribution. The problem here is that the methodology is applied to both of these situations indistinctly. Furthermore, the same fit statistics and likelihood ratio tests are used to select the best solution in both cases.

This article considers the structural equation modeling perspective introduced to the study of growth curves by Meredith and Tisak (1984, 1990). The parametrization of Bollen and Curran (2006) is used for consistency.

Growth Mixture Modeling With Normal Errors Assumption

When we have no knowledge of group membership in the analysis of growth patterns, we might hypothesize that there exist classes in the population that are unknown or latent. The analysis of latent curve models with unknown group membership allows us to explore the possibility that the data derive from a mixture of populations. The main interest here is finding out, with a small level of error, from which population a particular observation comes from. It is important to note that multiclass mixtures can result from a single, nonnormal population (Bauer & Curran, 2003a, 2003b). This methodology is commonly known in the literature as mixture modeling, and it makes use of mixed models and full information estimation (Arminger & Stein, 1997; Arminger, Stein, & Wittenberg, 1999; Bauer & Curran, 2003a, 2004; Jedidi, Jagpal, & DeSarbo, 1997; B. O. Muthén & Shedden, 1999). These techniques classify observations into groups and fit latent class models to those groups simultaneously.

The general equations for this approach, following the notation from Bollen and Curran (2006), are

$$y_{it}^{(g)} = \sum_{g=1}^{G} p_i^{(g)} \left[\alpha_i^{(g)} + \lambda_t^{(g)} \beta_i^{(g)} + \varepsilon_{it}^{(g)} \right] \tag{1}$$

$$\alpha_i^{(g)} = \mu_\alpha^{(g)} + \sum_{k=1}^{K} \gamma_{\alpha x_k}^{(g)} x_{ik} + \zeta_{\alpha i}^{(g)} \tag{2}$$

$$\beta_i^{(g)} = \mu_\beta^{(g)} + \sum_{k=1}^{K} \gamma_{\beta x_k}^{(g)} x_{ik} + \zeta_{\beta i}^{(g)} \tag{3}$$

where $p_i^{(g)}$ is the probability that the ith individual belongs to the gth group with all $p_i^{(g)} \geq 0$ and $\sum_{g=1}^{G} p_i^{(g)} = 1$, $y_{it}^{(g)}$ is the value of the response variable for individual i at time t in group g; $\alpha_i^{(g)}$ and $\beta_{1i}^{(g)}$ are the intercept and slope for individual i in group g, respectively, $\lambda_t^{(g)}$ is a constant that specifies the time of measurement in group g, and $\varepsilon_{it}^{(g)}$ is the disturbance (i.e., error) for individual i at time t in group g. The interpretation of the intercept is the implied value of the repeated measures at the first time of measurement ($\lambda_t^{(g)} = 0$) and the linear slope describes the rate of change in the repeated measures for unit increase in time. It is in the constant $\lambda_t^{(g)}$ that a researcher can incorporate linear and nonlinear trajectories in the model for each group g. Notice that Equation (1) allows for individual specific intercept and slope by including the subscript i for group g.

In Equations (2) and (3), the population means are given by the μs and random components, ζs, and the latter allows the αs and βs to be different for each individual. The γs are the covariate coefficients for the intercept and slopes. The covariate coefficients are interpreted in the same way as in multiple regression; a unit increase in the covariate implies a change of γ units in the intercepts or slopes, respectively. Since these are time invariant covariates (TICs), they may take a different value for each individual but remain the same across all times of measurement. The variance components for $\alpha_i^{(g)}$ and $\beta_{1i}^{(g)}$ are the variances of $\zeta_\alpha^{(g)}$ and $\zeta_\beta^{(g)}$, or $\psi_{\alpha\alpha}^{(g)}$ and $\psi_{\beta\beta}^{(g)}$, respectively. $\psi_{\alpha\beta}^{(g)}$ denotes the covariance between intercept and slope for each group g. The matrix representation of Equations (1) to (3) is

$$\mathbf{y}^{(g)} = \sum_{g=1}^{G} p_i^{(g)} \left[\Lambda^{(\mathbf{g})} + \eta^{(g)} + \varepsilon^{(g)} \right] \tag{4}$$

$$\eta^{(g)} = \mu_\eta^{(g)} + \Gamma^{(g)} \mathbf{x}^{(g)} + \zeta^{(g)} \tag{5}$$

where $\mathbf{y}^{(g)}$ is a $T \times 1$ vector of repeated observations for group g, $\Lambda^{(g)}$ is a $T \times m$ matrix of factor loadings for group g, $\eta^{(g)}$ is a $m \times 1$ vector of m latent factors for group g, and $\varepsilon^{(g)}$ is a $T \times 1$ vector of disturbances for group g. Time is included as a column in the $\Lambda^{(g)}$ matrix. Also, in Equation (5), $\mu_\eta^{(g)}$ is an $m \times 1$ vector of factor means for group g, and $\zeta^{(g)}$ is an $m \times 1$ vector of random components for group g.

$\Gamma^{(g)}$ is a $T \times T$ matrix of regression coefficients of the repeated measures on the covariates $\mathbf{x}^{(g)}$ for group g.

Jedidi et al. (1997) address the issue of mathematical identifiability of growth mixture models. The authors state that at least three occasions of measurement are needed to model a linear trend, and at least four and five occasions are needed for fitting quadratic and cubic trends, respectively. The authors also show that if the single group growth model (only one latent class) is mathematically identified, then the corresponding growth mixture model (more than one latent class) is identified as well, given that for each group the observed variables come from a multivariate normal distribution. The estimation of the model and its identification relies on this assumption of multivariate normality.

Issues in the Decision of the Number of Classes in GMM

There are several issues with the decision of how many latent classes to estimate in GMM. Local optimal solutions are more frequent in these models compared to those where the classes are known, and several authors have suggested that a large number of random starts are needed (see, e.g., Hipp & Bauer, 2006). Other authors have suggested the same approach to avoid nonconvergence of replications in simulation studies (Li, Harring, & Macready, 2014; Liu & Hancock, 2014). Moreover, the maximum likelihood (ML) estimator for these mixture models yield estimates that are not consistent when the multivariate normality assumption is violated (Arminger et al., 1999). Bauer and Curran (2003a, 2004) show that under situations where the multivariate normality assumption is violated, the number of latent groups can be overselected. The authors demonstrated, through a series of simulation studies, that in these cases multiple class solutions could be preferred even when a single group was used to simulate the data. Also, these spurious classes emerged even when the departure from multivariate normality was as small as a skewness and kurtosis equal to 1. The authors concluded that more classes were selected to better approximate the nonnormal distribution.

Most fit statistics currently used in GMM can be classified in one of three groups: information criteria for model selection, Bayesian-based information criteria, and classification-based information criteria (McLachlan & Peel, 2000). The authors explain this classification as follows: information criteria for model selection fit indices that include a measure of lack of fit and penalty for model complexity (e.g., AIC [Akaike's information criterion], EIC [empirical information criterion], and CVIC [cross-validation information criterion]); Bayesian-based information criteria are those developed within the Bayesian framework but can also be used in the frequentist framework (e.g., BIC [Bayesian information criterion], LEC [Laplace-empirical criterion]); and last, classification-based information criteria are those that use the complete-data likelihood within the expectation-maximization (EM) framework for the fitting of the mixture model (e.g., EN, NEC [Laplace-empirical criterion], and ICL-BIC [Integrated classification likelihood BIC]). The concept of entropy is central to these last types of fit statistics. Entropy is a value between 0 and 1 that

reflects the quality of classification; 0 represents randomness and 1 suggests perfect classification (Uher et al., 2009).

AIC is the most used information criteria fit index. AIC tends to favor models with more latent classes than the true model (Celeux & Soromenho, 1996; McLachlan & Peel, 2000; Soromenho, 1993). The most used Bayesian-based information criteria are BIC and LEC. BIC has been shown to not underestimate the true number of latent classes, asymptotically, for normal finite mixture models (Campbell, Fraley, Murtagh, & Raftery, 1997; Dasgupta & Raftery, 1998; Leroux, 1992; Roeder & Wasserman, 1997). However, other simulation research have found that BIC estimates fewer groups than the ''true'' model when, despite a valid component density, the sample size is not large (Celeux & Soromenho, 1996). BIC overestimates the number of latent classes when the model has been misspecified (Biernacki, Celeux, & Govaert, 1998) or the data are nonnormal (Bauer & Curran, 2003a, 2003b, 2004).

Tofighi and Enders (2008) conducted simulations studies that showed that the sample-adjusted BIC (SBIC) consistently favored the correct three-latent-classes solution over an erroneous four-latent-classes solution. Nylund, Asparouhov, and Muthén (2007) found that the BIC and SBIC perform best among fit indices in latent class analysis (LCA) and GMM analysis when all generated data follow normal distributions. Jung and Wickrama (2008) showed that the BIC correctly favors the ''true'' number of latent classes in GMM analysis. The authors considered a case where the outcome variables were slightly nonnormal (skewness and kurtosis equal to 1). Henson, Reise, and Kim (2007) also found that the SBIC performs best compared to other fit statistics.

Classification-based information criteria include a bias correction for entropy. CLC [Classification likelihood information criterion], NEC, and ICL-BIC are commonly used classification-based fit statistics. The CLC tends to overestimate the number of mixture components when the proportion of individuals in each group is not equal (Biernacki, Celeux, & Govaert, 1999). This limits its usefulness in most real data context where the ''true'' number of latent classes is unknown. NEC tends to favor models with multiple latent classes (Biernacki et al., 1999; McLachlan & Peel, 2000). In a simulation study by McLachlan and Peel (2000) with varying sample sizes, overlap and level of linear relation among observed outcome variables, the authors show empirically that of all these fits statistics, for mixtures of normal components, the LEC and ICL-BIC consistently identify the underlying number of latent classes. In general, Bauer and Curran (2003a, 2003b, 2004) conducted simulation studies that show that under certain conditions all fit indices tend to overestimate the number of components.

Bauer and Curran (2003a, 2003b, 2004) have exposed the general problem of using fit indices for the estimation of the number of mixture components. The authors found that in some instances all fit indices overestimate the number of groups. Bauer and Curran (2004) point out three possible conditions in which fit indices may lead to the estimation of spurious latent classes: misspecification of the model, continuous repeated observations that depart from normality, and nonlinear relationships between latent and observed variables. When we are interested in the first of these uses, fit

indices are used to select the optimal number of mixture components, but model coefficients lose their interpretation (e.g., slope parameters no longer describe individual developmental trajectories of the outcome variable) as compared to the second situation. The essential problem is that the same fit indices are used for model selection in both uses. The distinction between these two situations is one of the major problems in finite mixture model estimation. When a single class is present in the data and the model is misspecified (e.g., a linear growth is fitted when the growth is in fact curvilinear), the likelihood function for the single group solution is not a good representation of the data, and the likelihood function of a ''false'' multiple-group solution could yield a better approximation (Bauer & Curran, 2003a).

Bauer and Curran (2003a) note that in simulations in which a single ''true'' group with minor departures from normality is analyzed, fit indices tend to favor multilatent class solutions, with ICL-BIC being the most conservative. Nevertheless, ICL-BIC favored a two-class solution in about 70% of the replications when both skewness and kurtosis are equal to 1 and roughly 93% when skewness and kurtosis are 1.5 and 6, respectively.

B. O. Muthén (2003) developed and suggested the use of multivariate skewness tests (MST) and the multivariate kurtosis test (MKT). These are fit statistics that compered the multivariate skewness and kurtosis implied by the k-class growth mixture model with those obtained from the sample data. In these tests, a low probability value indicates departure of the model implied versus the sample calculated moments, and a large probability value indicates adequate model fit. Tofighi and Enders (2008) found, through simulation studies, that MST and MKT favor GMM solutions with less latent classes than that of the model used to simulate the data, making them overly conservative.

Preacher and Merkle (2012) propose that the major problem with fit indices is that researchers use them for model selection in structural equation models as if they have no error, namely, without considering the sample variability of the selection criteria. This point sheds light on two potential difficulties for model selection in finite mixture models: replicability of findings and comparing fit statistics for models with different numbers of mixture components. The authors show that, for the same set of models, the BIC favors different models as the sample size increases. Also noteworthy is their observation that without some measure of standard error or confidence intervals it is difficult to truly distinguish between models that have BIC scores that are close in value. They suggest comparing the BICs for two competing models, ΔBIC, and confidence intervals (CIs) for BIC and ΔBIC, but state that the overlap between their CIs ''does not guarantee that model selection decision will be stable over repeated samples'' (Preacher & Merkle, 2012, p. 12). Further research is needed in constructing confidence intervals for fit statistics as the authors suggest.

In a commentary to Bauer and Curran (2003a), both B. O. Muthén (2003) and Rindskopf (2003) suggest the use of the Lo–Mendell–Rubin likelihood ratio test (LMR-LRT; Lo, Mendell, & Rubin, 2001). This likelihood ratio test has been exposed as problematic in that the k-class model and the $k-1$-class model that it

compares are not nested, making it inappropriate (Jeffries, 2003). Nevertheless, Nylund et al. (2007) suggest that LMR-LRT may still be useful for class enumeration and base their statements on the Lo et al. (2001) simulation study. A clear competitor to the LMR-LRT is the bootstrap likelihood ratio test (BLRT) described by McLachlan and Peel (2000). The BLRT is a parametric bootstrap method for comparing nested models in LCA and GMM. The performance of this statistic test has not been widely explored in the literature. Our literature review reveled very few articles that perform simulation studies in GMM analysis that include BLRT. Nylund et al. (2007) showed that, with normally distributed outcome variables, BLRT outperforms the LMR-LRT. Tofighi and Enders (2008) found similar results and stated that BLRT performs better than MST and MKT in the selection of correct number of latent components in GMM analysis. Jung and Wickrama (2008) conducted the only study that we were able to find that explores the efficiency of BLRT in selecting the correct number of latent classes in GMM analysis when outcome variables are non-normal. Jung and Wickrama (2008) considered skewness and kurtosis values of 1 on the repeated measures and found that BLRT performed best among likelihood ratio tests and fit indices, with the exception of the BIC and SBIC.

We have seen the particular problems with each of the fit indices and likelihood ratio tests presented here as well as situations in which these statistics can lead to the extraction of spurious latent classes. The selection of the optimal number of latent classes in finite mixture models has to be preceded by a diligent investigation of the appropriateness of the assumptions about the data the researcher is making. The implementation of any model comes with a set of assumptions that the researcher, knowingly or unknowingly, makes about the data collected and the underlying growth curve of the outcome variables. Namely, if a linear growth is fitted, the researcher is assuming that the outcome variables increase or decrease linearly over time. Fit indices do not inform the researcher about erroneous specifications of the model, and they could overestimate the number of classes as a result of this misspecification. Thus, looking at fit statistics as those presented above, in the situations described as problematic, will confuse researchers in their model selection decisions.

Peugh and Fan (2015) conducted an extensive simulation study comparing a large variety of fit statistics and, following B. O. Muthén's (2003) recommendations, manipulated four design conditions: sample size, separation of latent class trajectories, membership proportions, and the amount of variance explained by covariates. The authors found that when sample size was small ($N = 500$), the indices considered were more likely to favor incorrect models. In the case of large sample size ($N = 3,000$), enumeration indices could identify the correct number of classes only when the latent class trajectories were well separated. Moreover, the inclusion of covariates helped only when sample size was large.

The recommendation of graphically inspecting the skewness and kurtosis of the data is appropriate (Bauer & Curran, 2003a; B. O. Muthén, 2003). While this is not a statistic, it provides insight about how reliable the solutions favored by the fit indices are. Another interesting recommendation is to inspect the latent class trajectories

against the observed data and the residuals to try to identify any of the three problematic situations described in Bauer and Curran (2004).

Goals of the Study

The goals of this study are (1) to explore the performance of the BIC, SBIC, and BLRT in GMM analysis with nonnormal distributed outcome variables and (2) to examine the effects of nonnormal TICs in the estimation of the number of latent classes when outcome variables are normally distributed. For both of these goals we will include nonnormal conditions not considered previously in the literature (e.g., Bauer & Curran, 2003a; Jung & Wickrama, 2008; Nylund et al., 2007; Tofighi & Enders, 2008).

In order to achieve these goals, two simulation studies were conducted. The first addresses the problem of spurious classes emerging as artifacts of the nonnormality of the dependent variables. The second simulation study explores the effects of nonnormal covariates in the estimation of the correct number of latent classes. The results from these simulations show the performance of these fit statistics and likelihood ratio tests and their usefulness as the correct number of classes' enumerators. Moreover, these findings could help applied researchers avoid conditions in which spurious classes could be selected in GMM analysis with real data and guide them in the interpretation of results.

Simulation Study I

The first simulation study builds on those performed by Bauer and Curran (2003a, 2004) and includes a more exhaustive exploration of how much departure from normality is too much. One of the goals of this article is to explore values of skewness and kurtosis not explored by the authors. Moreover, test statistics such as VLMR-LRT and LMR-adjusted LRT (Lo et al., 2001) are included in the analysis as suggested by B. O. Muthén (2003) and Rindskopf (2003). VLMR-LRT and LMR-adjusted LRT have been shown to perform poorly even in cases with normally distributed outcome variables (Jeffries, 2003; Jung & Wickrama, 2008; Nylund et al., 2007; Tofighi & Enders, 2008). Moreover, BLRT (McLachlan & Peel, 2000) was also included and the performance of this likelihood ratio test was explored in nonnormality conditions of outcome variables and covariates not previously presented in the literature (Jung & Wickrama, 2008). The results of this simulation will provide a more exhaustive comparison of the efficiency of BIC, SBIC, and BLRT under diverse, but often found in practice, nonnormality of outcome and covariates. Also, they could be a useful guide to the researcher interested in fitting finite mixture models to real-world data as it could provide more examples of situations in which spurious classes may emerge.

The main hypothesis of this simulation study is that for nonnormal data fit statistics and likelihood ratio tests will favor models with larger number of latent classes

than the ''true'' model. The reasoning behind this is that a normal one-class model would be a poor representation of the data and the finite mixture methodology will overcompensate for this lack of fit by fitting additional normal classes until the data are more closely approximated.

Simulation Design

The data for this simulation were generated using the five-occasion linear growth model described by Bauer and Curran (2003a). In this simulation, only a single ''true'' population existed in the data. The population mean trajectory was parametrized so that the average score would increase over time ($\mu_\alpha = 1.00$ and $\mu_\beta = 0.80$). The variance components were specified to allow individual variability in both intercepts and slopes ($\Psi_\alpha = 1.00$ and $\Psi_\beta = 0.20$). Moreover, intercepts and slopes were positively correlated ($\Psi_{\alpha\beta} = 0.11$). Also, the error variances for the dependent variables were specified to be increasing over time ($\Theta_\varepsilon = \text{diag}[1.00; 1.42; 2.25; 3.47; 5.09]$).

The distributional conditions are given by all possible combinations three values of skew (0, 1, and 1.6) and three values of kurtosis (0, 2, and 4). Five hundred samples were generated for each of the nine distributional conditions and three sample sizes, $N = 200$, $N = 400$, and $N = 800$. Thus, 27 sets of 500 samples were generated. The normal samples (i.e., skew 0, kurtosis 0) were generated and analyzed with Mplus 7.4. The remaining nonnormal conditions involved transformation of the normal distribution to the desired values of skew and kurtosis using the Fleishman (1978) method with the Vale and Maurelli (1983) extension as implemented in EQS 6.2 (Bentler, 1995). The values of kurtosis and skew used in this simulation, as well as those in Bauer and Curran (2003a), are often encountered in applied research and represent minor departures from normality that would not be of concern to the researcher (Micceri, 1989).

To test the hypotheses, we fit one- and two-class models to the data. The analysis was carried out in Mplus 7.4, specifying the EM algorithm with the MLR option for robust standard errors maximum likelihood estimation (L. K. Muthén & Muthén, 1998). Single population parameters were used as starting values for the EM algorithm, following the recommendations of L. K. Muthén and Muthén (1998) to avoid the local solutions. In addition, 100 random starts were employed to avoid local optimal solutions and nonconvergence issues (Hipp & Bauer, 2006; Li et al., 2014; Liu & Hancock, 2014). These starting values were applied to most parameters, with the exception of the growth factor means that were specified to have a high group and a low group ($\hat{\mu}_\alpha = 0.00$ and $\hat{\mu}_\beta = 0.00$ for Class 1, and $\hat{\mu}_\alpha = 1.50$ and $\hat{\mu}_\beta = 1.60$ for Class 2). This use of starting values is consistent with other simulation studies on finite normal mixtures (e.g., Bauer & Curran, 2003a; Biernacki et al., 1999; McLachlan & Peel, 2000).

The model was allowed 1,000 iterations to converge. Convergence patterns for growth mixture models were exhaustively explored by Bauer and Curran (2003a) and

are not of interest in the present research. Nevertheless, as in the research above, analyses were conducted both including and excluding nonconvergent solutions, and the results did not deviate in a meaningful way from those reported here.

The Effect of Nonnormality in the Decision of the Number of Latent Classes to Extract

The first goal of this study was to further investigate the conditions of nonnormality in which spurious classes emerge in finite mixture models when a one-class model was used to generate the data. Bauer and Curran (2003a) found that solutions with spurious classes are identified as optimal for even small departures of normality (i.e., skew 1 and kurtosis 1). Comparative fit for one- and two-class solutions were examined using AIC, BIC, and sample-corrected BIC. The hypothesis of interest is that fit statistics would favor the two-class solution, on average, for nonnormal data. Moreover, it was hypothesized that this preference would be greater the more the data departed from normality. Test statistics not used by Bauer and Curran (2003a) were included in the analysis: VLMR-LRT, LMR-adjusted LRT, and BLRT. Since these tests are based on the likelihood, it was expected that they would perform well for normal data but would have the same problems as fit statistics when the data are nonnormal.

   Tables 1 to 3 present the results of this simulation study with regard to fit statistics. Table 4 present the results for the likelihood ratio test comparing the one-class solution to the two-class solution. The fit statistics were on average higher for the one-class solution for normal data across all three sample sizes, as can be seen in Tables 1 to 3. The mean difference was calculated by subtracting the fit statistic of the two-class solution to that of the one-class solution. Thus, a negative value indicates that the statistic is better (lower) for the one-class solution. These results are similar to those reported by Bauer and Curran (2003a) and support the first part of the hypotheses.

   Table 1 shows that with small sample size ($N = 200$), BIC performs on average much better than AIC and SBIC in the normal condition, the no skewness and a kurtosis value of 4 and no kurtosis and a skewness value of 1. The reader can appreciate this by the negative mean difference and mean percentage change. These results corroborate the findings of Nylund et al. (2007), Jung and Wickrama (2008), and Tofighi and Enders (2008). Nevertheless, the BIC did not favor the GMM solution with the ''true'' number of classes. AIC and SBIC performed poorly in all nonnormal scenarios. These results provide empirical evidence in favor of the hypotheses. Once skew and kurtosis are introduced, the one-class, normal growth mixture model is no longer a good representation of the data and a spurious class is necessary to improve the fit. It is important to note that for sample sizes of 400 and 800 the BIC performed as poorly as the SBIC and the AIC, favoring the two-class solution. These results indicate that the BIC is not as useful to identify the correct number of latent

Table 1. Relative Fit of One-Class Versus Two-Class Models (of 500 Samples, N = 200).

Fit statistic	Mean difference[a]	Mean % change[a]	Class membership
Skewness 0, kurtosis 0			
AIC	−0.75	−0.02	
BIC	−10.65	−0.26	
SBIC	−1.14	−0.03	
Skewness 0, kurtosis 2			
AIC	14.45	0.35	Class 1 = 28 (14.10)
BIC	4.55	0.11	Class 2 = 172 (85.90)
SBIC	14.06	0.34	
Skewness 0, kurtosis 4			
AIC	14.52	0.36	Class 1 = 33 (16.65)
BIC	−11.87	−0.29	Class 2 = 167 (83.35)
SBIC	13.48	0.33	
Skewness 1, kurtosis 0			
AIC	16.93	0.41	Class 1 = 53 (26.65)
BIC	−9.45	−0.23	Class 2 = 147 (73.40)
SBIC	15.89	0.39	
Skewness 1, kurtosis 2			
AIC	17.98	0.44	Class 1 = 61 (30.26)
BIC	8.08	0.20	Class 2 = 139 (69.74)
SBIC	17.59	0.43	
Skewness 1, kurtosis 4			
AIC	21.42	0.53	Class 1 = 56 (28.03)
BIC	11.52	0.28	Class 2 = 144 (71.97)
SBIC	21.03	0.51	
Skewness 1.6, kurtosis 0			
AIC	49.13	1.20	Class 1 = 29 (14.50)
BIC	39.23	0.95	Class 2 = 171 (85.50)
SBIC	7.55	0.18	
Skewness 1.6, kurtosis 2			
AIC	31.55	0.77	Class 1 = 20 (10.20)
BIC	21.66	0.52	Class 2 = 180 (89.80)
SBIC	31.16	0.76	
Skewness 1.6, kurtosis 4			
AIC	31.52	0.77	Class 1 = 15 (7.57)
BIC	21.63	0.52	Class 2 = 180 (92.43)
SBIC	31.13	0.76	

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; SBIC = sample-corrected BIC.
[a]Mean difference = Fit1 − Fit2, and percentage change = (1 − Fit2/Fit1) × 100.

components as suggested by Nylund et al. (2007), Jung and Wickrama (2008), and Tofighi and Enders (2008) in moderate sample size nonnormal populations.

The column labeled *Class Membership* in Tables 1 to 3 show the classification of individuals according to the most likely class membership across replications. The percentage of individuals classified in Class 1 (no growth class) ranges from 10.20%

Table 2. Relative Fit of One-Class Versus Two-Class Models (of 500 Samples, N = 400).

Fit statistic	Mean difference[a]	Mean % change[a]	Class membership
Skewness 0, kurtosis 0			
AIC	−0.94	−0.01	
BIC	−12.91	−0.16	
SBIC	−3.40	−0.04	
Skewness 0, kurtosis 2			
AIC	25.18	0.31	Class 1 = 48 (12.08)
BIC	13.01	0.16	Class 2 = 352 (85.01)
SBIC	22.73	0.28	
Skewness 0, kurtosis 4			
AIC	33.66	0.41	Class 1 = 60 (14.99)
BIC	21.68	0.26	Class 2 = 340 (85.01)
SBIC	31.20	0.38	
Skewness 1, kurtosis 0			
AIC	37.83	0.46	Class 1 = 96 (24.00)
BIC	25.86	0.31	Class 2 = 304 (76.00)
SBIC	35.38	0.43	
Skewness 1, kurtosis 2			
AIC	40.94	0.50	Class 1 = 133 (33.30)
BIC	28.96	0.35	Class 2 = 267 (66.30)
SBIC	38.48	0.47	
Skewness 1, kurtosis 4			
AIC	49.93	0.61	Class 1 = 129 (32.21)
BIC	37.96	0.46	Class 2 = 271 (67.79)
SBIC	47.48	0.58	
Skewness 1.6, kurtosis 0			
AIC	104.37	1.27	Class 1 = 109 (27.36)
BIC	92.39	1.12	Class 2 = 291 (72.64)
SBIC	60.66	0.74	
Skewness 1.6, kurtosis 2			
AIC	72.46	0.89	Class 1 = 77 (19.15)
BIC	60.48	0.73	Class 2 = 323 (80.85)
SBIC	70.00	0.85	
Skewness 1.6, kurtosis 4			
AIC	71.95	0.88	Class 1 = 97 (24.26)
BIC	59.97	0.73	Class 2 = 303 (75.74)
SBIC	69.49	0.85	

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; SBIC = sample-corrected BIC.
[a]Mean difference = Fit1 − Fit2, and percentage change = (1 − Fit2/Fit1) × 100.

to 30.26% for *N* = 200, from 12.08% to 33.20% for *N* = 400, and from 7.85% to 34.18% for *N* = 800. This means that for some conditions of nonnormality about a third of the individuals were classified as having no growth even though a linear growth was present in the ''true'' population. The applied researcher using real data with similar conditions of nonnormality would be tempted to find substantively

Table 3. Relative Fit of One-Class Versus Two-Class Models (of 500 Samples, N = 800).

Fit statistic	Mean difference[a]	Mean % change[a]	Class membership
Skewness 0, kurtosis 0			
AIC	−1.04	−0.01	
BIC	−15.09	−0.04	
SBIC	−5.57	−0.03	
Skewness 0, kurtosis 2			
AIC	42.67	0.26	Class 1 = 63 (7.85)
BIC	28.61	0.17	Class 2 = 737 (92.15)
SBIC	38.14	0.23	
Skewness 0, kurtosis 4			
AIC	57.58	0.35	Class 1 = 76 (9.50)
BIC	43.53	0.27	Class 2 = 724 (90.50)
SBIC	53.05	0.32	
Skewness 1, kurtosis 0			
AIC	71.41	0.44	Class 1 = 149 (18.62)
BIC	57.35	0.35	Class 2 = 551 (81.38)
SBIC	66.88	0.41	
Skewness 1, kurtosis 2			
AIC	80.93	0.49	Class 1 = 273 (34.18)
BIC	66.87	0.41	Class 2 = 527 (65.82)
SBIC	76.40	0.47	
Skewness 1, kurtosis 4			
AIC	92.15	0.56	Class 1 = 232 (29.03)
BIC	78.09	0.48	Class 2 = 568 (70.97)
SBIC	87.62	0.53	
Skewness 1.6, kurtosis 0			
AIC	199.44	1.22	Class 1 = 264 (32.97)
BIC	185.38	1.13	Class 2 = 536 (67.04)
SBIC	194.91	1.19	
Skewness 1.6, kurtosis 2			
AIC	137.24	0.84	Class 1 = 154 (19.28)
BIC	123.19	0.75	Class 2 = 646 (80.72)
SBIC	132.72	0.81	
Skewness 1.6, kurtosis 4			
AIC	134.93	0.82	Class 1 = 216 (27.01)
BIC	120.88	0.74	Class 2 = 584 (72.99)
SBIC	130.40	0.80	

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; SBIC = sample-corrected BIC.
[a]Mean difference = Fit1 − Fit2, and percentage change = (1 − Fit2/Fit1) × 100.

meaningful explanations to these seemingly different developmental trajectories. The results presented here provide empirical evidence that the additional classes are an artifact of the finite mixture trying to approximate the nonnormality of the data with a mixture of normally distributed classes.

Table 4. Likelihood Ratio Tests of One-Class Versus Two-Class Models (of 500 Samples).

Distribution	VLMR-LRT	LMR-Adjusted LRT	BLRT
N = 200			
Skewness 0, kurtosis 0	67 (0.13)	64 (0.13)	25 (0.05)
Skewness 0, kurtosis 2	123 (0.25)	119 (0.24)	62 (0.12)
Skewness 0, kurtosis 4	138 (0.28)	132 (0.26)	72 (0.14)
Skewness 1, kurtosis 0	97 (0.19)	88 (0.18)	32 (0.06)
Skewness 1.6, kurtosis 0	118 (0.24)	109 (0.22)	35 (0.07)
N = 400			
Skewness 0, kurtosis 0	92 (0.18)	85 (0.17)	38 (0.08)
Skewness 0, kurtosis 2	169 (0.34)	158 (0.32)	57 (0.11)
Skewness 0, kurtosis 4	173 (0.35)	167 (0.33)	72 (0.14)
Skewness 1, kurtosis 0	148 (0.30)	137 (0.27)	45 (0.09)
Skewness 1.6, kurtosis 0	190 (0.38)	183 (0.37)	29 (0.06)
N = 800			
Skewness 0, kurtosis 0	89 (0.18)	82 (0.16)	25 (0.05)
Skewness 0, kurtosis 2	202 (0.40)	198 (0.40)	78 (0.16)
Skewness 0, kurtosis 4	245 (0.49)	238 (0.48)	127 (0.25)
Skewness 1, kurtosis 0	199 (0.40)	187 (0.37)	52 (0.10)
Skewness 1.6, kurtosis 0	343 (0.69)	338 (0.68)	153 (0.31)

Note. VLMR-LRT = frequency with which the Voung–Lo–Mendell–Rubin likelihood ratio test for 1 class versus 2 classes rejected the "true" one-class solution (Type I error rates across replications). LMR-adjusted LRT = frequency with which the Lo–Mendell–Rubin adjusted likelihood ratio test for 1 class versus 2 classes rejected the "true" one-class solution (Type I error rates across replications). BLRT = frequency with which the bootstrap likelihood ratio test rejected the "true" one-class solution (Type I error rates across replications).

Results of the likelihood ratio tests are presented in Table 4. All likelihood ratio tests worked better with normal data, but when nonnormality is introduced their performance becomes poorer. BLRT performed the best for this sample size with inappropriate results in just 5% to 31% of the replications. The best performance corresponded to normal conditions for sample sizes of 200 and 400, and the worse corresponded to sample size of 800, no kurtosis and skewness of 1.6. The second worse performance of the BLRT was for the sample size 800, skewness of zero and kurtosis of 4. The results for sample sizes 200 and 400 were practically the same for nonnormal conditions. Nevertheless, as sample size increased (*N* = 800) BLRT was more susceptible to departures from normality on the outcome variables. These results are in contrast with those of Jung and Wickrama (2008). VLMR-LRT and LMR-adjusted LRT perform poorly even when outcome variables are generated to be normal. These statistics failed to reject the true one-class solution from 13% to 69% of the replications.

The results in parentheses in Table 4 are Type I error rates or false positive rates. In Tables 1 to 3, we can see that fit statistics are on average smaller for one-class solutions than for two-class solutions for normal data. Moreover, Bauer and Curran

(2003a) found that fit statistics, such as BIC, favor the two-class solution 0% to 0.66% for $N = 200$ and $N = 600$ for normal data, respectively. These results are an improvement with respect to the average of times that fit statistics favor two-class models as reported by Bauer and Curran (2003a). Nevertheless, the Type I error rates corresponding to VLMR-LRT and LMR-adjusted LRT are 7 to 11 times larger than the conventional 5%. To illustrate the problem that this represents, consider the condition of nonnormality with a skewness value of 1.6, no kurtosis, and sample size of 800 in Table 4. Researchers would select the correct number of latent classes more often if they flip a fair coin to decide whether to select a one-class or a two-class model to their data than using these likelihood ratio tests. The argument can be made that a sample size of 800 is not large enough and that these likelihood ratio tests could perform better with larger statistical power. This is a limitation of this part of Simulation Study 1 and represents an opportunity of future research on the topic.

The best performance of the BLRT was a 5% error rate, as mentioned before. Considering that we would like Type I errors smaller than 5%, these results are marginally acceptable. The results for nonnormal outcome variables range from 6% to 31%, which range from barely larger up to six times greater than what any researcher would feel comfortable with. Nylund et al. (2007) point out that BLRT does not perform well when nonnormality is present in the data. This argument does not offer much consolation since normal data are improbable in real-life applications of GMM analysis (Micceri, 1989), and since all other fit statistics and likelihood ratio test perform significantly worse that the BLRT.

The results of Tables 1 to 4 will help researchers identify the risk of making Type I error when fitting normal growth mixture models to their nonnormal data. Specifically, these results inform about the dangers of using fit statistics and likelihood ratio tests to draw conclusions about the number of latent classes in the data. This is a particularly important problem in social sciences since extracting multiple classes often seems more interesting and allows researcher to assign substantive meaning to why the groups are different. For example, a researcher could hypothesize that the groups represent ''normal'' versus antisocial subgroups of children, or groups of individuals at different stages of alcohol dependence. The results presented in this simulation study suggest that the researcher should be skeptical about the extraction of multiple latent classes in growth mixture models when the data are nonnormal and when using fit statistics and likelihood ratio tests to draw conclusions.

The need persists for fit indices and test statistics that reliably help the researcher in choosing the ''true'' number of latent classes. The evidence suggests that such indices or statistics should not be based on the likelihood since it would carry many of the same limitations as those exposed here. Further research is needed on the development of more useful statistics for the estimation of the number of mixture components for nonnormal data. The problem is that the normal approach to GMM is not appropriate for most real data in the social sciences. Other distributions might be more appropriate such as zero inflated Poisson model for substance abuse or categorical distributions for Likert scale survey data. Further research is needed to empirically

show the robustness of fit statistics and likelihood ratio tests to departures of the data to any distribution used to derive the finite mixture model.

Simulation Study 2

This simulation explores the effect of nonnormally distributed TICs on growth mixture models. Arminger et al. (1999) showed analytically that the addition of TICs that depart from normality would aggravate the problem of the identification of spurious latent classes. The hypothesis in this simulation study is that the inclusion of nonnormal TICs would make it harder for fit statistics and likelihood ratio tests to identify the correct number of latent classes even when normally distributed outcome variables are present in the data. The results of this simulation study constitute empirical evidence that support the findings of Arminger et al. (1999). Moreover, it adds new nonnormality conditions to the exploration of the efficiency of BLRT and other likelihood ratio tests not explored in the literature (e.g., Bauer & Curran, 2003a; Jung & Wickrama, 2008; Nylund et al., 2007; Tofighi & Enders, 2008).

Simulation Design

The data for this simulation were generated using the same model specifications as for Simulation Study 1 with the addition of one and two TICs. The distributional conditions of the dependent variables were the following: normal (skew 0, kurtosis 0), positive kurtosis (skew 0, kurtosis 2), and positive skew (skew 1.6, kurtosis 0). One- and two-covariate conditional models were fitted to each distributional condition. The covariates were generated to have a 0.30 correlation to the slope factor (β), mean of zero, and standard deviation of one, and to be correlated with each other at 0.10. The TICs were transformed to have skew and kurtosis values of one. Five hundred samples of $N = 200$ were generated, transformed using EQS 6.2, and analyzed using Mplus 7.4. As for the previous simulation, only one class exists in the data.

Conditional Finite Mixture Model and the Identification of Spurious Latent Classes

The results presented in Tables 5 and 6 provide empirical evidence for the conclusions of Arminger et al. (1999). The positive mean difference indicates that the two-class models yield smaller (better) fit statistics than the one-class model across all distributional conditions and number of covariates. As shown in Tables 1 to 3, fit statistics favor the "true" one-class model, on average. Nevertheless, with slightly nonnormal TICs the fit statistics suggest a better fit of the two-class solution, even in the case of normally distributed dependent variables. In the normal dependent variable conditions, for both one and two TICs conditional models, the fit statistics favor, on average, the two-class solution and classify 34.00% and 40.84% individuals in Latent Class 1 (no growth class), respectively. This means that even with normally

Table 5. Relative Fit of One-Class Versus Two-Class Conditional Models (of 500 Samples, N = 200).

Fit statistic	Mean difference[a]	Mean % change[a]	Class membership[b]
One covariate, skewness 0, kurtosis 0			
AIC	39.84	0.85	Class 1 = 68 (34)
BIC	26.64	0.56	Class 2 = 132 (66)
SBIC	39.32	0.83	
One covariate, skewness 0, kurtosis 2			
AIC	35.62	0.76	Class 1 = 68 (34.00)
BIC	22.42	0.17	Class 2 = 132 (65.00)
SBIC	35.10	0.23	
One covariate, skewness 1.6, kurtosis 0			
AIC	51.81	1.10	Class 1 = 116 (58.18)
BIC	38.61	0.82	Class 2 = 84 (41.82)
SBIC	51.28	1.09	
Two covariates, skewness 0, kurtosis 0			
AIC	56.39	1.06	Class 1 = 82 (40.84)
BIC	39.90	0.74	Class 2 = 118 (59.16)
SBIC	55.74	1.05	
Two covariates, skewness 0, kurtosis 2			
AIC	45.50	0.86	Class 1 = 83 (41.35)
BIC	29.01	0.55	Class 2 = 117 (58.65)
SBIC	44.85	0.85	
Two covariates, skewness 1.6, kurtosis 0			
AIC	52.16	0.99	Class 1 = 115 (57.44)
BIC	35.66	0.67	Class 2 = 85 (42.56)
SBIC	51.51	0.98	

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; SBIC = sample-corrected BIC.
[a]Mean difference = Fit1 − Fit2, and percentage change = (1 − Fit2/Fit1) × 100.

Table 6. Likelihood Ratio Tests of One-Class Versus Two-Class Conditional Models When the Nonnormality is Only Present in the Covariate (of 500 Samples, N = 200, and 1 Covariate).

Distribution	VLMR-LRT	LMR-adjusted LRT	BLRT
Skewness 0, kurtosis 0	58 (0.12)	52 (0.10)	25 (0.05)
Skewness 0, kurtosis 2	155 (0.31)	148 (0.30)	76 (0.15)
Skewness 1.6, kurtosis 0	245 (0.49)	243 (0.49)	457 (0.91)

Note. VLMR-LRT = frequency with which the Voung–Lo–Mendell–Rubin likelihood ratio test for 1 class versus 2 classes rejected the "true" one-class solution (Type I error rates across replications); LMR-adjusted LRT = frequency with which the Lo–Mendell–Rubin adjusted likelihood ratio test for 1 class versus 2 classes rejected the "true" one-class solution (Type I error rates across replications); BLRT = frequency with which the bootstrap likelihood ratio test rejected the "true" one-class solution (Type I error rates across replications).

distributed dependent variables the slight positive skewness of the TICs introduced enough nonnormality into the data that over a third of the individuals in the samples were classified in the no-growth class. This misclassification was aggravated when the dependent repeated variables were also positively skewed.

The results for the positive kurtosis condition, for one covariate and the two covariates, were practically identical as shown in Table 5. In the positive skew conditions, over half the subjects were classified in Class 1 (no growth class). This means that applied researchers could conclude that the majority of individuals in their data are classified in a latent class of no growth when in fact a single group of linear growth is present in the data. It makes sense that about 60% of the individuals were classified in Class 1 since these data have more density near zero (skew 1.6) and a long right tail. These results are consistent with our hypothesis that nonnormality promotes the identification of spurious classes when using statistics based on the likelihood. Applied researchers can use Table 5 as a guide as to when to expect that fit statistics will suggest models with more latent classes than those that exist in the data and the percentage of individuals that would be classified in these spurious classes due to nonnormal TICs.

Table 6 shows the results for likelihood ratio tests for three conditions: normal; no skewness and kurtosis of 2 of the TIC; and no kurtosis, skewness of 1.6 of the TIC. As expected, the addition of a normal TIC produces identical results than in Table 4. All likelihood ratio tests performed slightly worse compared to the results of Table 4 when the TIC had no skewness and a kurtosis value of 2, and performed considerable worse in the no kurtosis and skewness of 1.6 condition. BLRT was affected the most in this last conditions, yielding a 91% false positive rate.

As noted by Bauer and Curran (2003a), nonnormality is a necessary and sufficient condition for the fit of more latent classes than the model used to simulate the data. When the data are nonnormal, the normal growth mixture model is not a good approximation and the finite mixture methodology overcompensates by fitting additional normal growth curves to the data. The two-class solution is a better approximation to the nonnormal data; thus, fit indices and test statistics based on the likelihood tend to favor it over the true one-class model.

Discussion

Fit statistics and likelihood ratio tests favor the ''true'' model when the data are normal. Nevertheless, when the data are nonnormal, fit indices tend to favor growth mixture models with more mixture components (latent classes) than the model used to simulate the data. Likelihood ratio tests also favor models with spurious classes. Further research is needed to develop test statistics that do not depend on the likelihood and that are robust to departures from normality in the population. This includes the case of normal dependent variables but nonnormal TICs. An interesting alternative to normal GMM is using distributions other than normal in fitting growth mixture models. This approach has been readily incorporated in statistical software such

as Mplus. Nevertheless, empirical studies such as the current one have not been conducted to show if those approaches have the same problems as the normal growth mixture models. It is important to note that using GMM with distributions other than the normal would not have been beneficial in the situations presented in the above simulations studies: situations in which the underlying distribution is almost normal. The reason for this is that the applied researcher would be unable to know firsthand if the underlying distribution is, for example, exponential or just positively skewed.

The results of the simulation studies presented above could be a useful guide to applied researchers in fitting growth mixture models to their data. Applied researchers may use Tables 1 to 6 to anticipate when the distributional conditions of their data would yield the extraction of more latent classes than exist in the data. This represents a major contribution to longitudinal data analysis within social sciences and could potentially provide explanation as to why some results obtained through the GMM cannot be replicated and others are ubiquitous. Given the nonnormal nature of data in the social sciences, researchers should be cautious when fitting growth mixture models and incorporate exhaustive data visualization before deciding on the number of latent classes. Also important is that the findings discussed here expose the need of better statistics for the estimation of the number of latent classes in GMM.

Some limitations of Simulation Study 2 are that the only conditions considered were sample size of 200, nonnormal TICs, and these covariates were correlated at 0.3 with the slope factor. Further research is needed that include time varying covariates, other sample sizes, and distributional conditions not explored here. Also interesting would be to explore if the linear relationship of the covariates and the growth factor has an effect on the amount of latent classes extracted.

Declaration of Conflicting Interests

Funding

References

Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors: Log likelihood functions, minimum distance estimation, fit indices, and a complex example. *Sociological Methods & Research*, *26*, 148-182.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, *64*, 475-494.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338-363. doi:10.1037/1082-989X.8.3.338

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henley (2003). *Psychological Methods*, 8, 384-393. doi:10.1037/1082-989X.8.3.384

Bauer, D. J., & Curran P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9, 3-29. doi:10.1037/1082-989X.9.1.3

Bentler, P. M. (1995). *EQS: Structural equations program manual* (Version 5.0) [Computer software manual]. Los Angeles, CA: BMDP Statistical Software.

Biernacki, C., Celeux, G., & Govaert, G. (1998). Assessing a mixture model for clustering with the integrated classification likelihood. Technical Report No. 3521.Rhone-Alpes: INRIA.

Biernacki, C., Celeux, G., & Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. Pattern Recognition Letters, 20, 267-272.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley.

Campbell, J. G., Fraley, C., Murtagh, F., & Raftery, A. E. (1997). Linear flaw detection in woven textiles using model-based clustering. *Pattern Recognition Letters*, *18*, 1539-1548.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Classification Journal*, *13*, 195-212.

Dasgupta, A., & Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, *93*, 294-302.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*, 202-226.

Hipp, J. R., & Bauer, J. B. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*, 36-53.

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.

Jeffries, N. (2003). A note on ''Testing the number of components in a normal mixture.'' *Biometrika*, *90*, 991-994.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302-317.

Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *Annuals of Statistics*, *20*, 1350-1360.

Li, M., Harring, J. R., & Macready, G. B. (2014). Investigating the feasibility of using Mplus in the estimation of growth mixture models. *Journal of Modern Applied Statistical Methods*, *13*, 484-513.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, *74*, 557-584.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767-778.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: John Wiley.

Meredith, W., & Tisak, J. (1984). *On ''Tuckerizing'' curves*. Presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*, 674-701.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 369-377.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463-469.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (Version 2) [Computer software manual]. Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (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.

Peugh, J., & Fan, X. (2015). Enumeration index performance in generalized growth mixture models: A Monte Carlo test of Muthén's (2003) hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 115-131.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*, 1-14.

Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 364-368.

Roeder, K., & Wasserman, L. (1997). Practical density estimation using mixtures of normals. *Journal of the American Statistical Association*, *92*, 894-902.

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology*, *120*, 322-335.

Soromenho, G. (1993). Comparing approaches for testing the number of components in a finite mixture model. *Computational Statistics*, *9*, 65-78.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent mixture models* (pp. 317-341). Charlotte, NC: Information Age.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., . . . McGuffin, P. (2009). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine*, *16*, 1-11. doi:10.1017/s0033291709991528

Vale, C. D., & Maurelli, V. A. (1993). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471.