A Comparison of Different Nonnormal Distributions in Growth Mixture Models

Educational and Psychological
Measurement

2019, Vol. 79(3) 577–597

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0013164418823865

journals.sagepub.com/home/epm

**Sookyoung Son¹, Hyunjung Lee¹,
Yoona Jang¹, Junyeong Yang¹ and Sehee Hong¹**

Abstract

The purpose of the present study is to compare nonnormal distributions (i.e., *t*, skew-normal, skew-*t* with equal skew and skew-*t* with unequal skew) in growth mixture models (GMMs) based on diverse conditions of a number of time points, sample sizes, and skewness for intercepts. To carry out this research, two simulation studies were conducted with two different models: an unconditional GMM and a GMM with a continuous distal outcome variable. For the simulation, data were generated under the conditions of a different number of time points (4, 8), sample size (300, 800, 1,500), and skewness for intercept (1.2, 2, 4). Results demonstrate that it is not appropriate to fit nonnormal data to normal, *t*, or skew-normal distributions other than the skew-*t* distribution. It was also found that if there is skewness over time, it is necessary to model skewness in the slope as well.

Keywords

nonnormal distribution, nonnormality, skew-*t* distribution, skew-normal distribution, growth mixture models

Introduction

Growth mixture models (GMMs) are generally applied to figure out if subgroups with qualitatively distinct growth trajectories exist within population (Bauer & Curran, 2003; Muthén, 2004). GMMs allow for heterogeneity in growth trajectories and accommodate for this heterogeneity by including latent classes (*C*) in addition to

¹Korea University, Seoul, Republic of Korea

Corresponding Author:

Sehee Hong, Department of Education, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea.

Email: seeehong@korea.ac.kr

continuous factors of a latent growth curve model (e.g., intercept and slope). Through this inclusion, a GMM fully captures both interindividual variation (i.e., within-class variation) and between-class variation (i.e., unobserved heterogeneity) (Jung & Wickrama, 2008). This variation requires a distributional assumption, which in general assumes a normal distribution. That is, a GMM assumes that the mixture of within-class normal distributions allows for the existence of within-class normality (Bauer & Curran, 2003; Feldman, Masyn, & Conger, 2009).

However, the distribution of nonnormality is very common in real data, particularly in certain areas such as juvenile delinquency, antisocial behavior, and diseases such as depression and attention deficit hyperactivity disorder. For example, body mass index in obesity studies (see Lee & McLachlan, 2014; Muthén & Asparouhov, 2014) and prostate-specific antigen scores in prostate cancer studies (see H. Lin, McCulloch, Turnbull, Slate, & Clark, 2000) have a long right tail. Youth alcohol use and marijuana use also have positively skewed distributions (D'Amico et al., 2016). Many researchers studying mixture modeling or structural equation modeling analyze models fitted to normal distribution with no treatment for nonnormal variables (e.g., Laird, Criss, Pettit, Dodge, & Bates, 2008). Also common are studies that convert nonnormal data into a normal distribution through recoding (e.g., Miller, Malone, & Dodge, 2010) or log transformation (e.g., Boers, Reinecke, Seddig, & Mariotti, 2010). If the nonnormality is not severe, it is common to employ a standard GMM model and correct only the standard error with robust estimates through robust maximum likelihood (MLR), consisting of ML estimation with robust standard errors. These strategies are undesirable for a number of reasons. First, recoding or transforming the data to create normal distribution causes inaccurate estimates due to a loss of data information (Manikandan, 2010). In addition, it is important to note that although some researchers believe that robust estimates are sufficient, MLR estimates the same parameters as ML, and only the standard errors are corrected. As a result, MLR still fails to fully take into account the information (i.e., skewness and kurtosis) in the data (Asparouhov & Muthén, 2016).

Above all, when normality assumptions within a class are violated, only spurious classes that are nonmeaningful latent subpopulations hold the heavy tails (Asparouhov & Muthén, 2016; Bauer & Curran, 2003; Bauer & Curran, 2004; Guerra-Peña & Steinley, 2016). However, when analyzing data showing nonnormality, it is necessary to consider appropriate assumptions regarding the distribution of the data. Among nonnormal distributions, skew- t distribution has attracted much attention. Skew- t distribution considers not only mean and covariance but also skewness and kurtosis (Azzalini & Capitanio, 2003; Gupta, 2003; T. I. Lin, Lee, & Hsieh, 2007). The multivariate skew- t distribution is defined as having a more flexible family of distributions with shape parameters representing skewness and kurtosis, in addition to having the properties of a multivariate t distribution (Gupta, 2003). The latent classes found by fitting the mixture model of the skew- t distribution represent a more meaningful subgroup. A specific example is presented in Muthén and Asparouhov's (2014) article that analyzed GMM using body mass index data with a

skewness of 1.5 and a kurtosis of 3.1. Comparisons between the number of classes with the skew- t distribution and the normal distribution revealed that one class is more appropriate in a skew- t distributed GMM, while three classes are appropriate in a normal distribution GMM. That is, although one class is sufficient in this example, additional latent classes are needed only to capture distributions for the observed variables. Therefore, when analyzing mixture models with nonnormal data, it is necessary to apply the skew- t distribution, which can reflect the properties of data.

Nonnormal distributions include three specific distributions (Wickrama, Lee, O'Neal, & Lorenz, 2016, p. 209): t distribution to account for excessive kurtosis, which indicates a heavy tail (McLachlan & Peel, 1998, 2000), skew-normal distribution to account for excessive skewness (Azzalini, 1985; Azzalini & Valle, 1996), and skew- t distribution to account for both distributions with excessive skewness and kurtosis (Lee & McLachlan, 2014). Accordingly, skew- t distribution is considered the most general distribution because normal, t , and skew-normal distribution correspond to a special form of skew- t distribution (Asparouhov & Muthén, 2016). The theoretical relevance of each distribution is presented in the next section.

Recent papers have used the skew- t distribution for structural equation models and mixture models when the data were nonnormal (Lee & McLachlan, 2014; T. Lin, Wu, McLachlan, & Lee, 2013). A few simulation studies on nonnormal distributions have also been published (Bauer & Curran, 2003; Guerra-Peña & Steinley, 2016; Muthén & Asparouhov, 2014). However, the number of such studies is still insufficient. Only one study has investigated parameter bias according to the assumption of distribution in a GMM (Muthén & Asparouhov, 2014), while a mere handful of other studies (i.e., Bauer & Curran, 2003; Guerra-Peña & Steinley, 2016) have focused on how to estimate the number of classes under the various conditions of nonnormality compared to a true model. Moreover, Muthén and Asparouhov's (2014) study involved a limited condition with only one true model. In other words, various factors that may influence the bias of the parameter estimates have not been considered as simulation conditions. Therefore, it is necessary to expand on previous simulation studies using diverse conditions that reflect empirical studies.

As such, the present study aims to provide a practical guideline for applied researchers by comparing the performance of different nonnormal distributions when estimating a GMM in nonnormal data. Specifically, the simulation study compares how much bias occurs in a GMM specified with different distributions (normal, t , skew-normal, and skew- t distributions). Although the model of the skew- t distribution as the true model is expected to have the smallest bias of the parameter estimates, this study attempts to ascertain the extent of bias in the models of other nonnormal distributions. The skew- t distribution was examined by dividing it into two types: skew- t with equal skew (hereafter called skew- t equal) and skew- t with unequal skew (hereafter called skew- t unequal). The skew- t equal distribution means zero skews for the slope of the growth model, that is, equal skews for the outcomes over time. On the other hand, the skew- t unequal distribution allows a skew parameter for both the intercept and slope random effects. According to Asparouhov and

Muthén (2016), if a sample size is insufficient, the additional skew parameter will not be statistically significant in order to keep the parsimony and to minimize the standard error for the remaining parameters. Hence, the present study also tests whether a more parsimonious model fitted to skew- t equal or skew-normal distribution performs better than a model fitted to skew- t unequal distribution under small sample size conditions.

In the present study, two simulation models with realistic conditions were used. The two models were (1) an unconditional GMM and (2) a GMM with a continuous distal outcome variable. A conditional model with outcome variables is more widely used in applied research to explain differences in the class-specific means across classes (e.g., Janosz, Archambault, Morizot, & Pagani, 2008). This conditional model also requires normality assumption of the outcome variables within each latent class (Asparouhov & Muthén, 2014; Bakk, Tekle, & Vermunt, 2013; Lanza, Tan, & Bray, 2013). If the distal outcome does not hold normally distributed within classes, the result is expected to be affected by violations of the normality assumption (Asparouhov & Muthén, 2014). Accordingly, the GMM with distal outcome variable was explored in Study 2. The research question is how accurately and reliably these two models will estimate parameters, depending on the assumption of distribution when a GMM with nonnormal data is analyzed under various conditions including a number of time points, sample sizes, and skewness for intercept.

Multivariate Continuous Nonnormal Distributions

Fitting the skew- t distribution to the data draw additional information on skewness and kurtosis, in addition to extracting the mean and covariance by fitting the normal distribution (Asparouhov & Muthén, 2016). This task requires a more complex process than simply fitting additional parameters to the model. Since modeling skew parameters are not done separately from modeling variance covariance matrix (i.e., skew parameters are entangled with variance covariance matrix), fitting a skew- t distribution is more than just matching skewness and kurtosis (Muthén & Asparouhov, 2014).

Observed variable Y_{it} for individual i at time t in the latent class c of the whole class variable C expressed as

$$Y_{it}|_{C_i=c} = \eta_{0i} + \eta_{1i}(a_t - a_0) + \epsilon_{it}, \quad (1)$$

where a_t is a time-related variable centered by a_0 . Random intercepts and random slopes are expressed as

$$\eta_{ji}|_{C_i=c} = \alpha_{jc} + \zeta_{ji}, \quad (2)$$

where j is 0 or 1 and the probability is expressed as logit

$$P(C_i = c) = \frac{\exp(a_c)}{\sum_s \exp(a_s)}. \quad (3)$$

GMMs generally assume that ϵ_{it} and ζ_{ji} have zero mean and within-class covariance matrices (i.e., normal distribution). However, if the data are nonnormal, the skew- t distribution can be applied to ϵ_{it} or ζ_{ji} . In this study, it is assumed that ϵ_{it} holds a normal distribution and skew- t distribution is applied to $\zeta_{ji} = (\zeta_0, \zeta_1)$ as in Muthén and Asparouhov (2014). This is because skewness cannot be modeled over time if ϵ_{it} is assumed to be a skew- t distribution (Lu & Huang, 2014). According to Lee and McLachlan (2014), there are two types of skew- t distributions: one is restricted and the other is unrestricted. These two are not nested within each other and are equivalent when they are univariate. The restricted skew- t distribution is able to apply explicit ML estimation in structural equation modeling, while the unrestricted skew- t distribution uses the Monte Carlo method, which is not easily generalized to structural equation modeling. Therefore, Muthén and Asparouhov's (2014) simulation studies examining the efficiency of the skew- t distribution have employed a restricted skew- t distribution. A restricted skew- t distribution was also used in the present study. A restricted multivariate skew- t distribution for observed variable Y can be expressed as

$$Y \sim \text{rMST}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta}, \nu), \quad (4)$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\delta}$ and ν denote a vector of intercepts, a variance covariance matrix, a vector of skew parameters, and a degrees of freedom parameter, respectively. If Y is the P dimensional variable, $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ are $P \times 1$ vector and $\boldsymbol{\Sigma}$ is the $P \times P$ variance covariance matrix (Asparouhov & Muthén, 2016). For a single class model, the density function of Y is expressed as

$$2t_{p,\nu}(Y, \boldsymbol{\mu}, \boldsymbol{\Omega}) T_{1,\nu+p}(y_1/\lambda, 0, 1), \quad (5)$$

where

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}^T,$$

$$d(Y) = (Y - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (Y - \boldsymbol{\mu}),$$

$$q = \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} (Y - \boldsymbol{\mu}),$$

$$y_1 = q \sqrt{\frac{\nu+p}{\nu+d(Y)}},$$

$$\lambda^2 = 1 - \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\delta},$$

and $t_{p,\nu}(Y, \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the multivariate t distribution density function expressed as

$$t_{p,\nu}(Y, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \frac{\Gamma(\frac{\nu+p}{2}) |\boldsymbol{\Omega}|^{-1}}{(\pi\nu)^{p/2} \Gamma(\frac{\nu}{2}) [1 + d(Y)/\nu]^{(\nu+p)/2}}, \quad (6)$$

and $T_{1,n}(z, 0, 1)$ stands for the standard t distribution function of univariate with n degrees of freedom. The skew- t distribution accounts for both excessive skewness and kurtosis by adding skew and df parameters (stronger skew possible than skew-normal). If ν increase to the infinity (∞), the skew- t distribution simplified to the skew-normal distribution. The skew-normal distribution accounts for excessive skewness in the observed distribution by adding a skew parameter to each variable. If δ becomes zero, the skew- t distribution simplified to the multivariate t distribution. The t distribution accounts for excessive kurtosis by including a degree of freedom parameter (thicker or thinner tails). If ν increase to the infinity (∞) and δ becomes zero, the skew- t distribution simplified to the normal distribution (Asparouhov & Muthén, 2016).

Assuming that Y is a vector of P dimensions, the restricted multivariate skew- t distribution rMST (μ, Σ, δ, ν) has the stochastic representation

$$Y = \mu + \delta |U_0| + U_1, \quad (7)$$

where U_1 is a p -dimensional vector with a multivariate t distribution with zero mean, Σ for covariance matrix, and ν for degree of freedom parameter. Whereas, U_0 is a one-dimensional variable with a standard t distribution with mean of 0, variance of 1, and degree of freedom of ν , where $|U_0|$ results in a half t distribution. The term $\delta|U_0|$ can be considered as a univariate skewness factor by multiplying δ which represents the skewness parameter, as a factor loading (Muthén & Asparouhov, 2014)

For the skew- t distribution rMST (μ, Σ, δ, ν), the mean and variance of Y can be obtained as follows:

$$E(Y) = \mu + \delta \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}}, \quad (8)$$

$$\text{Var}(Y) = \frac{\nu}{\nu-2} \left(\Sigma + \delta \delta^T \right) - \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{\nu}{\pi} \delta \delta^T. \quad (9)$$

The univariate skewness for Y can be obtained as follows.

$$\text{Skew}(Y) = \nu^{-3/2} \delta \sqrt{\frac{\nu}{\pi}} \left(\left(2\delta^2 + 3\sigma \right) \frac{\nu}{\nu-2} \frac{\Gamma(\frac{\nu-3}{2})}{\Gamma(\frac{\nu-2}{2})} - \delta^2 \frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^3 - 3 \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \nu \right), \quad (10)$$

where σ parameter denotes the diagonal element of Σ for each univariate variable. These equations indicate that δ and ν affect all three parameters of mean, variance, and skewness (Muthén & Asparouhov, 2014).

The mean, variance, and skew of Y for the skew-normal distribution simplify to

$$E(Y) = \mu + \delta \sqrt{\frac{2}{\pi}}, \quad (11)$$

$$\text{Var}(Y) = \Sigma + \left(1 - \frac{2}{\pi}\right) \delta \delta^T. \quad (12)$$

$$\text{Skew}(Y) = \nu^{-3/2} \delta^3 \sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1\right). \quad (13)$$

The mean, variance, and skew of Y for the t distribution simplify to

$$E(Y) = \mu, \quad (14)$$

$$\text{Var}(Y) = \frac{\nu}{\nu - 2} \Sigma. \quad (15)$$

$$\text{Skew}(Y) = 0, \quad (16)$$

although the kurtosis of Y is not formulated explicitly, since the kurtosis of the T distribution is $6 / (\nu - 4)$, and thus any level of kurtosis can be modeled using only the skew- t distribution.

For latent class c , observations for individual i at time T , Y is expressed as follows:

$$Y \sim \text{rMST}(\mu_c, \Sigma_c, \delta_{Yc}, \nu_c), \quad (17)$$

where

$$\mu_c = \Lambda \alpha_c, \quad (18)$$

$$\Sigma_c = \Lambda \Psi_c \Lambda' + \Theta_c, \quad (19)$$

$$\delta_{Yc} = \Lambda \delta_c, \quad (20)$$

and where for the linear growth model of (1)

$$\Lambda = \begin{pmatrix} 1 & a_1 - a_0 \\ 1 & a_2 - a_0 \\ \vdots & \vdots \\ 1 & a_{T-1} - a_0 \end{pmatrix} \quad (21)$$

the elements of α_c are shown in the formula (2) above, and Θ_c indicates the within-class covariance matrix.

Simulation Study

Data Generation

Study 1 generated the unconditional baseline model of a GMM with two latent classes fitted with skew- t unequal distribution. The reason for setting only two latent classes is that mixture models are known to have poorly behaved likelihood

Figure 1. Unconditional GMM of Study 1.

Note. GMM = growth mixture model; C = latent class; I = intercept; S = slope.

functions, that lead to false consequences. For example, a large number of local solutions can lead to seriously misleading results or singularities at the edges of the parameter space can lead to the non-convergence problem (McLachlan & Peel, 2000). In Study 1, the proportion of individuals was specified to be 50% in each latent class, respectively, and set to 0 by converting to logit. The residual variances for the observations are 0.5 at all the times. The means of latent variables I and S were 4 and 1 in Class 1 and 0 and 0 in Class 2, respectively. The variances for I and S were 1 and 0.7, respectively, for both classes, and the covariance was zero for both classes. Although the skew parameters for I varied depending on the conditions, the skew parameter for S was 2 and the degree of freedom parameter was 5 that represent kurtosis is 6 for both classes. Study 2 generated conditional GMM with the same parameters for continuous distal outcome as Study 1. In addition, the mean of U was 0 in Class 1 and 0.3 in Class 2, and the variance was 1 in both classes. The skew parameter for the variable of U was set to 2 only in Class 2. Figures 1 and 2 are simulation models for data generation in Study 1 and study 2, respectively, when the number of time points is 8.

Simulation Design Factors

Three manipulated conditions were common to Studies 1 and 2: the number of time points (NT), the sample size (SS), and the skewness for intercept (SI). In order to refer to the conditions value settings, we investigated the skewness, the kurtosis including a number of time points, and the sample size for GMMs or latent growth models using longitudinal data in areas such as juvenile delinquency, problem behavior, and alcohol use, situations where skewed distributions are frequently observed. The values set based on the found data and previous studies for nonnormality are as follows.

Two number of time points (4, 8) and three sample sizes (300, 800, 1,500) were manipulated. These values appeared most frequently in educational and psychological studies (also see Morgan, Hodge, & Baggett, 2016). For skewness of intercept

Figure 2. GMM with continuous distal outcome of Study 2.

Note. GMM = growth mixture model; C = latent class; I = intercept; S = slope; U = continuous distal outcome.

latent variable, three values were set as 1.2, 2, and 4. In the previous empirical study, the skewness of the observed variables was around ranged from 1.2 to 22.2 (e.g., alcohol use ranged from 1.99 to 10.69 in D’Amico et al., 2016; delinquent behavior ranged from 1.2 to 2.3 in Laird et al., 2008; Forgatch, Patterson, Degarmo, & Beldavs, 2009; antisocial behavior 0.8 to 2.82 in Walters & Ruscio, 2013). Extremely large values exceeding 10 observed less frequently. Bauer and Curran (2003) represented skews of 1 and 1.5 as a minor deviation from the normal distribution and Muthén and Asparouhov (2014) set the mild skewness to 2. Moreover, according to guidelines proposed by West, Finch, and Curran (1995), variables where skewness is greater than 2 are described as severely nonnormal. Therefore, this study considered skews of 1.2, 2, and 4 as a low, medium, and high value, respectively.

In summary, a total of 18 conditions ($3 \times 2 \times 3$) were considered in the data generation and 500 replications were generated for each condition. For each model that fitted to different distributions (i.e., normal, *t*, skew-normal, skew-*t* equal) were analyzed under the 18 conditions. The population model was a GMM fitted skew-*t* unequal distribution. *Mplus* 7.4 (Muthén & Muthén, 2015) was used to conduct data generation and subsequent analyses. For the GMM estimation, MLR was used.

Criteria of Evaluation

To evaluate the accuracy and reliability of the estimation, simulation outcomes included parameter bias (PB), mean squared error (MSE), and 95% coverage. In this study, parameter bias is indicated by a logit of probability that belongs to Class 1. Since the probability of belonging to Class 1 in the population model is 50%, the population parameter is zero ($0.5 = 1/(1 + \exp(0))$). Parameter bias was obtained by computing as the deviation of the parameter estimate ($\hat{\theta}_r$) from the population parameter (θ) on average across 500 replications. The criterion for the difference was the absolute value 0.1, of which is normally used (Bandalos, 2006). Standard errors of these parameter estimates (SE) were collected from each estimate and averaged across 500 replications.

$$\text{Bias}(\hat{\theta}) = \sum_{r=1}^R \left(\frac{(\hat{\theta}_r - \theta)}{\theta} \right) / R$$

MSE was defined as the sum of the square of the bias of parameter estimates and variance. The accuracy and consistency of the estimated parameter were evaluated through MSE. The smaller the values, the more accurate and reliable the estimation was.

$$MSE_{\hat{\theta}_r} = (\text{Bias}(\hat{\theta}_r))^2 + \text{Var}(\hat{\theta}_r)$$

The 95% coverage means the proportion of replications for which the 95% confidence interval includes a true value for the parameter. This indicates how accurately and precisely the parameters and standard errors are estimated. A range from .925 to .975 is considered a satisfactory standard of coverage (Bandalos, 2006).

Analysis of Variance

In addition to the inspection of parameter bias, MSE and coverage, an analysis of variance (ANOVA) was conducted to delineate the factors influencing parameter bias in GMM with nonnormal distribution. Independent variables were the three design factors (NT, SS, skew) and five different distributions. Eta-square (η^2) was measured to compare effect sizes of independent variables. The η^2 is defined as the ratio of variance explained by an effect in the total effects. Based on Cohen's (1988) criterion, if the eta-square is 0.01, it can be interpreted that the factor has a small effect size, while 0.06 is a medium effect size, and 0.15 is a large effect size.

Results

Study I

Admissible Solution Rates. Admissible solutions refer to the replications in which output was produced normally without any error message, whereas inadmissible solutions indicate replications in which the parameters were not estimated or were mathematically implausible, such as negative variance. The admissible solution rates (ASRs) for GMMs fitted to five different distributions are presented in Table 1. The ASR of normal and *t*-distribution were strikingly higher than those of skew-normal, skew-*t* equal, and skew-*t* unequal distribution under the condition in which the number of time points is 4. However, in cases where the number of time points and sample size were sufficient (i.e., the number of time points was 8 and the sample size was 800 or 1,500), there were no noticeable differences between and normal, *t*, skew-*t* equal, and skew-*t* unequal distribution. The model of skew-*t* unequal distribution reached over .90 when there was a combination of a small sample size (i.e., 300) and a weak skewness for intercept (i.e., 1.2). However, the ASR was noticeably lower as the skewness for intercept increased. The ASR of skew-normal distribution

Table 1. Admissible Solution Rates for a Growth Mixture Model Fitted to Five Different Distributions.

NT	SS	SI	Normal	<i>t</i>	Skew-normal	Skew- <i>t</i> equal	Skew- <i>t</i> unequal
4	300	1.2	0.99	0.99	0.59	0.80	0.90
4	300	2	0.98	0.99	0.40	0.79	0.74
4	300	4	0.97	0.96	0.25	0.86	0.47
4	800	1.2	1.00	1.00	0.79	0.92	0.99
4	800	2	1.00	1.00	0.58	0.95	0.97
4	800	4	1.00	1.00	0.33	0.98	0.81
4	1,500	1.2	1.00	1.00	0.86	0.98	1.00
4	1,500	2	1.00	1.00	0.66	0.99	1.00
4	1,500	4	1.00	1.00	0.51	1.00	0.95
8	300	1.2	0.99	1.00	0.66	0.95	0.96
8	300	2	0.99	1.00	0.46	0.93	0.85
8	300	4	0.97	0.99	0.24	0.88	0.54
8	800	1.2	1.00	1.00	0.87	1.00	1.00
8	800	2	1.00	1.00	0.61	0.99	1.00
8	800	4	1.00	1.00	0.36	0.99	0.88
8	1,500	1.2	1.00	1.00	0.93	1.00	1.00
8	1,500	2	1.00	1.00	0.72	1.00	1.00
8	1,500	4	1.00	1.00	0.51	1.00	0.98

Note. NT = number of time points; SS = sample size; SI = skewness for intercept.

was the lowest in all conditions, regardless of the number of time points and sample size.

Parameter Bias, SE, MSE, and Coverage Rate. Table 2 provides the simulation results of PB, SE, MSE, and a coverage rate of 95% confidence in Study 1. PB estimates were presented in absolute values. As expected, the skew-*t* unequal distribution estimated the parameter most accurately and reliably compared to other distributions. On the other hand, the parameter biases of normal, *t*, skew-normal, skew-*t* equal distribution were very high than the nominal value under most conditions.

Even though the GMMs fitted to skew-*t* unequal distribution indicated reasonable solutions under most conditions, it did not satisfy the evaluation criteria in some conditions. Specifically, the PB exceeded in a bad condition of a combination of a small number of time points (i.e., 4), small sample size (i.e., 300), and large skewness for intercept (i.e., 4). However, as the sample size increased, the bias approached within the standard value. This result indicates that more samples than 300 are needed to reliably estimate parameters in large skewness data. Besides the parameter bias, SE, and MSE were also enlarged in large skewness for intercept conditions. In addition, the 95% coverage rate was below the standard value in the condition of large skewness regardless of the number of time points and the sample size. Among large

Table 2. Bias, MSE, Coverage Rate of 95% CI, and Probability of Belonging to Class I for a Growth Mixture Model Fitted to Five Different Distributions.

NT			Normal						t						Skew-normal						Skew-t equal						Skew-t unequal											
			SS	SI	PB	SE	MSE	CV	P	PB	SE	MSE	CV	P	PB	SE	MSE	CV	P	PB	SE	MSE	CV	P	PB	SE	MSE	CV	P	PB	SE	MSE	CV	P				
4	4	300	1.2	.25	.51	2.41	.22	.56	.48	.29	.33	.56	.62	.15	.54	.60	.78	.54	.46	.48	.33	.62	.61	.02	.31	.07	.92	.50	.46	.48	.33	.62	.61	.02	.31	.07	.92	.50
4	4	300	2	.23	.58	3.34	.23	.44	.71	.44	.78	.46	.67	.06	.73	1.23	.71	.48	.40	.60	.41	.70	.60	.07	.41	.16	.87	.52	.40	.60	.41	.70	.60	.07	.41	.16	.87	.52
4	4	300	4	.57	.63	3.90	.25	.36	.66	.73	1.65	.56	.66	.72	.73	2.30	.64	.33	.00	.73	.72	.74	.50	.15	.64	.35	.81	.54	.00	.73	.72	.74	.50	.15	.64	.35	.81	.54
4	4	800	1.2	.51	.39	2.18	.08	.63	.44	.19	.23	.35	.61	.25	.44	.55	.77	.56	.61	.24	.41	.20	.65	.00	.20	.03	.95	.50	.61	.24	.41	.20	.65	.00	.20	.03	.95	.50
4	4	800	2	.70	.48	3.08	.12	.33	.74	.29	.64	.22	.68	.10	.54	1.59	.58	.52	.58	.30	.48	.48	.64	.01	.36	.07	.92	.50	.58	.30	.48	.48	.64	.01	.36	.07	.92	.50
4	4	800	4	1.39	.44	2.99	.07	.20	.30	.49	.98	.69	.58	.66	.64	2.70	.42	.34	-.03	.86	.49	.79	.49	.10	.57	.24	.84	.53	-.03	.86	.49	.79	.49	.10	.57	.24	.84	.53
4	4	1,500	1.2	.70	.24	1.98	.01	.67	.45	.14	.22	.09	.61	.38	.39	.63	.69	.59	.64	.15	.43	.04	.65	.00	.13	.01	.96	.50	.64	.15	.43	.04	.65	.00	.13	.01	.96	.50
4	4	1,500	2	1.01	.45	3.02	.05	.27	.76	.20	.61	.08	.68	.17	.63	2.09	.41	.54	.64	.24	.51	.30	.65	.01	.21	.04	.94	.50	.64	.24	.51	.30	.65	.01	.21	.04	.94	.50
4	4	1,500	4	1.55	.34	2.86	.01	.18	.09	.37	.56	.79	.52	.95	.52	3.09	.13	.28	-.06	.53	.33	.80	.49	.05	.81	.17	.86	.51	-.06	.53	.33	.80	.49	.05	.81	.17	.86	.51
8	8	300	1.2	.40	.45	1.91	.22	.60	.36	.26	.20	.66	.59	.14	.49	.37	.82	.54	.41	.31	.26	.66	.60	.01	.25	.06	.93	.50	.41	.31	.26	.66	.60	.01	.25	.06	.93	.50
8	8	300	2	-.04	.57	2.87	.23	.49	.65	.42	.59	.48	.66	.12	.75	.65	.75	.53	.54	.42	.48	.61	.63	.01	.35	.12	.88	.50	.54	.42	.48	.61	.63	.01	.35	.12	.88	.50
8	8	300	4	-.35	.70	3.88	.25	.41	.93	.60	2.17	.42	.72	.48	.68	2.05	.65	.38	.24	.73	.85	.71	.56	.09	.53	.36	.78	.52	.24	.73	.85	.71	.56	.09	.53	.36	.78	.52
8	8	800	1.2	.80	.30	1.81	.07	.69	.33	.15	.13	.39	.58	.25	.32	.39	.75	.56	.46	.21	.24	.33	.61	.00	.15	.02	.95	.50	.46	.21	.24	.33	.61	.00	.15	.02	.95	.50
8	8	800	2	.28	.45	2.79	.09	.43	.61	.23	.42	.21	.65	.33	.52	1.18	.60	.58	.58	.27	.40	.35	.64	.00	.23	.04	.93	.50	.58	.27	.40	.35	.64	.00	.23	.04	.93	.50
8	8	800	4	1.38	.43	2.97	.06	.20	.60	.67	1.40	.55	.65	.32	.64	2.37	.42	.42	.25	.64	.61	.72	.56	.07	.53	.21	.84	.52	.25	.64	.61	.72	.56	.07	.53	.21	.84	.52
8	8	1,500	1.2	1.05	.21	1.70	.02	.74	.32	.11	.11	.12	.58	.26	.38	.42	.75	.57	.44	.15	.21	.12	.61	.00	.10	.01	.95	.50	.44	.15	.21	.12	.61	.00	.10	.01	.95	.50
8	8	1,500	2	.52	.36	2.78	.03	.37	.59	.16	.38	.05	.64	.43	.51	1.51	.49	.61	.59	.20	.38	.15	.64	.00	.15	.02	.95	.50	.59	.20	.38	.15	.64	.00	.15	.02	.95	.50
8	8	1,500	4	1.56	.33	2.72	.01	.17	.38	.57	.90	.63	.59	.64	.57	2.84	.17	.34	.30	.51	.40	.74	.57	.02	.34	.11	.90	.51	.30	.51	.40	.74	.57	.02	.34	.11	.90	.51

Note. NT = number of time points; SS = sample size; SI = skewness for intercept; PB = parameter estimates bias; SE = standard error of the parameter estimate; MSE = mean squared error; CV = coverage rate of 95% confidence interval; P = probability of belonging to Class I. Acceptable results under criteria of PB, MSE, and coverage rates are shown in boldface.

Table 3. Analysis of Variance Results for Criteria According to Five Different Distributions.

Source	Sum of squares	df	Mean square	F	η^2
Number of time points (NT)	22.50	1.00	22.50	109.11***	0.00
Sample size (SS)	12.79	2.00	6.40	31.02***	0.00
Skewness for intercept (SI)	1146.70	2.00	573.35	2780.74***	0.12
Distribution (DIST)	8077.75	4.00	2019.44	9794.28***	0.50
NT $\times$ SS	1.54	2.00	0.77	3.74*	0.00
NT $\times$ SI	46.75	2.00	23.37	113.36***	0.01
NT $\times$ DIST	13.05	4.00	3.26	15.82***	0.00
SS $\times$ SI	57.61	4.00	14.40	69.85***	0.01
SS $\times$ DIST	111.47	8.00	13.93	67.58***	0.01
SI $\times$ DIST	451.29	8.00	56.41	273.60***	0.05

* $p < .05$. ** $p < .01$. *** $p < .001$.

skewness for intercept conditions, only a combination of 8 time points and a sample size of 1,500 reasonably satisfied all the criteria.

Under certain conditions, the model of skew- t equal distribution had a smaller bias than the model of skew- t unequal. This is a combination of 4 time points and a large skewness for intercept. For the sake of parsimony, some advantage of skew- t equal distribution was expected when the sample size was small, but this result was observed only when 4 time points and a large skewness for intercept. However, it is important to note that SE and MSE were considerably higher. This fact indicates that the estimates are fairly unstable and inaccurate in the model of skew- t equal distribution. Importantly, the skew parameter for slope should be specified if skewness is also observed over time. In summary, analysis using skew- t unequal distribution is recommended in unconditional GMM when the data were nonnormal.

Analysis of Variance. The ANOVA results for the bias of logit of probability in Study 1 are presented in Table 3. The main effects of the number of time points, sample size, skewness for intercept and five different distributions were all statistically significant. Two-way interactions were all statistically significant. Based on the square of eta, the explanatory power of distribution ($\eta^2 = 0.50$) was the highest and followed skewness for intercept ($\eta^2 = 0.12$) and two-way interaction term of skewness for intercept $\times$ distribution ($\eta^2 = 0.05$). This indicates that the choice of the appropriate distribution for the data is substantially important.

Study 2

Admissible Solution Rates. The ASRs for Study 2 are presented in Table 4. Overall, the ASR of Study 2 were higher than those of Study 1. When the skewed data were fitted to the normal and the t distribution, the solution was found in all replications, and when fitted to the skew- t distribution, the ASR was close to 1.00 except for the

Table 4. Admissible Solution Rates for a Growth Mixture Model With Continuous Distal Outcome Fitted to Five Different Distributions.

NT	SS	SI	Normal	<i>t</i>	Skew-normal	Skew- <i>t</i> equal	Skew- <i>t</i> unequal
4	300	1.2	1.00	1.00	0.84	0.99	0.98
4	300	2	1.00	1.00	0.77	1.00	0.94
4	300	4	1.00	0.99	0.63	0.96	0.77
4	800	1.2	1.00	1.00	0.97	1.00	1.00
4	800	2	1.00	1.00	0.95	1.00	1.00
4	800	4	1.00	1.00	0.86	1.00	0.98
4	1,500	1.2	1.00	1.00	1.00	1.00	1.00
4	1,500	2	1.00	1.00	0.99	1.00	1.00
4	1,500	4	1.00	1.00	0.94	1.00	0.99
8	300	1.2	1.00	1.00	0.88	1.00	0.99
8	300	2	1.00	1.00	0.78	1.00	0.98
8	300	4	0.99	1.00	0.64	0.99	0.84
8	800	1.2	1.00	1.00	1.00	1.00	1.00
8	800	2	1.00	1.00	0.96	1.00	1.00
8	800	4	1.00	1.00	0.87	1.00	0.99
8	1,500	1.2	0.99	1.00	1.00	1.00	1.00
8	1,500	2	0.99	1.00	1.00	1.00	1.00
8	1,500	4	0.99	1.00	0.95	1.00	1.00

Note. NT = number of time points; SS = sample size; SI = skewness for intercept.

condition of the 300 samples size and skewness for intercept was large. As in Study 1, the ASR was lowest when fitted to skew-normal distribution in Study 2.

Parameter Bias, SE, MSE, and Coverage Rate. Table 5 provides the simulation results of Study 2. Unlike those of Study 1, the results of Study 2 indicated more stable results not only in the model fitted to skew-*t* unequal distribution but also in some conditions of the model fitted to skew-normal and skew-*t* equal distributions. Specifically, the model fitted to skew-*t* unequal was found to be satisfactory in all conditions in Study 2, whereas proper results were not obtained under the conditions of large skewness and small sample size in Study 1. Moreover, the model of the skew-*t* equal distribution indicated satisfactory results in most conditions except for large skewness conditions. In addition, the model of skew-normal distribution obtained reasonable results under the conditions of a combination of a small sample size and small or moderate skewness. However, the models fitted to a normal or *t* distribution still yielded unacceptable results in all conditions. It is interpreted that the having the information of the distal outcome is more likely to estimate model parameter accurately and stably even if the model was complex.

The conditions that produced acceptable results in common for all three models of skew-normal, skew-*t* equal, and skew-*t* unequal distribution were combinations of a sample size of 300 and small or medium skewness. Hence, to further our

Table 5. Bias, MSE, Coverage Rate of 95% CI and Probability of Belonging to Class I for a Growth Mixture Model With Continuous Distal Outcome Fitted to Five Different Distributions.

NT	SS	SI	Normal						T						Skew-normal						Skew-t equal						Skew-t unequal																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
			PB			SE			MSE			CV			P			PB			SE			MSE			CV			P			PB			SE			MSE			CV			P																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								

Note. NT = number of time points; SS = sample size; SI = skewness for intercept; PB = parameter estimates bias; SE = standard error of the parameter estimate; MSE = mean squared error; CV = coverage rate of 95% confidence interval; P = probability of belonging to Class I. Acceptable results under criteria of PB, MSE, and coverage rates are shown in boldface.

Table 6. Parameter Estimates of a Growth Mixture Model With Continuous Distal Outcome When the Condition Is Middle Skewness.

Parameter for Class 2	Skew-normal				Skew-t equal				Skew-t unequal			
	PB	SE	MSE	CV	PB	SE	MSE	CV	PB	SE	MSE	CV
Number of time points = 4, sample size = 300, skewness for intercept = 2												
Mean u	0.74	0.30	0.18	0.79	0.68	0.29	0.16	0.89	0.05	0.19	0.03	0.96
Skew u	0.44	0.37	0.95	0.28	0.11	0.27	0.11	0.87	0.00	0.25	0.05	0.97
Number of time points = 8, sample size = 300, skewness for intercept = 2												
Mean u	0.79	0.26	0.15	0.78	0.43	0.25	0.09	0.90	0.07	0.18	0.03	0.94
Skew u	0.43	0.36	1.00	0.29	0.08	0.24	0.08	0.91	0.00	0.23	0.05	0.95

Note. PB = parameter estimates bias; SE = standard error of the parameter estimate; MSE = mean squared error; CV = coverage rate of 95% confidence interval.

investigation, we examined not only the bias on the probability of belonging to Class 1 but also whether the other parameters had been estimated accurately and reliably between these models of three distributions. Since the conditional model with outcome variable generally focuses on the interpretation of the outcome variable U, we estimated the bias, SE, and MSE for the mean and the skew parameter of U under a condition of a sample size of 300 and medium skewness (see Table 6). The mean of U in Class 1 confirmed reliable parameter estimates in all these three models, although these are not included in the table. As U was given skewness only in Class 2 in the population model, different results for the mean and skew parameter of U in Class 2 indicated between these three models. As expected, the model fitted to the skew-*t* unequal, which is a true model, had proper estimates. However, the bias for the models fitted to skew-normal or skew-*t* equal distributions exceeded standard values. In particular, results of the model fitted to skew-normal distribution indicated a considerably large bias for the skew parameter of U.

Analysis of Variance. The ANOVA results for the bias of the effect on distal outcome in Study 2 are presented in Table 7. The results are similar to the ANOVA results in Study 1. The main effects of the four conditions were all statistically significant. Two-way interactions were statistically significant in most cases, but interaction terms of number of time points $\times$ sample size was not significant. Based on the square of eta, the explanatory power of the distribution ($\eta^2 = .86$) was the highest, followed by interaction terms of skewness for intercept $\times$ distribution ($\eta^2 = .23$) and skewness for intercept ($\eta^2 = .21$). Compared with Study 1, the effect size of the main effects on distribution and skewness for intercept, and the interaction effect between these were much larger than in the unconditional model (i.e., $\Delta .09$ in skewness for intercept, $\Delta .36$ in distribution, and $\Delta .18$ in interaction terms of SI $\times$ DIST). This result suggests that using the appropriate distribution is even more important in the conditional model since the skewness is a feature of the data reflecting the reality.

Table 7. Analysis of Variance Results for Criteria According to Five Different Distributions.

Source	Sum of squares	df	Mean square	F	η^2
Number of time points (NT)	1.56	1.00	1.56	25.08***	0.00
Sample size (SS)	16.86	2.00	8.43	135.96***	0.01
Skewness for intercept (SI)	710.83	2.00	355.42	5730.68***	0.21
Distribution (DIST)	16475.18	4.00	4118.79	66410.99***	0.86
NT $\times$ SS	0.21	2.00	0.11	1.70	0.00
NT $\times$ SI	1.21	2.00	0.61	9.79***	0.00
NT $\times$ DIST	4.07	4.00	1.02	16.39***	0.00
SS $\times$ SI	0.99	4.00	0.25	4.00***	0.00
SS $\times$ DIST	10.27	8.00	1.28	20.70***	0.00
SI $\times$ DIST	810.25	8.00	101.28	1633.05***	0.23

* $p < .05$. ** $p < .01$. *** $p < .001$.

Conclusion and Discussion

Growth mixture models are often analyzed under the assumption of normality of the data, which rarely holds in practice, especially in areas such as juvenile delinquency, problem behavior, diseases, or disorders. The multivariate skew- t distributions of interest in these waves can be very useful for applied researchers in that they provide a flexible and robust framework for controlling skewness and kurtosis. However, apart from Muthén and Asparouhov (2014), simulation studies on the skew- t distribution are very limited, and research on the parameter bias in various conditions considering the number of time points, sample size, and skewness has not been attempted. In addition, skew- t distribution is not actively applied even in applied studies with seriously asymmetric data. Therefore, the present study emphasizes the importance of applying the skew- t distribution by examining how the models of normal distribution and other nonnormal distributions (i.e., t , skew-normal, skew- t equal, and skew- t unequal) cause bias in parameter estimation. This study also provides guidelines for the number of time points and sample sizes when analyzing GMMs with nonnormal data.

A number of conclusions can be drawn from this study. First, with even a small skewness in analyzing unconditional GMMs, it is inappropriate to fit to the distributions of normal, t , skew-normal, except to the skew- t distribution. Also, when skewness occurs over time, the skew parameter should be set to the slope, as well (i.e., the skew- t unequal distribution in this study). However, even if a model is fitted to skew- t unequal distribution, the bias was larger than 0.1 or coverage rates dropped to less than 0.90 in the large skewness conditions. Therefore, it should be noted that even under skew- t unequal distribution, the number of time points should be at least 8 and a sample size of more than 1,500 is recommended when the skewness of the data is large.

Second, the application of the skew- t unequal distribution is also recommended irrespective of the sample size, skewness, and number of time points in the conditional model. The PB and MSE were lowest in the model fitted to skew- t unequal distribution in all conditions, compared to models fitted to other distributions in a GMM

with a continuous outcome variable. Even in the condition whereby the sample size was so small that the more parsimonious model (i.e., skew- t equal distribution) was more robust, the performance of the model fitted to skew- t unequal distribution was the best, although the models of skew- t equal and skew-normal distribution met the criteria standards. Specifically, the models of skew-normal and skew- t equal distribution were satisfactory in the class classification parameters in some conditions (i.e., combination of when small or medium skewness and small sample size); however, the means of the continuous outcome variable U and the skew parameter exceeded the standard bias value. These three models were also compared using a likelihood ratio test (LRT). As these three models were nested, an LRT is possible. In the case of using MLR as an estimation method, a Satorra–Bentler LRT (SB LRT) is preferred to general LRT, according to prior research (Satorra & Bentler, 2001). Although the results are not presented in detail in the text of this article, it is confirmed that the model of skew- t unequal distribution is statistically significantly better than the models of the other distributions. Therefore, the model of skew- t unequal distribution should be used because a GMM generally describes not only class classification but also trajectory by classes.

Third, certain principles must be considered when analyzing a model fitted to a skew- t distribution. As can be seen from the results of the ASR, it should be noted that as the skewness increases, the convergence rate of the model fitted to skew- t distribution becomes lower and requires much more computational demand than in a model of normal distribution.

To summarize, both of the simulation results of Studies 1 and 2 indicate that when you analyze GMM for nonnormally distributed data, it is important to fit the model with appropriate distribution that reflects the properties of the data. It should be noted, however, that there are substantial differences between the results of the two study models (i.e., Study 1 of unconditional GMM and Study 2 of conditional GMM with continuous outcome variables). Specifically, in the models with the distal outcome variable (Study 2), skew- t unequal distribution models performed well under all simulation conditions. In contrast, in the unconditional model (Study 1), the bias of skew- t unequal distribution models was large when the sample size and the number of time points were not sufficient. Generally, first perform classification a latent class in an unconditional model and then verify a conditional model when analyzing GMM. Therefore, sufficient sample size and number of time points should be considered for an accurate classification in an unconditional model.

This study was designed as equal proportions in two class sizes. To extend our findings to even unbalanced class size conditions, additional simulation was conducted with unequal class proportions (.75 for Class 1 and .25 for Class 2) under conditions of number of time points 8, sample size 1,500, and skewness for intercept 2 in both studies. As same with the above results, only the model fitted to of skew- t unequal distribution satisfied the evaluation criteria, and the parameter biases of the models fitted to other distributions were unacceptably large (ranged from 0.14 to 9.43). This indicates the skew- t unequal distribution is robust to unequal class size data.

The limitations and recommendations for future research based on the results of this study are as follows. Although, this study considered diverse simulation conditions based on the feature of real data, there were limitations in reflecting all the conditions. Therefore, a future simulation study is needed for considering the conditions not included in this study. For example, skewness for the slope may be included as a condition. Also, since the df parameter can be separately set for each class in a mixed model, it is possible to set various conditions for the kurtosis. If population data have a kurtosis of 0, the expected results of a model fitted to skew-normal distribution would not differ from those of the present study, thereby enabling a meaningful comparison with the model of skew- t unequal distribution. In addition, on the basis of these differences between the two studies as mentioned above, it may be necessary to confirm that the same results are supported in other models, including GMM with multiple covariates, latent growth models, and latent profile analyses.

We only analyzed GMMs with the correct number of class that is equal to the number of classes in generating data, because this study focused on the comparison of parameter bias for five different distributions (i.e., normal, t , skew-normal, skew- t equal, and skew- t unequal) when the data are nonnormality. On the other hand, detecting the correct number of classes is another critical issue for an applied researcher as they usually do not know the number of classes a priori. According to Muthén and Asparouhov's (2014) study, fitting to skew- t unequal distribution than other distributions detected the correct number of classes using Bayesian information criterion (BIC) with nonnormal data. Nevertheless, enumeration issues with nonnormal data have not thoroughly reviewed. Therefore, further studies detecting the correct number of classes will be demanding. For example, comparison of performances for information criteria such as Akaike information criterion (Akaike, 1987), BIC (Schwarz, 1978), and aBIC (sample-size adjusted BIC; Sclove, 1987) under various data conditions with nonnormal data can be investigated.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-322.
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, 21, 329-341.
- Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling*, 23, 1-19.

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367-389.
- Azzalini, A., & Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715-726.
- Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43, 272-311.
- Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In R. C. Serlin (Series Ed.), G. R. Hancock & R. O. Mueller (Vol. Eds.), *Structural equation modeling: A second course* (pp. 385-462). Greenwich, CT: Information Age.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
- 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.
- Boers, K., Reinecke, J., Seddig, D., & Mariotti, L. (2010). Explaining the development of adolescent violent delinquency. *European Journal of Criminology*, 7, 499-520.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- D'amico, E. J., Tucker, J. S., Miles, J. N., Ewing, B. A., Shih, R. A., & Pedersen, E. R. (2016). Alcohol and marijuana use trajectories in a diverse longitudinal sample of adolescents: Examining use patterns from age 11 to 17 years. *Addiction*, 111, 1825-1835.
- Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology*, 45, 652-676.
- Forgatch, M. S., Patterson, G. R., Degarmo, D. S., & Beldavs, Z. G. (2009). Testing the Oregon delinquency model with 9-year follow-up of the Oregon Divorce Study. *Development and Psychopathology*, 21, 637-660.
- Guerra-Peña, K., & Steinley, D. (2016). Extracting spurious latent classes in growth mixture modeling with nonnormal errors. *Educational and Psychological Measurement*, 76, 933-953.
- Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 37, 359-363.
- Janosz, M., Archambault, I., Morizot, J., & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues*, 64, 21-40.
- 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.
- Laird, R. D., Criss, M. M., Pettit, G. S., Dodge, K. A., & Bates, J. E. (2008). Parents' monitoring knowledge attenuates the link between antisocial friends and adolescent delinquent behavior. *Journal of Abnormal Child Psychology*, 36, 299-310.
- Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, 20, 1-26.

- Lee, S., & McLachlan, G. J. (2014). Finite mixtures of multivariate skew t-distributions: Some recent and new results. *Statistics and Computing*, 24, 181-202.
- Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. H., & Clark, L. C. (2000). A latent class mixed model for analyzing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine*, 19, 1303-1318.
- Lin, T. I., Lee, J. C., & Hsieh, W. J. (2007). Robust mixture modeling using the skew t distribution. *Statistics and Computing*, 17, 81-92.
- Lin, T., Wu, P. H., McLachlan, G. J., & Lee, S. X. (2013). *The skew-t factor analysis model*. Retrieved from <http://arxiv.org/abs/1310.5336>
- Lu, X., & Huang, Y. (2014). Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. *Statistics in Medicine*, 33, 2830-2849.
- Manikandan, S. (2010). Data transformation. *Journal of Pharmacology and Pharmacotherapeutics*, 1, 126-127.
- McLachlan, G. J., & Peel, D. (1998, August). Robust cluster analysis via mixtures of multivariate t-distributions. In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR) (pp. 658-666). Berlin, Germany: Springer-Verlag.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models* (Wiley series in probabilities and statistics). New York, NY: Wiley-Interscience.
- Miller, S., Malone, P. S., & Dodge, K. A. (2010). Conduct problems prevention research group. Developmental trajectories of boys' and girls' delinquency: Sex differences and links to later adolescent outcomes. *Journal of Abnormal Child Psychology*, 38, 1021-1032.
- Morgan, G. B., Hodge, K. J., & Baggett, A. R. (2016). Latent profile analysis with nonnormal mixtures: A Monte Carlo examination of model selection using fit indices. *Computational Statistics & Data Analysis*, 93, 146-161.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Thousand Oaks, CA: Sage.
- Muthén, B., & Asparouhov, T. (2014). Growth mixture modeling with non-normal distributions. *Statistics in Medicine*, 34, 1041-1058.
- Muthén, B. O., & Muthén, L. K. (2015). *Mplus 7.4 [Computer software]*. Los Angeles, CA: Muthén & Muthén.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Walters, G. D., & Ruscio, J. (2013). Trajectories of youthful antisocial behavior: Categories or continua? *Journal of Abnormal Child Psychology*, 41, 653-666.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). *Structural equation models with nonnormal variables: Problems and remedies*. Thousand Oaks, CA: Sage.
- Wickrama, K. K., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). An introduction to growth mixture models (GMMs). In *Higher-order growth curves and mixture modeling with Mplus: A practical guide* (pp. 209-226). New York, NY: Routledge.