CrossMark

An evaluation of the use of covariates to assist in class enumeration in linear growth mixture modeling

Jinxiang Hu[1] · Walter L. Leite[2] · Miao Gao[3]

Abstract This study examined whether the inclusion of covariates that predict class membership improves class identification in a growth mixture modeling (GMM). We manipulated the degree of class separation, sample size, the magnitude of covariate effect on class membership, the covariance between the intercept and the slope, and fit two models with covariates and an unconditional model. We concluded that correct class identification in GMM requires large sample sizes and class separation, and that unconditional GMM performs better than GMM with covariates if the sample size and class separation are sufficiently large. With small sample sizes, GMM with covariates outperformed unconditional GMM, but the percentage of correct class enumeration was low across different fit criteria.

Keywords Growth mixture modeling · Class enumeration · Predictors of class membership · Information indices · Likelihood ratio tests

Growth mixture modeling (GMM; Muthén, 2001, 2002; Muthén & Muthén, 2000; Muthén & Shedden, 1999) has been increasingly used to separate individuals into unknown subgroups characterized by their distinct growth trajectory over time

✉ Walter L. Leite
  Walter.Leite@coe.ufl.edu

  Jinxiang Hu
  Fran.hu@nih.gov

[1] National Institute of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

[2] College of Education, University of Florida, 1215 Norman Hall, Gainesville, FL 32611, USA

[3] Nanjing Normal University, Jiangsu, China

(Bauer, 2007). GMM is similar to latent class growth modeling (LCGM; Nagin, 1999), but LCGMs set the variances and covariances of growth factors within class to zero (Muthén, 2004). Regardless of its popularity, class enumeration in GMM has remained an active topic of research (Li & Hser, 2011; Liu & Hancock, 2014; Peugh & Fan, 2012; Peugh & Fan, 2014; Tofighi & Enders, 2007;). One topic still needing additional research is in which conditions including covariates in GMM facilitates the identification of the correct number of classes (Li & Hser, 2011; Liu & Hancock, 2014, Peugh & Fan, 2014; Tofighi & Enders, 2007).

There have been a few recommendations in the literature about incorporating covariates in GMM for class enumeration: Muthén (2004) suggested that properly selected covariates may facilitate GMM class enumeration, estimation of class proportions and class membership; Lubke and Muthén (2007) studied covariate effects in factor mixture modeling and suggested that covariate incorporation is beneficial for parameter recovery and class assignment in the case of covariates only predicting latent class membership. Using continuous normally-distributed covariates, Tofighi and Enders (2007) concluded that covariates hampered GMM class enumeration unless the sample size was 200. Peugh and Fan (2014) added more nuance to the knowledge about inclusion of covariates in GMM by finding that the effect of adding covariates into GMM varied across model selection criteria with smaller sample sizes (i.e., 300, 500), but with larger sample sizes (i.e., N=3,000) the addition of covariates decreased correct class enumeration rates with most fit criteria. However, using binary and non-normally distributed covariates, Li and Hser (2011) found it is beneficial to include covariates for GMM class enumeration. Therefore, there is evidence that the effect of including covariates on class enumeration in GMM is complex. The purpose of this paper is to add to current knowledge about GMM by examining whether the way covariates are included in the model affects

class enumeration. More specifically, we examine GMM with covariates added as predictors of intercept and slope and GMM with covariates as direct predictors of the outcome as compared with unconditional GMM. In particular, direct-effect GMM, which is an extension of the direct-effect latent growth model discussed by Stoel, Van den Wittenboer, and Hox (2004), has never been examined to improve class enumeration in GMM. These model specifications are compared across variations of class separation, sample size, covariate effect, and magnitude of covariance between intercept and slope. The structure of this paper is the following: We first describe the three GMM specifications that are the focus of this paper, then present model fit indexes commonly used for class enumeration. Next, we review previous studies of class enumeration in mixture modeling. Finally, we present two Monte Carlo simulation studies examining class enumeration with three GMM specifications with several fit indices across a variety of conditions.

Growth mixture models

GMM can be used to describe a variety of linear and nonlinear growth trajectories. In this paper, we will focus on the linear growth pattern that has been widely used in applied research (e.g., Abroms, et al, 2005; Greenbaum, et al, 2005; McDonough, Sacker, & Wiggins, 2005; Stoolmiller, Kim, & Capaldi, 2005). We assume that covariates are antecedents of class membership (Muthén, 2004), which implies that they may affect class membership and can be used as auxiliary variables for class enumeration.

Unconditional growth mixture model

An unconditional growth mixture model (U-GMM, Muthén, 2002) can be represented by the following equations:

$$y_{ij}^k = \eta_{iI}^k + \eta_{iS}^k t_j + \varepsilon_{ij}^k \tag{1}$$

$$\eta_{iI}^k = \gamma_{00}^k + \xi_{i0}^k \tag{2}$$

$$\eta_{iS}^k = \gamma_{01}^k + \xi_{i1}^k \tag{3}$$

Equation 1 presents the measurement part of the U-GMM where y_{ij}^k is the value of the outcome variable for participant i at occasion j in class k, t_j is a time code assumed here to be the same for all individuals, η_{iI}^k is the random intercept, η_{iS}^k is the random slope, and ε_{ij}^k is the residual. Equations 2 and 3 are the structural part of the growth mixture model, where γ_{00}^k and γ_{01}^k are the mean intercepts and slopes as latent factors across all participants and occasions ξ_{i0}^k and ξ_{i1}^k are residuals of the latent growth factors at the between level.

$$\ln\left[\frac{P(c_{ik}=1)}{P(c_{iK}=1)}\right] = \lambda_k$$
$$c_{ik} = \begin{cases} 1 & if\ subject\ i\ belongs\ to\ class\ k \\ 0 & otherwise \end{cases} \tag{4}$$

Equation 4 is the unconditional multinomial logit part of the U-GMM where c_{ik} is one if participant i belongs to the kth class or zero otherwise, and λ_k is the log odds of belonging to class k instead of reference class K (Lubke & Muthén, 2007).

Growth mixture models with covariates

Previous studies about the effects of covariates on class enumeration in GMM (i.e., Li & Hser, 2011; Tofighi & Enders, 2007) investigated a conditional GMM where covariate(s) predict the growth factors and class membership. With such a model, it is assumed that the growth factors fully mediate the relationship between predictors and the outcome variables (Stoel, Van den Wittenboer, & Hox, 2004). We will refer to this conditional GMM as growth predictor growth mixture model (GP-GMM), because it is an expansion of the growth predictor latent growth model discussed by Stoel, Van den Wittenboer, and Hox (2004) to characterize models with covariates predicting the latent growth factors. The GP-GMM can be represented by the following equations:

$$y_{ij}^k = \eta_{iI}^k + \eta_{iS}^k t_j + \varepsilon_{ij}^k \tag{5}$$

$$\eta_{iI}^k = \gamma_{00}^k + \sum_{p=1}^{P} \gamma_{p0}^k x_{ip} + \xi_{i0}^k \tag{6}$$

$$\eta_{iS}^k = \gamma_{01}^k + \sum_{p=1}^{P} \gamma_{p1}^k x_{ip} + \xi_{i1}^k \tag{7}$$

While the GP-GMM assumes that the growth factor fully mediates the relationships between predictors and outcomes, the direct effect growth mixture model (DE-GMM) relaxes this assumption, and is an expansion of the direct effect latent growth model (Stoel, Van den Wittenboer, & Hox, 2004) to characterize GMM with covariates predicting the observed variables directly. The DE-GMM can be represented by the following equations:

$$y_{ij}^k = \eta_{iI}^k + \eta_{iS}^k t_j + \sum_{p=1}^{P} \beta_{jp}^k x_{ip} + \varepsilon_{ij}^k \tag{8}$$

$$\eta_{iI}^k = \gamma_{00}^k + \xi_{i0}^k \tag{9}$$

$$\eta_{iS}^k = \gamma_{01}^k + \xi_{i1}^k \tag{10}$$

In Equation 8, β_{jp}^k is the effect of predictor x_{ip} on the outcome at time j for class k. Equation 11 can be used with both

the GP-GMM and DE-GMM to describe the relationship between the log odds of class membership and predictors:

$$\ln\left[\frac{P\left(c_{ik}=1\big|x_i\ldots x_p\right)}{P\left(c_{iK}=1\big|x_i\ldots x_p\right)}\right] = \lambda_k + \sum_{p=1}^{P} \delta_p^k x_{ip} \qquad (11)$$

where x_{ip} is the pth covariate used as the predictor of the latent class. γ_{p0}^k and γ_{p1}^k are direct effects of x_{ip} on the intercept and slope factors, respectively. In Equation 8, the covariates x_{ip} predict the log odds of the probability of belonging to class k with effects equal to δ_p^k (Lubke & Muthén, 2007).

Both the GP-GMM and the DE-GMM assume that the within-class model is correctly specified. Because the full mediation assumption of the GP-GMM may not hold, the use of the DE-GMM could protect against the effect of misspecification of the relationship between predictors and outcomes in the within-class model.

Research about the effect of covariates on class enumeration in GMM has focused solely on the GP-GMM. Muthén (2004) suggested taking a first glance at the number of classes without covariates then adding covariates to help make a final decision about the number of classes. However, Muthén (2006) warned that researchers should not be surprised if the class formation changed after adding covariates. This is because adding covariates might change the normality of the latent intercept and the latent slope factor within class. The task of selecting the number of classes is further complicated by the availability of several model selection indices and statistics, which differ in the sensitivity to model complexity, sample size, the presence of covariates and the presence of covariance between the intercept and the slope in the model. No agreement has yet been reached regarding which index works best as indicator of the number of classes (Lubke & Muthén, 2007; Tofighi & Enders, 2007; Yang, 2006). In the next section, we review fit indices and statistics commonly used for class enumeration in mixture modeling as well as studies comparing their performance.

Class enumeration in growth mixture modeling

In this study, we will focus on the following model information-based fit indices: Akaike's Information Criterion (AIC; Akaike, 1987), Bayesian information criterion (BIC; Schwartz, 1978), adjusted BIC (ABIC, Sclove, 1987). We will also examine the following likelihood ratio tests: Lo-Mendell-Rubin likelihood ratio test (LMR; Lo, Mendell, & Rubin, 2001), and the bootstrapped likelihood ratio test (BLRT; McLachlan, 1987; McLachlan & Peel, 2000).

The AIC, the BIC and the ABIC can be calculated with Equations 12 to 14, and they are all functions of the log-

likelihood (LL) of the fitting model and number of parameters (p):

$$AIC = -2LL + 2p \qquad (12)$$

$$BIC = -2LL + pLn(N) \qquad (13)$$

$$ABIC = -2LL + pLn\left(\frac{n+2}{24}\right) \qquad (14)$$

As an alternative to information-based indices, the likelihood ratio test (LRT) compares the fit of models that are nested in each other (Bollen, 1989).

$$LR = -2(LL_{k-1} - LL_k) \qquad (15)$$

where LL_{k-1} is the log-likelihood of the model with k-1 classes and LL_k is the log-likelihood of the model with k classes. LR is a function of the difference of these two likelihoods and assumes a chi square difference distribution. However, with mixture models, the standard LRT presented above cannot be used because the LR is not asymptotically distributed as chi-square (McLachlan & Peel, 2000). Lo, Mendell, and Rubin (2001) solved this problem by proposing an approximation of the chi-square distribution in the mixture context that is a weighted sum of chi-squares. The LMR statistic compares the fit of the k-1 class model versus the k class model. A significant p value rejects the k-1 class model in favor of the k class model. A bootstrap version of the LRT (BLRT) is to use bootstrap samples as the distribution of the log-likelihood ratio test statistic rather than the assumed chi-square distribution (Nylund et al., 2007). The BLRT statistic is similar to LMR, but only compares the statistics against the bootstrap empirical sampling distribution. A significant p value of BLRT rejects the k-1 class model in favor of the k class model (Tofighi & Enders, 2007). Yang (2006) examined the class enumeration performance of the information indices with latent class analysis models. The results showed that ABIC was more accurate in detecting the correct number of latent classes than the AIC. The BIC was very sensitive to sample size and performed well at a sample size of 1,000, while the AIC was consistent but sample size did not improve its accuracy substantially. Tofighi and Enders (2007) recommended the ABIC. Nylund et al. (2007) found that the BLRT had the best performance, followed by BIC and ABIC, for class enumeration. However, Nylund et al. (2007) only investigated unconditional GMM. Lubke and Muthén (2007) concluded that covariates helped correct assignment of classes especially when class separation was large. However, in their study, the covariate was only predicting class membership and had no effect on the factor mean, which was not common in GMM. Li and Hser (2011) found that it was beneficial to include covariates in GMM to assist in the identification of correct number of class when sample size was equal to or larger than 400.

They also found that the BIC performed best across sample sizes, the ABIC performed well when sample size was larger than 400, and the AIC had the poorest performance. BIC outperformed LRTs in general.

Peugh and Fan (2014) found that a sample size of 3,000 was needed to optimize class identification. Under conditions of largest sample size (N = 3,000), large class separation (MD = 2), maximum variance explained (25 %), and equal class proportion, BIC, LMR, and BLRT were able to identify the correct number of classes. With a sample size of 500, all fit criteria performed poorly. By comparing their results with a previous study by Peugh and Fan (2012) that did not include covariates in the GMM, Peugh and Fan (2014) concluded that adding covariates decreased the detection rate of BIC and BLRT under large sample size, large class separation, and large variance explained.

Li and Hser (2011), Peugh and Fan (2014), and Tofighi and Enders (2007) have studied the effect of covariate inclusion in GMM and class enumeration performance of fit indices and statistics when the generating model had covariates, but these studies used the GP-GMM for class enumeration with covariates. The current study expands the existing GMM literature by further investigating the performance of two different types of GMM with covariates and GMM without covariates on class enumeration. Our research questions are: (a) Is there any difference in performance between the U-GMM, GP-GMM and DE-GMM with respect to accuracy of class enumeration? (b) How do commonly used indices and statistics for class enumeration perform across different GMM specifications?

GMM can differ with respect to which parameters are allowed to vary between classes. We conducted two studies to address the research questions above, each with a different degree of model complexity with respect to allowing parameters to vary between classes. Study 1 compared versions of the U-GMM, GP-GMM and DE-GMM where intercept means were allowed to vary between classes. Study 2 compared versions of these models where intercept and slope means, residual variances and covariance were allowed to vary between classes. By comparing the results of Study 1 and Study 2, we were able to examine whether class enumeration performance depended on which parameters were allowed to vary between classes.

Method

Population model

Both Study 1 and Study 2 used a two-class linear GP-GMM with one covariate as a population model, because it is popular in applied research (e.g., Abroms et al, 2005; Greenbaum et al, 2005; McDonough, Sacker, & Wiggins, 2005; Stoolmiller, Kim, & Capaldi, 2005) and has also been used in previous simulation studies (e.g., Lubke & Muthén, 2007; Nylund

et al., 2007):

$$\gamma_{ij}^k = \eta_{iI}^k + \eta_{iS}^k t_j + \xi_{ij}^k \tag{16}$$

$$\eta_{iI}^k = \gamma_{00}^k + \gamma_{10}^k x_{i1} + \xi_{i0}^k \tag{17}$$

$$\eta_{iS}^k = \gamma_{01}^k + \gamma_{11}^k x_{i1} + \xi_{i1}^k \tag{18}$$

Population parameters for Equations 16 to 18 used in the Monte Carlo simulations in Studies 1 and 2 are shown in Table 1. Data were simulated for a normally distributed outcome measured at five equally spaced time intervals and one normally distributed standard covariate. In a simulation study with 4 and 7 waves, Tofighi and Enders (2007) found that the number of repeated measures had only a minor impact on class enumeration. We chose five waves because it is within the range of waves examined by Tofighi and Enders (2007). The number of classes was set at two through reviewing previous LCA and GMM simulation studies (e.g., Li & Hser, 2011; Lubke & Muthén, 2007; Nylund, et al., 2007). Lubke and Muthén (2007) stated that unequal class proportions do not have a noticeable effect on model selection. Therefore, we chose to set population class proportions at .50. Following previous studies (e.g., Li & Hser, 2011; Lubke & Muthén, 2007; Nylund, et al., 2007; Yang, 2006), we defined effects of the covariate on the growth factors that were invariant across classes. We selected population coefficients for the covariate such as that it explains 15 % of the variance of the intercept and 5 % of the slope. In turn, population residual variances for the intercept, slope and outcome were chosen so that the intercept and the slope together account for 20 % of the variance of the outcome at each measurement occasion. These targets are similar to Li and Hser (2011), who simulated a covariate that explained 20 % of the growth factors' variance, and Tofighi and Enders (2007), who generated covariates accounting for 15 % of the variance of the intercept and slope.

Manipulated conditions in Study 1

In study 1, we manipulated three conditions: class separation (multivariate Mahalanobis Distance (MD) of 1.3, 2.6, and 3.9), sample size (N = 400, 1,000, and 2,000) and covariate's effect on class membership (pseudo R^2 = 0.1 and 0.3). The variables manipulated resulted in a total of $3 \times 3 \times 2 = 18$ unique conditions. For each condition, 100 datasets were generated using the R statistical package.

The three levels of class separation (MD = 1.3, 2.6, and 3.9) manipulated correspond to differences of 1, 2, and 3 standard deviations between the means of the intercept factors of the two classes. Lubke and Muthén (2007) simulated class separation of 0, 1, 1.5, and 2 standard deviations but mentioned that real data could have higher levels of class separation than simulated in their study. In Study 1, the separation between classes was

Table 1 Population values for Monte Carlo simulation studies

Parameter	MD	Study 1		Study 2	
		Class 1	Class 2	Class 1	Class 2
Mean of intercept (γ_{00})	1.3	0	1.3	0	0.79
	2.6	0	2.6	0	1.13
	3.9	0	3.9	0	1.28
Mean of slope (γ_{01})	1.3	1	1	1	1.79
	2.6	1	1	1	2.13
	3.9	1	1	1	2.38
Effect of covariate on intercept (γ_{10})		0.15	0.15	0.15	0.15
Effect of covariate on slope (γ_{11})		0.05	0.05	0.05	0.05
Residual variances:					
Intercept (var(ξ_0))		1.42	1.42	1.42	1.42
Slope (var(ξ_1))		0.19	0.19	0.19	0.19
Y1 (var(ε_1))		6.68	6.68	6.68	6.68
Y2 (var(ε_2))		7.48	7.48	7.48	7.48
Y3 (var(ε_3))		9.88	9.88	9.88	9.88
Y4 (var(ε_4))		13.88	13.88	13.88	13.88
Y5 (var(ε_5))		19.48	19.48	19.48	19.48
Residual covariance ($cov(\xi_1\xi_2)$)		0	0	0	0

Note. Notation in parenthesis refers to Equations 16 to 18; MD = Mahalanobis Distance

determined solely by the population mean difference on the intercept factor. Tofighi and Enders (2007) performed a review of GMM applications and found a large range of sample sizes (i.e., 110–5,833) in applied studies. In this simulation, we generated datasets with sample sizes of 400, 1,000, and 2,000, which are approximately between the 25th and 75th percentiles of the sample sizes in studies identified by Tofighi and Enders (2007).

None of the previous studies about class enumeration in GMM manipulated the strength of the effect of covariates on class membership. The McKelvey and Zavoina Pseudo-R^2 (Hu et al. 2006; McKelvey & Zavoina, 1975) can be used to manipulate this effect. The McKelvey and Zavoina Pseudo-R^2 can be interpreted as the proportion of the variance of a hypothetical continuous variable underlying the latent class variable that is explained by the covariate. Covariate effects corresponding to pseudo-R^2 of 0.1 and 0.3 were manipulated in this study.

Manipulated conditions in Study 2

We varied the intercept mean and slope mean differences in the population model so that the two classes have unique mean intercept and mean slope. In addition to the three conditions (class separation, sample size, and covariate effect) manipulated in Study 1, we manipulated whether the residual covariance matrix of intercept and slope were allowed to vary between classes. The variables manipulated resulted in a total of 3 × 3 × 2 × 2 =36 unique conditions. For each condition, 200 datasets were generated using the R statistical package.

Analysis procedure

For each simulated dataset, the U-GMM as shown in Equations 1–4, the GP-GMM shown in Equations 5, 6, 7, and 11 and DE-GMM shown in Equations 8–11 with one to three latent classes were fit using the MPLUS 7.1 software with maximum likelihood estimation. In Study 1, these models were fitted once to each simulated dataset, allowing means of intercepts to vary between classes. In Study 2, these models were fit twice to each simulated dataset, once allowing factor means to vary between classes, and another time allowing factor means, residual variances and covariance of growth factors to vary between factors. The covariate's effects on the growth factors in the GP-GMM and on the outcomes in the DE-GMM were set to be invariant across classes. Similarly to Peugh and Fan (2014), we fit GMM to each simulated dataset with the correct number of classes (i.e., two classes) and incorrect but neighboring number of classes (i.e., one and three classes), because a review of previous GMM studies indicated that class enumeration procedures resulted in class counts that were close to the correct number of classes regardless of simulated condition. For example, in Li and Hser (2011), the number of classes selected using different criteria across all conditions fell mostly within one to three classes even though the authors fit models with one to five classes.

Class enumeration was compared across information indices (i.e., AIC, BIC, and ABIC) and likelihood ratio tests (i.e.,

LMR and BLRT). The outcome of the Monte Carlo simulation study was a dichotomous indicator of whether the correct number of classes, two classes in this case, was identified using each class enumeration criterion. For the AIC, BIC, and ABIC, if the model with two classes produced a smaller value than the models with one and three classes, an outcome of 1 was assigned to indicate correct class enumeration. For the LMR and BLRT, a significant p value ($p \leq .05$) indicates that the model with k classes fits better than the model with k-1 classes. Therefore, the number of classes was determined by the following criterion: if $p \leq .05$ for the comparison between models with one and two classes and $p > .05$ for the comparison between models with two and three classes, a value of 1 was assigned to the outcome and regarded as correct class enumeration. This is a strict criterion given that the LMR and BLRT not only had to indicate better fit of two class model over three class model (as in Nylund et al., 2007), but also had to indicate better fit of the two class model over the one class model.

Mixture models are known for high frequency of convergence problems, local maxima, and improper solutions (Li & Hser, 2011; Lubke & Muthén, 2007; Nylund et al., 2007). To eliminate these problems, we set starting values at 400 random sets in the initial stage and 100 optimizations in the final stage. These numbers are much larger than the default MPLUS settings of 20 and 4 for the initial and final stages respectively, and previous research has obtained adequate results by just doubling the MPLUS defaults (e.g., Peugh & Fan, 2014).

Results

In this section, the performance of model fit indices and statistics for selecting correct number of classes with three different GMM are reported for both Study 1 and Study 2. To aid in the interpretation of results of the Monte Carlo simulation study (Bandalos & Leite, 2013) a mixed-design ANOVA table was calculated with the dichotomous indicator of each of the fit indexes selecting the correct model as the outcome. This is a linear probability model (Agresti, 2002) that is in general not recommended for dichotomous outcomes, but it is useful here because our goal is to simply compare effects across manipulated conditions rather than to interpret parameter estimates. Furthermore, as Agresti (2002) pointed out, this model is valid over a restricted range of predictors, and in our simulation all of our predictors are categorical variables. Class separation, sample size, and covariate effect on class membership were between-iteration effects in both Studies 1 and 2, because these factors were manipulated across different simulated datasets. GMM type was a within-iteration effect in Study 1, both GMM type and whether the residual covariance matrix of the intercept and the slope was allowed to vary between classes were within-iteration effects in Study 2. Generalized eta

squared (η^2) (Olejnik and Algina 2003) was used to quantify the effects of the manipulated factors. We considered effects equal or larger than 0.01 to be substantial and focused interpretation on them, because the large number of replications in simulation studies results in high levels of power and only those effects reaching some pre-specified level of effect size should be interpreted for practical significance (Bandalos & Leite, 2015). These effects are shown in Table 2 (all results were available upon request to the first author).

Study 1

Table 3 presents a comparison of class enumeration with AIC, BIC, ABIC, LMR, and BLRT across simulated conditions. With correct classes enumeration with the AIC as the outcome, GMM type ($\eta^2 = .018$), and the interaction of GMM type by class separation ($\eta^2 = .018$) had the largest effect sizes, followed by the interaction of class separation and sample size ($\eta^2 = .010$). GMM type, class separation and sample size had significant main effects. The two GMM with a covariate performed better than the GMM without covariate only when class separation was small ($MD = 1.3$) and sample size was 1,000 or less. For these conditions, the DE-GMM outperformed the GP-GMM.

For class enumeration with the BIC, we observed significant main effects of class separation, sample size, and covariate effect. Class separation ($\eta^2 = .604$) had the largest effect size, followed by sample size ($\eta^2 = .131$) and covariate effect ($\eta^2 = .027$). However, there were two three-way interactions present: Interaction of class separation by sample size ($\eta^2 = .110$) and interaction of class separation by covariate effect ($\eta^2 = .010$). The two GMM with a covariate only outperformed GMM without covariates when the covariate effect was stronger ($R^2 = .3$) and class separation was 2.6. We found that class enumeration with the BIC was very sensitive to class separation and sample size. In these conditions, the GP-GMM performed slightly better than the DE-GMM when the pseudo-R^2 was 0.1 but slightly worse when the pseudo R^2 was 0.3.

For class enumeration with the ABIC, we observed a substantial main effect of all four manipulated factors: class separation, sample size, covariate effect, and GMM type. Class separation ($\eta^2 = .530$) had the largest effect size, followed by the interaction of class separation by sample size ($\eta^2 = .041$), sample size ($\eta^2 = .014$), and covariate effect ($\eta^2 = .014$). However, three-way interactions were also present: class separation by sample size by GMM type, sample size by covariate effect by GMM type. The two GMM with a covariate performed better than GMM without a covariate when class separation was 1.3, or when class separation was 2.6 in combination with a sample size of 1,000 or smaller. The DE-GMM resulted in higher accuracy of class enumeration than the GP-GMM in these conditions.

Table 2 Effect sizes of manipulated factors in Study 1 and Study 2

	AIC	BIC	ABIC	LMR	BLRT
Study 1					
MD	-	.604	.530	.328	.385
N	-	.131	.014	.043	.010
R²	-	.027	.014	-	-
GMM type	.018	-	-	-	-
MD × N	.010	.110	.041	.015	.023
MD × R²		.010	-	-	-
MD × GMM type	.018	-	-	.025	.011
Study 2					
MD	.057	.574	.473	.184	.441
N	.015	.325	.055	.079	.193
R²	-	.034	.014	-	.016
Cov	-	.028	.013	-	.014
GMM type	.058	.025	.012	-	.024
MD × N	-	-	.075	.022	-
Cov × GMM type	-	.019	.012	-	-
MD × N × GMM type	-	.017	.021	-	.018
Cov × GMM type × MD × N	-	.011	-	-	-

Note. MD = Mahalanobis Distance class separation. N = sample size. R² = class separation pseudo R² . Cov= whether covariance matrix of growth factors was allowed to vary between classes. Only η^2 equal to or greater than 0.01 are shown

For class enumeration with the LMR, there were substantial main effects of class separation and sample size, as well as two-way interactions. Class separation ($\eta^2 = .328$) had the largest effect size, followed by sample size ($\eta^2 = .043$), the interaction of class separation by GMM type ($\eta^2 = .025$), the interaction of class separation by sample size ($\eta^2 = .015$). The two GMM with a covariate performed better than GMM without covariate when class separation was 1.3, or when class separation was 2.6 in combination with a sample size of 1,000 or smaller. In most conditions, the DE-GMM outperformed the GP-GMM.

The BLRT performed similarly to LMR. Class separation had the largest effect size ($\eta^2 = .385$), followed by interaction of class separation by sample size ($\eta^2 = .023$), interaction of class separation by GMM type ($\eta^2 = .011$), and sample size main effect ($\eta^2 = .010$). The GP-GMM and DE-GMM performed better than the U-GMM when class separation was 1.3, or when class separation was 2.6 in combination with a sample size of 1,000 or smaller, and DE-GMM outperformed GP-GMM except when the sample size was 400.

In summary, the GMM with covariates were the best models for class enumeration, with the DE-GMM outperforming the GP-GMM most of the time, when sample sizes and Mahalanobis Distances were at the smallest levels simulated, but the U-GMM works better than both GMM with

Table 3 Percentage of correct model identification of fit indices in Study 1

Study 1				U-GMM					GP-GMM					DE-GMM				
Cov	R²	MD	N	AIC	BIC	ABIC	LMR	BLRT	AIC	BIC	ABIC	LMR	BLRT	AIC	BIC	ABIC	LMR	BLRT
No	0.1	1.3	400	0.15	0	0.07	0.03	0.03	0.26	0	0.21	0.2	0.11	0.31	0	0.27	0.29	0.17
			1000	0.2	0	0.02	0.07	0.02	0.15	0	0.09	0.18	0.17	0.23	0.01	0.11	0.21	0.2
			2000	0.28	0	0.02	0.09	0.18	0.1	0	0.03	0.17	0.15	0.09	0	0.02	0.19	0.16
		2.6	400	0.32	0.04	0.21	0.09	0.07	0.27	0.02	0.35	0.31	0.2	0.3	0	0.31	0.23	0.14
			1000	0.53	0.07	0.2	0.25	0.22	0.32	0.02	0.18	0.27	0.27	0.49	0.02	0.36	0.37	0.38
			2000	0.84	0.08	0.48	0.68	0.75	0.19	0.04	0.2	0.26	0.22	0.32	0.06	0.29	0.35	0.34
		3.9	400	0.8	0.37	0.81	0.69	0.43	0.54	0.29	0.78	0.49	0.49	0.51	0.24	0.65	0.45	0.38
			1000	0.87	0.85	0.93	0.89	0.93	0.64	0.9	0.93	0.76	0.84	0.68	0.75	0.96	0.81	0.89
			2000	0.89	1	0.99	0.93	0.97	0.7	1	0.99	0.84	0.91	0.75	1	0.99	0.83	0.9
	0.3	1.3	400	0.16	0	0.1	0.11	0.06	0.22	0	0.25	0.27	0.16	0.27	0.01	0.29	0.17	0.12
			1000	0.18	0	0.03	0.07	0.04	0.16	0	0.05	0.18	0.15	0.33	0	0.19	0.28	0.25
			2000	0.12	0	0.02	0.07	0.1	0.2	0	0.05	0.21	0.2	0.15	0	0.06	0.22	0.15
		2.6	400	0.42	0.05	0.25	0.19	0.11	0.34	0.06	0.45	0.27	0.25	0.3	0.03	0.4	0.28	0.23
			1000	0.62	0.06	0.29	0.36	0.35	0.31	0.09	0.3	0.32	0.35	0.49	0.08	0.52	0.45	0.47
			2000	0.75	0.24	0.69	0.74	0.79	0.44	0.42	0.62	0.52	0.6	0.57	0.47	0.83	0.69	0.74
		3.9	400	0.84	0.71	0.89	0.81	0.71	0.38	0.5	0.67	0.56	0.57	0.49	0.45	0.72	0.52	0.56
			1000	0.86	0.99	0.97	0.9	0.96	0.69	0.95	0.93	0.81	0.83	0.61	0.95	0.92	0.75	0.77
			2000	0.89	1	0.99	0.95	0.97	0.64	1	0.97	0.74	0.83	0.65	1	0.99	0.75	0.83

Note. MD = Mahalanobis Distance class separation, N = sample size, GMM = growth mixture models, R² = class separation pseudo R² , U-GMM = unconditional GMM, GP-GMM = growth predictor GMM, DE-GMM = direct effect GMM, Cov= whether the residual covariance matrix of growth factors was allowed to vary between intercept and slope

covariates when sample sizes and distances increase. Among fit indices and statistics, the AIC was consistently the best index for class enumeration when sample sizes and Mahalanobis distances were at the smallest levels simulated. The BIC showed very a high detection rate but only at high class separation (MD = 3.9) and larger sample size (1,000 or larger). ABIC had a similar pattern with BIC but performed better than BIC at smaller class separation of 2.6 or smaller, or at class separation of 3.9 and sample size of 400. At the most favorable condition of class separation of 3.9 and sample size of 1,000 or larger, ABIC and BIC had similar detection rates. LMR and BLRT shared similar performances. Comparing across ABIC and the LRTs, under favorable conditions like class separation of 3.9 and sample size of 1,000 or above, ABIC performed better than LRTs in that ABIC had a higher detection rate and ABIC performed well in both GMM with and without covariates. Under the same conditions, the LRTs worked better for GMM without a covariate.

Study 2

Table 4 presents a comparison of class enumeration with AIC, BIC, ABIC, LMR, and BLRT across simulated conditions. For class enumeration with the AIC, GMM type (η^2 =.058), class separation (η^2 =.057), and sample size (η^2 =.015) had effect sizes larger than .01. In general GMM without covariates performed better than GMM with a covariate, and the DE-GMM outperformed the GP-GMM by small differences.

For BIC, class separation (η^2 =.574) had the largest effect size, followed by sample size (η^2 =.325), covariate effect (η^2 =.034), whether the residual covariance of intercept and slope was allowed to vary between classes (η^2 =.028), model (η^2 =.025), interaction of model and residual covariance allowed to vary between classes (η^2 =.019), interaction of model, class separation, and sample size (η^2 =.017), and four-way interaction of model, class separation, sample size, and whether the residual covariance was allowed to vary between classes (η^2 =.011). The BIC did not achieve a correct enumeration rate above 0.8 until class separation was at 2.6 in combination with a sample size of 2,000, or class separation at 3.9 with a sample size of 1,000 or larger. With the BIC, the GMM without covariates performed better than the two GMM with a covariate. When class separation and sample size were at the largest levels (MD=3.9, N=2,000), all growth mixture models had the same 100 % detection rate on BIC. GMM not allowing the residual covariance matrix to vary between classes resulted in slightly better class enumeration than when the residual covariance matrix was allowed to vary.

For ABIC, class separation (η^2 =.473) had the largest effect size, followed by the interaction of class separation by sample size (η^2 =.075), sample size (η^2 =.055), the interaction of class separation by sample size by model (η^2 =.014), covariate effect (p <.001,η^2 =.014), whether the residual covariance

matrix of growth factors was allowed to vary between classes (η^2 =.013), type of model (η^2 =.012), and the interaction of model and whether the residual covariance varied between classes (η^2 =.012). The U-GMM resulted in higher accuracy of class enumeration than both the DE-GMM and GP-GMM except when class separation and sample size were small (MD=1.3, N=400). When class separation was 3.9 and sample size was 2,000, detection rate for ABIC was almost all 100 %.

For the LMR, class separation (η^2 =.184) had the largest effect size, followed by sample size (η^2 =.079), the interaction of class separation by sample size (η^2 =.022). In general, GMM without covariates performed better than GMM with a covariate except when class separation and sample size were really small (MD = 1.3, N = 400). The DE-GMM resulted in better or equal class enumeration than the GP-GMM in most conditions. When class separation is large (MD=3.9) in combination with a sample size of 1,000 or larger, all three models detected the correct number of classes in at least 72 % of iterations.

For BLRT, class separation had the largest effect size (η^2 = .441), followed by sample size, interaction of class separation by sample size (η^2 = .023), interaction of class separation by model (η^2 = .011), and sample size (η^2 = .010). The U-GMM produced better class enumeration than the two models with covariates, but the DE-GMM outperformed the GP-GMM. However, when class separation is 3.9 in combination with a sample size of 1,000 or larger, all models had correct class enumeration of 92 % or higher.

In summary, the AIC only worked better than the other fit indices and statistics with small sample sizes and small distances. BIC was very sensitive to sample size and class separation, and showed very high detection rate at only high class separation (MD = 3.9) and larger sample size (1,000 or above). ABIC performed better than BIC especially when class separation was small (MD = 1.3). LMR performed slightly better than BLRT at small class separation (MD = 1.3) and small sample size (N = 400), while BLRT showed much higher and accurate detection rate than LMR at large class separation and sample size (MD = 3.9, N = 1,000 or above). The ABIC performed slightly better than BLRT at small sample size of 400. The GMM without covariate performed better than the GMM with covariates in general, and only when class separation and sample size were small (MD = 1.3, N = 400) do GMM with a covariate slightly outperform GMM without covariates. Allowing the residual covariance matrix of intercept and slope to vary between classes had no or little effect on the accuracy of class enumeration. In both studies, an interaction of sample size and class separation was observed, which supported the finding of Lubke and Neale (2006) that the smaller class separation can be partially compensated for by increasing the sample size to get a correct model. There were only small differences between GP-

Table 4 Percentage of correct model identification of fit indices in Study 2

Study 2				U-GMM					GP-GMM					DE-GMM				
Cov	R²	MD	N	AIC	BIC	ABIC	LMR	BLRT	AIC	BIC	ABIC	LMR	BLRT	AIC	BIC	ABIC	LMR	BLRT
No	0.1	1.3	400	0.24	0	0.16	0.12	0.02	0.28	0	0.22	0.2	0.07	0.24	0	0.26	0.23	0.07
			1000	0.42	0	24	0.33	0.3	0.24	0	0.04	0.26	0.08	0.24	0	0.1	0.25	0.13
			2000	0.64	0.04	0.42	0.49	0.57	0.25	0	0	0.24	0.08	0.26	0	0.04	0.33	0.12
		2.6	400	0.6	0.01	0.52	0.37	0.28	0.34	0.02	0.52	0.26	0.18	0.35	0.04	0.49	0.23	0.21
			1000	0.73	0.52	0.9	0.75	0.92	0.31	0.12	0.32	0.34	0.32	0.41	0.14	0.73	0.53	0.73
			2000	0.78	0.96	1	0.9	0.99	0.4	0.62	0.75	0.63	0.72	0.5	0.6	0.94	0.77	0.94
		3.9	400	0.68	0.56	0.85	0.73	0.74	0.38	0.25	0.7	0.52	0.58	0.4	0.28	0.7	0.44	0.6
			1000	0.8	0.99	0.97	0.88	0.96	0.52	0.9	0.98	0.8	0.95	0.47	0.78	0.98	0.76	0.97
			2000	0.78	1	1	0.86	0.99	0.53	1	1	0.84	0.96	0.52	1	1	0.86	1
	0.3	1.3	400	0.34	0.02	0.2	0.16	0.08	0.28	0	0.32	0.34	0.13	0.31	0	0.3	0.22	0.12
			1000	0.58	0.04	0.42	0.4	0.52	0.2	0	0.04	0.2	0.06	0.28	0	0.21	0.28	0.21
			2000	0.74	0.16	0.7	0.7	0.79	0.22	0	0.04	0.25	0.1	0.32	0.02	0.16	0.36	0.23
		2.6	400	0.57	0.21	0.66	0.5	0.44	0.3	0.1	0.52	0.34	0.32	0.36	0.08	0.56	0.34	0.39
			1000	0.68	0.76	0.94	0.83	0.94	0.35	0.3	0.63	0.54	0.6	0.42	0.3	0.79	0.58	0.79
			2000	0.77	1	1	0.88	0.98	0.4	0.9	0.98	0.76	0.91	0.46	0.9	1	0.84	0.98
		3.9	400	0.77	0.77	0.94	0.85	0.88	0.4	0.4	0.72	0.55	0.64	0.42	0.38	0.77	0.58	0.68
			1000	0.78	1	0.97	0.88	0.95	0.48	0.96	0.98	0.79	0.94	0.56	0.96	0.98	0.8	0.96
			2000	0.78	1	0.99	0.88	0.98	0.55	1	1	0.78	0.96	0.68	1	0.99	0.84	0.98
Yes	0.1	1.3	400	0.22	0	0.13	0.14	0.04	0.28	0	0.22	0.2	0.07	0.22	0	0.25	0.24	0.08
			1000	0.32	0	0.06	0.2	0.08	0.24	0	0.04	0.26	0.08	0.23	0	0.09	0.25	0.08
			2000	0.42	0.01	0.04	0.22	0.12	0.25	0	0	0.24	0.08	0.22	0	0.03	0.31	0.1
		2.6	400	0.42	0	0.36	0.24	0.14	0.34	0.02	0.52	0.26	0.18	0.32	0.02	0.34	0.22	0.15
			1000	0.65	0.12	0.58	0.55	0.6	0.31	0.12	0.32	0.34	0.32	0.4	0.04	0.52	0.44	0.5
			2000	0.77	0.44	0.82	0.78	0.94	0.4	0.62	0.75	0.63	0.72	0.48	0.24	0.76	0.68	0.79
		3.9	400	0.57	0.22	0.66	0.53	0.5	0.38	0.25	0.7	0.52	0.58	0.38	0.12	0.59	0.35	0.42
			1000	0.74	0.76	0.95	0.81	0.94	0.52	0.9	0.98	0.8	0.95	0.46	0.5	0.92	0.72	0.92
			2000	0.82	0.79	1	0.86	0.98	0.53	1	1	0.84	0.96	0.52	0.94	1	0.88	0.98
	0.3	1.3	400	0.22	0.01	0.11	0.12	0.06	0.28	0	0.32	0.34	0.13	0.32	0	0.3	0.22	0.12
			1000	0.29	0	0.1	0.18	0.12	0.2	0	0.04	0.2	0.06	0.3	0	0.16	0.28	0.18
			2000	0.48	0	0.1	0.28	0.24	0.22	0	0.04	0.25	0.1	0.28	0.02	0.14	0.32	0.22
		2.6	400	0.44	0.06	0.42	0.32	0.23	0.3	0.1	0.52	0.34	0.32	0.34	0.06	0.56	0.3	0.26
			1000	0.7	0.25	0.71	0.64	0.77	0.35	0.3	0.63	0.54	0.6	0.38	0.18	0.69	0.57	0.66
			2000	0.76	0.74	0.95	0.85	0.97	0.4	0.9	0.98	0.76	0.91	0.48	0.74	0.98	0.79	0.98
		3.9	400	0.74	0.4	0.82	0.59	0.7	0.4	0.4	0.72	0.55	0.64	0.36	0.26	0.73	0.51	0.58
			1000	0.75	0.92	0.98	0.88	0.96	0.48	0.96	0.98	0.79	0.94	0.49	0.86	0.98	0.78	0.95
			2000	0.75	1	0.99	0.86	0.98	0.55	1	1	0.78	0.96	0.66	1	0.99	0.84	0.98

Note. MD = Mahalanobis Distance class separation, N = sample size, GMM = growth mixture models, R^2 = class separation pseudo R^2, U-GMM = unconditional GMM, GP-GMM = growth predictor GMM, DE-GMM = direct effect GMM, Cov= whether the residual covariance matrix of growth factors was allowed to vary between intercept and slope

GMM and DE-GMM in terms of identifying the correct number of classes. Comparing the data simulating in Study 1 and Study 2, we found that allowing intercept and slope residual covariance matrix differences between two classes increased the detection rate of the AIC and ABIC under unfavorable conditions (*MD* = 1.3 *N* = 400) only slightly.

Discussion

While some methodological studies about class enumeration in growth mixture models did not include covariates in the generating model (e.g., Liu & Hancock, 2014; Nylund et al. 2007; Peugh & Fan, 2012; Yang, 2006), other studies such as

Li and Hser (2011) and Tofighi and Enders (2007) studied the GMM performance when the generating model had a covariate and compared performance of different growth mixture models. However, these papers only examined the GP-GMM but not the DE-GMM. Therefore, the purpose of this study was to conduct further evaluation on whether the inclusion of covariates in GMM was beneficial to class enumeration. Two GMM with covariates (i.e., the GP-GMM and DE-GMM) and an unconditional GMM were fit to the data generated based on the GP-GMM. Commonly used fit indices and test statistics for class enumeration were compared.

In our studies, we found that the inclusion of a covariate only helps with class enumeration when class separation and sample size are small (i.e., $MD = 1.3$ and $N = 400$) and in this case, it is better to fit the DE-GMM and the GP-GMM. However, the inclusion of a covariate in this case makes little practical difference because the percentages of correct class identification are low across all fit criteria. We consider that the difference between models is only of practical significance in conditions where using the unconditional model allows a percentage of correct identification above 70 % but the other models do not. Therefore, we conclude that under favorable conditions, such as class separation of 3.9 or class separation of 2.6 with a sample size of 2,000, GMM without covariates performs better than GMM with covariates. We found that our results are supported by the existing literature: Peugh and Fan (2014) concluded that adding covariates decreased the detection rate of BIC and BLRT under large sample size (i.e., $N = 3000$), large class separation ($MD = 2$), and large variance explained. Yet Peugh and Fan (2014) did not compare the performance of different GMM with and without covariates in the same paper with the same population model. Tofighi and Enders (2007) simulated a GP-GMM with covariates and concluded that the inclusion of covariates hampered GMM class extraction capability. However, it is worth noting that Tofighi and Enders (2007) reached this conclusion under the most favorable conditions in their study (i.e., sample size =1,000, largest class separation of 3 standard deviations which was equivalent to $MD = 3.9$ in our study; normal distribution within class; mixing percentages of 20: 33: 47 %), and therefore their conclusion is similar to the results of our study.

Li and Hser (2011) concluded that inclusion of a covariate was beneficial to GMM class enumeration when GMM misspecification was minor (i.e., defined as GMM without covariate fit to data generated by GMM with covariate and the same covariate mean). They simulated GMM with covariates and fixed the class separation as a class mean difference equivalent to 3 standard deviations, yet they varied the covariate mean difference between classes. Their results were different from our results, and this difference might be due to the fact that in Li and Hser's study (2011) covariates were binary or non-normally distributed, while the covariate was normally distributed in our study. Lubke and Muthén (2007) concluded

that including a covariate helped the class assignment of factor mixture models. They concluded that correct class assignment increased as covariate difference between classes increased (MD of covariate increased from .5 to 2). In our study, the covariate effect was defined as the variance that covariate explained the intercept and the slope as well as the effect on class membership, which was manipulated through the pseudo R^2. In Lubke and Muthén's study (2007) the covariate in the simulation model only predicted the class membership but not the intercept and the slope. These differences may explain why our results contrast with Li and Hser (2011) and Lubke and Muthén (2007), but agree with Tofighi and Enders (2007) and Peugh and Fan (2014).

On the performance of the fit indexes, the ABIC was the best indicator followed by the LRTs when class separation was large and sample size was large ($MD = 3.9$, $N = 1,000$ or larger). Peugh and Fan (2014) also found that sample size adjusted BIC performed better than LMR and BLRT when the class-mixing proportion was equal. Specific to our studies, ABIC showed higher detection rate and worked well under both GMM with covariate and the GMM without covariate. Under the same conditions, LMR and BLRT performed better in GMM without covariates. This conclusion is consistent with the findings of Peugh and Fan (2014), who found that adding covariates decreased the detection rate of BLRT comparing to GMM without covariates. We found that the BIC was sensitive to sample size and class separation, and only achieved satisfactory performance when class separation was 3.9 in combination with a sample size of 1,000 or larger. Under these conditions, GMM type did not impact the BIC's performance. Liu and Hancock (2014) also found BIC performed well in linear GMM with $MD = 2$ or larger, and large sample size. The AIC performed consistently especially in unconditional GMM, but did not show substantial improvement with increase of class separation conditions, as other indices did.

Our comparison of fit indices resulted in conclusions that are similar to previous studies: On the performance of ABIC, Yang (2006) also concluded that the ABIC was accurate in the context of latent class analysis. Tofighi and Enders (2007) concluded that ABIC detected the correct number of classes even when sample size was 400. Nylund et al. (2007) also concluded that the ABIC performed well in GMM when sample size was 500 or larger. On the performance of BIC, Yang (2006), Tofighi and Enders (2007), Nylund et al. (2007), and Liu and Hancock (2014) all reached the same conclusion that the BIC performed adequately when sample size was above 1,000. These conclusions about the sensitivity of BIC to sample size are not surprising since BIC is a function of sample size. On the performance of AIC, our results mirrored those of Yang (2006) and Nylund et al. (2007), who concluded that AIC performs poorly because it does not improve much as sample size increases.

The LRTs performed very well in class enumeration particularly with unconditional GMM when class separation was 3.9. LMR performed almost as well in class identification as BLRT. Our results about the performance of LMR supported the conclusion of Tofighi and Enders (2007), Nylund (2007), and Liu and Hancock (2014). Our results about the BLRT's performance are consistent with Nylund (2007). The poor performance of LRTs on GMM with covariates might be due to the model complexity of GMM with covariates.

Conclusion

The evaluation of inclusion of covariates in GMM under various specifications, the difference between DE-GMM and GP-GMM, and the performance of model fit indexes was the focus of this paper. Consistent with previous studies, we concluded that GMM depends on larger sample sizes and is difficult to implement with small sample sizes. When class separation is large and a sufficient sample size is available (e.g., 1,000), the ABIC, LMR, and BLRT can be employed together to detect the correct number of classes. Based on the findings of this study, the practice we recommend is to assess the class separation without covariates if the sample size is large. However, if the sample size is small (e.g., 400) and the class separation is not clear, we recommend including a covariate to improve class enumeration accuracy, with the awareness that class enumeration is frequently inadequate with small sample sizes. Covariates should be selected based on substantive theory and empirical analysis (Muthén, 2004). Also, when selecting covariates, care must be given to whether the predictor merely predicts the latent class membership or whether the predictor predicts the latent growth factor within class as well, or both (Bauer, 2007). If a predictor is to be included, the DE-GMM is recommended because it performs slightly better than GP-GMM on most occasions.

The findings of this study are limited by the fact that only linear growth models were examined. Furthermore, only one level of class proportions and number of waves was examined, and we simulated data where the covariate had effects on the intercept and slope but not direct effects on the outcome (i.e., the full mediation assumption discussed by Stoel, Van den Wittenboer, and Hox (2004) was held in the simulated data). Future research could focus on population models with nonlinear growth and multilevel levels of number of waves. We also restricted our study to only the most commonly used fit criteria, but a comprehensive evaluation of fit criteria, including the consistent AIC (C-AIC), Hannan and Quinn (HQ), Hurvich and Tsai AIC, normalized entropy criterion (NEC), classification likelihood criterion (CLC), integrated completed likelihood criterion (ICL), and adjusted Lo-Mendell-Rubin (aLMR), which were not examined in this study, can be found in Peugh and Fan (2014). We only analyzed

conditions where growth trajectory was correctly specified, so it remains an open question whether class identification with large sample sizes using different fit criteria are robust to some degree of misspecification of the growth trajectory.

A separate but related area of research is about how to include covariates in GMM when the interest is in either estimating the effect of covariates on class membership or estimating the effects of class membership on distal outcomes. In this area of research, several multiple-step alternatives to fitting a single GMM with covariates have been proposed, such as the pseudo class (PC) method (Clark & Muthén, 2009), Vermunt's three-step method (Vermunt, 2010), and the Lanza, Tan, and Bray (2013) method, and comparisons across these methods, such as the recent study by Asparouhov and Muthén (2014), is an area of active research.

References

Abroms, L., Simons-Morton, B., Haynie, D. L., & Chen, R. (2005). Psychosocial predictors of smoking trajectories during middle and high school. *Addiction, 100,* 852–861.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling, 21,* 329–341.

Bandalos, D. L., & Leite, W. (2013). The role of simulation in structural equation modeling. In G. R. Hancock & R. Mueller (Eds.), *A Second Course in Structural Equation Modeling* (2nd ed.). Greenwich, CT: Information Age Publishing.

Bandalos, D. L., & Leite, W. L. (2015). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 625–666). Greenwich: Information Age Publishing.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42,* 757–786.

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

Clark, S., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis.* Retrieved from https://www.statmodel.com/download/relatinglca.pdf

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology, 73,* 229–238.

Hu, B., Shao, J., & Palta, M. (2006). Pseudo-R^2 in logistic regression model. *Statistica Sinica, 16*(3), 847.

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.

Li, L., & Hser, Y. I. (2011). On Inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research, 46,* 266–302.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identificaiotnin 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.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14,* 26–47.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41,* 499–532.

McDonough, P., Sacker, A., & Wiggins, R. D. (2005). Time on my side? Life course trajectories of poverty and health. *Social Science and Medicine, 61,* 1795–1808.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36,* 318–324.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4,* 103–120.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling (pp. 1-33)*. Lawrence Erlbaum Associates.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29,* 81–117.

Muthén, B. O. (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). Newbury Park: Sage.

Muthén, B. (2006). The potential of growth mixture modelling. *Infant and Child Development, 15,* 623–625.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882–891.

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

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4,* 139–177.

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling, 19,* 204–226.

Peugh, J., & Fan, X. (2014). Enumeration index performance in generalized growth mixture models: A Monte Carlo test of Muthén's (2003) hypothesis. *Structural Equation Modeling, 21,* 1–17.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434–447.

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Stoel, R. D., Van den Wittenboer, G., & Hox, J. J. (2004). Including time-invariant covariates in the latent growth curve model. *Structural Equation Modeling, 11,* 155–167.

Stoolmiller, M., Kim, H. K., & Capaldi, D. M. (2005). The course of depressive symptoms in men from early adolescence to young adulthood: identifying latent trajectories and early predictors. *Journal of Abnormal Psychology, 114*(3), 331.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock (Ed.), *Mixture models in latent variable research* (pp. 317–341). Greenwich: Information Age.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18,* 450–469.

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics and Data Analysis, 50*(4), 1090–1104.