The Impact of Total and Partial Inclusion or Exclusion of Active and Inactive Time Invariant Covariates in Growth Mixture Models

Thierno M. O. Diallo*, Ph.D.

Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, Australia
& Statistiques & M.N., Sherbrooke, Canada

Alexandre J.S. Morin *, Ph.D.

Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, Australia
&

HuiZhong Lu, Ph.D.

Scientific Computer Center
University of Sherbrooke, Sherbrooke, Canada

* The authors (TMOD & AJSM) contributed equally to this article and both should be considered first authors.

This is the prepublication version of the following manuscript:

Diallo, T.M.O, Morin, A.J.S. & Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, *22*, 166-190.

©American Psychological Association, 2017. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: http://dx.doi.org/10.1037/met0000084

Acknowledgements
TMOD acknowledges Christian Tremblay for his technical support during the data analysis process.

Corresponding author:
Alexandre J.S. Morin,
Institute for Positive Psychology and Education
Australian Catholic University, Strathfield campus
Locked Bag 2002
Strathfield NSW 2135, Australia
Email: Alexandre.Morin@acu.edu.au

Abstract

This article evaluates the impact of partial or total covariate inclusion or exclusion on the class enumeration performance of growth mixture models (GMM). Study 1 examines the effect of including an inactive covariate in GMM when the population model is specified without covariates. Study 2 examines the case where the population model is specified with two covariates influencing only the class membership. Study 3 examines a population model including two covariates influencing the class membership and the growth factors. In all studies, we contrast the accuracy of various indicators (AIC, CAIC, SCAIC, BIC, SBIC, LMR/ALMR, BLRT) to correctly identify the number of latent classes in the data as a function of different design conditions (sample size, mixing ratio, invariance or non-invariance of the variance-covariance matrix, class separation and correlations between the covariates in Studies 2 and 3) and covariate specification (exclusion, partial or total inclusion as influencing class membership, partial or total inclusion as influencing class membership and the growth factors in a class-invariant or class-varying manner). The accuracy of the indicators shows important variation across studies, indicators, design conditions, and specification of the covariates effects. However, the results suggest that GMM class enumeration process should be conducted without covariates and should rely mostly on the BIC and CAIC as the most reliable indicators under conditions of high class separation (as indicated by higher entropy), versus the SBIC, BLRT, and SCAIC under conditions of low class separation (as indicated by lower entropy).

Keywords: Growth mixture modeling, covariate, partial and total inclusion, model complexity, class enumeration.

Growth Mixture Models (GMM; Muthén, & Shedden, 1999) overcome the main limitation of latent growth modeling, which assumes that all individuals under study are drawn from a single population for which common population parameters are estimated (e.g., intercepts, slopes, and error variances). Whereas latent growth models estimate the average trajectory for all individuals in the sample as well as inter-individual variability around this average trajectory, GMM allows for the identification of unobserved subpopulations through the combination of categorical and continuous latent variables. GMM are specifically designed to explain longitudinal heterogeneity by separating a general population into subgroups following qualitatively and quantitatively distinct trajectories (e.g., Morin, Maïano, Nagengast, Marsh, Morizot, & Janosz, 2011). These subgroups are usually referred to as latent profiles, or latent classes, and more closely related to prototypes than to physical groupings of persons. GMM estimate probabilities of profile membership for each individual in all of the latent classes (McLachlan & Peel, 2000; Muthén & Shedden, 1999), based on individual similarity with the specific configuration of variables represented by the latent classes. This characteristic makes GMM more naturally suited to research anchored into a person-centered perspective of longitudinal development than classical growth models (e.g., Bergman, Magnusson, & El-Khouri, 2003).

Like other types of mixture models (e.g., latent profile analyses, factor mixture analyses), GMM can incorporate covariates and estimate relationships between these covariates and individuals probabilities of membership into each class. One way to incorporate covariate effects into GMM is to directly conduct all analyses in a single step, incorporating covariates to the models from the start (the so-called one-step procedure). The one-step approach has many practical disadvantages, which Vermunt (2010, p. 451) summarizes in the following manner:

"The first is that it may sometimes be impractical, especially when the number of potential covariates is large, as will typically be the case in a more exploratory study. Each time that a covariate is added or removed not only the prediction model but also the measurement model needs to be reestimated. A second disadvantage is that it introduces additional model building problems, such as whether one should decide about the number of classes in a model with or without covariates. Third, the simultaneous approach does not fit with the logic of most applied researchers, who view

introducing covariates as a step that comes after the classification model has been built. Fourth, it assumes that the classification model is built in the same stage of a study as the model used to predict the class membership, which is not necessarily the case. "

An alternative three-step procedure has been recommended (Vermunt, 2010; Asparouhov & Muthen, 2014). In this alternative procedure, the first step consists of estimating the GMM using only the repeated measures from which the growth process is estimated. It is in this stage that the optimal number of latent classes to retain (for example, a two-class solution) is determined (i.e., the class enumeration process). Once the optimal number of latent classes has been determined, the second step aims to classify individuals into these latent classes, based on their probability of belonging into each class. Covariates are then integrated to the model in the third step, aiming to assess how they relate to class membership and other GMM parameters, while taking into account individuals' probabilities of membership into each of the latent classes.

The key question that remains is how well these two approaches perform in terms of statistical accuracy. In GMM, covariates can be used to predict latent class membership only, or also be allowed to predict growth parameters over and above class membership. In the first condition, covariate omission is unlikely to result in biased parameter estimates, as long as the within-class normality assumption is not violated. In the second case, the omission of active covariates is likely to result in biased parameter estimates of within-class parameters. However, keeping in mind that these covariates may be included in a later step, to what extent is this omission likely to bias the class-enumeration process? What about the inclusion of inactive covariates, or the partial omission of active covariates? A second element to consider is that the addition of covariates involves an increase in the number of parameters. Given that the class enumeration process is typically conducted based on the comparison of statistical indices which penalize model complexity, covariate inclusion may also reduce the statistical power of the class enumeration process. The current series of studies aims to bring some clarification to the possible role of covariates in the GMM class enumeration process.

Previous Studies on the Performance of GMM

Despite their extensive use, GMM still pose substantial challenges. One of these challenges is the extraction of the correct number of latent classes in the data (i.e., the class enumeration procedure). For instance, Bauer and Curran (2003) found that in the presence of nonnormality GMM often tended to overextract latent classes as a way to take this nonnormality into account (for proposed solutions see Muthén & Asparouhov, 2009, 2015). Similarly, model misspecification in the within class-covariance structure (Bauer & Curran, 2004) or unmodelled complex nonlinear trajectories (Peugh & Fan, 2012) can also lead to incorrect inferences regarding the number of classes in the data. These observations lead to the recommendation that GMM should be built carefully, supported by theory, and validated externally through examination of associations with meaningful covariates (Bauer & Curran, 2004; Marsh, Lüdtke, Trautwein, & Morin, 2009; Morin, Maïano et al., 2011; Muthén, 2003).

In this regard, the impact of including or excluding covariates in the class enumeration process has been considered in two studies. Tofighi and Enders (2007) studied the performance of nine different indicators used in GMM to help in the class enumeration procedure (e.g., AIC, BIC) in relation to the number of repeated measures, sample size, class separation, class proportions and the within-class distribution. Data were generated according to the following process (Scheme 1). First, covariates were generated from a known distribution. Second, class membership was generated from a multinomial distribution conditional on the covariate. Third, the repeated measures were generated from a conditional normal growth model (given covariate and class membership). Using Scheme 1, a population model with covariates predicting the latent class as well the growth factors, and class specific effect of the covariates, was considered. All models were estimated twice, with and without covariates (specified as predictors). Two within-class distributions (multivariate normal; nonnormal) were considered. Thus, estimating a GMM without covariate could result in a misspecification of the within-class distribution, particularly in the nonnormal scenario. Tofighi and Enders found that the ability of the indicators to recover the true number of latent classes increased when the normality condition was satisfied and when covariates were excluded– even though this last condition represented a misspecification as the covariates were part of the population generating model. The negative impact of covariates inclusion started to decrease when sample size reached 1000. As noted by the authors, one of the limitations of their study is related to the complexity of the population model where residual variances as well as covariates effects on the growth factors were class-specific.

More recently, Li and Hser (2011) also examined the impact of covariates inclusion on the GMM class enumeration procedure, using a different data generation scheme (Scheme 2) and a simpler population model (the growth factor variances and covariances as well as covariates effect on the growth factors were specified as invariant across classes, and the residual variances of the outcomes was specified as homoscedastic). Scheme 2 first generates class membership from a multinomial distribution, then generates the covariate given the class membership, and finally generates the repeated measures from a conditional normal growth model. Scheme 2 is more convenient for the manipulation of the within-class distribution of the covariates. As in Scheme 1, data generated with Scheme 2 and analyzed without covariates results in a misspecified within-class distribution when covariates are categorical or nonnormal. As shown in Li and Hser's (2011) Appendix B, the model implied by the two data generation schemes can be converted to one another in model form. However, except for special cases, the class membership of a model generated by Scheme 1 cannot be fully expressed by a model resulting from Scheme 2. For instance, the estimation of models including covariates effects on the growth factors (Scheme 1) provides a correctly-specified model for data generated using Scheme 2 in Li and Hser's (2011) Study 1. However, the models estimated by Tofighi and Enders (2007) remain in general misspecified for data generated by Scheme 2 in most of the other scenarios considered by Li and Hser (2011). Thus, in Li and Hser's (2011) Studies 2 and 3, when δ (the mean of covariate x in class 2) is equal to 1, the model estimated by Tofighi and Enders (2007) would be correctly-specified for data generated with Scheme 2, but would lack parsimony. In contrast, for data generated according to Scheme 2 when $\delta \neq 1$ in Li and Hser's (2011) Studies 2 and 3, a model similar to Tofighi and Enders' (2007) would not be correctly specified. In the situation where a model including covariates is exactly specified or correctly specified but unparsimonious in relation to the population model, Li and Hser's (2011) results showed that GMM tended to result in a more accurate class enumeration when covariates were included in the models than when they were excluded (producing minor to severe misspecifications). Conversely, in situations where a model including covariates is misspecified in relation to the population model, GMM tended to result in a more accurate class enumeration when covariates were excluded from the estimated models. Furthermore, the negative impact of covariates inclusion in the class enumeration process tended to increase with the degree of misspecification induced by the covariate inclusion.

Because in Tofighi and Enders (2007) a model with covariates is always correctly specified, the results of the two studies are comparable only when the model with covariates is exactly specified or correctly specified but unparsimonious in relation to the population model in Li and Hser's (2011). In this situation, Tofighi and Enders's (2007) results shows that GMM performs better without covariates, and that the negative impact of covariate inclusion is weakened with larger sample sizes. Conversely, in similar situations, Li and Hser (2011) results show that GMM tended to perform better with covariates. However, additional results by Li and Hser (2011) who also considered situations where including the covariate resulted in a misspecified model are far more nuanced. Unfortunately, these additional results cannot be compared to those reported by Tofighi and Enders (2007) who did not consider similar situations. This is worrisome as, in the few cases where a comparison of results is possible across studies, the conclusions proved different.

One explanation for these discrepancies in results between these two studies (when comparable) is potentially related to the complexity of the population model considered. Although one might argue that a simpler model is more parsimonious, one may also argue that complexity is necessary when the object of study is as complex as humans can be (e.g., Morin, Maïano et al., 2011; Morin & Marsh 2014). These contradictory findings, coupled with the limited alternative specification of covariates effects considered, suggest a need for a more comprehensive examination of these issues.

One area where both of studies might have erred on the side of too much complexity is in considering only the case where the covariates have an impact on the growth factors over and above their impact on the likelihood of membership in the latent classes. In GMM, the latent categorical variable representing the subpopulations aims at providing an efficient summary of the information imparted by the latent growth factors summarising the trajectories. Thus, effects of covariates on the growth factors over and above their impact on class membership suggest that this summary may not be fully efficient. Interestingly, the few applied studies in which this possibility was investigated (e.g., Morin, Maïano et al., 2011; 2014; Morin, Rodriguez et al., 2012) showed that covariates effects on the growth factors over and above their effects on class membership was the exception. Although these results, based on real data where no knowledge of the population model was available, remain anecdotic at best, they suggest that population models used in simulations should consider various forms of covariates effects as one of the manipulated factor – a main objective of the present study.

In contrast, an area where previous studies may have been too simplistic is by considering only conditions where all covariates were included or excluded. Realistically, partial inclusion of covariates is likely to be far more frequent in applied research given the unrealism of undergoing a data collection including all likely covariates, or even of being aware of all likely covariates. Indeed, in the real world, researchers often have no clear a priori understanding of which covariate(s) can significantly influence the latent class membership or/and the latent growth factors and even when they do, these covariates have not always been assessed in the available data set. Thus, simulations studies would do well to consider situations where important covariates are omitted from the model or where an inactive covariate (i.e., having no effect on the model) is erroneously included in the model.

In the present series of studies, our objective is to investigate the impact of partial or complete inclusion or exclusion of active or inactive covariates, and misspecifications of their effects, in the GMM class enumeration process. In doing so, we consider multiple design conditions related to model complexity typically ignored by previous studies. This study thus substantially extends on the conditions considered by Tofighi and Enders (2007) and Li and Hser (2011), in an effort to provide a more definitive answer to the question of covariate inclusion in the GMM class enumeration process.

The Specification of GMM

Muthén and Shedden (1999) introduced GMM to model heterogeneity in growth trajectories. A categorical latent variable is used to capture heterogeneity in the growth model parameters (i.e., intercept, linear and quadratic slopes). Here, we assume a quadratic model for the repeated measures (within-class, level-1). Thus, for individual i in class k $(k = 1, 2, ..., K)$ at time t $(t = 1, 2, ..., T)$:

$$y^k{}_{it} = \eta^k{}_{0i} + \eta^k{}_{1i}.\lambda_t + \eta^k{}_{2i}.\lambda^2{}_t + \varepsilon^k{}_{it} \quad (1.1)$$

where $y^k{}_{it}$ is the outcome variable, $\lambda_t = t - 1$ is the time score at time t, $\eta^k{}_{0i}$ is the random intercept, $\eta^{k}{}_{1i}$ the random linear slope, $\eta^{k}{}_{2i}$ the quadratic slope, and $\varepsilon^{k}{}_{it} \sim N(0,\psi_t)$ is the normally distributed residual for $y^{k}{}_{it}$ which is assumed to be independent across time. In this study, we specify the residual variance, ψ_t, to be equal across time points and latent classes. We also specify a quadratic growth process where the time scores are fixed and invariant across classes. For class k, the person-level (within class, level 2) model is specified as:

$$\begin{aligned}\eta^{k}{}_{0i} &= \gamma^{k}{}_{00} + \gamma^{k}{}_{10}x_{i1} + \ldots + \gamma^{k}{}_{P0}x_{ip} + \xi^{k}{}_{i0} \\ \eta^{k}{}_{1i} &= \gamma^{k}{}_{01} + \gamma^{k}{}_{11}x_{i1} + \ldots + \gamma^{k}{}_{P1}x_{ip} + \xi^{k}{}_{i1} \\ \eta^{k}{}_{2i} &= \gamma^{k}{}_{02}\end{aligned} \tag{1.2}$$

where $x_{i1},\ldots,x_{ip}$ are level 2 covariates, $\gamma^{k}{}_{00}, \gamma^{k}{}_{01}, \gamma^{k}{}_{02}$ are the intercepts (or means when the covariates are centered at zero or omitted) of the growth factors η_i in k th class, all the other $\gamma^{k}{}_{mn}$ s are the linear regression coefficients of the growth factors η_i on the level 2 covariates in the k th class; and $\xi^{k}{}_{i0}$ and $\xi^{k}{}_{i1}$ are the multivariate normal distributed residuals of growth factors with a zero mean vector, $Var(\xi^{k}{}_{i0}) = \sigma^{k}{}_{00}, Var(\xi^{k}{}_{i1}) = \sigma^{k}{}_{11},$ and $Cov(\xi^{k}{}_{i0}, \xi^{k}{}_{i1}) = \sigma^{k}{}_{01}$. To follow Tofighi and Enders (2007) and Li and Hser (2011), we specify no random error and no covariate effect for the quadratic slope factor in Equation (1.2), although such effects can be included in GMM.

The prediction of the latent class variable by the level 2 covariates is done through a multinomial logistic regression model for the *K* classes. The probability that individual i is a member of class k is thus specified as conditional on the covariates and modeled as:

$$P(c_{ik} = 1 \mid x_1,\ldots,x_P) = \frac{\exp(a_k + \sum_{s=1}^{P} b_{ks}.x_{is})}{\sum_{k=1}^{K}\exp(a_k + \sum_{s=1}^{P} b_{ks}.x_{is})} \tag{1.3}$$

with $a_K = 0, b_{Ks} = 0,$ and a_k and b_{ks} are logit intercept and slope, respectively.

Method: Overall Presentation of the Three Studies

Population Models

To investigate the impact of total and partial inclusion or exclusion of active and inactive covariates in GMM, three studies with different population models are considered. The full set of population parameters used in these studies is provided in Tables S1-S5 of the online supplements.

Study 1: The population model includes no covariate.

Study 2: The population model includes 2 covariates influencing only class membership.

Study 3: The population model includes 2 covariates influencing class membership, the intercept factor, and the linear slope factor.

The parameters for the population model were chosen after an extensive review of previous simulation studies and substantive applications of GMM. First, the number of classes in simulated data for mixture models vary generally between 2 (Chen, Kwok, Luo, &Willson, 2010; Li & Hser, 2011; Liu & Hancock, 2014; Nylund, Asparouhov & Muthén, 2007; Tovanen, 2007) and 3 (Tein, Coxe, & Cham, 2013; Peugh & Fan, 2012, 2013; Tofighi & Enders, 2007). As noted by Tofighi and Enders (2007) the number of extracted classes in applied research similarly typically varies between 2 and 4. Given the number of conditions considered in the current series of studies, and the computational intensive nature of GMM, we selected a 2 class population model. This choice is aligned with what has been done in the majority of previous simulations studies of GMM.

Second, to maximise comparability of our results with Li and Hser (2011), we choose the same mean growth trajectories parameters in the two latent classes. Therefore, our three studies assume a quadratic population model with two classes presenting the same quadratic slope mean. The growth trajectories variance-covariance parameters were varied depending on simulation conditions.

Third, the number of time points in simulation studies of GMM vary between 4 and 7 (Chen et al., 2010; Li & Hser, 2011; Liu & Hancock, 2014; Nylund et al., 2007; Peugh & Fan, 2012; Tofighi & Enders, 2007; Tovanen, 2007). Based on Tofighi and Enders' (2007) results showing that this factor only had a minimal impact on class enumeration, this factor was not allowed to vary and set to 7 time points to maximise comparability with Li & Hser (2011) and Tofighi and Enders (2007). In the studies presented here, times values are specified as ranging from 0 to 6 with unit increments.

Manipulated Factors

On the basis of previous results showing that these factors had an impact on the ability to identify the true number of latent classes in GMM (Enders & Tofighi, 2008; Li, & Hser, 2011; Lubke &

Muthén, 2007; Peugh & Fan, 2012; Tofighi & Enders, 2007), five factors were manipulated: the sample size, the mixing ratio, the within-class variance-covariance matrices, the class separation, and the correlation between the two covariates. A challenge in simulation studies is to put a limit on the design conditions that are considered in order to keep the studies within reasonable computational boundaries. In particular, even though nonnormality has been shown to impact on the performance of GMM (Bauer & Curran, 2003; Tofighi & Enders, 2007), we elected to focus here on conditions where the normality condition was satisfied, but acknowledge this as an important direction for future study.

Sample Size. Previous research has shown sample size to be critical in the ability to recover the true number of classes present in the data (Li & Hser, 2011; Nylund et al., 2007; Tofighi & Enders, 2007; Yang, 2006). In our studies, data was generated with 4 different sample sizes: 200, 400, 1000 and 2000. These values replicate those used by Li and Hser (2011) reflecting small, moderate, large, and very large sample sizes typically observed in applied GMM research. Indeed, when one looks at published GMM studies, these sample sizes roughly cover a range values corresponding to the 25th to 75th percentiles of the sample size distribution found in these studies (Tofighi & Enders, 2007).

Mixing Ratio. Previous research shows that it is easier to recover the true number of classes present in the data when the relative size of the classes is similar than widely different, particularly when the global sample size is smaller (e.g., Nylund et al., 2007; Tofighi & Enders, 2007). Furthermore, previous simulation studies with 2-class population models generally used mixing proportions of 0.50/0.50 and 0.25/0.75 (e.g., Nylund et al., 2007; Liu & Hancock, 2014). In this study, data was simulated according to balanced (.50/.50) and unbalanced ratios. In this study we selected the mixing ratio of 0.30/0.70 for the unbalanced condition, which is slightly higher than .25/.75 but gave us more latitude to explore more extreme conditions which were finally not retained[1].

Within-Class Variance-Covariance Matrices. Following previous results showing that recovering the true number of classes present in the data in GMM was facilitated when the latent variance-covariance matrix of the growth factors was invariant across classes (e.g., Bauer & Curran, 2003) two distinct population models were simulated[2]. In the first condition, the latent growth factors variance-covariance matrix was simulated as invariant across classes. In the second condition, the latent growth factors variance-covariance matrix was specified as non-invariant across classes. In the presence of covariates, the (residual) variance components also depend on the size of the regression coefficients. Thus, the regression coefficients were held equal across classes in the invariance condition, and were simulated as class-specific for the non-invariance condition. The population parameters used for these conditions are presented in Tables S1 to S5 in the online supplements. In summary, in the non-invariance condition, the latent growth factors were simulated with a slightly higher level of variability in one class than in the other, and with a slightly lower level of covariance. The variance-covariance matrices were always correctly specified in the estimated models.

Class Separation. Another factor that has been found to influence class enumeration accuracy (Lubke & Muthén, 2007; Peugh & Fan, 2012; Tofighi & Enders, 2007) is the degree of separation between the latent classes, with classes showing a lower level of overlap being easier to identify. Like others (Nylund et al. 2007; Tofighi & Enders, 2007; Tein et al., 2013) we used the Cohen's d measure to define class separation based on manipulating the intercept (or initial time point) of the latent growth factors in the different classes. Following in part from these previous studies, but wanting to consider slightly more extreme conditions, we selected class separation levels corresponding approximately to 0.50 SD (low class separation condition) and 3.00 SD (high class separation condition) between the intercept means of the two classes. These values roughly correspond to the minimum and maximum class separation conditions used in previous studies. In the current series of studies, conditions of high and low class separation had to be implemented in a slightly different manner in Studies 1-2 versus 3, due to the effects of covariates on the growth factors in Study 3. Importantly, when covariates are included in the population model, their omission may result in a decrease in class separation – and thus in class enumeration accuracy. We more fully address the implementation of the class separation conditions across the three studies, and the impact of different specifications of covariates effects on class separation, in the online supplements.

Correlation Between Covariates. Although this is the first study to consider this condition, data from studies 2 and 3 were simulated according to two distinct values for the correlations between the covariates. The assumption behind this decision is that more highly correlated covariates bring less unique information and, as such, the exclusion of one covariate may be more easily compensated by the inclusion of the other when this correlation is higher. Corresponding to low and moderate effect sizes (Cohen, 1988), the two population values chosen for this correlation are 0.20 and 0.50.

Alternative Specifications of Covariates Effects

To assess the impact of covariate inclusion/exclusion for various forms of misspecifications, each design condition was analyzed using multiple alternative models. In Study 1 we first analyzed the data without including any covariates, in accordance with the population model. Then an inactive covariate was introduced. In a second model, this inactive covariate was specified as influencing latent class membership. In a third model, it was specified as influencing latent class membership and the growth factors, and these effects were specified as equal (invariant) from one class to the other. In a fourth model, this inactive covariate was specified as influencing latent class membership and the growth factors, and this effect was allowed to differ across classes (non-invariant). In this study 1, models excluding covariates were correctly specified, while models including covariates are over-specified through the estimation of parameters equal to zero in the population model. A two-class model included 10 (without covariate), 11 (covariate influencing class membership only), 13 (covariate influencing class membership and growth factor with invariance), and 15 (covariate influencing class membership and growth factor with non-invariance) free parameters.

In Studies 2 and 3, a model including none of the active covariates from the population model was first analyzed. Second, a model including the first (X1) covariate was estimated, and this covariate was specified as influencing class membership. Third, a model was estimated in which X1 was specified as influencing class membership and the growth factors (invariant). Fourth, a model was specified in which X1 was specified as influencing class membership and the growth factors (non-invariant). Steps two, three, and four were then replicated including only the second covariate (X2) and then including X1 and X2, for a total of 10 distinct models for Studies 2 and 3. For these studies, a two-class model included 10 (without covariates), 11 (one covariate influencing class membership only), 12 (two covariates influencing class membership only), 13 (one covariate influencing class membership and growth factors with invariance), 16 (two covariates influencing class membership and growth factors with invariance), 15 (one covariate influencing class membership and growth factors with non-invariance) and 20 (two covariates influencing class membership and growth factors with non-invariance) free parameters.

In Study 2, the model including X1 and X2 influencing only class membership is correctly specified. In contrast, models including X1 and X2 influencing class membership and the growth factors (invariant or non-invariant) are consistent with the population model but over-specified (i.e., including parameters equal to zero in the population). The model without covariates is also consistent with the population model, although not exactly-specified (i.e., omitting active covariates unlikely to influence class enumeration, see Li & Hser, 2011). Models of partial inclusion of X1 or X2 are inconsistent with the population model and thus misspecified (see online supplements for details).

In Study 3, models without covariates, with partial covariate inclusion, or with both covariates specified as influencing only class membership are misspecified since they omit necessary parameters from the population model. Models including X1 and X2 influencing class membership and the growth factors are consistent with the population model and correctly-specified when corresponding to the simulation condition for the within-class variance-covariance matrices, misspecified when their effects are constrained to invariance when the variance-covariance matrix is not invariant, or over-specified when their effects are estimated as not-invariant but invariant in the population model.

These three studies make this stimulation different from previous studies. First, when Tofighi and Enders (2007) included covariates, their effects were always correctly specified. In contrast, the current study and Li and Hser (2011) consider alternative misspecified models including covariates. The current study also considers over-parameterized models including inactive covariates. However, whereas the current study considers misspecification in the form of partial covariate inclusion, in Li and Hser (2011) the misspecification stemmed from the incorrect class membership prediction. In addition, whereas we consider the impact of mixing ratio and invariant/non-invariant variance-covariance matrices in the present study, Li and Hser (2011) only considered situations of equal mixing ratio and invariant variance-covariance matrices. However, Li and Hser (2011) also considered the situation where the model distribution conditional on class membership was non-normal, whereas the current study only considers the normality condition. Overall, we believe that the incorporation of these new considerations offers an important complement to the results from previous studies in terms of providing a richer set of recommendations for applied research.

Partial Factorial Design

GMM are complex, computer intensive, and time consuming models. In total, Study 1 considered 128 design cells: 4 sample sizes x 2 class separations x 2 configurations of the growth factor variance-covariance matrix x 2 mixing ratio x 4 types of analyses. Study 2 and 3 each considered 640 design cells: 4 sample sizes x 2 class separations x 2 configurations of the growth factor variance-covariance matrix x 2 mixing ratio x 10 types of analyses x 2 levels of correlations between X1 and X2. Given the total number of design cells (1408 in total), we relied on a partial factorial design (Beauchaine & Beauchaine, 2002; Tofighi & Enders, 2007). The design required the definition of a normative condition that was more likely to represent the best case scenario. This normative scenario was then used as a constant from which each specific design condition was systematically varied. The normative condition corresponded to a sample size of 1000, a mixing ratio of 0.50/0.50, an invariant within-class variance-covariance matrix, and a high level of class separation. The basic set of seven conditions considered in each study is presented in Table 1. Study 1 thus considers 28 conditions (7 basic conditions x 4 types of analyses) while studies 2 and 3 each consider 140 conditions (7 basic conditions x 2 correlations values x 10 types of analyses). Thus, this partial factorial design allow us to consider a total of 308 (28+140+140) conditions, rather than 1408.

Data Generation and Evaluation

All models were simulated and estimated using Mplus 7.0 Monte Carlo feature (Muthén & Muthén, 2012). A total of 1000 replications were generated within each analyzed cell, using a multivariate normal distribution (Muthén & Muthén, 2001). To verify the ability of the models to identify the correct number of latent classes in the data (*k* = 2 in all conditions considered here) models including 1 to 4 latent classes were systematically evaluated for all of the design conditions considered here. For 2 classes or more, we used 400 sets of random starting values and 40 iterations for each sets of the random starting values (e.g., Hipp & Bauer, 2006). The 20 solutions with the highest likelihood values were selected for final stage iteration. In model evaluation, models that failed to converge on a solution, or that converged on an improper solution, were discarded and all results presented here are based on proper solutions only[3]. Still, we briefly discuss rates of convergence associated with each study, and report exact rates of nonconvergence and improper solutions in Tables S67 to S69 at the end of the online supplements. Given the complexity and computer-intensive nature of the models considered here, this study was only made possible through Compute Canada high performance computing facilities (https://computecanada.ca/).

For model evaluation the following information criteria (IC) were used: the Bayesian Information Criterion (BIC; Schwartz, 1978), the sample size adjusted BIC (SBIC; Sclove, 1987), the Akaike Information Criterion (AIC; Akaike, 1987), the consistent AIC (CAIC; Bozdogan, 1987), and a sample size adjusted CAIC (SCAIC; Tofighi & Enders, 2007). Additional likelihood-ratio based tests (LRT) were also used: the Lo-Mendell-Rubin likelihood ratio test (LMR; Lo, Mendell, & Rubin, 2001), the adjusted LMR likelihood ratio test (ALMR; Lo et al., 2001; Vuong, 1989), and the bootstrapped likelihood ratio test (BLRT; McLachlan, 1987; McLachlan & Peel, 2000). The default method was used to calculate the BLRT and 40 draws were requested for the model with k classes in the initial stage and were followed by 10 final stage optimisations. For BIC, SBIC, AIC, CAIC, SCAIC the model with the lowest value among the four competing models specified for the sample data ($k = 1$, $k = 2$, $k = 3$, or $k = 4$) indicates the best fitting model. For the LMR, ALMR, and BLRT, p-values were used to choose between a k and a $k + 1$ class model. Following Nylund et al. (2007), a process starting from a one-class model and increasing sequentially the number of classes to select the optimal number of classes was followed. Thus, among the four competing models specified for the sample data, as soon as a significant p-value for a model with k classes was followed by a non-significant p-value for a model with $k + 1$ classes, the model with k classes was retained as the best fitting model. Recent simulation studies indicate that the BIC, SBIC, CAIC, and BLRT are particularly effective in choosing the model which best recovers the population's true parameters in mixture models (Henson, Reise, & Kim, 2007; McLachlan & Peel, 2000; Nylund et al., 2007; Peugh & Fan, 2013; Tein et al., 2013; Tofighi & Enders, 2007; Yang, 2006). Furthermore, when the indicators fail to retain the optimal model, the AIC, ABIC, and BLRT tend to overestimate the number of classes, whereas the BIC, LMR, and CAIC tends to underestimate it.

Outcomes

The outcomes considered here are the performance of the ICs and LRTs to correctly identify the presence of 2 latent classes in the data. This is reflected by the proportion of replications in which each indicator correctly identified the 2-class model as the correct solution.

Study 1: Population Model Excluding Covariates

In Study 1 we examined the effect of over-parameterization (induced by the inclusion of inactive covariates) versus exact parameterization on the class enumeration performance of GMM.

Sample Size

Tables S6-S7 of the online supplements provide the class enumeration performance of each indicator as a function of sample size. For the correctly specified model – without covariates – the influence of sample size remained small. Among the ICs, the AIC clearly did not perform as well than the other ICs. On average the AIC accurately enumerated the number of classes approximately 74% of the time across sample size conditions (72.10% to 77.70%) and pointed to the three-class solution more than 20% of time. The performance of the CAIC, SCAIC, BIC and SBIC was better, and comparable to one another. The CAIC and the BIC selected the 2 class solution 100% of the time across sample sizes. The performance of the SCAIC ranged from 95.80% for N=200 to 100% for N=2000 and that of the SBIC from 81.50% for N=200 to 99.50% for N=2000. Among the LRTs, the BLRT performed slightly better than the LMR/ALMR, while these two indicators had a performance that was undistinguishable from one another and similar to that of the AIC. The BLRT accurately identified the correct number of classes 89% of the time across sample size conditions (86.20%-90.40%) and pointed to the three class solution 10% of the time. Conversely, the LMR/ALMR accurately identified the correct number of classes approximately 76%/77% of the time across sample size conditions (73.60%-82.90%) and pointed to the three class solution more than 20% of time.

When an inactive covariate was incorrectly specified as influencing class membership, the impact of sample size on class enumeration differed across indicators. For small sample sizes of N=200, all ICs measures systematically supported the three-class solutions (69.80% for the AIC, 100% for the CAIC and the BIC, 95.30% for the SCAIC, and 76.30% for the SBIC) whereas the LRTs still supported the two-class solution more than 80% of time (80.50% for the LMR, 81.20% for the ALMR, and 87.20% for the BLRT). When sample size was $\geq$ 400, all indicators accurately identified the correct number of latent class with the same pattern of results as those observed in the correctly-specified model without covariates. Finally, when the covariate was incorrectly specified as influencing class membership as well as the growth factors (with or without invariance), the performance of the indicators was comparable across sample sizes and the same pattern of results as those obtained in the correctly-specified model without covariates was observed.

Class Separation

Table 2 provides the class enumeration performance for the low and high class separation conditions. As noted previously, the other design factors were held constant at: N = 1000, invariant variance-covariance matrix, and mixing ratio of 0.50/0.50. The influence of class separation on the ability to correctly identify the real number of classes varied importantly across indicators. Table 2 showed that it was easier to identify the correct number of classes when classes were well separated. For the correctly specified model without covariates and high class separation, the AIC selected the two class solution 72.90% of the time and the three class solution 21.20% of the time, the CAIC, SCAIC, BIC and SBIC selected the two class solution 100% of the time, the LMR/ALMR selected the two class solution 74% of the time and the three class solution more than 20% of the time, and the BLRT selected the two class solution 90.20% of the time. When the classes were not well separated, the AIC and LMR/ALMR performed as in the high separation condition and selected the correct number of classes more than 70% of time. The performance of the SCAIC showed a substantial decrease in this condition and selected the one class solution 21.90% of the time (versus 0% in the high separation condition). The class separation condition had more impact on the BIC and CAIC. These indicators underestimated the number of classes by selecting the one class solution more than 50% of the time (58.40% for BIC and 72.10% for CAIC) in the low separation condition, and the correct number of classes less than 45% of the time (41.60% for BIC, 27.90% for CAIC).

When an inactive covariate was incorrectly specified as influencing class membership, the impact of class separation on the class enumeration process was attenuated. In this condition, all indicators correctly detected the number of classes present in the data most of the time. The performance of the ICs was comparable to one another in the two conditions, with the exception of the AIC which did not perform as well in the low separation condition (59.80%) as in the high separation condition (70.90%). The AIC selected the three class solution more than 20% of the time (26.40% for the low and 21.00% for the high separation conditions). The CAIC, SCAIC, BIC and SBIC correctly identified the two class solution almost 100% of the time across separation conditions. In the high separation condition, all ICs performed equally well (86%), while the BLRT proved more efficient (75.40%) than the LMR/ALMR (67.30%/68.10%) in the low separation condition. The CAIC, SCAIC, BIC and SBIC tended to outperform the BLRT in models including an inactive covariate, particularly when the class separation was low.

When the inactive covariate was specified as influencing the class membership and the latent growth factors in a similar manner in all classes (invariance), class separation had a dramatic impact on the class enumeration process, particularly in relation to the ICs. Among the ICs only the SBIC proved robust to this misspecification and detected the correct number of classes in the data 99.10% of the time when the class separation was high, and 82% of the time when the class-separation was low. In contrast, in the low separation condition, the CAIC and BIC incorrectly supported the one class solution more than 70% of time (88.20% for CAIC, 77.10% for BIC) and supported the two class solution less than 25% of the time (11.80% for CAIC, 22.90% for BIC), the SCAIC incorrectly supported the one class solution 35.60% of time and the two class solution 64.20% of the time, while the AIC incorrectly supported the three class solution 27.50% of time and supported the two class solution 52.20% of the time. In the high separation condition, all of these indicators performed better and correctly selected the two class solution 75.40% for the AIC, 100% for the CAIC, SCAIC and the BIC, and 99.10% for the SBIC. The performance of the LRTs was comparable to that of the SCAIC with a slight advantage to the BLRT which correctly identified the two class solution 68.10% and 91.30% of the time in the low and high separation conditions respectively.

Finally, when the covariate was incorrectly specified as influencing the class membership and the latent growth factors freely across classes (non-invariance), the impact of the class separation was even more pronounced. For instance in the low separation condition, the accuracy of the SBIC dropped to 61.10% whereas the CAIC incorrectly retained the one class solution 98.30% of the time, the SCAIC 65% of the time, and the BIC 94.70% of the time. The AIC and LRTs performed slightly better in the low separation condition. The AIC selected the two class solution 51.40% of the time in the low separation condition versus 69.70% of the time in the high separation condition. The LMR/ALMR selected two classes 59.60%/59.30% of the time (versus 23.20%/24.20% for one class) in the low separation condition versus 84.70%/85.70% in the high separation condition (and three classes 13.20%/12.40% of the time). The BLRT selected two classes 69.20% of the time, one class 13.80% and three classes 13.30% in the low separation condition, and correctly selected two classes 92.60% of the time in the high separation condition.

Mixing Ratio

Table S8 of the online supplements shows that the mixing ratio conditions had a negligible impact on the class enumeration performance of the indicators. The accuracy of the various indicators, as well the difference in performance between the indicators, remained approximately the same across the two set of mixing ratio considered. Across the two mixing ratio, all indicators correctly selected the two class solution more than 75 % of the time, except the AIC which retained the correct model less than 70% of the time. Across mixing ratios, the CAIC, SCAIC, BIC and the SBIC performed comparably and slightly better than the BLRT which itself performed better than other LRTs.

Variance-Covariance Matrix

Table S9 of the online supplements gives the class enumeration performance for the conditions of invariant and non-invariant variance-covariance matrices. The influence of this design factor in the ability to correctly enumerate the number of classes present in the data varied across indicators. First, this condition apparently had no effect on the performance of the ICs, which selected the correct two class solution almost 100% of the time across conditions and models, with the exception of the AIC which had a slightly better performance in the presence of a non-invariant variance-covariance matrix (90% versus 72% for the invariant condition). In contrast, LRTs performed better in the invariant condition by correctly selecting the two class solution more than 80% of times, versus less than 60% in the non-invariant condition. The BLRT seemed particularly sensitive to the structure of the variance-covariance matrix. In particular its accuracy dropped from 90.20% for the correctly specified model to 13.60% for the model with the inactive covariate specified as influencing class membership and latent growth factor (with invariant effects) in the condition where the variance-covariance matrix was non-invariant. Conversely, the performance of the LMR/ALMR increased from 30% for the correctly specified model to 50% for the model with the inactive covariate specified as influencing class membership and latent growth factor (with invariant effects) in the condition where the variance-covariance matrix was non-invariant. Comparable figures for the invariant variance-covariance matrix were 90.20% and 92.60% for the BLRT, and 75% and 85% for the LMR/ALMR. Generally, when they failed to select the proper two class model, the LRTs tended to over-extract latent classes.

Nonconvergence and Improper Solutions

As noted above, all results reported so far are based on models which converged on proper solutions. In Study 1, most models converged on proper solution, with rates of nonconvergence and improper solutions ranging respectively from 0 to 3.60%, and from 0 to 1.60% across conditions (see Table S67 of the online supplements for details).

Summary and Discussion

A summary of the results from Study 1 is presented in Table 3. In Study 1, most indicators presented high rates of accuracy to correctly identify the two class solution. The CAIC, SCAIC, BIC and SBIC correctly identified the number of classes present in the data at least 80% of the time across conditions. The performance of the AIC was significantly lower with an accuracy of 53.29% across conditions and a tendency to over extract classes 30.04% of the time. The CAIC, SCAIC, BIC and SBIC performed better than the LRTs. The BLRT identified the correct number of classes 70% of the time whereas the LMR/ALMR had an accuracy of 67% across conditions. The LRTs tended to overextract classes when they failed to correctly identify the two class solution: Across conditions, the BLRT selected the three class solution 22.69% of the times; the LMR/ALMR did so 27% of the time.

The performance of the indicators was generally stable across the range of sample sizes and mixing ratio considered, with one main exception. Indeed, the ICs – but not the LRTs – seemed to be more sensitive to minor misspecification of inactive covariate effects (i.e., covariate influencing only the class membership) with small sample size (N = 200) than to more severe misspecifications (i.e., covariate influencing the class membership and the growth factors with or without invariance). This observation suggests that, based on the results from Study 1, it may be safer to conduct the class enumeration process without including covariates when the available sample size is small, or at least to devote more attention to the BLRT when covariates need to be included. Similarly, the variance-covariance matrix also had a negligible impact on the performance of the ICs (with the exception of the AIC, which became more efficient in the presence of non-invariant variance-covariance matrices). However, this condition had a considerable impact on the performance of the LRTs. In particular, with non-invariant variance-covariance matrices, the LRTs tended to over-extract latent classes, suggesting that the ICs should be trusted more than the LRTs in this condition. In contrast, the class separation condition had an important impact on class enumeration, with the identification of the correct number of classes made considerably harder in a low separation condition. In this situation, the most accurate indicators proved to be the SBIC and BLRT.

In comparison with the estimation of correctly specified model including no covariates, the results show that the incorrect inclusion of an inactive covariate in the model generally resulted a slight improvement in performance for the CAIC (90.43% versus 86.79% across conditions), SCAIC (87.88% versus 85.21% across conditions), and BIC (90.40% versus 87.57% across conditions), in a slight deterioration of performance for the AIC (49.47% versus 64.71% across conditions), LMR/ALMR (64.38%/65.26% versus 71.71%/73.07% across conditions), and BLRT (67.95% versus 76.14% across conditions) and in no changes for the SBIC (80.71% versus 81.14% across conditions). The improvement in the accuracy of the CAIC, SCAIC, and BIC associated with incorrectly allowing an inactive covariate to influence class membership is particularly marked in the low class separation condition. In this condition, the inclusion of an inactive covariate allowed these indices to reach a level of performance comparable to the SBIC and BLRT. However, the results also show that more severe forms of misspecifications were clearly harmful to the performance of the indicators, especially in conditions of low class separation, such that including direct effects from the inactive covariates to the growth factors resulted in an important decrease in the ability to correctly identify the true number of latent classes present in the data. Thus, based on Study 1, it appears that allowing covariates, even inactive ones, to influence class membership in situations where $N > 200$ may help to recover the true number of latent classes in the data. Given the impact of the class separation condition on the efficacy of the various indicators, it may be useful to examine the entropy (as an indicator of class separation[4]) in order to select which indicators should be privileged in the class enumeration process. Our results suggest that the SBIC and BLRT should receive more attention at low levels of entropy.

Finally, the results show that there is a risk of using the AIC and LMR/ALMR class enumeration process (also see Henson et al., 2007; Nylund et al., 2007; Peugh & Fan, 2013; Tofighi & Enders, 2007; Yang, 2006). In fact, across design conditions, the performance of the AIC, and to a lesser extent the LMR/ALMR, proved to be suboptimal, inferior to that of the other ICs and to the BLRT, and more sensitive to design conditions. This pattern of results was also fully replicated in Studies 2 and 3. Thus, to avoid making this paper longer than necessary, we avoid further discussions of these indicators in Studies 2 and 3, although all numerical results are still provided.

Study 2: Population Model with Covariates Predicting Class Membership

In Study 2 we examined the effect of exact, over and under-parameterization on the class enumeration performance GMM through the use of population models including covariates specified as influencing class membership only.

Sample Size

Tables S10-S15 of the online supplements present the class enumeration performance of the indicators by sample size for a correlation between X1 and X2 of 0.20, whereas tables S22-S27 provide results for a correlation of 0.50. The same pattern of results was observed for the two correlation values and thus we do not further discuss this condition here.

Exclusion. For models without covariates, the performance of the indicators was approximately the same across sample sizes. The performance of the SCAIC ranged from 95% for N=200 to 99.8% for N=2000 and that of the SBIC ranged from 75% for N=200 to 99.50% for N=2000. The CAIC and BIC selected the 2 class solution 100% of the time for all samples sizes. The CAIC, SCAIC, BIC and SBIC performance was comparable when $N \geq 400$. The BLRTs accurately enumerated the number of classes 86% of the time across sample size conditions (85.70%-88.80%).

Partial Inclusion. When a single covariate was included in the model, the same pattern of results as what was observed without covariates was seen for the IC and LRT. The results remained unchanged when X1, rather than X2, was included

Total Inclusion. When the two covariates were included, a different pattern of results was observed across indicators. For the SBIC and BLRT, a more important deterioration in accuracy was observed when the covariates were allowed to influence class membership only (i.e., correctly specified models) or class membership and the growth factors in a class-invariant manner, but less marked when the covariates were allowed to influence the growth factors in a non-invariant manner. For instance, across sample sizes the accuracy of the BLRT dropped from 86.83% (without covariates) to 74.58% with covariates influencing the growth factors in a non-invariant manner, and to 66.13% with covariates influencing class membership and growth factors (invariance).

Class Separation

Table 4 present the results for the low and high class separation conditions for models without covariates and for models including X1 and X2 for a correlation of 0.20. Tables S16-S17 of the online supplements include the remaining results for a correlation of 0.20 and Tables S30-S32 present all results for a correlation of 0.50. Class separation had an important impact on class enumeration, being associated with a decrease in accuracy in the low separation condition.

Exclusion. In the high separation condition, the CAIC, SCAIC, BIC, and SBIC picked the two class solution close to 100% of the time when no covariate was included in the model. In the low separation condition, deterioration in the accuracy of most of the ICs (especially the CAIC and the BIC), which showed a tendency for underextraction. For instance, the CAIC picked the two class solution 31.90% of the time and the one class solution 68.10% of the time, the SCAIC picked the two class solution 79.70% of the time and the one class solution 20.10% of the time, the BIC picked the two class solution 46.20% of the time and the one class solution 53.80% of the time, and the SBIC picked the two class solution 89.80% of the time. The BLRT correctly identified the two class solution more than 86% of the time in both conditions. Thus, the BLRT, SBIC and SCAIC proved the most efficient indicators in the low class separation condition.

Partial Inclusion. With a single covariate included in the model, the ICs had similar accuracy as in models without covariates for the high class separation condition. In the low separation condition, the ICs had a higher accuracy in the models with a covariate influencing the class membership only than in models without covariates. However, their accuracy decreased when the covariate was also specified as having an impact on the growth factors. For example, in the low separation condition, the CAIC had an accuracy of 60.80% with X1 influencing only class membership, 17% with X1 influencing class membership and the growth factors (invariance), and 1.80% with X1 influencing the class membership and the growth factors (non-invariance). Similar figures for the BIC are 76.70%, 32.40%, and 6%. Compared to the model without a covariate, in models including a covariate the accuracy of the BLRT also tended to deteriorate. In the high separation condition, the BLRT had roughly a performance similar to that observed in models without covariates. In the low separation condition, the accuracy of the BLRT dropped from 88% in models without covariates to 72.70% in a model where X1 influenced class membership and the growth factors (non-invariant). However, its performance remained comparable in all models including a covariate. Finally, the partial inclusion of X2 (simulated as having a stronger effect on class membership) outperformed that of X1.

Total Inclusion. With both covariates included, the ICs performed better in models where the covariates influenced the class membership only than in models where the two covariates influenced class membership and the growth factors (non-invariant). For the BLRT, conditions of partial inclusion resulted in higher rates of classification accuracy than total inclusion. For example, across class separation conditions, the BLRT had an accuracy of 79.60% with X1 influencing only class membership, 78.70% with X1 influencing class membership and the growth factors (invariant), and 77.15% with X1 influencing class membership and the growth factors (non-invariant). In contrast, it had an accuracy of 70.90% when the two covariates influenced class membership, 58.85% when they also influenced the growth factors (invariant), and 66.30% when this influence was non-invariant.

***Correlation*.** Generally the same pattern of results was seen when the correlation between X1 and X2 was equal to 0.20 or 0.50. However, in some situations the accuracy increased with the size of the correlation. For example, in models with one covariate (X1) influencing class membership and the growth factors (invariant), the ability of the CAIC/SCAIC to pick the two class solution went from 17%/73% to 28.40%/82.50% in the low class separation condition as the correlation increased from 0.20 to 0.50. Similarly, in model including one covariate (X1) influencing class membership only, the ability of the BIC to pick the two class solution went from 76.70% to 99.70% in the low class separation condition. In models including one covariate (X1) influencing class membership and the growth factors (non-invariant), the ability of the SBIC to pick the two class solution went from 73.20% to 81.80% in the low class separation condition. Finally, for models with one covariate (X1) influencing class membership only, the ability of the BLRT to pick the two class solution went from 74.90% to 83.40% in the high class separation condition.

Mixing Ratio

Tables S18-S20 of the online supplements present the results as a function of mixing ratio conditions when the X1-X2 correlation was 0.20, whereas Tables S33-S35 present the results when this correlation was 0.50. The same pattern of results was observed for both conditions, and remained approximately the same across the two set of mixing ratio. For the models without covariates, the CAIC, SCAIC, BIC and SBIC selected the two class solution 100% of the time, and the BLRT selected the two class solution 86.95% of the time across the two set of mixing ratio. When a single covariate was included in the model, the accuracy of the CAIC, SCAIC, BIC, SBIC, and BLRT remained stable across mixing ratio and models differing in type of covariate inclusion. Finally, when both covariates were included, the pattern of results was similar to what was observed in the partial inclusion condition for all indicators except the BLRT, which decreased to 73% of the time across the two mixing ratio conditions (and picked the three class solution more than 20% of the time).

Variance-Covariance

Tables S21-S23 of the online supplements present the results associated with the conditions of invariant and non-invariant variance-covariance matrices for models without covariates and for models including both covariates (X1 and X2) for correlation of 0.20. Tables S36-S38 present all results for a correlation of 0.50. Generally, the accuracy of the indicators showed few variations as a function of the variance-covariance conditions.

Exclusion. Without covariates, this condition had no effect on the performance of the indicators. The ICs selected the correct two class solution almost 100% of the time across conditions, while the BLRT selected the two class solution 86% of the time in the two variance covariance conditions.

Partial Inclusion. With a single covariate in the model, the accuracy of the CAIC, the SCAIC, and the BIC remained stable across variance-covariance conditions and type of covariate inclusion. The performance of the SBIC was also generally unaffected by the variance-covariance conditions, but differed across models. Thus, the accuracy of the SBIC was highest (100%) in the non-invariant variance-covariance condition with one covariate (X1) influencing class membership and the growth factors (invariant and non-invariant effects), and lowest (83.30%) in the non-invariant variance-covariance condition with the other covariate (X2, specified as having a stronger effect) influencing class membership and the growth factors (invariant and non-invariant). The performance of the BLRT increased when the variance-covariance-matrix was non-invariant with partial covariate inclusion, and reached 100% when one covariate (X1) was specified as influencing the class membership and the growth factors (invariant and non-invariant).

Total Inclusion. With two covariates in the model, the accuracy of the CAIC, SCAIC, BIC, and SBIC remained stable across variance-covariance conditions and type of covariate inclusion. When the variance-covariance matrix was invariant, the BLRT had a higher accuracy in models of partial, versus total, inclusion. For instance, the accuracy of the BLRT ranged from 65% to 74.70% in models including both covariates versus 81.60% to 84.90% in models including only X2 when the variance-covariance matrix was invariant. A similar pattern of results was observed for the non-invariant variance-covariance-matrix, with a single exception: With both covariates specified as influencing class membership and the growth factors (non-invariant), the efficacy of the BLRT reached 100%.

Correlation**.** Generally the same pattern of results was seen when the X1-X2 correlation was equal to 0.20 or 0.50. However, the accuracy of the indicators varied with the size of the correlation. The accuracy of the CAIC, the SCAIC and BIC tended to be constant across correlations and variance-covariance conditions. However, the accuracy of the SBIC and BLRT varied as a function of the correlation and type of covariate inclusion when the variance-covariance matrix was non-invariant. More precisely, the accuracy of the SBIC decreased with the size of the correlation in conditions of partial inclusion but increased with the size of the correlation in condition of total inclusion. For example, for non-invariant the variance-covariance matrices, with X1 influencing class membership only, the SBIC had an accuracy of 100% when the correlation was equal to 0.20, versus 86.70% when it was equal to 0.50. Similar figures for models including X2 were respectively 90.50% and 77.80%. Conversely, when both covariates were specified as influencing the class membership and the growth factors (non-invariant), its accuracy was 90.50% when the correlation was equal to 0.20, versus 100% when the correlation was equal to 0.50. Similarly, when the variance-covariance matrix was non-invariant, the accuracy of the BLRT decreased with the size of the correlation when covariates were specified as influencing only the class membership. For example, with X1 influencing the class membership, the BLRT had a rate of accuracy of 96% when the correlation was equal to 0.20, versus 80% when the correlation was equal 0.50. When covariate(s) were specified as influencing class membership and the growth factors (invariant), the accuracy of the BLRT remained stable across correlations when X1, or both X1 and X2, were included in the model, but increased with the size of the correlation when only X2 was included. When the covariate(s) were specified as influencing class membership and the growth factors (non-invariant) the accuracy of the BLRT was stable across correlations in conditions of partial or total inclusion of the covariates when the variance-covariance matrix was invariant, but (a) increased across correlations in conditions of partial inclusion of X2 and (b) decreased in conditions of partial inclusion of X1 and of total inclusion of both covariates.

Nonconvergence and Improper Solutions

As for Study 1, all results discussed so far are based on models which converged on proper solutions. In Study 2, rates of non-convergence remained generally low, ranging from 0 to 2.83% across conditions (see Table S68 of the online supplements). However, improper solutions were more frequent, ranging from 0 to 30.92% across conditions, with the highest rates being associated with the 3 and 4 class solutions when the covariates influenced class membership and the growth factors in a non-invariant manner (particularly in conditions of low sample size, different within-class variance-covariance matrices, low class separation, and high correlations). Interestingly, 0% of improper solutions where observed for models without covariates, for models with covariates influencing class membership only or class membership and growth in an invariant manner, and for the 1-2 class models when the covariates influenced class membership and growth in a non-invariant manner.

Summary and Discussion

A summary of the results from Study 2 is presented in Table 5. In Study 2, across conditions and models, the CAIC, SCAIC, and BIC constantly identified the correct number of classes more than 87% of the time on average. The SBIC identified the correct number of classes 78.49% of the time across conditions and models. The ICs generally performed better than the BLRT, which identified the correct number of classes 67.30% of the time across conditions and models and had a tendency to over extract classes (22.12% of the time). The performance of the ICs (CAIC, SCAIC, BIC) was roughly comparable across conditions of total or partial inclusion or exclusion of covariates in the model. Still, the results showed an improvement in the accuracy of the CAIC (98.59%), SCAIC (91.36%), and BIC (97.95%), when covariate(s) were included in the model and specified as influencing only class membership, and a slight decrease in the accuracy of these indicators (CAIC:

83.36%; SCAIC: 87.26%; BIC: 86.72%) when covariate(s) were specified as having an impact on the growth factors over their impact on class membership. The accuracy of the SBIC and BLRT was also relatively stable across conditions, selecting the correct solution on average 74.40% of the time with no covariate and 78.95% of the time with covariates for the SBIC, versus 68.68% and 67.15% for the BLRT. However, these differences have limited practical significance as it is typically impossible in applied research to know beforehand the specific nature of the covariates effects. Thus, the estimation of models without covariates seems a reasonable compromise in terms of accuracy.

The performance of the indicators was approximately the same across the range of sample sizes and mixing ratio considered. The size of the correlation also had a minimal, yet inconsistent, impact on the performance of some indicators. For instance, a larger correlation increased the performance of the CAIC, SCAIC, BIC, and SBIC under conditions of low class separation, but increased the performance of the BLRT under conditions of high class separation. The impact of the invariance/non-invariance of the variance-covariance matrix also remained minimal. The two indicators with the greatest sensitivity to the variance-covariance matrix – in conjunction to the size of the correlation between the covariates – are again the SBIC and BLRT. More precisely, when the variance-covariance matrix was non invariant, the SBIC and BLRT proved particularly sensitive to the size of the correlation, with a larger correlation associated with a decrease in accuracy under conditions of partial covariate inclusion and a higher accuracy under conditions of total covariate inclusion. Clearly, on the basis of these results, in the presence of different variance-covariance matrices, the CAIC, SCAIC and BIC should be trusted more than the SBIC and BLRT.

The class separation condition had an important impact on the class enumeration performance of the indicators, with their accuracy showing a marked improvement in the high separation condition. Among the ICs, the low separation condition had the largest effect on the CAIC and BIC which tended to underextract latent classes in that situation. For these indicators, in the low separation condition, models with covariate(s) influencing only the class membership outperformed other models. The indicators showing the highest accuracy in conditions of low class separation, and the lowest level of sensitivity to the type of covariates specification proved to be the SCAIC, SBIC and BLRT. This is interesting as two of these indicators (SBIC and BLRT) are clearly the least efficient, and most reactive, across other the design conditions. These results reinforce the conclusions from Study 1 in suggesting that the entropy associated with the estimated models should first be examined. In conditions of low class separation (entropy values close to .50 here), then the SBIC, BLRT, and SCAIC should be given more importance than AIC, CAIC, BIC, and LMR/ALMR. The results clearly show that a risk is associated with the BIC and CAIC in conditions of low class separation.

The results show that most indicators, and particularly the CAIC, SCAIC and BIC, performed relatively well across the various conditions of covariates exclusion/inclusion considered here, although these conditions still had an impact on the accuracy of the class enumeration process. More precisely, these results show that, unless the effects of the covariates are correctly specified, there are advantages to conducting the class enumeration process without including covariates in the model. For most research settings were a priori knowledge of expected covariates effects is limited, our results suggest that conducting the class enumeration without covariates represents a reasonable compromise in terms of accuracy. Finally, our results also suggest a two stage process, starting with an examination of the entropy. At high entropy values, the CAIC, SCAIC, and BIC should be used in priority. At low entropy values, the SBIC, BLRT, and SCAIC appear more appropriate.

Study 3: Population Model with Covariates Predicting Class Membership and Growth Factors

In Study 3 we examined the class enumeration performance of GMM considering misspecifications related to a partial or a total exclusion of actives covariates, or to an incorrect specification of their effects in the model.

Sample Size

Tables S39-S44 of the online supplements present the results as a function of sample size for models without covariates or including X1 and X2 for a correlation of 0.20, while Tables S52-S57 present the results for a correlation of 0.50.

Exclusion. Without covariates, the accuracy of the ICs and LRT is approximately the same across sample sizes. Among the ICs, the accuracy of the SCAIC ranged from 95.10% for N=200 to 99.80% for N=2000, that of the SBIC from 74.70% for N=200 to 99.50% for N=2000. Both the CAIC and the BIC selected the 2 class solution 100% of the time for all samples size conditions. The CAIC, SCAIC, and BIC performance was comparable when $N \geq 400$ and the SBIC had a performance similar to that of these indicators when $N \geq 1000$. The BLRT also accurately identified the number of latent classes, but did not performed as well as the ICs, ranging from 87.30% for *N*=2000 to 89.60% for *N*=400.

Partial Inclusion. When a single covariate was included, the accuracy of the indicators varied across specifications of the effects of the covariates. When X1 influenced only the class membership, the CAIC and the BIC had a level of accuracy that remained unaffected by sample size and similar to results without covariates. Conversely, the accuracy of the SCAIC decreased slightly with very large sample sizes, dropping from 95.70% for *N*=1000 to 88.80% for *N*=2000. A similar pattern was observed for the SBIC, with an increase in accuracy as sample size increased up to *N*=1000 (69.10% for *N*=200 to 86.10% for *N*=1000), followed by a decrease (72.80% at *N*=2000). Furthermore, the accuracy of the BLRT also decreased with sample sizes in this condition, going from 85.70% for *N*=200 to 24.80% for *N*=2000 (where it picked the four class-solution 65.80% of the time). These indicators had a higher level of accuracy in the models without covariates. When X2 was included in the model, the accuracy of all the indicators importantly decreased with sample sizes. For instance, when *N*=2000 all indicators selected the four class solution more than 92% of the time.

When X1 was allowed to influence the class membership and the growth factors (invariant), the accuracy of CAIC, SCAIC, and BIC was close to 100% across sample sizes for $N \leq 1000$ and comparable with that of the model without covariates. For *N*=2000 the accuracy of these indicators dropped to 89.90% for the CAIC, to 79.20% for the SCAIC, to 86.50% for the BIC. The accuracy of the SBIC increased with sample sizes from 69.10% for *N*=200 to 98.50% for *N*=1000 and dropped to 74.70% for *N*=2000. Similarly, the accuracy of BLRT increased from 84% for *N*=200 to 90.10% to *N*=2000. With X2, the accuracy of the CAIC, SCAIC, and BIC was close to 100% across all sample sizes conditions, that of the SBIC increased with sample sizes (65.90% for *N*=200 to 99.50% for *N*=2000), whereas that of the BLRT was stable across sample sizes (83.80% for *N*=200 to 84.80% to *N*=2000). In the situation where the covariate influenced the class membership and the growth factors in a non-invariant manner, a similar pattern of results was observed, although the accuracy of the indicators was slightly higher with X1 (and comparable to models without covariates).

Total Inclusion. When both covariates were included, the same pattern of results as previously described when only X2 was included was observed. Importantly, the accuracy was generally greater or equivalent in models without covariates than in models including both covariates for CAIC, SCAIC, BIC, SBIC, and BLRT.

Correlation. Without covariates, the size of the correlation had no impact on the accuracy of the indicators. Generally the impact of the correlation was unrelated to the type of covariate inclusion (partial or total) and to the specification of their effect. The only exception was observed in conditions of partial inclusion of X2 specified as influencing only class membership. In this context, the accuracy of the indicators was generally weaker and tended to decrease when the correlation increased. For example, across sample sizes, the CAIC had an accuracy of 66.18% when the correlation had a value of 0.20 and of 48.88% when it had a value of 0.50. Similar figures across sample sizes were of 29.25% and 13.55% for the SCAIC, 58.80% and 41.75% for the BIC, 10% and 2.45% for the SBIC, and 17.95% and 8.58% for the BLRT. In contrast, when X1 is specified as influencing only class membership, similar values were unchanged and respectively of 99.90% and 99.95% for the CAIC, 93.40% and 96.25% for the SCAIC, 99.63% and 99.85% for the BIC, 78.05% and 85.35% for the SBIC, and 62.95% and 69.95% for the BLRT. Similar results were observed in the other conditions.

Class Separation

Table 6 presents the results for the class separation conditions for models without covariates and including both covariates for a correlation of 0.20. Tables S45-S46 of the online supplements present the remaining results for a correlation of 0.20, while Tables S58-S60 present the results for a correlation of 0.50. Class separation had a substantial impact on class enumeration. Generally, the accuracy of the indicators decreased when the trajectories were not well separated.

Exclusion. Without covariates, the accuracy of the CAIC, SCAIC, BIC, and SBIC was close to 100% in the high class separation condition. In the low separation condition, a deterioration of accuracy was seen for most of ICs, especially for the CAIC and the BIC which tended to under extract classes. In this condition, the CAIC picked the two class solution 16% of the time and the one class solution 84% of the time and the BIC picked the two class solution 26.30% of the time and the one class solution 73.70% of the time. In contrast, the SCAIC picked the two class solution 65.20% of the time and the one class solution 34.50% of the time, and the SBIC picked the two class solution 80% of the time. Although the BLRT proved less accurate than the ICs in the high separation condition, it had an accuracy of 84.70% across conditions, and seemed unaffected by the class separation.

Partial Inclusion. When a single covariate was included, the ICs had a similar level of accuracy as in models without covariates in the high class separation condition. These results were unaffected by the specification of the effects of the covariates. In the low separation condition, the ICs had a higher level accuracy in the model where the covariate was specified as influencing only the class membership than in the other conditions where their accuracy decreased. For example, the CAIC had an accuracy of 96.70% with X1 influencing only the class membership, 5.30% with X1 influencing class membership and the growth factors (invariant), and 0.70% with X1 influencing class membership and the growth factors (non-invariant). Similar figures were 99.10%, 11.80% and 1.50% for the BIC, 99.20%, 47.60% and 21.40% for the SCAIC, and 94.50%, 71.40% and 46.80% for the SBIC. In contrast, the accuracy of the BLRT was lower when the covariate was only allowed to influence class membership than in the other conditions, and tended to be higher in the high class separation condition and lower in the low class separation condition. Thus, in the high separation condition, the BLRT had an accuracy of 59.30% when X1 influenced only the class membership, 91% when X1 influenced class membership and the growth factors (invariant), and 90.50% when X1 influenced class membership and the growth factors (non-invariant). In the low separation condition, these rates were respectively 74%, 69.50%, and 65.20%. This phenomenon was more accentuated in models including X2, with rates of accuracy of 0.10%, 84.70%, and 88.90% in the high separation condition versus 15.70%, 80%, and 77.50% in the low separation condition. Across conditions, the accuracy of the indicators tended to be higher with X1 than X2.

Total Inclusion. When both covariates were included in the model, the effect of the class separation conditions on the accuracy of the indicators varied as a function of the specification of the effect of the covariates. In the high separation condition, when the covariates influenced only class membership, the accuracy of the indicators when both covariates were included (CAIC: 84.10%; SCAIC: 10.30%; BIC: 63.70%; SBIC: 1.30%; BLRT: 0.20%) was higher than when only X2 was included (CAIC: 65.40%; SCAIC: 4.50%; BIC: 43.50%; SBIC: 0.60%; BLRT: 0.10% for the partial inclusion of X2), but lower than when only X1 was included (CAIC: 99.90%; SCAIC: 95.70%; BIC: 99.60%; SBIC: 86.10%; BLRT: 59.30% for X1). However, when the covariates influenced the class membership and the growth factors (invariant and non-invariant) the indicators had similar rates of accuracy in the total inclusion condition (close to 100% for the CAIC, SCAIC, BIC, and SBIC, and to 85% for the BLRT) as in the partial inclusion of X2. In the low separation condition, the accuracy of the indicators in the total inclusion condition was comparable or lower to that observed when only X2 was included independently of the type of covariate inclusion. For example, the CAIC had an accuracy of 98.40% when X2 influenced class membership only, 64.20% when it influenced class membership and the growth factors (invariant), and 25.10% when it influenced class membership and the growth factors (non-invariant). Comparable figures are 63.90%, 96.30%, 86.40% for the SCAIC; 95.20%, 78.40%, 43.70% for the BIC; 36.10%, 97.80%, 95.70% for the SBIC; and 15.70%, 80%, 77.50% for the BLRT. When both covariates are included, these rates are, respectively, 92.20%, 44.40%, and 1.90% for the CAIC, 61.70%, 91.90%, 56% for the SCAIC; 95.50%, 62.80%, 8.50% for the BIC; 29.90%, 94.60%, 84.90% for the SBIC, and 11.60%, 54.60%, 59% for the BLRT.

Correlation. Without covariates, the size of the correlation had no impact on the accuracy of the indicators in the high separation condition, but slightly improved their accuracy in the low separation condition. Generally, the impact of the correlation was related to the type of covariate inclusion (partial or total) and to the specification of their effect. The same pattern of results was seen for the partial inclusion of X2 and the total inclusion of both covariates, but a different pattern of results was observed for the partial inclusion of X1. When the covariate influenced only class membership, the accuracy of the indicators decreased as a function of the size of the correlation across separation conditions for the condition where X2, or both covariates, were included, but was stable across correlations and separation conditions when X1 was included. For example, with high separation and X2 influencing class membership only, the CAIC had an accuracy of 65.40% when the correlation was 0.2, but only 10.40% when it was 0.5. Similar figures were respectively of 4.50% and 0.10% for the SCAIC, 43.50% and 2.90% for the BIC, 0.60% and 0% for the SBIC, and 1% and 0% for the BLRT. In the low separation condition, similar figures were respectively 98.40% and 16.60% for CAIC, 63.90% and 0.20% for SCAIC, 95.20% and 7.20% for BIC, 36.10% and 0% for SBIC, and 15.70% and 0% for BLRT. Similar rates were observed for models where both covariates were allowed to influence class membership. In contrast, when X1 was specified as influencing only class membership similar figures (across separation conditions) were respectively 100% and 100% for CAIC, SCAIC, and BIC, 90.30% and 94% for SBIC, and 66.65% and 72.50% for BLRT.

When the covariate influenced both the class membership and the growth factors, the accuracy of the indicators was stable across correlations and class separation conditions for models including X2 or both covariates. In contrast, models including X1, the accuracy of the indicators increased as a function of the correlation in the low separation condition but remained stable across correlations in the high separation condition. For example, for models including X1, CAIC, SCAIC, BIC, and SBIC had an accuracy close to 100%, and BLRT had an accuracy greater than 70% across correlation values in the high separation condition. In contrast, in the low separation condition and models where X1 is allowed to influence the growth factors (invariant), CAIC had an accuracy of 5.30% when the correlation is 0.20, but of 67.50% when it is 0.50. Similar figures are respectively 47.60% and 97.60 % for SCAIC; 11.80% and 81.20% for BIC; 71.40% and 98.50% for SBIC; and 69.50% and 82.80% for BLRT. When the effects of X1 are specified as non-invariant, these values further decrease to 0.70% and 29.10% for CAIC, 21.40% and 88.60% for SCAIC, 1.50% and 47.50% for BIC, 46.80% and 95.80% for SBIC, and 65.20% and 80.60% for BLRT.

Mixing Ratio

Tables S47-S49 of the online supplements present the results as a function of the mixing ratio for a correlation of 0.20, while Tables S61-S63 present these results for a correlation of 0.50.

Exclusion. For the models without covariates, accuracy was similar across indicators and mixing ratio conditions. More precisely, across mixing ratio, the CAIC, SCAIC, BIC and SBIC selected the two class solution 100% of the time, and the BLRT 87.80% of the time.

Partial Inclusion. When a single covariate (X1) was included, the accuracy of the CAIC, SCAIC, and BIC remained stable and close to 100% across mixing ratio conditions and type of covariate inclusion. Compared to the model without covariates, the accuracy of the SBIC dropped from 98.20% to 86.10% with a mixing ratio 0.50/0.50 and from 98.70% to 76.20% with the mixing ratio of 0.30/0.70 when X1 was specified as influencing class membership only. However, its accuracy was comparable to that observed in models without covariates when X1 was specified as influencing both class membership and the growth factors (invariant and non-invariant) for mixing ratio of 0.50/0.50. In contrast, its accuracy dropped to 2.70% when the mixing ratio was equal to 0.30/0.70 for models where the effect of X1 on the growth factor was invariant. The BLRT showed a similar pattern of performance across conditions. Compared to the model without covariates, the inclusion of X1 specified as influencing only the class membership resulted in a drop in accuracy from 88.10% to 59.30% for a mixing ratio of 0.50/0.50 and from 87.50% to 47.50% for a mixing of 0.30/0.70. However, its accuracy was comparable to that observed in models without covariates when X1 was specified as influencing both class membership and the growth factors. When the covariate X2 influenced only class membership, the accuracy of all indicators dropped drastically for the two mixing ratio. In this condition, the CAIC selected the two class solution 53.30% of the time and the four class solution 23.90% of the time across mixing ratio, while the SCAIC selected the four class solution 89.20% of the time, the BIC selected the two class solution 33.25% of the time and the four class solution 44.40% of the time, the SBIC selected the four class solution 96.65% of the time, and the BLRT selected the four class solution 98.45% of the time. When X2 was specified as influencing class membership and the growth factors, the results were similar with models including X1.

Total Inclusion. When both covariates were included, the pattern of result was similar to what was observed in models in which X2 was specified as influencing only the class membership for the SCAIC, SBIC, and BLRT. Across the two set of mixing ratio, the accuracy of the CAIC and BIC was higher in the total inclusion condition than in the models where only X2 was included but lower than in the models where only X1 was included or in models including no covariates. In addition, similar to what was observed in models including a single covariate, the performance of the indicators tended to be higher when the mixing ratio was equal to 0.50/0.50. For example, when the covariates influenced only class membership, the accuracy of the CAIC was 84.10% when the mixing ratio was 0.50/0.50 but 58.90% when it was 0.30/0.70. Similar figures where 10.30% and 4.20% for SCIC, 63.70% and 39.30% for BIC, 1.30% and 0.40% for SBIC, and 0.20% and 0% for BLRT.

Correlation. Without covariates, the size of the correlation had no impact on the accuracy of the indicators across mixing ratio conditions. Generally, the impact of the correlation was related to the type of covariate inclusion (partial or total) and the specification of their effect. Similar results were seen for the partial inclusion of X2 and the total inclusion of both covariates, while a different pattern was observed for the partial inclusion of X1. When X1 was specified as influencing only class membership, or class membership and the growth factors (invariant), the accuracy of the SBIC and BLRT increased as a function of the correlation across the two mixing ratio, while the accuracy of the CAIC, SCAIC, and BIC was close to 100% across conditions. For example, when X1 only influenced class membership, the accuracy of the SBIC increased from 52.95% when the correlation was 0.20 to 87.75% when it was 0.50 across mixing ratio, while the BLRT increased from 53.40% to 63.40%. In contrast, when X1 was specified as influencing class membership and the growth factors (non-invariant), the accuracy of all indicators was constant across correlations and mixing ratio. When X2, or both covariates, were specified as influencing only class membership, the accuracy of the SCAIC, SCBIC and BLRT was constant across correlations and mixing ratio, while that of the CAIC and BIC decreased as a function of the correlation across mixing ratio. When X2, or both covariates, were specified as influencing class membership and the growth factors (invariant or non-invariant), the accuracy of all indicators remained constant across correlations and mixing ratio. For example, with X2 specified as influencing only class membership, the SCAIC, SBIC and BLRT had an accuracy of almost 0% across mixing ratio and correlations. In contrast, the BIC had an accuracy of 33.25% across mixing ratio when the correlation was 0.20 and of 1.65% when the correlation was 0.50.

Variance-Covariance

Table 7 presents the results for the variance-covariance matrices conditions for models without covariates and including both covariates with a correlation of 0.20. Tables S50-S51 of the online supplements present the remaining results for a correlation of 0.20, while Tables S64-S66 present results for a correlation of 0.50. Generally, the accuracy of the indicators decreased when the variance-covariance matrix was non-invariant, especially when covariates were included.

Exclusion. Without covariates, this condition had no effect on class enumeration. The ICs selected the two class solution almost 100% (versus 90% for the BLRT) of the time across conditions.

Partial Inclusion. With a single covariate, important differences were observed across variance-covariance conditions. When the variance-covariance matrix was invariant, the accuracy of the CAIC, SCAIC, BIC remained comparable to that observed in models without covariates, while an important decrease was observed in the accuracy of the BLRT, which dropped to 59.30%. In contrast, when the variance-covariance matrix was non-invariant, the accuracy of the ICs and BLRT dropped drastically. For instance, CAIC picked the two class solution 19.20% of the time and the three class solution 80.60% of the time, SCAIC picked the two class solution 0.30% of the time and the three class solution 79.60% of the time, BIC picked the two class solution 8.20% of the time and the three class solution 91.20% of the time, SBIC picked the two class solution 0% of the time and the four class solution 57.50% of the time, and BLRT picked the four class solution 91.40% of the time.

The results also differed according to the specification of covariate effects. When X1 was allowed to influence class membership and the growth factors (invariant and non-invariant), the accuracy of the indicators increased to the level observed in models without covariates across variance-covariance conditions. In contrast, when X2 was only allowed to influence class membership, an important deterioration in accuracy was observed across the two variance-covariance conditions. In particular, for the invariant variance-covariance matrix condition, CAIC picked the two class solution 65.40% of the time and the four class solution 23.60% of the time. In the non-invariant condition, CAIC picked the two class solution 73.90% of the time and the three class solution 26.10% of the time. Rates of accuracy were respectively 4.50% (invariant variance-covariance) versus 5.10% (non-invariant variance-covariance condition) for SCAIC, 43.50% versus 52.70% for BIC, 0.60% versus 0.30% for SBIC, and 0.10% versus 0.30% for BLRT. Most of these indicators showed a clear tendency to overextract latent classes in these conditions. However, when X2 was also allowed to influence the growth factors, the results were comparable to that observed for models including X1.

Total Inclusion. With both covariates, the same pattern of result was seen as for models including only X2, with few exceptions related to the SBIC and BLRT which proved less accurate and more sensitive to covariate specification with a non-invariant variance-covariance matrix.

Correlation. Without covariates and an invariant variance-covariance matrix, the correlation did not impact class enumeration accuracy. The BLRT accuracy increased as a function of the correlation with a non-invariant variance-covariance matrix (90.20% for a correlation of 0.20 to 94.45% when it was 0.50 across conditions). The other ICs had an accuracy close to 100% across correlations.

With covariate(s) in the model, the impact of the correlation was related to the type of covariate inclusion (partial or total) and to the specification of their effects. When X1 influenced only class membership: (a) the SCAIC accuracy was constant across correlations values (48% to 52% for correlations of 0.20 and 0.50); (b) the CAIC and BIC accuracy was constant and close to 100% across correlations values in the invariant variance-covariance condition but increased across correlations values in the non-invariant condition (from 19.20% to 68.50% for the CAIC and from 8.20% to 45.60% for the BIC); (c) the SBIC (86.10% to 93%) and BLRT (59.30% to 68.20%) accuracy increased in the invariant condition but was constant across correlations in the non-invariant condition (from 0% to 0.70% for both SBIC and BLRT). Similarly, when X1 was specified as influencing class membership and the growth factors in a non-invariant manner, the accuracy of the ICs remains close to 100% across conditions, while the BLRT accuracy tends to be higher when the correlation was 0.20 (90.5% when the variance-covariance matrix is invariant versus 100% when it is non-invariant) rather than 0.50 (88.80% versus 86.70%). However, when X1 effects on the growth factors is specified as invariant across classes: (a) the accuracy of the CAIC (95.65% to 97.60% for correlations of 0.20 and 0.50, across conditions), SBIC (50.15% to 51.45%) and BLRT (47.20% to 46.10%) was constant across correlations values and conditions; (b) the accuracy of the BIC and SCAIC was constant and close to 100% across correlations values in the invariant condition but increased across correlations values in the non-invariant condition (BIC: 75.70% to 83.80%; SCAIC : 13.2% to 22.90%).

The performance of the indicators tended to be worse when X2 was included in the model. In the situation where X2 influenced only class membership: (a) the SBIC and BLRT accuracy was constant across correlations values and close to 0%; (b) the CAIC (from 65.40% to 10.40% for correlations of 0.20 and 0.50 when the variance-covariance matrix was invariant and from 73.90% to 56.60% when the variance-covariance matrix was non-invariant), SCAIC (4.50% to 0.10% in the invariant condition and 5.10% to 0.90% in the non-invariant condition), and BIC (43.50% to 2.90% in the invariant condition and 52.70% to 31.30% in the non-invariant condition) accuracy decreased across correlation values, particularly when the variance-covariance matrix was specified as non-invariant. When X2 influenced class membership and the growth factors in an invariant manner, the indicators' accuracy was constant across correlations values in the variance-covariance invariance condition (close to 100% for the ICs, and 84.70% to 85.20% for the BLRT for correlations of 0.20 and 0.50) but decreased across correlations values for non-invariant variance-covariance (100% to 20.80% for the CAIC, 82.70% to 0% for the SCAIC, 100% to 6.70% for the BIC, and 40.40% to 0% for the BLRT). Finally, when X2 was specified as influencing class membership and the growth factors in a non-invariant manner: (a) the CAIC, SCAIC and BIC accuracy was constant and close to 100% across correlations values and conditions; (b) the SBIC and BLRT accuracy was constant across correlations values in the variance-covariance invariance condition (close to 100% for the SBIC and 88.90% to 88.40% for the BLRT) but increased across correlations values in the variance-covariance non-invariance condition (76.90% to 100% for both SBIC and BLRT). Globally, with both covariates included, the overall pattern of results remained similar to that observed when only X2 was included.

Nonconvergence and Improper Solutions

As for Study 1 and 2, all results discussed so far are based on models which converged on proper solutions. In Study 3, rates of non-convergence remained generally low, ranging from 0 to 3.58% across conditions (see Table S69 of the online supplements). However, improper solutions were more frequent, ranging from 0 to 46.69% across conditions, with the highest rates being associated with the 3 and 4 class solutions when the covariates influenced class membership and the growth factors in a non-invariant manner (particularly in conditions of low sample size, unequal mixing ratio, different within-class variance-covariance matrices, low class separation, and high correlations). As for study 2, 0% of improper solutions where observed for models without covariates, for models with covariates influencing class membership only or class membership and growth in an invariant manner, and for the 1-2 class models when the covariates influenced class membership and the growth factors in a non-invariant manner.

Summary and Discussion

A summary of the results from Study 3 is presented in Table 8. In Study 3, the class enumeration accuracy differed importantly across indicators and conditions. The CAIC, SCAIC, and BIC correctly identified the number of classes more than 74% of the time across conditions, while the SBIC did so 66.41% of the time, and the BLRT 57.58% of the time. The BLRT showed a marked tendency for overextraction, retaining the four class solution 23.87% of the time.

A main conclusion of this study is that, although the population models included covariates, the class enumeration performance of the indicators presented a higher level of accuracy in models excluding covariates. There were only few exceptions to this observation and these exceptions remained of a negligible magnitude and limited to specific design conditions. Thus, in models excluding covariates, the CAIC, SCAIC, BIC, and SBIC constantly identified the correct number of classes more than 91% of the time across conditions, and the BLRT did so 83.22% of the time. In contrast, in models including one or both covariates the accuracy of the indicators was lower, particularly in models including X2 (rather than X1) or both covariates, and in models where the covariates were incorrectly specified as having an influence limited to the class membership.

Overall, in models excluding covariates, the accuracy of the indicators remained approximately the same across the range of design conditions considered, while the CAIC, SCAIC, BIC, SBIC, and BLRT proved to be much more sensitive to the design conditions when covariate(s) were included in the models. In particular, the results seemed particularly sensitive to the size of the population correlation between covariates in models of partial or total inclusion of covariates and tended to show a decrease in the accuracy of the indicators when the correlations was equal to 0.50 (a commonly observed correlation in the social sciences) versus 0.20. Similarly, models including covariate(s) (especially models including only X2, or both covariates) tended to be much more sensitive to unequal mixing ratio, non-invariant variance-covariance structure, and low class separation. It is noteworthy that the results from this study, at least when the models are estimated without covariates, also show that the SBIC, BLRT, and SCAIC outperform the BIC and CAIC in conditions of low class separation. In other conditions however, the BIC and CAIC clearly outperform the remaining indicators. This observation thus supports the recommendations made in Studies 1 and 2 to use the entropy in a first step to guide the decision to mainly focus on either the BIC and CAIC (in high class separation conditions) or the SBIC, BLRT, and SCAIC (in low class separation conditions).

The results clearly suggest that applied research, in which the nature of covariate effects is unknown beforehand, would do best to conduct the class enumeration process without covariates. The results even show that when covariates are known to have an impact on class membership and growth factors, researchers would still be safer to conduct the class enumeration process without covariates.

General Discussion

The purpose of this series of studies was to evaluate the impact of complete or partial inclusion or exclusion of active or inactive covariates on the accuracy of the class enumeration procedure in growth mixture models. To this end, three complementary studies were considered. In Study 1 we examined the effect of including an inactive covariate in GMM when the correctly specified model did not include covariates. In Study 2, we considered a population model in which two covariates were specified as influencing only the class membership, and investigated the impact of completely, or partially, including, partially including, or excluding these covariates from the class enumeration process. Finally, in Study 3, we considered a population model in which two covariates were specified as influencing class membership and the growth factors, and investigated the impact of completely, or partially, including or excluding these covariates in the class enumeration process. Across all studies, we contrasted models where the covariates were: (a) excluded; (b) only allowed to influence class membership; (c) allowed to influence class membership and the growth factors in a class-invariant manner; and (d) allowed to influence class membership and the growth factor in a class-varying manner. The accuracy of multiple indicators (AIC, CAIC, SCAIC, BIC, SBIC, LMR/ALMR, and BLRT) to correctly identify the number of latent classes in the data was assessed across these multiple specifications and design conditions (sample size, class separation, mixing ratio, variance-covariance matrix, and size of the correlation between the covariates).

Our results showed very low rates of nonconvergence, remaining under 4% across studies and conditions. Similarly, in most conditions, improper solutions approached 0% for most conditions, with the sole exception of 3- and 4-class models in which covariates were allowed to influence class membership and the growth factors in a class-varying manner. In this specific condition, rates of improper solutions (exclusively related to negative estimates of the growth factor variances or out-of-bound estimates of the growth factor correlations) could reach 20% to 50%. These results are promising, suggesting that convergence problems may be relatively rare in optimal conditions. However, these results also suggest that conducting the class enumeration process while including covariates caries important risks of nonconvergence, when these covariates are allowed to predict class membership and the growth factors in a non-invariant manner. Unfortunately, it is impossible to contrast our results with previous research as the few previous studies of covariate effects did not report rates of nonconvergence and improper solutions (Tofighi & Enders, 2010; Li & Hser, 2011).

In terms of class enumeration accuracy, under conditions where their results are comparable, previous studies yielded different recommendations regarding the potential usefulness of including covariates in the model as part of the class enumeration process, with one study concluding that covariates should be excluded from the class enumeration process (Tofighi, & Enders, 2007), and the other leading to the recommendation that covariate inclusion was beneficial in many situations (Li & Hser, 2011). However, the bulk of results from these studies remain hard to compare as they relied on different population models, models misspecifications, and design conditions, whereas the models considered in the current study encompass and substantially expand those covered in these previous studies in an attempt to provide a more definitive answer. Interestingly, although this simulation study is the first to systematically examine the effect of misspecification due to partial covariate inclusion, our findings regarding the class enumeration performance of over-parametrized GMM are similar to those of Li and Hser (2011). Indeed, in both studies (ours and Li & Hser, 2011), a similar pattern of results was observed for the IC. With small sample sizes of 200, the BIC, CAIC, SCAIC tend to perform well. In contrast, the SBIC performs poorly with small samples, but much better with larger samples. However, our results showed important variations in the accuracy of the indicators across studies, covariate specification, and design conditions. This clearly suggests that the reality is more complex than what was previously suggested (Li & Hser, 2011; Tofighi, & Enders, 2007). However, common conclusions and useful practical guidelines emerged from the current results.

First, a common observation across studies, design conditions, and specification of covariate effects is that neither the AIC, nor the LMR/ALMR, could be considered as reliable indicators in the class enumeration process. As already shown in previous studies (e.g., Li & Hser, 2011; Nylund et al., 2007; Tofighi & Enders, 2007), these indicators tended to perform substantially less accurately than the other indicators (such as the BIC, CAIC, and SCAIC), and also presented a very high level of sensitivity to design conditions.

Second, although the BLRT possesses strong statistical foundations, and was previously found to be a reliable indicator of class enumeration in previous studies (e.g., Li & Hser, 2011; Nylund et al., 2007), our results showed that this indicator performed poorly in some design conditions, such as when the estimated model is exactly specified, and in conditions of high class separation. For example, when the fitting model is exactly specified in Study 2, the BLRT failed to reach an accuracy rate greater than 75% even with large samples sizes. These observations thus call for further examination of the BLRT performance across a wider range of conditions.

Third, both the BIC and CAIC tended to outperform the SCAIC, SBIC, and the BLRT in the class enumeration process, and did so across most design conditions. However, the performance of these indicators was substantially reduced in conditions of low class separation. In these conditions, the SBIC, the BLRT, and to a lesser extent the SCAIC, tended to outperform the BIC and CAIC. Based on these results, we recommend that the class enumeration process be conducted in two-steps, starting with the examination of the entropy (which is routinely provided as part of most statistical packages providing mixture modeling capabilities) as an indicator of class separation. For models resulting in a low level of class separation (corresponding to an entropy value close to 0.50 in this study), the SBIC, BLRT, and SCAIC should be preferred. Conversely, when the entropy suggests a higher level of class separation (closer to a value of 0.90 in this study), then the BIC and CAIC should be preferred. It should be noted that an important limitation of our recommendations is that our results emerged from a partial factorial design in which the high class separation condition, but not the low class separation condition, was specified as a normative condition around which the other design factors were varied. Thus, we can confidently conclude that both the BIC and CAIC are quite robust across many design conditions when class separation is high. However, the performance of the indicators in the low class separation condition was only evaluated within the normative condition of $N = 1000$ with a mixing ratio of 0.50/0.50 and an invariant within-class variance-covariance matrix. Future studies would do well to more systematically examine whether the SBIC, BLRT, and SCAIC remain as accurate in conditions of low class separation across other design conditions.

Fourth, the results suggested that the class enumeration performance of GMM tended to be better – or at least not markedly reduced – when no covariate was included in the model across most studies, conditions, and types of covariates specifications. Indeed, although both Studies 2 and 3 include examples of design conditions where including covariate(s) could improve the accuracy of the class enumeration process, these improvements remained small and of limited practical significance. Furthermore, these few examples are limited to models where the effect of the covariate is correctly specified in line with the population model. In contrast, misspecifications of covariates effects tend to be related to often drastic decreases in accuracy. In typical applied research settings, it is often impossible to ascertain beforehand the exact type of covariate effect that is present in the population model. Theoretical expectations may lead one to expect some degree of associations between covariates and the GMM model, but theory and research are rarely advanced enough to suggest whether these effects should be limited to the prediction of class membership, extended to the growth factors, or allowed to differ across classes, and to ensure that no meaningful covariate has been excluded from the study. In this context, our results suggest that the risks associated with the misspecification of these effects tend to be substantially greater than the risk of simply omitting important covariates from a model during the class enumeration process. Furthermore, the size of the correlations between the covariates has a substantial impact on the accuracy of the class enumeration process and the partial inclusion of active covariates (especially those with smaller effects) tends to outperform their complete inclusion. In this context, and taking into account the fact that applied data sets are likely to correspond to complex population models involving multiple covariates, our results call for caution before the decision is made to include covariates in the class enumeration process.

A few other elements deserve additional attention. First, as expected, we found that non-invariant variance-covariance matrices and low class separation reduced the class enumeration accuracy (Bauer & Curran, 2003; Lubke & Muthén, 2007; Peugh & Fan, 2012; Tofighi & Enders, 2007). Similarly, although this is the first study in which this condition was considered, we found that the size of the correlation between the covariates had a non-negligible impact on the class enumeration performance when covariates were included in the model. However our results show that the sample size and the mixing ratio only had limited effects on the class enumeration accuracy. In the case of the sample size, this result can likely be related to the fact that our studies, following Li and Hser (2011), only considered models with seven measurement points. Previous GMM studies in which a more substantial effect of the sample size was observed only considered sample size variations across a normative condition including four measurement points (Nylund et al., 2007; Peugh & Fan, 2012; Tofighi & Enders, 2007). In contrast, Li and Hser (2011), using a seven measurement points normative condition, found sample sizes effects similar to the effects observed here. This observation is in line with the results from simulation studies of latent growth models showing that sample size limitations can be compensated by increasing the number of measurements points (Diallo & Morin, 2015; Diallo, Morin & Parker, 2014). However, the weak effects of the mixing ratio suggest that this condition could be dropped from future studies of GMM to make room for other design factors. For instance, important considerations that were not considered here have to do with the possible impact of different degrees of deviation from the within-class normality assumptions of GMM (Tofighi & Enders, 2007), or within-class variations in the shape of the trajectories (Peugh & Fan, 2012).

Limitations

Unfortunately, it is very seldom possible to consider all plausible scenarios in a single study, especially one focusing on a type of model as complex as GMM. Therefore, generalization of the results beyond the range of conditions considered in the present study should be done with caution. In particular, three limitations of the present study are worth considering. First, population models were estimated as meeting within-class multivariate normality assumptions and the latent trajectories were simulated as following curvilinear trends in all classes so that the impact of deviation from these conditions remains unknown. Second, the population model assumed the class-invariance of the residual variances of the repeated measures and of the covariates effects on the growth factors (when present). However, as discussed by Tofighi and Enders (2007), GMM can be much less restrictive and the generalizability of our results to an even wider range of conditions is worth considering in the context of future studies. Third, in studies 2 and 3 we only considered time-invariant covariates and ignored time-varying covariates as well distal outcomes. Like time-invariant covariates, time-varying covariates and distal outcomes may also influence the class enumeration process. Pending further research, we thus cannot assume that the current results will generalize to these other situations.

Recommendations for Practice

Based on the results from the current study, we advise researchers to conduct the GMM class enumeration process without covariates. Consistent with the three-step procedure (Vermunt, 2010;

Asparouhov & Muthen, 2014), the current results, and our own experience of working with these models, we propose the following sequence for modeling the relationships between the GMM model parameters (class membership, and growth factors) and covariates. In presenting this sequence, we assume that readers are already familiar with the GMM estimation process. Less familiar readers should consult Morin (2016) and Morin and Wang (2016), which provide (particularly in their online supplements) an extensive pedagogical coverage of the estimation of GMM and mixture models more generally, as well as practical suggestions regarding how to deal with the challenges that may face researchers working with these models.

(1) The class enumeration process should be conducted without covariates. At this stage, applied researchers will be facing four different alternatives.

   a. The model without covariates converges on a proper solution with well-replicated log-likelihood. In this situation the estimated model is retained for the second step.

   b. The model converges on an interpretable solution with well-replicated log-likelihood and no empty classes, but includes out-of-bound parameter estimates (negative residual or variance estimates, out-of-bound correlation estimates, etc.). In this situation, these out-of-bound parameter estimates may simply reflect random sampling variations or the presence of a slight over-parameterization in a subset of classes (e.g., Tolvanen, 2007). In this situation, researchers should re-estimate the model after increasing the number of starting values. If the situation is unchanged, the model can be re-estimated after constraining the out-of-bound parameter estimate (e.g., negative variances or residuals) to be equal or higher than the boundary value (e.g., zero). If these suggestions do not work, then researchers can move on to 1c.

   c. The model converges on a solution with a well-replicated log-likelihood, but includes multiple severely out-of-bound parameter, or result in a degenerate solution including empty classes. In this situation, researchers should simplify the model in hand. The simplest way of doing so is to fix some model parameters (e.g., time specific residuals, growth factor variances-covariances) to be invariant across classes, and then restart the process from 1a.

d. The model fails to converge, or to converge on a solution with a well-replicated log-likelihood. In this situation, users should increase the number of starting values first. If the situation persists researchers should simplify the model in hand using strategies similar to those suggested in 1c, and then restart the process from 1a.

(2) Once the optimal model has been retained, and the final number of latent classes on the data has been determined, researchers should include covariates in the model to establish relationship between covariates and the optimal number of latent while taking into account individuals' probabilities of membership into each of the latent classes. In this step it is preferable to contrast models in which the covariates are allowed to influence: (a) the class membership; (b) the class membership and the growth factors in a class-invariant manner; (c) the class membership and the growth factors in a class-specific manner. The relative fit of these models can be contrasted using classical Likelihood ratio tests, as well as the ICs covered in this study (e.g., Lubke & Neale, 2006, 2008). Examples of this procedure are available in the literature (e.g., Morin, Maïano et al., 2011; 2014; Morin, Rodriguez et al., 2012). The efficacy of this procedure in correctly identifying the nature of covariates effects should be the object of future studies in which models including the correct number of classes but different covariate specification are systematically contrasted. Given that our results show that GMM best recovers the true number of latent classes without covariates, future studies would thus do well to investigate how to best recover the optimal specification of covariates in models including the correct number of latent classes.

Conclusion

The key objective of the current series of studies was to provide recommendations as to whether the GMM class enumeration process should be conducted with, or without, having covariates directly included in the model. Our answer is that generally, the class enumeration process should be conducted without the inclusion of covariates in the model, and is not likely to result in biased results regarding the number of latent classes in the data even when key covariate effects are present in the population model. However, although comprehensive, it is important to note than this study did not address the perhaps even more important issue of how to assess covariate effects in the context of GMM, and how to ensure that model results, in terms of parameter estimates, remain unbiased by covariate inclusion, or exclusion. On this regard, our result simply suggests that covariates, when appropriate, should be included in the model after the class enumeration process has been completed. Substantively, it has previously been argued that theory and caution should guide the decision to include, or not, covariates in a model (Marsh et al., 2009; Morin, Maïano, et al., 2011; Morin, Morizot et al., 2011). In the current study, we only considered the case where covariates were time-invariant predictors. However, among the key substantive considerations that should be taken into account is the anticipated nature of the covariates themselves as either predictors of the growth process under consideration, correlates of the growth process, or simply outcomes of the growth process (see e.g., Morin, Maïano, et al., 2011). For interested readers, previous research provides guidelines regarding the key issues to consider, and recommended methods, when covariate effects need to be investigated in the context of mixture models (e.g., Asparouhov & Muthén, 2014; Lanza, Tan & Bray, 2013; Petras & Masyn, 2010; Vermunt, 2010).

Endnotes

[1] We also simulated data according to a mixing ratio of .10/.90. Comparison between the unbalanced scenarios (.30/.70; .10/.90) only resulted in very modest changes for all outcomes considered, with a single exception showing that the mixing ratio of .10/.90 tended to result in slightly more nonconverging or improper solutions with sample size of 200 than the mixing ratio of .30/.70.

[2] The idea that model estimation tends to result in estimates that are closer to the population values when the population generating model involves class-invariant variance-covariance matrices does not suggest that practical applications to real data should rely on models that implicitly assume such constraints without first verifying their adequacy (Morin, Maïano et al., 2011; Peugh & Fan, 2013).

[3] A replication was labelled nonconvergent when it failed to converge on a solution with any of the sets of starting values (Bauer & Curran, 2003; McLachlan & Peel, 2000). A solution was labelled improper when it included an out-of-bound parameter estimates (e.g., negative variances, correlations ≥ 1 or ≤ -1), resulted in empty classes, or produced warnings that the solutions may not be trustworthy

(e.g., non-positive definite matrix), etc. Because improper solutions are rarely interpreted (Chen, Bollen, Paxton, Curran, & Kirby, 2001) they were excluded from the reported results. Additional analyses including them (e.g., treating them as supporting solutions with fewer latent classes; Tofighi & Enders, 2007) did not lead to substantially different conclusions than those reported here.

[4] The entropy captures the fuzziness, or accuracy, of the GMM classification. The fuzziness of the classification is measured at the individual level by the posterior probabilities of class membership. A high level of classification uncertainty (i.e., fuzziness) is present when individuals have similar posterior probabilities of membership into the various classes. In contrast, a high level of classification accuracy is present when individuals have highly differentiated posterior probabilities of membership into the various classes. The normalized entropy is measured on a zero to one scale and is commonly used to examine the degree to which the classification resulting from a specific model is accurate. Entropy is known to be negatively influenced by the level of within-class variability, and positively influenced by the degree of class separation (e.g., Lubke & Muthén, 2007; Muthén, 2004). Given the clear relation between entropy and the degree of class separation, we consider entropy values > 0.80 to reflect a high level of class separation, and whereas ≤ 0.50 to reflect a poor level of class separation. The calculation of the entropy is described in the online supplements.

References

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

Asparouhov, T., & Muthén, B.O. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, *21*, 1-13.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent 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.

Beauchaine, T. P., & Beauchaine, R. J. (2002). A comparison of maximum covariance and k-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*, *7*, 245-261.

Bergman, L.R., Magnusson, D., & El-Khouri, B.M. (2003). *Studying individual development in an interindividual context: A person-oriented approach*. Mahwah, NJ: Erlbaum.

Bozdogan, H. (1987). Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370.

Chen, F., Bollen, K., Paxton, P., Curran, P., & Kirby, J. (2001). Improper solutions in structural equations: Causes, consequences, strategies. *Sociological Methods & Research, 29,* 468–508.

Cohen, J. (1988). *Statistical Power for the Behavioral Sciences* (second ed.). Hillsdale, NJ: Erlbaum.

Diallo, T.M.O, & Morin, A.J.S. (2015). Power of Latent Growth Curve Models to Detect Piecewise Trajectories. *Structural Equation Modeling*. Early view DOI: 0.1080/10705511.2014.935678

Diallo, T.M.O., Morin, A.J.S., & Parker, P.D. (2014). Statistical Power of Latent Growth Curve Models to Detect Quadratic Growth. *Behavior Research Methods, 46*, 357-371.

Enders, C. K., & Tofighi, D. (2008). The impact of miss-specifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*, 75–95.

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

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

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.

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

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

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. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models? *Multivariate*

Behavioral Research, *43*, 592-620.

Marsh, H.W., Lüdtke, O., Trautwein, U., & Morin, A.J.S. (2009). Latent Profile Analysis of Academic Self-concept Dimensions: Synergy of Person- and Variable-centered Approaches to the Internal/External Frame of Reference Model. *Structural Equation Modeling, 16*, 1-35.

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

Morin, A.J.S. (2016, In Press). Person-centered research strategies in commitment research. In J.P. Meyer (Ed.), The Handbook of Employee Commitment. Cheltenham, UK: Edward Elgar.

Morin, A.J.S., Maïano, C., Marsh, H.W., Nagengast, B., & Janosz, M. (2014, In press). School life and adolescents' self-esteem trajectories. *Child Development*. DOI: 10.1111/cdev.12089

Morin, A.J., Maïano, C., Nagengast, B., Marsh, H.W., Morizot, J. & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: Impact of untested invariance assumptions on interpretations. *Structural Equation Modeling, 18,* 613-648.

Morin, A.J.S., & Marsh, H.W. (2014, in press). Disentangling Shape from Levels Effects in Person-Centred Analyses: An Illustration Based University Teacher Multidimensional Profiles of Effectiveness. *Structural Equation Modeling.*

Morin, A.J.S., Morizot, J., Boudrias, J.-S., & Madore, I., (2011). A multifoci person-centered perspective on workplace affective commitment: A latent profile/factor mixture Analysis. *Organizational Research Methods, 14*, 58-90.

Morin, A.J.S., Rodriguez, D., Fallu, J.-S., Maïano, C., & Janosz, M. (2012). Academic Achievement and Adolescent Smoking: A General Growth Mixture Model. *Addiction*, *107*, 819-828.

Morin, A.J.S., & Wang, J.C.K. (2016). A gentle introduction to mixture modeling using physical fitness data. In N. Ntoumanis, & N. Myers (Eds.), *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists* (pp. 183-210). London, UK: Wiley

Muthén, B.O. (2003). Statistical and Substantive Checking in Growth Mixture Modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 369-377.

Muthén, B. (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, CA: Sage.

Muthén, B. & Asparouhov, T. (2009). *Growth mixture modeling: Analysis with non-Gaussian random effects*. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 143-165. Boca Raton: Chapman & Hall/CRC Press.

Muthén, B. & Asparouhov, T. (2015). Growth mixture modeling with non-normal distributions. *Statistics in Medicine*, *34*, 1041-1058.

Muthén, L., & Muthén, B. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

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

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

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. Piquero & D. Weisburd (Eds.), Handbook of Quantitative Criminology (pp. 69-100). New York, NY: Springer.

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

Peugh, J. & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling*, *20*, 616-639.

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

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

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*, *20*, 640-657.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age.

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

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*, 307–333.

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*, 339-361.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*, 1090–1104.

Table 1.

Summary of the Basics Simulation Design Conditions used on All Studies.

Condition	Sample Size	Mixing Ratio	Var.-Covar. Matrix	Class Separation
1 (normative condition)	**1000**	**0.5/0.5**	**Invariant**	**High**
2	**200**	0.5/0.5	Invariant	High
3	**400**	0.5/0.5	Invariant	High
4	**2000**	0.5/0.5	Invariant	High
5	1000	**0.3/0.7**	Invariant	High
6	1000	0.5/0.5	**Variant**	High
7	1000	0.5/0.5	Invariant	**Low**

Table 2.

Percentages of Times Enumerated by the Indicators for Different Number of Classes by Class Separation for Study 1.

		Model without Covariates				Model with C on Inactive Covariate				Model with C, I S invariant on Inactive Covariate				Model with C, I S variant on Inactive Covariate			
# of Classes		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
High	AIC	0	72.90	21.20	5.90	0	70.90	21	8.20	0	75.40	15.80	8.90	0	69.70	20.1	10.2
	CAIC	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	SCAIC	0	99.60	0.40	0	0	99.90	0.10	0	0	99.90	0.1	0	0	100	0	0
	BIC	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	SBIC	0	97.70	2.30	0	0	98.50	1.50	0	0	99.10	0.90	0	0	99.60	0.40	0
	LMR	0	74	23.70	2.30	0	85	13.10	1.90	0	88.70	9	2.30	0	84.70	13.20	2.10
	ALMR	0	75.40	22.80	1.80	0	85.50	12.60	1.80	0	89.30	8.90	1.90	0	85.70	12.40	1.90
	BLRT	0	90.20	9.10	0.70	0	87.10	11.20	1.70	0	91.30	7.70	0.90	0	92.60	6.80	0.60
Low	AIC	1.40	71.60	20.30	6.70	0	59.80	26.40	13.80	0.90	52.20	27.50	19.40	1.20	51.40	27.20	20.20
	CAIC	72.10	27.90	0	0	0	100	0	0	88.20	11.80	0	0	98.30	1.70	0	0
	SCAIC	21.90	77.50	0.60	0	0	99.60	0.40	0	35.60	64.20	0.20	0	65	35	0	0
	BIC	58.40	41.60	0	0	0	100	0	0	77.10	22.90	0	0	94.70	5.30	0	0
	SBIC	9.40	88.50	2	0.10	0	97.60	2.40	0	16	82	2	0	37.90	61.10	0.90	0
	LMR	8.60	73.30	16.10	2	10.10	67.30	19.10	3.60	17.60	59.10	19.60	3.80	23.30	59.60	14.20	2.90
	ALMR	9.30	73.80	15.20	1.70	10.50	68.10	17.80	3.60	18.40	59.20	18.80	3.60	24.20	59.30	13.70	2.80
	BLRT	3.90	85.10	10.30	0.70	7.50	75.40	15	2.10	8	68.10	19.80	4.10	13.80	69.20	13.30	3.70

Note. Class separation was varied while holding other design factors constant as follows: Sample size N = 1000, same matrix variance of covariance, and mixing ratio of 50%:50%. C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor; AIC = Akaike's Information Criterion; CAIC = Consistent AIC; SCAIC = Sample Size Adjusted CAIC; BIC = Bayesian Information Criterion; SBIC = Sample Size Adjusted BIC; LMR = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR = Adjusted LMR; BLRT = Bootstrapped Likelihood Ratio Test.

Table 3.

Summary of Results for Study 1

Covariate Condition	Sample Size	Class Separation	Mixing Ratio	Variance-Covariance
(A) Without Covariates	Small impact of sample size.	**High Separation:** ICs/LRTs perform well. **Low Separation**: Underextraction for CAIC, SCAIC, BIC, SBIC and BLRT (severe for BIC and CAIC)	Small impact of the mixing ratio conditions.	Small impact on the ICs. **Invariant:** LRTs perform well. **Non-Invariant**: Overextraction for LRTs
(B) With inactive covariate effects on C	***N*=200**: Overextraction for ICs. ***N* ≥ 400**: (B) ≈ (A).	Small impact of class separation on ICs and LRTs; (B) ≥ (A) in the low separation condition.	(B) ≈ (A)	(B) ≈ (A) for ICs **Invariant:** (B) ≈ (A) for LRTs **Different**: (B) ≥ (A) for LRTs
(C) With inactive covariate effects on C, I, S invariant	(C) ≈ (A).	**High Separation:** ICs/LRTs perform well. **Low Separation**: Underextraction for CAIC, SCAIC, BIC, and LRTs (severe for BIC and CAIC); (A) ≥ (C)	(C) ≈ (A)	(C) ≈ (A) for ICs **Invariant:** (B) ≈ (A) for LRTs **Non-Invariant**: (C) ≥ (B) for LRTs
(D) With inactive covariate effects on C, I, S variant	(D) ≈ (A).	**High Separation:** ICs/LRTs perform well. **Low Separation**: Underextraction for CAIC, SCAIC, BIC, SBIC and LMR/ALMR (severe for CAIC, SCAIC, BIC); Overextraction for AIC; (C) ≥ (D).	(D) ≈ (A)	(D) ≈ (C)
Summary of covariate effects	**ICs:** ***N* = 200**: (A) ≈ (C) ≈ (D) ≥ (B) ***N* ≥ 400**: (A) ≈ (B) ≈ (C) ≈ (D) **LRTs**: (A) ≈ (B) ≈ (C) ≈ (D)	**High Separation:** (A) ≈ (B) ≈ (C) ≈ (D) **Low Separation:** (B) ≥ (A) ≥ (C) ≥ (D).	(A) ≈ (B) ≈ (C) ≈ (D)	**ICs:** (A) ≈ (B) ≈ (C) ≈ (D) **LRTs:** **Invariant: (**A) ≈ (B) ≈ (C) ≈ (D) **Non-Invariant:** (D) ≈ (C) ≥ (B) ≥ (A)

Note. IC = Information Criteria; LRT = Likelihood Ratio Test; CAIC = Consistent Akaike's Information Criterion; SCAIC = Sample Size Adjusted CAIC; BIC = Bayesian Information Criterion; SBIC = Sample Size Adjusted BIC; LMR = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR = Adjusted LMR; BLRT = Bootstrapped Likelihood Ratio Test; C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor.

Table 4.

Percentages of Times Enumerated by the Indicators for Different Number of Classes by Class Separation for Study 2 with an X1-X2 Correlation of 0.20 and with X1-X2 centered at 0 and with unit variance

		Model without Covariates				Model with C on Covariates X1 X2				Model with C, I S invariant on Covariates X1 X2				Model with C, I S variant on Covariate X1 X2			
# of Classes		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
High	AIC	0	72.80	24.20	3	0	32.20	46.30	21.40	0	30.30	56.50	13.20	0	43.70	46.30	10
	CAIC	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	SCAIC	0	99.70	0.30	0	0	99.80	0.20	0	0	99.70	0.30	0	0	100	0	0
	BIC	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	SBIC	0	98.30	1.60	0.10	0	97.40	2.50	0.10	0	96.50	3.50	0	0	99.50	0.50	0
	LMR	0	71.40	25.30	3.30	0	52.20	34.60	13.20	0	48.80	39.20	12	0	57.30	39.10	3.60
	ALMR	0	72.70	24.30	3	0	53	34.20	12.80	0	50	38.70	11.30	0	57.70	38.80	3.50
	BLRT	0	86.10	13.10	0.80	0	73	22	5	0	65	30	5	0	74.70	22.90	2.40
Low	AIC	0.70	75.10	21.80	2.40	0	33.60	47.20	19.30	0	30	47.20	22.80	0.10	41.30	46.70	11.90
	CAIC	68.10	31.90	0	0	0	100	0	0	54.70	45.30	0	0	98.10	1.90	0	0
	SCAIC	20.10	79.70	0.20	0	0	99.80	0.20	0	7.60	91.90	0.40	0	44	56	0	0
	BIC	53.80	46.20	0	0	0	100	0	0	36.90	63.10	0	0	91.40	8.60	0	0
	SBIC	9	89.80	1.20	0	0	96.80	3.20	0	1.60	95.10	3.20	0.10	14.90	84.80	0.30	0
	LMR	8	73.10	15.70	3.20	0	49	36.20	14.70	5.90	45.50	40.90	7.60	8.60	53.70	32.40	5.30
	ALMR	8.70	73.80	14.80	2.70	0	49.90	35.60	14.50	6.40	46	40.50	7.10	9	54.10	31.60	5.30
	BLRT	2.10	88	9.80	0.10	0	66.30	27.30	6.40	0.90	52.70	34.80	11.70	3.20	57.90	31.90	7

Note. Class separation was varied while holding other design factors constant as follows: Sample size *N* = 1000, same matrix variance of covariance, and mixing ratio of 50%:50%. C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor; AIC = Akaike's Information Criterion; CAIC = Consistent AIC; SCAIC = Sample Size Adjusted CAIC; BIC = Bayesian Information Criterion; SBIC = Sample Size Adjusted BIC; LMR = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR = Adjusted LMR; BLRT = Bootstrapped Likelihood Ratio Test.

Table 5.

Summary of Results for Study 2

Covariate Condition	Sample Size	Class Separation	Mixing Ratio	Variance-Covariance	Correlation
(A) Without Covariates	Small impact of sample size	**High Separation:** ICs/LRTs perform well. **Low Separation**: Underextraction for ICs (severe for BIC and CAIC)	Small impact of mixing ratio conditions	Small impact of variance-covariance conditions	No impact
(B) Partial inclusion of covariate effects on C	(B) ≈ (A); X1 ≈ X2	**High Separation:** (B) ≈ (A). **Low Separation**: (B) ≥ (A) for ICs; (A) ≥ (B) for BLRT. X2 ≥ X1.	(B) ≈ (A)	**Invariant:** (B) ≈ (A) **Non-Invariant**: (B) ≥ (A) for SBIC, BLRT; (B) ≈ (A) for other ICs. X1 ≈ X2	Small impact
(C) Total inclusion of covariate effects on C	**SBIC; BLRT**: (B) ≥ (C) **Other IC**: (B) ≈ (C)	**High Separation:** (C) ≈ (B). **Low Separation**: (C) ≈ (B) for ICs; (B) ≥ (C) for BLRT	**ICs:** (C) ≈ (B) **BLRT:** (B) ≥ (C)	**Invariant:** (C) ≈ (B) **Non-Invariant**: (B) ≥ (C) for SBIC, BLRT; (C) ≈ (B) for other ICs;	Small impact
(D) Partial inclusion of covariate effects on C, I, S invariant	(D) ≈ (B); X1 ≈ X2	**High Separation:** (D) ≈ (B). **Low Separation**: (B) ≥ (D) for ICs; (D) ≈ (B) for BLRT. X2 ≥ X1.	(D) ≈ (B)	(D) ≈ (B)	Small impact
(E) Total inclusion of covariate effects on C, I, S invariant	**SBIC; BLRT**: (C) ≥ (E) **Other IC**: (C) ≈ (E)	**High Separation:** (E) ≈ (C). **Low Separation**: (C) ≥ (E) for ICs; (E) ≈ (C) for BLRT.	(E) ≈ (C)	(E) ≈ (C)	Small impact
(F) Partial inclusion of covariate effects on C, I, S variant	(F) ≈ (D); X1 ≈ X2	**High Separation:** (F) ≈ (D). **Low Separation**: (D) > (F) for ICs; (F) ≈ (D) for BLRT. X2 ≥ X1.	(F) ≈ (D)	(F) ≈ (D)	Small impact
(G) Total inclusion of covariate effects on C, I, S variant	**SBIC; BLRT**: (G) ≥ (E) **Other IC**: (G) ≈ (E)	**High Separation:** (G) ≈ (E). **Low Separation**: (E) > (G) for ICs; (G) ≈ (E) for BLRT.	(G) ≈ (E)	**Invariant:** (G) ≈ (E) **Non-Invariant**: (G) ≥ (E) for SBIC, BLRT; (G) ≈ (E) for other ICs;	Small impact
Summary of Covariate Effects	**SBIC; BLRT:** (A) ≈ (B) ≈ (D) ≈ (F) ≥ (C) ≈ (G) ≥ (E) **Other IC:** (A) ≈ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G)	**High:** (A) ≈ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G) **Low, ICs**: (C) ≈ (B) ≥ (A) ≥ (D) ≈ (E) > (F) ≈ (G) **Low, BLRT**: (A) ≥ (B) ≈ (D) ≈ (F) ≥ (C) ≈ (E) ≈ (G)	**ICs:** (A) ≈ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G). **BLRT:** (A) ≈ (B) ≈ (D) ≈ (F) ≥ (C) ≈ (E) ≈ (G)	**SBIC/BLRT (Invariant), Other ICs (all conditions):** (A) ≈ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G). **SBIC/BLRT (Non-Invariant)**: (B) ≈ (D) ≈ (F) ≈ (G) ≥ (A) ≥ (C) ≈ (E).	Small inconsistent impact in models with covariates.

Note. IC = Information Criteria; CAIC = Consistent Akaike's Information Criterion; BIC = Bayesian Information Criterion; SBIC = Sample size adjusted BIC; BLRT = Bootstrapped Likelihood Ratio Test; C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor.

Table 6.

Percentages of Times Enumerated by the Indicators for Different Number of Classes by Class Separation for Study 3 with an X1-X2 Correlation of 0.20 and with X1-X2 centered at 0 and with unit variance

		Model without Covariates				Model with C on Covariates X1 X2				Model with C, I S invariant on Covariates X1 X2				Model with C, I S variant on Covariate X1 X2			
# of Classes		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
High	AIC	0	71.60	23.20	5.20	0	0	0	100	0	48.50	20.60	30.90	0	52.30	25.60	22.10
	CAIC	0	100	0	0	0	84.10	8.30	7.60	0	100	0	0	0	100	0	0
	SCAIC	0	99.70	0.30	0	0	10.30	3.80	85.90	0	99.40	0.60	0	0	100	0	0
	BIC	0	100	0	0	0	63.70	11.30	25	0	100	0	0	0	100	0	0
	SBIC	0	98.20	1.70	0.10	0	1.30	0.60	98.10	0	97.40	2.50	0.10	0	99.60	0.40	0
	LMR	0	71	26.40	2.60	0	31.20	28.50	40.30	0.10	76.40	17.90	5.60	0	82.20	14.80	3
	ALMR	0	72.90	25	2.10	0	33.30	28.30	38.40	0.10	77.20	17.40	5.30	0	82.90	14.20	2.90
	BLRT	0	88.10	11	0.90	0	0.20	0.20	99.60	0	81.10	13.50	5.40	0	85.50	11.20	3.30
Low	AIC	2.20	74	20.10	3.70	0	0.10	12.40	87.50	0	31.30	48	20.70	0.10	41.80	48.20	9.90
	CAIC	84	16	0	0	0	99.20	0.80	0	55.60	44.40	0	0	98.10	1.90	0	0
	SCAIC	34.50	65.20	0.30	0	0	61.70	24.10	14.20	7.60	91.90	0.50	0	44	56	0	0
	BIC	73.70	26.30	0	0	0	95.50	4.30	0.20	37.20	62.80	0	0	91.50	8.50	0	0
	SBIC	18.60	80	1.40	0	0	29.90	29.80	40.30	1.60	94.60	3.70	0.10	14.90	84.90	0.20	0
	LMR	14.90	67.10	15.60	2.40	2.20	56.50	30.70	10.60	5.90	45.40	39.90	8.80	8.50	54	32	5.50
	ALMR	16.10	67.10	14.70	2.10	2.30	57.90	29.90	9.90	6.10	46	39.50	8.40	8.80	54.60	31.10	5.50
	BLRT	8.90	81.30	9.10	0.70	0	11.60	25.80	62.60	1.20	54.60	32.60	11.60	3.50	59	32.10	5.40

Note: Class separation was varied while holding other design factors constant as follows: Sample size *N* = 1000, same matrix variance of covariance, and mixing ratio of 50%:50%. C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor; AIC = Akaike's Information Criterion; CAIC = Consistent AIC; SCAIC = Sample Size Adjusted CAIC; BIC = Bayesian Information Criterion; SBIC = Sample Size Adjusted BIC; LMR = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR = Adjusted LMR; BLRT = Bootstrapped Likelihood Ratio Test.

Table 7.

Percentages of Times Enumerated by the Indicators for Different Number of Classes by Matrix of Variance-Covariance for Study 3 with an X1-X2 Correlation of 0.20 and with X1-X2 centered at 0 and with unit variance

		Model without Covariates				Model with C on Covariates X1 X2				Model with C, I S invariant on Covariates X1 X2				Model with C, I S variant on Covariate X1 X2			
# of Classes		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Same	AIC	0	71.60	23.20	5.20	0	0	0	100	0	48.50	20.60	30.90	0	52.30	25.60	22.10
	CAIC	0	100	0	0	0	84.10	8.30	7.60	0	100	0	0	0	100	0	0
	SCAIC	0	99.70	0.30	0	0	10.30	3.80	85.90	0	99.40	0.60	0	0	100	0	0
	BIC	0	100	0	0	0	63.70	11.30	25	0	100	0	0	0	100	0	0
	SBIC	0	98.20	1.70	0.10	0	1.30	0.60	98.10	0	97.40	2.50	0.10	0	99.60	0.40	0
	LMR	0	71	26.40	2.60	0	31.20	28.50	40.30	0.10	76.40	17.90	5.60	0	82.20	14.80	3
	ALMR	0	72.90	25	2.10	0	33.30	28.30	38.40	0.10	77.20	17.40	5.30	0	82.90	14.20	2.90
	BLRT	0	88.10	11	0.90	0	0.20	0.20	99.60	0	81.10	13.50	5.40	0	85.50	11.20	3.30
Different	AIC	0	0	7.70	92.30	0	0	0	100	0	0	0	100	0	0	0	100
	CAIC	0	100	0	0	0	19.30	80	0.70	0	91.10	8.90	0	0	100	0	0
	SCAIC	0	100	0	0	0	0	35.70	64.30	0	8.10	86.10	5.80	0	100	0	0
	BIC	0	100	0	0	0	5.10	89.20	5.70	0	72.90	27.10	0	0	100	0	0
	SBIC	0	100	0	0	0	0	4.20	95.80	0	0.60	58.30	41.10	0	100	0	0
	LMR	0	69.20	15.40	15.40	0	30.60	44.40	25	0	26.10	61.60	12.30	0	64.30	35.70	0
	ALMR	0	69.20	15.40	15.40	0	31	44.30	24.70	0	27	61.60	11.50	0	64.30	35.70	0
	BLRT	0	92.30	7.70	0	0	0	0	100	0	0.60	26.60	72.80	0	92.90	0	7.10

Note: Matrix variance-covariance structure was varied while holding other design factors constant as follows: Sample size N = 1000, high class separation, and mixing ratio of 50%:50%. C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor; AIC = Akaike's Information Criterion; CAIC = Consistent AIC; SCAIC = Sample Size Adjusted CAIC; BIC = Bayesian Information Criterion; SBIC = Sample Size Adjusted BIC; LMR = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR = Adjusted LMR; BLRT = Bootstrapped Likelihood Ratio Test.

Table 8.
Summary of Results for Study 3

Covariate Condition	Sample size	Class separation	Mixing ratio	Variance-covariance	Correlation
(A) Without Covariates	Small impact of sample size	**High Separation:** ICs/BLRT perform well. **Low Separation**: Underextraction for ICs (severe for CAIC and BIC).	Small impact of mixing ratio conditions	Small impact of variance-covariance conditions	Small to moderate impact varying as a function of the separation and variance-covariance conditions
(B) Partial inclusion of covariate effects on C	Accuracy decreases for *N* ≥ 1000. (A) ≥ (B); X1 ≥ X2	**High Separation:** (B) ≈ (A) ICs; (A) ≥ (B) BLRT. **Low Separation**: (B) ≥ (A) ICs; (A) ≥ (B) BLRT. X1 ≥ X2.	(A) ≥ (B) X1 ≥ X2	**Invariant:** (B) ≈ (A) for ICs; (A) ≥ (B) for BLRT. **Non-Invariant**: (A) ≥ (B). X1 ≥ X2	Small to large impact varying as a function of the sample size, separation, mixing ratio, and variance covariance conditions
(C) Total inclusion of covariate effects on C	(C) ≈ (B)	**High Separation:** (B_{x1}) ≥ (C) ≥ (B_{x2}). **Low Separation**: (C) ≈ (B).	(C) ≈ (B)	(C) ≈ (B)	Moderate to large impact varying as a function of the separation, mixing ratio, and variance covariance conditions.
(D) Partial inclusion of covariate effects on C, I, S invariant	***N* < 1000**: (D) ≈ (B) ***N* ≥ 1000**: (D) ≥ (B) X1 ≥ X2	**High Separation:** (D) ≈ (B) ICs; (D) ≈ (A) BLRT. **Low Separation**: (B) ≥ (D) ICs; (D) ≥ (B) BLRT. X1 ≥ X2.	(D) ≈ (A)	(D) ≈ (A).	Small to moderate impact varying as a function of the sample size separation, mixing ratio, and variance covariance conditions
(E) Total inclusion of covariate effects on C, I, S invariant	(E) ≈ (C)	**High Separation:** (E) ≈ (B_{x2}). **Low Separation**: (E) ≈ (C)	(E) ≈ (C)	(E) ≈ (C)	Small impact varying as a function of the sample size, separation, mixing ratio, and variance covariance conditions
(F) Partial inclusion of covariate effects on C, I, S variant	(F) ≈ (D)	**High Separation:** (F) ≈ (D). **Low Separation**: (D) ≥ (F) for ICs; (F) ≈ (D) for BLRT.	(F) ≈ (D)	(F) ≈ (D).	Small to moderate impact varying as a function of the separation, mixing ratio, and variance covariance conditions
(G) Total inclusion of covariate effects on C, I, S variant	(G) ≈ (E)	(G) ≈ (E)	(G) ≈ (E)	(G) ≈ (E)	Small to moderate impact varying as a function of the separation, mixing ratio, and variance covariance conditions
Summary of Covariate Effects	***N* < 1000**: (A) ≥ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G) ***N* ≥ 1000**: (A) ≥ (D) ≈ (F) ≥ (B) ≈ (C) ≈ (E) ≈ (G)	**High, ICs:** (A) ≈ (D) ≈ (F) ≈ (B_{x1}) ≥ (C) ≥ (B_{x2}) ≈ (E) ≈ (G) **Low, ICs:** (B) ≈ (C) ≈ (E) ≥ (D) ≥ (A) ≥ (F) ≈ (G) **High, BLRT:**(A) ≈ (D) ≈ (F) ≥ (B_{x1}) ≥ (C) ≥ (B_{x2}) ≈ (E) ≈ (G) **Low, BLRT:**(A) ≥ (D) ≈ (F) ≥ (B) ≈ (C) ≈ (E) ≈ (G)	(A) ≈ (D) ≈ (F) ≥ (B) ≈ (C) ≈ (E) ≈ (G)	**Invariant, ICs:** (A) ≈ (B) ≈ (C) ≈ (D) ≈ (E) ≈ (F) ≈ (G) **Invariant (BLRT); Non Invariant:** (A) ≈ (D) ≈ (F) ≥ (B) ≈ (C) ≈ (E) ≈ (G)	Small to large, generally negative but inconsistent, impact across conditions, varying as function of the other conditions.

Note. IC = Information Criteria; LRT = Likelihood Ratio Test; CAIC = Consistent Akaike's Information Criterion; BIC = Bayesian Information Criterion; BLRT = Bootstrapped Likelihood Ratio Test; C = Latent Class Membership; I = Intercept Factor; S = Linear Slope Factor.