Impact of Misspecifications of the Latent Variance–Covariance and Residual Matrices on the Class Enumeration Accuracy of Growth Mixture Models

Thierno M. O. Diallo, Alexandre J. S. Morin & HuiZhong Lu


To link to this article: http://dx.doi.org/10.1080/10705511.2016.1169188

Published online: 09 May 2016.
Impact of Misspecifications of the Latent Variance–Covariance and Residual Matrices on the Class Enumeration Accuracy of Growth Mixture Models

Thierno M. O. Diallo, Alexandre J. S. Morin, and HuiZhong Lu

1Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, Australia, and Statistiques & M.N., Sherbrooke, Canada
2Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, Australia
3Calcul Québec, University of Sherbrooke, Sherbrooke, Canada

This series of simulation studies was designed to assess the impact of misspecifications of the latent variance–covariance matrix (i.e., $\Psi_k$) and residual structure (i.e., $\Theta_k$) on the accuracy of growth mixture models (GMMs) to identify the true number of latent classes present in the data. Study 1 relied on a homogenous (1-class) population model. Study 2 relied on a population model in which the latent variance–covariance matrix is constrained to be 0 ($\Psi_k = 0$). Study 3 relied on a population model in which the latent variance–covariance matrix was specified as invariant across classes ($\Psi_k = \Psi$). Finally, Study 4 relied on a more realistic specification of the latent variance–covariance matrix as different across classes ($\Psi_k$). In each of these studies, we assessed the class enumeration accuracy of GMMs as a function of different types of estimated model (6 models corresponding to the 3 types of population models used to simulate the data and involving the free estimation of the residual structure across latent classes or not) and 4 design conditions (within-class residual matrix, sample size, mixing ratio, class separation). Overall, our results show the advantage of relying on models involving the free estimation of the $\Psi_k$ and $\Theta_k$ matrices within all latent classes. However, based on the observation that inadmissible solutions occur more frequently in these models than in more parsimonious models, we propose a more comprehensive sequential strategy to the estimation of GMM.

Keywords: accuracy, class enumeration, growth mixture, latent class growth analysis, latent variance–covariance, misspecification, residual

INTRODUCTION

Over the past decade, growth mixture modeling (Muthén & Shedden, 1999) has become increasingly popular within the social (e.g., Wagmiller, Lennon, Kuang, Alberti, & Aber, 2006), educational (e.g., Boscardin, Muthén, Francis, & Baker, 2008), and health (e.g., Morin, Rodriguez, Fallu, Maïano, & Janosz, 2012) sciences. Part of this popularity stems from the ability of growth mixture models (GMMs) to overcome the main limitation of conventional growth models, which assume that all individuals are drawn from a single population for which common population parameters are estimated. GMMs represent longitudinal heterogeneity by identifying subgroups (latent classes) following qualitatively and quantitatively distinct trajectories (Morin et al., 2011).
Another reason for the increasing popularity of these models is the availability of user-friendly statistical packages, such as Latent GOLD (Vermunt & Magidson, 2005), the MMLCR library for Splus/R (Buyske, 2001), Mplus (Muthén & Muthén, 1998-2014), OpenMx (Boker et al., 2011), the GLMAM package for STATA (Rabe-Hesketh, Skrondal, & Pickles, 2004), and the PROC TRAJ macro for SAS (Jones, Nagin, & Roeder, 2001). However, these packages are not equivalent in terms of the level of flexibility they allow in the specification of GMMs, and regarding the default parameterization that they use to specify the within-class variance–covariance components, two characteristics that could have important implications but are seldom discussed explicitly. Perhaps the most widely known parameterization of growth mixture models comes from Nagin’s (1999) latent class growth analysis (LCGA) approach as implemented in SAS, which estimates no individual variability within the latent classes. This approach thus assumes that individual heterogeneity is completely captured by the latent classes (i.e., that all class members follow the exact same trajectory) so that deviations from the average within-class trajectories are assumed to random error. This restriction is not inherent to GMMs, which are fully suited to the examination of within-class variability.

However, other more flexible common parameterizations also rely on restrictive assumptions. For instance, the Mplus default specification of GMMs (which can be relaxed by the user) assumes that the within-class variance–covariance parameters (representing within-class variability around the estimated average growth trajectory, as well as the time-specific residuals) are invariant across latent classes. Finally, some programs rely on no default parameterization, but leave to the user the complex decision of which parameters to freely estimate, or constrain to equality, across classes. In practice the free estimation of all GMM parameters within each latent class often results in convergence on improper solutions or nonconvergence of the expectation maximization (EM) algorithm, leaving the user with the decision of where to include constraints. There has been a long-standing debate in the literature regarding the relative superiority of each possible approach (e.g., Nagin, 2010; Nagin & Tremblay, 2005; Sampson & Laub, 2005; for a review, see Morin et al., 2011). Unfortunately, over and above logical arguments favoring specific approaches, very few empirically derived guidelines exist to help users in this complex decision process. Providing such guidelines is the objective of the present series of studies. Before reviewing the limited evidence available to date, we provide a more detailed presentation of GMM specifications.

Latent Curve Models

When population homogeneity is assumed with respect to the estimated model parameters and global shape of the growth trajectories, then GMMs are reduced to latent curve models (LCMs; Bollen & Curran, 2006; Meredith & Tisian, 1990). In the structural equation modeling (SEM) approach to LCM, for p repeated measures, the unconditional LCM can be expressed as

$$y_i = \Lambda \eta_i + \varepsilon_i,$$

where $y_i$ is a $p \times 1$ vector of repeated measures for individual $i$, $\eta_i$ is a $q \times 1$ vector of latent variables representing the growth parameters, $\Lambda$ is a $p \times q$ matrix of factor loadings relating the growth factors to the observed variables and specified to reflect the passage of time, and $\varepsilon_i$ is a $p \times 1$ vector of time-specific residuals. The implied mean and covariance of the LCM are

$$\mu = \Lambda \alpha,$$

$$\Sigma = \Lambda \Psi \Lambda' + \Theta,$$

where $\alpha$ is the mean vector for the growth factors, $\Psi$ is a $q \times q$ matrix representing the latent growth factors’ variance–covariance, and $\Theta$ is a $p \times p$ matrix of time-specific residuals. An LCM generally assumes that the latent growth factors and the time-specific residuals are mutually independent, that they follow multivariate normal distributions, and that $\Theta$ is diagonal. Under these assumptions, the marginal distribution of the repeated measure can be represented as

$$f(y_i) = \phi(\Lambda \alpha, \Lambda \Psi \Lambda' + \Theta),$$

where $\phi$ is a $p$-dimensional normal probability density function.

Growth Mixture Models

GMMs relax the assumption of population homogeneity by relying on a finite mixture approach to handle population heterogeneity in the analysis of growth trajectories. In GMMs, a categorical latent variable is used to model population heterogeneity so that each extracted latent class can be characterized by its own LCM. Under multivariate normality distributions, the distribution of the repeated measures is represented by a finite mixture of normal distributions with a moment structure similar to that of the LCM. For the unconditional GMM, the density function is expressed as

$$f(y_i) = \sum_{k=1}^{K} p_k \phi_k(\Lambda_k \alpha_k, \Lambda_k \Psi_k \Lambda_k' + \Theta_k),$$
where the categorical latent variable contains \( k = 1, \ldots, K \) latent classes, and each latent class is defined by its own density function \( \phi_k \) with a mean of \( \Lambda_k \alpha_k \), and a covariance \( \Psi_k \Lambda_k' \Lambda_k + \Theta_k \) (these parameters have the same meaning as in an LCM). The mixing proportion parameter \( p_k \) defines the probability that an individual belongs to class \( k \) with all \( p_k \geq 0 \) and \( \sum_{k=1}^{K} p_k = 1 \). All parameters may vary across classes but, to reduce complexities during model estimation, restrictions can be imposed on the model. For instance, the functional form of the growth process is typically constrained to be the same across latent classes (i.e., \( \Lambda_k = \Lambda \)).

The variance–covariance of the latent growth factors, as well residual variance–covariance of the repeated measures may also be held invariant across classes (i.e., \( \Psi_k = \Psi \) and \( \Theta_k = \Theta \)). Taken together, these three constraints form the default setting in Mplus, which assumes that latent classes differ only in their mean trajectories (i.e., \( \Lambda_k \)). In the LCGA parameterization (as implemented in SAS), additional restrictions are imposed so that \( \Psi_k = 0 \).

For GMMs, the implied first- and second-order moments of the aggregate distribution of the repeated measures can be expressed as (Bauer & Curran, 2004):

$$
\mu = \sum_{k=1}^{K} p_k \Lambda_k \alpha_k,
$$

(6) $$
\Sigma = \sum_{k=1}^{K} \sum_{j=k+1}^{K} p_k p_j (\Lambda_k \alpha_k - \Lambda_j \alpha_j)(\Lambda_k \alpha_k - \Lambda_j \alpha_j)' + \sum_{k=1}^{K} p_k (\Lambda_k \Psi_k \Lambda_k' + \Theta_k).
$$

(7)

where the aggregate mean of the growth factors is a weighted sum of the class-specific means, and the aggregate variance–covariance can be decomposed into a between-class component and a within-class component. The between-class component (the first term) includes the mean difference across classes on the growth factors, whereas the within-class component (the second term) is a weighted sum of the variance–covariance components within each class. These two terms are interdependent.

**Misspecifications of the Latent Variance–Covariance and Residual Matrices in GMMs**

As noted previously, many restrictions are typically imposed on the latent variance–covariance (\( \Psi_k \)) and residual (\( \Theta_k \)) matrices during the estimation of GMMs. Although these restrictions are sometimes related to the statistical package that is used, they can also often be user-defined to simplify the estimation process. However, there has yet to be a comprehensive empirical investigation of the impact of these specifications on the performance of GMM. To date, what is known is that similar constraints could potentially result in the overextraction of spurious latent classes and biased parameter estimates in other types of mixture models (Bauer & Curran, 2004; Heggeseth & Jewell, 2013; Lubke & Muthén, 2007; Lubke & Neale, 2006, 2008; Magidson & Vermunt, 2004; McLachlan & Peel, 2000; Peugh & Fan, 2013). Similarly, when they are systematically tested in practice, these constraints are often empirically rejected (e.g., Bauer & Curran, 2003; Morin et al., 2011).

LCGA has also been previously shown to potentially result in the overextraction of latent trajectory classes (e.g., Bauer & Curran, 2004; Muthén & Muthén, 2000). However, with few exceptions (e.g., McLachlan & Peel, 2000; Muthén et al., 2000) in which it is quickly mentioned that relying on a class-varying latent variance–covariance matrix could possibly make a substantive difference, the global assumption is that relying on restricted parameterizations of GMMs is unlikely to result in substantially biased results. Rather, it is argued that such restricted parameterizations might simply result in the extraction of a greater number of latent classes, differing from one another only quantitatively and thus having a similar interpretation, as a way to absorb the unmodeled within-class variability. However, Morin et al. (2011) recently showed that restricted specifications of the latent variance–covariance matrix in GMMs were likely to result in more severe biases than previously expected. These authors contrasted three models based on the LCGA parameterization, the Mplus default parameterization, and a third model in which the latent variance–covariance components were freely estimated in all classes (\( \Psi_k \)) but not the time-specific residuals (\( \Theta_k = \Theta \)). Although their results are based on real data and thus provide no clue as to the true population model, they showed that the three parameterizations resulted in the extraction of the same number of latent classes, but in drastically different solutions in terms of within-class average trajectories and sizes of the latent classes. Our objective here is to extend Morin et al.’s (2011) results using simulated data.

Three types of misspecifications of the variance–covariance structure have been considered in the context of growth modeling: underspecification, overspecification, and general misspecification (e.g., Ferron, Dailey, & Yi, 2002; Kwok, West, & Green, 2007). These three types of misspecification can be considered separately for \( \Psi_k \) and \( \Theta_k \). For instance, underspecification can occur when the true model includes class-specific matrices (i.e., \( \Psi_k \) or \( \Theta_k \)) while the model specification selected for the analysis constrains the within-class matrices to be equal across classes (i.e., \( \Psi_k = \Psi \) or \( \Theta_k = \Theta \)) or equal to zero (i.e., \( \Psi_k = 0 \)). Alternatively, overspecification occurs when the true model involves matrices that are equal across classes on either the latent growth factors (i.e., \( \Psi_k = \Psi \)) or the time-specific residuals (i.e., \( \Theta_k = \Theta \)) while the model selected for the analysis requests the estimation of class-specific matrices (i.e., \( \Psi_k \) or \( \Theta_k \)). Similarly, overspecification also occurs when the true model has no variability on the latent
growth factors (i.e., $\Psi_k = 0$) and is analyzed while requesting within-class estimates of variability. Finally, general misspecification can occur when the true model is homogenous, corresponding to an LCM (i.e., there is a single population), but the analysis is conducted assuming population heterogeneity (using LCFA or GMM).

Bauer and Curran (2003, 2004) have probably devoted the most attention to the critical evaluation of GMMs as a way to model population heterogeneity in growth trajectories. Although these authors mainly focused on the effects of violating critical assumptions (linearity, normality, etc.) of various parameterizations of mixture models (latent profile analyses, GMM, etc.), they also demonstrated that a general misspecification of the covariance structure when GMMs are estimated based on data generated from a homogenous population model could result in the extraction of spurious latent classes and, thus, biased parameter estimates. Peugh and Fan (2012) more recently showed that GMMs were able to correctly identify a one-class solution when applied to data generated from a homogenous population when model assumptions were reasonably met. To our knowledge, a single study has investigated the consequences of underspecifying the residual structure for the class enumeration process and parameter bias of GMMs (Enders & Tofghi, 2008). In this study, the true $\Theta_k$ was class-specific but underspecified as constant across classes ($\Theta_k = \Theta$). The findings revealed that such underspecification of $\Theta_k$ resulted in biased estimates of the within-class growth trajectories ($\Lambda_{k}$) and latent variance–covariance estimates ($\Psi_k$), as well as a lack of precision for most parameter estimates. However, only a modest impact on class enumeration was observed. The effects of overspecifications and general misspecifications were not considered in this study, though. To our knowledge, no study has yet systematically considered the impact of overspecifications or underspecifications of the latent variance–covariance matrix ($\Psi_k$), or examined the consequences of the three different types of misspecifications while considering both $\Psi_k$, $\Theta_k$, and their interaction.

**Method**

For this series of simulation studies, a quadratic LCM (Study 1), a three-class LCGA (Study 2: $\Psi_k = 0$, specified either with unequal $\Theta_k$ or equal $\Theta_k = \Theta$ across latent classes), and a three-class GMM (specified with equal $\Psi_k = \Psi$ in Study 3 or unequal $\Psi_k$ in Study 4, and with unequal $\Theta_k$ or equal $\Theta_k = \Theta$ across latent classes) were used as population models. For Studies 2 to 4, two of the classes were specified according to a quadratic growth function, whereas the remaining class was specified according to a linear growth function. The population parameters used in this study, as a function of design conditions presented here, are reported in Table 1. For the LCM (Study 1), the standardized effect sizes (for details of calculation, see Kwok, Luo, & West, 2010; Kwok et al., 2007; Raudenbush & Liu, 2001) of the slope factors were chosen to reflect a large effect size for the linear slope factor and a medium effect size for the second growth factor. For the LCGA (Study 2) and the GMM (Study 3 and 4) models, the intercept, linear slope, and quadratic slope means, as well as the size of the time-specific residual parameters, are all taken from Enders and Tofghi’s (2008) study to extend on their results in an easy-to-integrate manner for users. Data were generated and analyzed with the variance and covariance parameters of the quadratic slope fixed to zero. The number of repeated measures was not varied following Tofghi and Enders (2007) who showed that this factor had a limited impact on GMM class enumeration performance. As in Li and Hser’s (2011)
TABLE 1
Population Parameters for the Population Models

<table>
<thead>
<tr>
<th>Latent Curve Models (Study 1)</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept mean, (a_0)</td>
<td>0.16</td>
</tr>
<tr>
<td>Linear slope mean, (a_1)</td>
<td>0.10</td>
</tr>
<tr>
<td>Quadratic slope mean, (a_2)</td>
<td>0.05</td>
</tr>
<tr>
<td>Intercept variance, (\psi_{a_{i1}})</td>
<td>0.20</td>
</tr>
<tr>
<td>Linear slope variance, (\psi_{a_{i2}})</td>
<td>0.10</td>
</tr>
<tr>
<td>Intercept/linear slope covariance, (\psi_{a_{i11}})</td>
<td>0.05</td>
</tr>
<tr>
<td>Residual variance ((\theta_0; \theta_1; \theta_2; \theta_3; \theta_4; \theta_5))</td>
<td>0.10, 0.24; 0.69; 1.67; 3.40; 6.10; 9.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Growth Mixture Models (Studies 2–4)</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept mean, (a_0)</td>
<td>7</td>
<td>2.50</td>
<td>1</td>
</tr>
<tr>
<td>Linear slope mean, (a_1)</td>
<td>0.35</td>
<td>1.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Quadratic slope mean, (a_2)</td>
<td>– 0.03</td>
<td>– 0.07</td>
<td>0</td>
</tr>
<tr>
<td>High separation—invariance (Study 3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept variance, (\psi_{a_{i1}})</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Linear slope variance, (\psi_{a_{i2}})</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Intercept/linear slope covariance, (\psi_{a_{i11}})</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Low separation—invariance (Study 3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept variance, (\psi_{a_{i1}})</td>
<td>2.50</td>
<td>1.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Linear slope variance, (\psi_{a_{i2}})</td>
<td>0.50</td>
<td>0.30</td>
<td>0.10</td>
</tr>
<tr>
<td>Intercept/linear slope covariance, (\psi_{a_{i11}})</td>
<td>0.28</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>Low separation—noninvariance (Study 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept variance, (\psi_{a_{i1}})</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Linear slope variance, (\psi_{a_{i2}})</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Intercept/linear slope covariance, (\psi_{a_{i11}})</td>
<td>1.80</td>
<td>1.80</td>
<td>1.80</td>
</tr>
<tr>
<td>Residual variance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>Ratio 2:1</td>
<td>0.50</td>
<td>0.38</td>
<td>0.25</td>
</tr>
<tr>
<td>Ratio 5:1</td>
<td>1.25</td>
<td>0.75</td>
<td>0.25</td>
</tr>
</tbody>
</table>

study, the number of repeated measures was fixed to seven (with values ranging from 0–6 with unit increments).

**Manipulated Factors**

Because our focus is on studying the impact of misspecifications of the within-class latent variance–covariance matrix (\(\Psi_k\)) and residual structure (\(\Theta_k\)) in a series of studies in which the latent variance–covariance specification used in the population model was systematically manipulated, it appeared critical to also systematically consider variations in the within-class time-specific residual matrix used to generate these population models. Alternative specifications of the residual structure as equal, slightly unequal, and severely unequal across latent classes was thus the key manipulated factor considered in this study. However, based on prior results showing that these factors had an impact on the ability to identify the true number of latent classes in a GMM (Enders & Tofghi, 2008; Li & Hser, 2011; Lubke & Muthén, 2007; Peugh & Fan, 2012; Tofghi & Enders, 2007), three additional factors were also manipulated for the heterogeneous population models (Studies 2–4): the sample size, the mixing ratio, and the degree of class separation. For the homogenous population models (Study 1) only sample size was varied. Our key objective in manipulating these factors was mainly to assess the extent to which our conclusions generalized across these key designs conditions.

**Within-class residual matrix.** For the homogenous population model, time-specific residuals were specified as uncorrelated with one another and unequal across time periods, as is typical in LCM studies and applications based on the SEM framework (e.g., Cheong, 2011; Diallo, Morin, & Parker, 2014; Yu, 2002). In heterogeneous population models, three conditions were considered. In the first condition, time-specific residuals were held constant across classes and measurement points. In the second condition, these residuals were varied across classes but held constant across time. In this condition, the largest to smallest within-class residual was specified according to a ratio of 2:1. In the third condition, the residuals were held constant across time but specified as differing between classes according to a 5:1 ratio. These
three conditions were taken from Enders and Tofghi’s (2008) study. Although setting the time-specific residuals to be homoscedastic over time is not typical in practical applications of GMMs, this is a condition that is often used in simulation studies (e.g., Li & Hser, 2011; Tofghi & Enders, 2007) to simplify the estimation process and reduce the computational time. Our objective was to study overspecifications (estimating $\Theta_k$ when the population model corresponds to $\Theta_k = \Theta$) and underspecifications (estimating $\Theta_k = \Theta$ when the population model corresponds to $\Theta_k$) of the residuals, making it unnecessary to integrate additional complexity by allowing $\Theta_k$ to differ across time points. Finally, although $\Theta_k$ can also incorporate nonzero covariances (Heggieseth & Jewell, 2013), the typical specification of GMM and LCM that we followed here constrains these covariances to be zero given that their free estimation changes the meaning of the model.

**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 (e.g., Li & Hser, 2011; Nylund, Asparouhov, & Muthén, 2007; Tofghi & Enders, 2007; Yang, 2006). We generated data according to three different sample sizes (300, 750, and 1,500), chosen to reflect small, medium, and large sample sizes typically seen in GMM studies (e.g., Fantt & Henrich, 2010; Schaeffler, Petras, Ialongo, Poduska, & Kellam, 2003; Sterba, Prinstein, & Cox, 2007) and simulations (e.g., Li & Hser, 2011; Peugh & Fan, 2012, 2013).

**Mixing ratio.** Previous research showed that it was easier to recover the true number of classes present in the data when the relative size of the classes is similar to widely different, particularly when the global sample size is smaller (e.g., Nylund et al., 2007; Tofghi & Enders, 2007). In this study, data were thus simulated so that the size of the three latent classes forming the population model was specified either as equal (33%, 33%, and 33%) or unequal (17%, 33%, and 50%). The unequal condition was based on Enders and Tofghi’s (2008) results showing this condition to have a greater impact than an alternative condition of 50%, 33%, and 17%.

**Class separation.** The degree of separation between classes has also been found to influence class enumeration accuracy (Lubke & Muthén, 2007; Peugh & Fan, 2012; Tofghi & Enders, 2007), with classes showing less overlap being easier to identify. In this study, the degree of separation between classes was varied by manipulating the variances of the latent growth factors within each latent class. In the high separation condition the choice of the within-class latent variances for the growth factors resulted in intercept differences of (a) 3.06 standard deviation (SD) units between Class 1 and 2, 4.49 SD between Class 1 and 3, and 2.12 SD between Class 2 and 3 when $\Psi_k = \Psi$ (Study 3); and (b) 3.18 SD between Class 1 and 2, 4.90 SD between Class 1 and 3, and 1.50 SD between Class 2 and 3 when $\Psi_k$ (Study 4). In the low separation condition, selected parameter values resulted in intercept differences of (a) 1.42 SD between Class 1 and 2, 1.90 SD between Class 1 and 3, and 0.47 SD between Class 2 and 3 when $\Psi_k = \Psi$ (Study 3); and (b) 0.67 SD between Class 1 and 2, 1.70 SD between Class 1 and 3, and 0.45 SD between Class 2 and 3 when $\Psi_k$ (Study 4). These values are similar to those used by Enders and Tofghi (2008) and Tofghi and Enders (2007).

Furthermore, to ensure that these values did indeed result in high and low levels of class separation, preliminary analyses based on the retained population values were conducted on 100,000 cases with one replication. First, these analyses revealed average posterior probabilities of class membership (based on modal assignment) greater than .85. Second, one way to see how well individuals are classified into classes is to look at the entropy of the GMM. The entropy captures the fuzziness, or accuracy, of the classification. 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), and can be measured on a zero to one scale with high values indicating better classification of individuals. Given the clear relation between entropy and the degree of class separation, entropy values >.80 can be taken to reflect a high level of class separation whereas values ≤.50 can be taken to reflect a poor level of class separation. For population models corresponding to Model 3 ($\Psi_k = \Psi; \Theta_k = \Theta$) or Model 4 ($\Psi_k = \Psi; \Theta_k$), these analyses resulted in entropy values ranging from (a) .79 and .90 for the condition of high class separation, and (b) .43 and .60 for the conditions of low class separation. For population models corresponding to Model 5 ($\Psi_k; \Theta_k = \Theta$) or Model 6 ($\Psi_k; \Theta_k$), these analyses resulted in entropy values ranging from (a) .79 and .94 for the condition of high class, and (b) .43 and .66 for the conditions of low class separation. Similar values are frequently seen in simulation studies (e.g., Asparouhov & Muthen, 2014; Bakk, Oberski, & Vermunt, 2014; Vermunt, 2010).

**Data Generation, Verification, and Evaluation**

Three distinct sample size conditions were considered for the homogenous (LCM) population model used in Study 1. Because the class separation condition was manipulated in part through the manipulation of the growth factors variance, which was fixed to be 0 in the population model ($\Psi_k = 0$) used in Study 2, class separation was not manipulated in this study. This resulted in a total of 18 design cells for Study 2 (3 residuals × 3 sample sizes × 2 mixing ratio). The population models used in Studies 3 and 4 were associated with a total of 36 design cells for each study (3 residuals ×
3 sample sizes × 2 mixing ratio × 2 class separations). A total of 1,000 replications were generated within each of these 93 design cells, for a total of 93,000 replications (i.e., data sets to analyze). All models were simulated and estimated using Mplus Monte Carlo features (Muthén & Muthén, 1998–2014).

Each data set was analyzed according to six different specifications: (a) the default LCA approach (Model 1: \( \Psi_k = 0 \) and \( \Theta_k = \Theta \)); (b) a LCGA approach with class-specific residuals (Model 2: \( \Psi_k = 0 \) and \( \Theta_k \)); (c) the Mplus default parameterization (Model 3: \( \Psi_k = \Psi \) and \( \Theta_k = \Theta \)); (d) a Mplus default parameterization with class-specific residuals (Model 4: \( \Psi_k = \Psi \) and \( \Theta_k \)); (e) a GMM with class-specific latent variance–covariance matrices but class-invariant residuals (Model 5: \( \Psi_k \) and \( \Theta_k = \Theta \)); and (f) a GMM with class-specific latent variance–covariance matrices and residuals (Model 6: \( \Psi_k \) and \( \Theta_k \)). For each parameterization, models including one to four latent classes were estimated. All models including two classes or more were estimated using 400 sets of random starting values, 40 iterations for each of these sets, and the 20 solutions with the highest likelihood values were retained for final stage optimization. Given the number of design conditions and replications (93,000), the number of models estimated within each condition (six models, each with one to four classes, resulting in \( 6 \times 4 \times 93,000 \) analyses), and the computer intensiveness of GMM, this study was made possible through Compute Canada high-performance computing facilities (https://compuccanada.ca/).

Moreover, to demonstrate that the data generated from the population models could reasonably be assumed to be samples from that population model, the average probability of belonging to each class as well as the accuracy of parameter estimates were investigated using a single data set of 100,000 cases (one replication) for all models across conditions using correctly specified models. The accuracy of parameter estimates was assessed through standardized bias [i.e., standardized bias = \( 100^\ast(\text{average estimate--parameter})/SD \), where SD is the standard deviation of the estimates]. Following Collins, Schafer, and Kam (2001), standardized bias exceeding 40% to 50% in absolute value is considered influential. Standardized bias estimates (using a single simulated data set of 100,000 cases with one replication for all models across conditions) were all in the fully acceptable range, with the exception of a few of the variance parameters in models with a freely estimated latent variance–covariance matrix (Model 5 and Model 6) where the maximum standardized bias reached 59% in a few extreme situations. Fixing the model parameters to their population values stabilized the models but resulted in significant standard deviation bias for the few fixed effect parameters. Furthermore, additional sensitivity analyses showed that both strategies (fixing the parameters to their population values or freely estimating them using an exactly specified model) were able to correctly identify the population generating model. Given that our main interest in this study is to adequately identify the population generating GMM rather than focusing in individual parameter estimates, we elected to freely estimate all parameters in all models for consistency with practical applications of GMMs. Finally, as seen from the preliminary analyses presented earlier, the class separation resulted in a range of entropy values that properly define high and low class separation conditions.

To guide the selection of the optimal number of latent classes, 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 (SACAIC; 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; Vuong, 1989), and the bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000). For the IC, the model with the lowest value among the four competing models (i.e., including one to four latent classes within each specification) indicates the best fitting model. For the LRT, \( p \) 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 (from one to four) was followed. As soon as a significant \( p \) value for a model with \( k \) classes was followed by a nonsignificant \( p \) value for a model with \( k + 1 \) classes, the model with \( k \) classes was retained as the best fitting model. Our decision to focus on these specific indicators, relative to other possible indicators, stems from a desire to remain well aligned with current applied research practice, which tends to focus on these easily available indicators. However, the class enumeration performance of the entropy, normalized entropy criterion (NEC), classification likelihood criterion (CLC), integrated completed likelihood criterion (ICL–BIC), and the sample-size-adjusted ICL–BIC (ICL–SBIC) are provided at the end of the supplemental materials accompanying this article and available upon request from the corresponding author.

In this study, the outcome variable is the performance of the IC and LRTs in identifying the correct number of latent classes present in the data. We thus examined the proportion of replications in which each of the indicators correctly identified the three-class solution for heterogeneous population models used in Studies 2 to 4, and the one-class solution for homogenous population models used in Study 1. Recent simulation studies indicate that the BIC, SBIC, CAIC, and BLRT are particularly effective in choosing the model that best recovers the true parameters in mixture
models (Henson, Reise, & Kim, 2007; McLachlan & Peel, 2000; Nylund et al., 2007; Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofghi & Enders, 2007; Tolvanen, 2007; Yang, 2006). Furthermore, when the indicators fail to retain the optimal model, the AIC, SBIC, and BLRT tend to overestimate the number of classes, whereas the BIC, LMR, and CAIC tend to underestimate it. As shown in these previous simulation studies, the AIC generally tends to be the least accurate of these indicators, and results based on the LMR and ALMR tend to be indistinguishable from one another. This was also observed here. Thus, although we report detailed results for all of these indicators, we do not discuss the specific results associated with the AIC, and do not differentiate between the LMR and ALMR. Similarly, our results showed that the main variations observed in terms of class enumeration accuracy were related to the nature of the estimated model (Model 1 to Model 6) and to the residual structure condition. For this reason, we focus our presentation on the effects of these specific conditions and simply note, when relevant, the global effects of the other design conditions. Interested readers can find these additional results in a set of supplements available on request from the corresponding author.

RESULTS

Study 1: Homogenous Population Model

Table 2 presents the class enumeration performance of the indicators as a function of model specifications and sample sizes for the homogenous (LCM) population model used in Study 1. Across sample sizes, a marked and constant tendency to overextract spurious latent classes was observed for all indicators when the data were analyzed with LCGA (Model 1: \( \Psi_k = 0; \Theta_k = \Theta \)) or LCGA with class-specific residuals (Model 2: \( \Psi_k = 0; \Theta_k \)). In fact, across sample sizes, the accuracy of all indicators was less than 10%, with all indicators identifying the three-class or four-class solutions most of the time. However, when latent variability was allowed within the extracted latent classes (Model 3 to Model 6), the accuracy of all indicators drastically improved, and kept on increasing as a function of sample size. More precisely, for Models 3 to 6, the accuracy of the IC and LRT proved to be higher than 90% at \( N = 1,500 \), and to vary between 65% and 100% across model specifications and sample size conditions. These indicators have been previously shown to present a sample size dependency (e.g., Marsh, Lüdtke, Trautwein, & Morin, 2009). For Model 3 to Model 6, this tendency was evident, but remained minimal, when the sample size increased from \( N = 300 \) (with accuracy rates over 90%) to \( N = 750 \) (with accuracy rates from 65%–90%), but disappeared at \( N = 1,500 \) (with accuracy rates over 90%).

Study 1: Summary and Discussion

The results found for data generated according to a homogenous (one-class) population model are fully in line with those obtained by Bauer and Curran (2003) when the data were analyzed using a LCGA parameterization (Model 1, with \( \Psi_k = 0 \) and \( \Theta_k = \Theta \), or Model 2 with \( \Psi_k = 0 \) and \( \Theta_k \)). More precisely, these results show that relying on a LCGA parameterization (i.e., \( \Psi_k = 0 \)) tends to erroneously suggest the presence of population heterogeneity (overextract latent classes) when in fact there is none. This overextraction tendency appears to be quite severe, and was observed over 90% of the time across indicators and conditions. In contrast, and in support of Peugh and Fan’s (2012) results, the results reported here showed that models assuming within-class variance and covariance on the latent growth factors (Model 3–Model 6: \( \Psi_k = \Psi \) or \( \Psi_k \)) correctly identified the correct one-class solution more than 70% of the time across indicators and conditions, with rates of accuracy that reached 90% to 100% when the sample size reached \( N = 1,500 \). In contrast, allowing for between-class differences in the estimation of the time-specific residuals or in the variance–covariance of the latent growth factors did not seem to affect the performance of the model in recovering the true population model homogenous structure.

Study 2: Heterogeneous Population Model With no Within-Class Variability

Table 3 presents the class enumeration performance of the indicators as a function of model specification and the population residual ratio for the heterogeneous population model including no within-class variability (\( \Psi_k = 0 \)) used in Study 2. Specific results associated with all design conditions are reported in Tables S1 to S6 of the supplements (available from the corresponding author).

Exactly Specified Models

When the estimated model was exactly specified (i.e., when Model 1 [\( \Psi_k = 0; \Theta_k = \Theta \)] is estimated for population models in which the residuals are equal across classes, or when Model 2 [\( \Psi_k = 0; \Theta_k \)] is estimated for population models in which the residuals are unequal across classes), the accuracy of the CAIC, SCAIC, BIC, and the SBIC was close to 100% across conditions. The LRTs were less accurate than the IC across conditions, with accuracy rates in the range of 60% to 70%. Interestingly, for models based on equal residuals, the BLRT (72%) was slightly more accurate than the LMR (62.3%), whereas the opposite was observed for models based on unequal residuals (69.2% for LMR vs. 55.5% for BLRT). When they failed to correctly identify the three-class solution, the LRT showed a clear tendency to overextract latent classes. For these exactly specified models, the IC and the LRT proved
## TABLE 2
Class Enumeration Performance: Homogenous Population Models (Study 1)

<table>
<thead>
<tr>
<th>No. of classes</th>
<th>Model 1: ( \Psi_k = 0 ) and ( \Theta_k = \Theta )</th>
<th>Model 2: ( \Psi_k = 0 ) and ( \Theta_k )</th>
<th>Model 3: ( \Psi_k = \Psi ) and ( \Theta_k = \Theta )</th>
<th>Model 4: ( \Psi_k = \Psi ) and ( \Theta_k )</th>
<th>Model 5: ( \Psi_k ) and ( \Theta_k = \Theta )</th>
<th>Model 6: ( \Psi_k ) and ( \Theta_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 300 )</td>
<td>AIC 0 0 0 100 0 0 0 100 99.2 0.8 0 0 93.5 4.8 1.5 0.2 93.5 4.8 1.5 0.2 92.8 5 1.4 0.8</td>
<td>CAIC 0 0 1 100 0 99 1 0 0 99.5 0.5 0 0 99.5 0.5 0 0 99.7 0.3 0 0</td>
<td>SCAIC 0 0 0 100 0 99.3 0.7 0 0 94.4 5 0.6 0 94.4 5 0.6 0 94.5 5.1 0.4 0</td>
<td>BIC 0 0 100 0 0 0 0 0 99.5 0.5 0 0 98.6 1.4 0 0 98.6 1.4 0 0 99.1 0.9 0 0</td>
<td>SBIC 0 0 0 100 0 99.3 0.6 0 0 93.7 5.5 0.6 0 93.7 5.5 0.6 0 93.5 5.1 0.4 0</td>
<td>LMR 8 40.4 31.7 19.9 5.9 29.3 31.2 33.6 99.5 0.5 0 0 96.3 3.5 0.1 0.1 96.3 3.5 0.1 0.1 98.7 1.3 0.3 0</td>
</tr>
<tr>
<td>( N = 750 )</td>
<td>AIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>CAIC 0 0 0.1 99.9 3.2 96.8 0 70.2 17 6.2 6.6 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 8.7 5.5</td>
<td>SCAIC 0 0 0 100 0 82.5 17.5 70.2 17.2 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>BIC 0 0 0 100 0 1.1 98.3 0 70.2 17.2 6.6 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 8.2 6.0</td>
<td>SBIC 0 0 0 100 0 49.1 50.9 70.2 17.2 5.9 6.7 88.9 9.5 1.3 0.3 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>LMR 11.3 45.7 43 20.6 70 9.4 70 20 7 3 88.9 10.0 1.0 0.1 88.9 10 1 0.1 67 21.8 9.4 1.8</td>
</tr>
<tr>
<td>( N = 1,500 )</td>
<td>AIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>CAIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>SCAIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>BIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>SBIC 0 0 0 100 0 0 0 0 1.5 98.5 70 17.4 5.9 6.7 88.9 9.5 1.4 0.2 88.9 9.6 1.3 0.2 65.8 20 7.8 6.4</td>
<td>LMR 11.9 47.2 41 20.8 70.1 9.1 70.2 19.6 7.3 2.9 88.9 10.0 1.0 0.1 88.9 10 1 0.1 67 21.9 9.3 1.8</td>
</tr>
</tbody>
</table>

Note. \( \Psi_k \) = Latent variance-covariance matrix; \( \Theta_k \) = residual variance-covariance matrix; 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 likelihood ratio test; BLRT = bootstrapped likelihood ratio test.
### TABLE 3

Class Enumeration Performance: Heterogeneous Population Models With No Within-Class Variability (Study 2)

| Θₖ = Θ | AIC | CAIC | SCAIC | BIC | SBIC | LMR | BLRT | Θₖ 2:1 | AIC | CAIC | SCAIC | BIC | SBIC | LMR | BLRT | Θₖ 5:1 | AIC | CAIC | SCAIC | BIC | SBIC | LMR | BLRT |
| 1 | 0 | 94.2 | 5.8 | 0 | 51.5 | 48.5 | 0 | 30 | 67.5 | 2.5 | 47.5 | 33.1 | 16.6 | 2.8 | 0 | 55.3 | 44.7 | 0 | 52.9 | 47.1 | 0 | 100 | 0 | 0 | 100 | 0 |
| 2 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 30.3 | 68.9 | 0.8 | 47.5 | 35.1 | 17.4 | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 |
| 3 | 0 | 100 | 0 | 0 | 99.8 | 0.2 | 0 | 30.3 | 68.9 | 0.8 | 47.7 | 35 | 17.3 | 0 | 0 | 99.5 | 0.5 | 0 | 99.6 | 0.4 | 0 | 100 | 0 | 0 | 100 | 0 |
| 4 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 30.3 | 68.9 | 0.8 | 47.7 | 35 | 17.3 | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 |
| 1 | 0 | 99.6 | 0.4 | 0 | 95.5 | 4.5 | 0 | 30.3 | 68.9 | 0.8 | 47.5 | 35 | 17.4 | 0 | 0 | 93.8 | 6.2 | 0 | 94.1 | 5.9 | 0 | 100 | 0 | 0 | 100 | 0 |
| 2 | 0 | 62.3 | 37.7 | 0 | 69.2 | 30.8 | 0 | 27.6 | 49.2 | 23.2 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 67.2 | 32.8 | 0 | 60.7 | 39.3 | 0 | 100 | 0 | 0 | 100 | 0 |
| 3 | 0 | 63.2 | 36.8 | 0 | 69.4 | 30.6 | 0 | 27.7 | 49.5 | 22.8 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 68.1 | 31.9 | 0 | 61.1 | 38.9 | 0 | 100 | 0 | 0 | 100 | 0 |
| 1 | 0 | 99.6 | 0.4 | 0 | 95.5 | 4.5 | 0 | 30.3 | 68.9 | 0.8 | 47.5 | 35 | 17.4 | 0 | 0 | 93.8 | 6.2 | 0 | 94.1 | 5.9 | 0 | 100 | 0 | 0 | 100 | 0 |
| 2 | 0 | 62.3 | 37.7 | 0 | 69.2 | 30.8 | 0 | 27.6 | 49.2 | 23.2 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 67.2 | 32.8 | 0 | 60.7 | 39.3 | 0 | 100 | 0 | 0 | 100 | 0 |
| 3 | 0 | 63.2 | 36.8 | 0 | 69.4 | 30.6 | 0 | 27.7 | 49.5 | 22.8 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 68.1 | 31.9 | 0 | 61.1 | 38.9 | 0 | 100 | 0 | 0 | 100 | 0 |
| 1 | 0 | 99.6 | 0.4 | 0 | 95.5 | 4.5 | 0 | 30.3 | 68.9 | 0.8 | 47.5 | 35 | 17.4 | 0 | 0 | 93.8 | 6.2 | 0 | 94.1 | 5.9 | 0 | 100 | 0 | 0 | 100 | 0 |
| 2 | 0 | 62.3 | 37.7 | 0 | 69.2 | 30.8 | 0 | 27.6 | 49.2 | 23.2 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 67.2 | 32.8 | 0 | 60.7 | 39.3 | 0 | 100 | 0 | 0 | 100 | 0 |
| 3 | 0 | 63.2 | 36.8 | 0 | 69.4 | 30.6 | 0 | 27.7 | 49.5 | 22.8 | 47.5 | 34.6 | 14.8 | 3.1 | 0 | 68.1 | 31.9 | 0 | 61.1 | 38.9 | 0 | 100 | 0 | 0 | 100 | 0 |

**Note.** Model 1 to Model 6 = type of estimated model; Ψₖ = latent variance-covariance matrix; Θₖ = residual variance-covariance matrix; 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 likelihood ratio test; BLRT = bootstrapped likelihood ratio test.
to be quite stable across sample sizes and mixing ratio conditions, with the exception of the SBIC, which became slightly more accurate as a function of sample size (from 82.1%–100%) in the unequal residual condition.

Models With Correctly Specified Variance–Covariance

When the latent variance–covariance matrix of the estimated model was correctly specified ($\Psi_k = \Psi$), results showed important differences based on whether the residuals were underspecified (Model 1 [$\Psi_k = 0; \Theta_k = \Theta$] when $\Theta_k$) or overspecified (Model 2 [$\Psi_k = 0; \Theta_k$] when $\Theta_k = \Theta$).

Underspecified residuals. For models based on an underspecified residual structure (estimating $\Theta_k = \Theta$ when $\Theta_k$), the accuracy of the IC and BLRT decreased (leading to an even more pronounced tendency toward overextraction) compared to that observed with exactly specified models, and the accuracy of the LMR showed an improvement (76.4% and 80.5% for residual ratios of 2:1 and 5:1, respectively). More precisely, when the residuals were specified as unequal by a ratio of 2:1, the accuracy of the CAIC decreased to 90.6%, the SCAIC to 68.9%, the BIC to 54.3%, and the BLRT to 18.8%. The situation was even worse when the residuals were specified as unequal by a ratio of 5:1, a situation in which the IC and BLRT tended to decrease by approximately 10% as sample size increased (by approximately 23% for residual ratios of 2:1 and 5:1, respectively). The accuracy of the IC and BLRT remained similar to that observed in the previous condition. Across these conditions, these rates of accuracy tended to decrease by approximately 10% as sample size increased from $N = 300$ to $N = 1,500$, whereas that of the LRT remained essentially unchanged.

Overspecified residuals. Models based on an overspecified residual structure (estimating $\Theta_k$ when $\Theta_k$) had no effect on the accuracy of the CAIC, SCAIC, BIC, SBIC, and LMR when compared to that observed with exactly specified models, but resulted in a slight increase in the accuracy of the BLRT (55.5%). These results showed few variations as a function of sample size and mixing ratio.

Models With a Moderately Overspecified Variance–Covariance Matrix

When the latent variance–covariance matrix of the estimated model was moderately overspecified according to a class-invariant parameterization ($\Psi_k = \Psi$), results differed based on whether the residuals were freely estimated in all classes (using Model 4: $\Psi_k = \Psi; \Theta_k$), or constrained to be equal across classes (using Model 3: $\Psi_k = \Psi; \Theta_k = \Theta$), and according to whether these models resulted in the correct specification of residuals, underspecification (estimating Model 3 when $\Theta_k$), or overspecification (estimating Model 4 when $\Theta_k = \Theta$).

Models with correctly or underspecified equally estimated residuals (Model 3). For models estimated according to Model 3 ($\Psi_k = \Psi; \Theta_k = \Theta$), the class enumeration performance of the IC and LRT showed an important decrease when compared to the exactly specific model. In particular, for population models based on class-invariant residuals, Model 3 (i.e., using a correct specification of the residuals) resulted in a decrease in the performance to a level of 68.9% for the IC, and of 49.2% for the LRT. These accuracy rates remained generally similar for the BIC, CAIC, and LMR when the population model relied on an unequal residual structure based on a 2:1 ratio, although the accuracy of the SBIC and SCAIC slightly decreased to 58.4%, and that of the BLRT to 25.9%. When the population models relied on a more pronounced unequal residual structure (5:1), the accuracy of the IC reached a level closer to 56%, whereas the performance of the LRT remained similar to that observed in the previous condition. Across these conditions, these rates of accuracy tended to decrease by approximately 10% as sample size increased from $N = 300$ to $N = 1,500$, whereas that of the LRT remained essentially unchanged. An interesting variation was observed regarding the effects of the mixing ratio condition. Thus, when the residual structure was correctly specified (Model 3 when $\Theta_k = \Theta$), the accuracy of all indicators was higher when the mixing ratio was equal (85% for the IC and 60% for the LRT) than unequal (50% for the IC and 37% for the LRT). In contrast, when the residual structure was underspecified (Model 3 when $\Theta_k$), although some minor variations were noted as a function of the residual structure (2:1 vs. 5:1), mixing ratio had only limited effects on the indicators’ accuracy.

Models with correctly specified freely estimated residuals (Model 4). For population models based on class-varying residuals, the estimation of models where these residuals were correctly specified (using Model 4: $\Psi_k = \Psi; \Theta_k$) resulted in a clear improvement in the accuracy of the IC and LRT when compared with models estimated according to Model 3 ($\Psi_k = \Psi; \Theta_k = \Theta$), and this improvement was a direct function of the size of residual differences in the population model. Thus, when the residuals differed based on a 2:1 ratio, rates of accuracy were closer to 85% for the IC, 57.3% for the LMR, and 50.7% for the BLRT. When the residuals differed based on a 5:1 ratio, these figures increased to 97.7% for the IC, 67% for the LMR, and 59.5% for the BLRT. Across these conditions, a slight decrease in the accuracy of the IC (but not the LRT) was also noted as a function of sample size, albeit far less pronounced than for the previous model (i.e., Model 3), with decreases in accuracy of about 5% to 10% as sample
size increased from 300 to 1,500. Finally, when the residuals were specified based on a 2:1 ratio, the accuracy of the indicators proved to be higher (by approximately 5% to 10%) in the equal mixing ratio condition, whereas mixing ratio conditions had a negligible impact on accuracy rates when the residuals differed by a 5:1 ratio.

**Models with overspecified freely estimated residuals (Model 4).** Models based on a moderately overspecified latent variance–covariance matrix \((\Psi_k = \Psi)\), relying on models in which the residuals were also overspecified \((\Theta_k = \Theta)\), resulted in the lowest rates of accuracy so far. All indicators presented a marked tendency for underextraction, selecting the one-class model close to 50% of the time across indicators, and the two-class model close to 35% of the time. These results showed few variations as a function of sample size or mixing ratio conditions.

**Models with a severely overspecified variance–covariance matrix.** With a severely overspecified \((\Psi_k)\) variance–covariance matrix, a different pattern of results emerged depending on whether the residuals were underspecified (using Model 5 \((\Psi_k; \Theta_k = \Theta)\)) when \(\Theta_k\), versus correctly (using Model 6 \((\Psi_k; \Theta_k)\) when \(\Theta_k\)) or overspecified (using Model 6 when \(\Theta_k = \Theta\)).

**Models with correctly specified equally estimated residuals (Model 5).** Compared to the exactly specified model, models based on a severely overspecified latent variance–covariance matrix and correctly specified residuals (using Model 5 \((\Psi_k; \Theta_k = \Theta)\) had no effect on the accuracy of the IC and the LRT. The CAIC, SCAIC, BIC, and the SBIC had comparable accuracy rates close to 100%, whereas the LRT had an accuracy rate closer to 70% (with a rate of overextraction of 30%). Moreover, similar and stable accuracy was seen for the IC and the LRT across sample size and mixing ratio conditions.

**Models with underspecified equally estimated residuals (Model 5).** Severely overspecifying the variance–covariance along with underspecifying the residuals (using Model 5: \(\Psi_k; \Theta_k = \Theta\)) resulted in a clear deterioration in the accuracy of the IC (more severe for the SCAIC and the SBIC) and the LRT. When compared to the exactly specified model, the accuracy of the IC decreased as a direct function of the degree of underspecification of the residuals. However, the LRT’s accuracy remained stable across residual conditions. More precisely, when the residuals differed by a 2:1 ratio, no deterioration was seen in the accuracy of the CAIC and the BIC (100%), a small deterioration in accuracy was noted for the SCAIC (91.9%), a moderate deterioration was observed for the SBIC (71.4%), and a large deterioration was noted for the LRT (43.6% for the LMR and 30.5% for the BLRT). In contrast, when residuals differed by 5:1, a systematic overextraction tendency was observed for the IC (73.6% for the CAIC, 90.3% for the SCAIC, 78% for the BIC, and 97.1% for the SBIC). Overall, these results only showed minimal variations as a function of sample size and mixing ratio conditions.

**Models with correctly or overspecified freely estimated residuals (Model 6).** For models in which the latent variance–covariance matrix was severely overspecified \((\Psi_k)\), freely estimating the residuals \((\Theta_k)\) resulted in similar results irrespective of whether this free estimation of residuals corresponded to a correct specification or to an overspecification. More precisely, these models generally resulted in high rates of accuracy, comparable to those observed for the exactly specified models: close to 100% for the CAIC, SCAIC, BIC, and SBIC; 55% for the BLRT; and 60% for the LMR. Interestingly, the accuracy of all indicators remained generally stable across sample size, residual ratio, and mixing ratio conditions, apart from the SBIC, which showed a slight increase in accuracy across sample size conditions (80.8%–99.9%).

**Study 2: Summary and Discussion**

When we relied on a population model corresponding to restrictive LCGA specifications \((\Psi_k = 0)\), the results showed important variations as a function of the estimated model. In particular, exactly specified models (i.e., relying on an LCGA with an exactly specified residual structure) resulted in rates of accuracy close to 100% for the CAIC, SCAIC, BIC, and SBIC, and closer to 60% to 70% for the LRT across design conditions. However, relying on an underspecified residual structure in conjunction with constraining the latent variance–covariance matrix to be zero resulted in a decrease in the accuracy of all indicators (except for the LMR), which tended to overextract latent classes, whereas overspecifying the residual structure resulted in rates of accuracy equivalent to those observed for exactly specified models. Similar results were observed for models relying on a moderately overspecified class-invariant latent variance–covariance matrix \((\Psi_k = \Psi)\), for which relying on a correct specification of the residual structure resulted in rates of accuracy approaching (but still lower) that observed in the exactly specified models, whereas relying on an underspecified or overspecified residual structure resulted in a drastic decrease in rates of accuracy. However, for models relying on a severely overspecified latent variance–covariance matrix \((\Psi_k)\), models relying on correctly specified, or overspecified, residual structures resulted in rates of accuracy comparable to that observed in the exactly specified models (except for the BLRT, which did not perform as well), whereas relying on an underspecified residual structure resulted in a decrease in the accuracy of all indicators that was a direct function of the degree of underspecification (worse in the 5:1 than in the 2:1
condition). Across most of these situations, indicators tended to overextract latent classes when they failed to correctly identify the three-class solution, with the sole exception of models involving a moderately overspecified latent variance–covariance matrix (Ψ_k = Ψ) and an over-specified residual structure for which indicators tended to underextract latent classes. A particularly important result is that although the design conditions (sample size, mixing ratio, and class separation) also had an impact, this impact was much smaller than the effects of model misspecifications and more pronounced for models relying on an underspecified residual structure and moderately over-specified (class-invariant) latent variance–covariance matrix.

Study 3: Heterogeneous Population Model With Class Invariant Within-Class Variability

Table 4 presents the class enumeration performance of the various indicators as a function of model specification and the population residual ratio for the heterogeneous population model including class invariant within-class variability (Ψ_k = Ψ) used in Study 3. More specific results associated with all design conditions are reported in Tables S7 to S18 of the supplements (available from the corresponding author).

Exactly Specified Models

When the true model was estimated (i.e., Model 3 [Ψ_k = Ψ; Θ_k = Θ] when Θ_k = Θ, or Model 4 [Ψ_k = Ψ; Θ_k] when Θ_k), the accuracy of the IC and the LRT was higher in conditions where the residuals were freely estimated (particularly when the residuals differed by a ratio of 5:1) than when they were constrained to be equal across classes. More precisely, for exactly specified models with class-invariant residuals, the SCAIC and the SBIC had a higher rate of accuracy (63.5% and 64.4%) than the remaining indicators (with approximately 50.0%). These rates of accuracy substantially improved for the exactly specified models with freely estimated residuals, reaching 83.4% for the CAIC, 93.0% for the SCAIC, 92.0% for the BIC, 82.8% for the SBIC, 79.0% for the LMR, and 64.3% the BLRT when residuals differed by a ratio of 5:1. Furthermore, when the indicators failed to pick the correct three-class solution, they tended to underextract latent classes for models based on class-invariant residuals, but to overextract latent classes for models based on freely estimated residuals. Across all of these exactly specified conditions, the accuracy of the IC and LRT was positively related to sample size and was higher in the equal (vs. unequal) mixing ratio condition. For models with class-invariant residuals, the IC and the LRT had a lower accuracy rates in the high class separation condition with rates of underextraction reaching 50.0% for the one-

class solution. For models with freely estimated residuals (2:1), the accuracy of the CAIC was stable across class separation conditions but the SCAIC, BIC, SBIC, LMR, and BLRT had lower rates of accuracy in the high class separation condition than in the low class separation condition. However, for models with freely estimated residuals (5:1), the IC and the LRT had higher rates of accuracy in the high class separation condition.

Models With an Underspecified Variance–Covariance Matrix

When the variance–covariance matrix of the estimated model was underspecified (Ψ_k = 0), the results showed few differences based on whether the residuals were correctly specified, underspecified (estimating Model 1 [Ψ_k = 0; Θ_k = Θ] when Θ_k) or overspecified (estimating Model 2 [Ψ_k = 0; Θ_k] when Θ_k = Θ). More precisely, estimating underspecified variance–covariance matrices (Ψ_k = 0 when Ψ_k = Ψ) resulted in an important deterioration of the accuracy of the IC and BLRT, with a tendency for overextraction approaching 100%. However, when the residuals were underspecified (using Θ_k = Θ when Θ_k), the LMR only tended to overextract latent classes 26.3% of the time with ratio 2:1 and 22.6% with ratio 5:1, versus 55.2% of the time when the residuals where overspecified (using Θ_k when Θ_k = Θ). A similar pattern of results was observable across sample sizes, mixing ratio, and class separation conditions.

Models With Correctly Specified Variance–Covariance

With a correctly specified variance–covariance matrix (Ψ_k = Ψ), overspecified residuals (Model 4 [Ψ_k = Ψ; Θ_k] when Θ_k = Θ) resulted in an improvement in the accuracy of the IC and the LRT when compared to the exactly specified model. In contrast, the effects of underspecified residuals (Model 3 [Ψ_k = Ψ; Θ_k = Θ] when Θ_k) were related to the ratio of the residuals (2:1 vs. 5:1).

Models with underspecified equally estimated residuals (Model 3). Compared to the exactly specified models, models based on a correct specification of the latent variance–covariance matrix (Ψ_k = Ψ) with an underspecified residual structure (Θ_k = Θ) resulted in an increase in the accuracy of the IC and a small (for LMR) to important (for BLRT) decrease in the accuracy of the LRT under a residual ratio condition of 2:1. Specifically, for a residual ratio condition of 2:1, accuracy rates were 95.5% for the CAIC, 90.3% for the SCAIC, 96.1% for the BIC, 83.8% for the SBIC, 70.2% for the LMR, and 56.9% for the BLRT. However, these rates decreased under a residual ratio of 5:1 to 64.7% for the CAIC, 50.7% for the SCAIC, 62.1% for the BIC, 43.1% for the SBIC, 47.2% for the LMR, and 25.5% for the BLRT. All indicators tended to overextraction latent classes when they failed to pick the correct three-class
Table 4: Class Enumeration Performance: Heterogeneous Population Models With Equal Within-Class Variability (Study 3)

<table>
<thead>
<tr>
<th>No. of classes</th>
<th>Model 1: ( \Psi_k = 0 ) and ( \Theta_k = \Theta )</th>
<th>Model 2: ( \Psi_k = 0 ) and ( \Theta_k = \Theta )</th>
<th>Model 3: ( \Psi_k = \Psi ) and ( \Theta_k = \Theta )</th>
<th>Model 4: ( \Psi_k = \Psi ) and ( \Theta_k = \Theta )</th>
<th>Model 5: ( \Psi_k ) and ( \Theta_k = \Theta )</th>
<th>Model 6: ( \Psi_k ) and ( \Theta_k = \Theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>CAIC</td>
<td>SCAIC</td>
<td>BIC</td>
<td>SBIC</td>
<td>LMR</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. Model 1 to Model 6 = type of estimated model; \( \Psi_k \) = latent variance-covariance matrix; \( \Theta_k \) = residual variance-covariance matrix; 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 likelihood ratio test; BLRT = bootstrapped likelihood ratio test.
solution. Accuracy rates showed a slight (CAIC and BIC) to important (SCAIC, SBIC, and LRT) decrease as a function of sample size. Furthermore, with a residual ratio of 2:1, the IC and LRT had comparable rates of accuracy across mixing ratio conditions, whereas the CAIC and BIC had lower accuracy in the unequal mixing ratio condition when the residual ratio was of 5:1. Finally, the accuracy of all indicators seemed to be higher in the high (vs. low) class separation condition.

Models with overspecified freely estimated residuals (Model 4). When the variance–covariance matrix of the estimated model was correctly specified (Ψ_k = Ψ) but the residual structure was overspecified (Θ_k), the results fell in between those obtained for an underspecified residual structure with a residual ratio of 2:1 versus 5:1. More precisely, rates of accuracy were 52.2% for the CAIC, 68.9% for the SCAIC, 57.1% for the BIC, 73.3% for the SBIC, 65.7% for the LMR, and 49.6% for the BLRT. However, when compared to models based on an underspecified residual structure, models based on an overspecified residual structure resulted in the underextraction of latent classes when the indicators failed to correctly retain the three-class solution. The accuracy of the IC and the LRT appeared to increase across sample size conditions, reaching a more satisfactory level when the sample size was 1,500. Similarly, the IC and the LRT had higher rates of accuracy in the equal (vs. unequal) mixing ratio condition, as well as high (vs. low) class separation conditions.

Models with an overspecified variance–covariance matrix. When the variance–covariance matrix was overspecified and the residual structure was constrained to be invariant (using Model 5: Ψ_k; Θ_k = Θ), the results differed depending on whether the estimated model resulted in a correct specification (when Θ_k = Θ) or an underspecification (when Θ_k) of the residual structure. Compared to these models, models relying on the free estimation of the residual structure across classes (using Model 6: Ψ_k; Θ_k) resulted in generally greater rates of accuracy, irrespective of whether these models resulted in a correct specification (when Θ_k) or overspecification (when Θ_k = Θ) of the residual structure.

Models with correctly specified equally estimated residuals (Model 5). Compared to the exactly specified model, models based on an overspecified latent variance–covariance matrix (Ψ_k) and underspecified residuals (Θ_k = Θ) showed a different pattern of results depending on the residual ratio (2:1 vs. 5:1). For models based on a residual ratio of 2:1, rates of accuracy for the CAIC, SCAIC, and BIC improved when compared to those observed for the exactly specified model, attaining close to 95%. However, rates of accuracy decreased for the SBIC (77.9%), the LMR (63.1%), and the BLRT (49.5%). In this situation, accuracy rates were roughly constant across sample size conditions for the SBIC, slightly improved as a function of sample size for the CAIC, SCAIC, and the BIC, and decreased for the SCAIC, LMR, and BLRT. However, all indicators had comparable rates of accuracy across mixing ratio conditions. In contrast, for a residual ratio of 5:1, the accuracy of all indicators deteriorated when compared to that observed for the exactly specified models, to 57.9% for the CAIC, 54.3% for the BIC, 36.9% for the SCAIC, 26.1% for the LMR, 25.3% for the SBIC, and 16.8% for the BLRT, with all indicators presenting a tendency for overextraction. In this situation, the accuracy of all IC and LRTs deteriorated as a function of sample size, and was lower in the equal (vs. unequal) mixing ratio condition. Finally, across both residual ratios (2:1 and 5:1), the accuracy of the SBIC and BLRT was lower in the high (vs. low) class separation condition, that of the CAIC and BIC was higher in the high (vs. low) class separation condition, and that of the SCAIC was constant across these two conditions when the residual ratio was 2:1, but lower in the high (vs. low) class separation condition when the residual ratio was 5:1 (LMR showed the opposite pattern).

Models with correctly or overspecified freely estimated residuals (Model 6). Compared to the exactly specified model, models based on an overspecified latent variance–covariance matrix (Ψ_k) and freely estimated residuals (Θ_k) were associated with a small to large increase in the accuracy of the IC and the LRT, irrespective of whether this free estimation resulted in a correct specification or an overspecification of the residual structure. Thus, across conditions, accuracy rates reached 94.0% and more for the IC, 82.7% for the LMR, and 73.0% for the BLRT. However, with overspecified residuals, a tendency to underextract was observed for the IC and LMR when they failed to select the

Models with underspecified equally estimated residuals (Model 5). Compared to the exactly specified model, models based on an underspecified latent variance–covariance matrix (Ψ_k) and overspecified residuals (Θ_k = Θ) showed a different pattern of results depending on the residual ratio (2:1 vs. 5:1). For models based on a residual ratio of 2:1, rates of accuracy for the CAIC, SCAIC, and BIC improved when compared to those observed for the exactly specified model, attaining close to 95%. However, rates of accuracy decreased for the SBIC (77.9%), the LMR (63.1%), and the BLRT (49.5%). In this situation, accuracy rates were roughly constant across sample size conditions for the SBIC, slightly improved as a function of sample size for the CAIC, SCAIC, and the BIC, and decreased for the SCAIC, LMR, and BLRT. However, all indicators had comparable rates of accuracy across mixing ratio conditions. In contrast, for a residual ratio of 5:1, the accuracy of all indicators deteriorated when compared to that observed for the exactly specified models, to 57.9% for the CAIC, 54.3% for the BIC, 36.9% for the SCAIC, 26.1% for the LMR, 25.3% for the SBIC, and 16.8% for the BLRT, with all indicators presenting a tendency for overextraction. In this situation, the accuracy of all IC and LRTs deteriorated as a function of sample size, and was lower in the equal (vs. unequal) mixing ratio condition. Finally, across both residual ratios (2:1 and 5:1), the accuracy of the SBIC and BLRT was lower in the high (vs. low) class separation condition, that of the CAIC and BIC was higher in the high (vs. low) class separation condition, and that of the SCAIC was constant across these two conditions when the residual ratio was 2:1, but lower in the high (vs. low) class separation condition when the residual ratio was 5:1 (LMR showed the opposite pattern).
correct three-class solution. The accuracy of the IC and the LRT improved across sample sizes and residual ratio conditions, except for the BLRT, which became less accurate across sample sizes when the residual ratio was 5:1. Similarly, the IC and the LRT had roughly constant rates of accuracy across mixing ratio conditions when the residual structure was equal across latent classes. The SBIC and the LRT also had constant rates of accuracy across unequal residual conditions, whereas the CAIC and BIC had higher rates of accuracy in the equal (vs. unequal) mixing ratio condition. The SCAIC tended to have constant accuracy rates across mixing ratio conditions when the residual ratio was 2:1 but higher accuracy in the unequal mixing condition when the residual ratio was 5:1. Finally, the accuracy of the CAIC, BIC, and LMR was higher in the high (vs. low) class separation condition, whereas that of the SCAIC, the SBIC, and the BLRT showed no systematic pattern of variation across class separation conditions.

Study 3: Summary and Discussion
The results showed that for population models based on a class-invariant latent variance–covariance matrix ($\Psi_k = \Psi$), the rates of accuracy observed for the exactly specified models were the lowest observed so far, but tended to increase substantially when these models relied on freely estimated residuals. Furthermore, for these exactly specified models, the accuracy of the IC and the LRT showed substantial variations as a function of the design conditions, showing more accuracy in conditions of larger sample sizes, equal mixing ratio, low class separation for equal and unequal (2:1) residual conditions, and high class separation for the unequal (5:1) residual condition. Using these exactly specified models as the benchmark for the other types of model specifications, the results first showed that models relying on an LCGA parameterization ($\Psi_k = \Theta$), irrespective of the residual specification, resulted in very low rates of accuracy across indicators and conditions, with rates of overextraction approaching 100%. Second, for models relying on a correctly specified ($\Psi_k = \Psi$) or overspecified ($\Psi_k$) latent variance–covariance matrix, the results showed that relying on models assuming class-invariant residuals helped to increase the accuracy of the indicators, unless this specification resulted in a severe underspecification of the residual structure (when the residual ratio corresponded to 5:1). Third, for models relying on a correctly specified ($\Psi_k = \Psi$) latent variance–covariance matrix and overspecified residuals, rate of accuracy are in line with those observed in the exactly specified models. Fourth, the highest rates of accuracy observed for this population model ($\Psi_k = \Psi$) were associated to the estimation of a model involving an overspecified ($\Psi_k$) latent variance–covariance matrix and a freely estimated residual structure ($\Theta_1$) across latent classes. This condition is also the one in which the indicators showed the lowest levels of reactivity to the design conditions. As in Study 2, the results also showed that, although the design conditions (sample size, mixing ratio, and class separation) had some effect on the accuracy of the indicators, these effects remained much smaller than those of the main specification of the estimated model. Furthermore, models relying on underspecified residual structures appear to be associated with a higher level of reactivity to these other design conditions.

Study 4: Heterogeneous Population Model With Class-Varying Within-Class Variability
Table 5 presents the class enumeration performance of the various indicators as a function of model specification and the population residual ratio for the heterogeneous population model including class-varying within-class variability ($\Psi_k$) used in Study 4. More specific results associated with all design conditions are reported in Tables S19 to S30 of the supplements (available from the corresponding author).

Exactly Specified Models
When the true population model was estimated (i.e., using Model 5 ($\Psi_k; \Theta_k = \Theta$) when $\Theta_k = \Theta$, or using Model 6 ($\Psi_k; \Theta_k$) when $\Theta_k$), the accuracy of the IC and the LMR remained higher than 70% across conditions, and remained similar across residual ratio conditions, showing a tendency to overextract latent classes when the indicators failed to pick the correct three-class solution. More precisely, with class-invariant residuals, rates of accuracy of 87.1% were noted for the CAIC, 99.0% for the SCAIC, 91.6% for the BIC, 96.6% for the SBIC, 73.4% for the LMR, and 78.8% for the BLRT. With class-varying residuals, comparable figures were respectively 91.4%, 94.0%, 93.1%, 90.1%, 77.1%, and 67.1% for population models corresponding to a residual ratio of 2:1, and 91.8%, 93.5%, 93.0%, 90.9%, 79.1%, and 62.7% for population models corresponding to a residual ratio of 5:1.

Models With a Severely Underspecified Variance–Covariance Matrix
When the variance–covariance matrix of the estimated model was underspecified as a LCGA ($\Psi_k = 0$), the results again showed few differences based on whether the residuals were correctly specified, underspecified (estimating Model 1 ($\Psi_k = 0; \Theta_k = \Theta$) when $\Theta_k$), or overspecified (estimating Model 2 ($\Psi_k = 0; \Theta_k$) when $\Theta_k = \Theta$). More precisely, estimating models according to a LCGA ($\Psi_k = 0$) specification while the population model included a class-specific latent variance–covariance matrix ($\Psi_k$) resulted in rates of overextraction of 100% for the IC and the BLRT. The LMR performed better in these conditions with accuracy rates ranging between 26.3%
| No. of classes | AIC | CAIC | BIC | SBIC | LMR | ALMR | BLRT | CAIC | SCAIC | BIC | SBIC | LMR | ALMR | BLRT | CAIC | SCAIC | BIC | SBIC | LMR | ALMR | BLRT |
|----------------|-----|------|-----|------|-----|------|------|------|-------|-----|------|-----|------|------|------|-------|-----|------|-----|------|------|------|
| 1 | 0.1 | 0.2 | 0.2 | 0.2 | 5.1 | 5.3 | 5.2 | 0.9 | 3.3 | 6.4 | 6.8 | 6.5 | 6.7 | 6.5 | 6.1 | 6.5 | 6.7 | 6.8 | 6.5 | 6.7 | 6.8 |
| 2 | 0.4 | 0.7 | 0.7 | 0.7 | 9.8 | 10.6 | 9.6 | 1.9 | 3.1 | 5.2 | 6.1 | 5.1 | 5.4 | 5.2 | 5.3 | 5.2 | 5.4 | 5.1 | 5.3 | 5.4 | 5.2 |
| 3 | 0.9 | 1.5 | 1.5 | 1.5 | 18.9 | 20.6 | 17.9 | 3.3 | 5.5 | 7.6 | 7.3 | 6.4 | 6.5 | 6.3 | 6.4 | 6.4 | 6.5 | 6.6 | 6.7 | 6.5 | 6.6 |
| 4 | 1.4 | 2.1 | 2.1 | 2.1 | 34.8 | 36.5 | 31.9 | 5.4 | 7.7 | 9.8 | 9.1 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 |

Note: Model 1 to Model 6 = type of estimated model; \( \Psi_k \) = latent variance-covariance matrix; \( \Theta_k \) = residual variance-covariance matrix; 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 likelihood ratio test; BLRT = bootstrapped likelihood ratio test.
and 34.0% (and a tendency to underextract the two-class solution 8.1% to 42.7% of the time) across residual conditions. A similar pattern of results was observable across sample size, mixing ratio, and class separation conditions.

Models With a Moderately Underspecified Variance–Covariance Matrix

When the variance–covariance matrix of the estimated model was moderately underspecified ($\Psi_k = \Psi$), the IC and the LRT had higher rates of accuracy than models relying on a severely underspecified variance–covariance matrix ($\Psi_k = 0$), but lower rates of accuracy than the exactly specified models. Across models, the results differed depending on whether the residuals were freely estimated (Model 4: $\Psi_k = \Psi; \Theta_k = \Theta$) or not (Model 3: $\Psi_k = \Psi; \Theta_k = \Theta$), with few differences based on whether these models involved a correct specification or a misspecification of the residuals.

Models with correctly or underspecified equally estimated residuals (Model 3). Compared to the exactly specified model, estimating models based on a moderate underspecification of the variance–covariance matrix and a class-invariant residual structure (Model 3: $\Psi_k = \Psi; \Theta_k = \Theta$) resulted in an important decrease in the accuracy of the IC and the LRT across residual ratio conditions with important rates of overextraction. Rates of accuracy also decreased as a function of the degree of misspecification of the residual structure. When the population model involved class-invariant residuals, rates of accuracy were 61.1% for CAIC, 60.7% for the BIC, 49.4% for the SCAIC, 36.2% for the SBIC, 45.2% for the LMR, and 17.2% for the BLRT. Similar figures were respectively 39.2%, 32.9%, 16.5%, 8.8%, 49.4%, and 5.7% for population models corresponding to a residual ratio of 2:1, and 27.3%, 23.0%, 10.0%, 4.9%, 52.6%, and 3.2% for population models corresponding to a residual ratio of 5:1. Furthermore, in unequal residual ratio conditions, the accuracy of all indicators decreased as a function of sample size, and was slightly lower in the unequal mixing ratio condition across residual ratio conditions. In the equal residual ratio condition, the CAIC, BIC, and LRT had higher rates of accuracy in the high (vs. low) class separation condition, whereas the SCAIC and SBIC had lower rates of accuracy in the high class separation condition. However, all indicators had higher rates of accuracy in the high (vs. low) class separation condition when the residual ratio was 2:1, whereas in the unequal residual of ratio 5:1, the SCAIC, SBIC, and BLRT had lower accuracy in the high (vs. low) class separation condition, whereas the CAIC, BIC, and LMR had higher accuracy in the high class separation condition.

Models With Correctly Specified Variance–Covariance Matrix

Models estimated based on a correctly specified variance–covariance matrix ($\Psi_k$) performed better than all models considered based on either a severely ($\Psi_k = 0$) or moderately misspecified ($\Psi_k = \Psi$) variance–covariance matrix. However, rates of accuracy differed based on whether the residuals were underspecified (Model 5: $\Psi_k; \Theta_k = \Theta$) or overspecified (Model 6: $\Psi_k$).

Models with underspecified equally estimated residuals (Model 5). Compared to the exactly specified model, estimating models based on a correctly specified variance–covariance matrix and an underspecified residual structure (Model 5: $\Psi_k; \Theta_k = \Theta$) resulted in greater accuracy rates for all indicators than all of the previously reviewed conditions based on Model 1 ($\Psi_k = \Psi; \Theta_k = \Theta$), Model 2 ($\Psi_k = 0; \Theta_k$), Model 3 ($\Psi_k = \Psi; \Theta_k = \Theta$), and Model 4 ($\Psi_k; \Theta_k$). However, the rates of accuracy
associated with this specific condition depended on the residual ratio condition. Thus, for population models based on a residual ratio of 2:1, rates of accuracy for the IC were similar to those observed in the exactly specified model (CAIC, 96.5%; SCAIC, 92.3%; BIC, 97.7%; SBIC, 79.4%), although a decrease in the accuracy of the LRT (LMR, 61.1%; BLRT, 49.1%) was also observed. In this condition (2:1), the accuracy of the BIC and the SBIC stayed constant across sample size conditions, whereas that of the CAIC increased as a function of sample size, and that of the SCAIC, LMR, and BLRT decreased. Similarly, the accuracy of the CAIC, SCAIC, and BIC remained generally comparable across mixing ratio and class separation conditions, whereas that of the SBIC, LMR, and BLRT tended to be lower in the equal (vs. unequal) mixing ratio condition as well as in the high (vs. low) class separation condition (except for the LMR, which were more accurate in the high class separation condition). However, rates of accuracy decreased substantially for all indicators for population models based on a residual ratio of 5:1, with rates of overextraction varying between 34.4% and 85.9% across indicators. In this condition (5:1), the accuracy of all indicators further decreased as a function of sample size, and in the equal (vs. unequal) mixing ratio condition. Finally, whereas the SCAIC, SBIC, and LMR had similar rates of accuracy across class separation conditions, the BLRT had a lower accuracy in the high class separation condition, whereas the CAIC and the BIC had a higher accuracy in the high class separation condition.

**Models with overspecified freely estimated residuals (Model 6).** Estimating models based on a correctly specified variance–covariance matrix and an overspecified residual structure (Model 6: \( \Psi_k; \Theta_k \)) resulted in accuracy rates that are similar to those observed for the exactly specified models for all indicators. In other words, this specification resulted in the greatest rates of accuracy observed so far when compared to models estimated using Model 1 (\( \Psi_k = 0; \Theta_k = \Theta \)), Model 2 (\( \Psi_k = 0; \Theta_k \)), Model 3 (\( \Psi_k = \Psi; \Theta_k = \Theta \)), Model 4 (\( \Psi_k = \Psi; \Theta_k \)), and Model 5 (\( \Psi_k; \Theta_k = \Theta \)). In this condition, the CAIC (85.3%), SCAIC (98.7%), BIC (98.4%), and SBIC (96.6%) were more accurate than the LRT (75.0% for the LMR and 77.3% for the BLRT), who presented a slight tendency for overextraction when they failed to correctly pick the three-class model. Across sample sizes, rates of accuracy remained constant for the SCAIC; increased for the CAIC, BIC, SBIC, and LMR; and decreased for the BLRT. In contrast, the accuracy of all IC and the LRT remained unaffected by mixing ratio conditions. Finally, the accuracy of the SCAIC, SBIC, and BLRT remained constant across the two class separation conditions, whereas that of the CAIC, BIC, and LMR was higher in the high class separation condition.

**Study 4: Summary and Discussion**

For population models involving a class-specific latent variance–covariance matrix (\( \Psi_k \)), relying on exactly specified models resulted in generally high rates of accuracy for the indicators (85%–98% for the IC, and 62%–80% for the LRT) and showed only very limited effects of the other design conditions. Interestingly, when the latent variance–covariance matrix was correctly specified (\( \Psi_k \)), relying on an overspecification of the residual structure had no detrimental effect on the accuracy of the various indicators and again showed only limited effects of the other design conditions. Similarly, when the latent variance–covariance matrix was correctly specified, relying on a moderate underspecification of the residual structure (2:1) also had only a limited impact on the accuracy of the IC, but was associated with a decrease in the accuracy of the LRT, whereas relying on a severe underspecification of the residual structure (5:1) resulted in a substantial decrease in the accuracy of all indicators and a greater level of sensitivity to the other design conditions. In contrast, models relying on a moderately (\( \Psi_k = \Psi \)) or severely (\( \Psi_k = 0 \)) underspecified latent variance–covariance matrix systematically resulted in a decrease in the accuracy of the indicators, irrespective of the residual specification. In particular, relying on a LCGA parameterization (\( \Psi_k = 0 \)) resulted in rates of overextraction approaching 100% for the IC and BLRT, across conditions. Models relying on a moderately underspecified latent variance–covariance matrix (\( \Psi_k = \Psi \)) proved to be slightly more accurate, especially when they relied on a freely estimated residual structure, but remained substantially less accurate than models relying on a correctly specified latent variance–covariance matrix (\( \Psi_k \)), and far more sensitive to the other design conditions.

**GENERAL DISCUSSION**

This series of simulation studies was designed to assess the impact of different forms of misspecifications of the latent variance–covariance matrix (i.e., \( \Psi_k \)) and residual structure (i.e., \( \Theta_k \)) in the context of the GMM class enumeration process. To this end, a series of four simulation studies relying on distinct population models was conducted to assess the ability of the IC and LRT to correctly identify the true number of latent classes present in the data. The first study relied on a homogenous (one-class) population model, whereas Studies 2 to 4 relied on three-class population models differing in the way the latent variance–covariance matrix was specified. Study 2 relied on Nagin’s (1999) LCGA specification (\( \Psi_k = 0 \)), Study 3 on the Mplus default specification (\( \Psi_k = \Psi \)), and Study 4 on a class-varying specification (\( \Psi_k \)). In each of these studies, we assessed the class enumeration performance of GMMs as a function of the type of estimated model: (a) Model 1: \( \Psi_k = 0 \) and \( \Theta_k = \Theta \); (b) Model 2: \( \Psi_k = 0 \) and \( \Theta_k \); (c) Model 3:
$\Psi_k = \Psi$ and $\Theta_k = \Theta$; (d) Model 4: $\Psi_k = \Psi$ and $\Theta_k$; (e) Model 5: $\Psi_k$ and $\Theta_k = \Theta$; and (f) Model 6: $\Psi_k$ and $\Theta_k$. We also considered the impact of the true within-class residual matrix (simulated to be invariant across classes, or to differ across classes based on a ratio of 2:1 or 5:1), and assessed the extent to which the results generalized across three additional design conditions (sample size, mixing ratio, class separation).

Very limited research had been conducted in this area, and no study has ever systematically contrasted all possible types of model misspecifications at the level of the latent variance–covariance matrix, the residual structure, and their combination. So far, previous research has shown that restricted GMM parameterizations (e.g., LCGA) could result in the overextraction of spurious latent classes as a way to compensate for the misspecification of the within-class variance–covariance component (Bauer & Curran, 2003). However, Peugh and Fan (2012) showed that GMMs allowing for within-class heterogeneity in the latent variance–covariance matrix (but assuming the invariance across classes of both $\Psi_k = \Psi$ and $\Theta_k = \Theta$) were able to correctly recover a homogenous population model. Finally, Enders and Tofghi (2008) showed that when $\Psi_k$ was correctly specified, underspecifying $\Theta_k$ had only a minor impact on the class enumeration accuracy for heterogeneous population models. However, none of those studies systematically considered the impact of overspecifications or underspecifications of $\Psi_k$, of overspecifications of $\Theta_k$, or the joint impact of simultaneous misspecifications of both $\Psi_k$ and $\Theta_k$.

The results obtained in the current series of studies are important for many reasons. First, they replicated Peugh and Fan’s (2012) results, showing that models involving within-class variability in the latent variance–covariance matrix were able to successfully recover a homogenous population model, irrespective of whether this variability was constrained to be invariant across classes ($\Psi_k = \Psi$) or not ($\Psi_k$), and irrespective of the residual specification ($\Theta_k = \Theta$ or $\Theta_k$). In contrast, in line with Bauer and Curran (2003), our results showed that LCGA systematically resulted in the overextraction of spurious latent classes, thus clearly arguing against the value of such a parameterization.

Second, across studies and conditions, our results showed that models in which $\Psi_k$ was underspecified resulted in a substantial decrease in class-estimation accuracy when compared to the exactly specified model. In contrast, almost no risk was associated with the overspecification of $\Psi_k$, particularly when it was freely estimated across classes. Third, the results revealed that the advantages associated with models including a within-class latent variance–covariance matrix ($\Psi_k = \Psi$ or $\Psi_k$) were also largely dependent on the specification of the residual structure. More precisely, the results show that models relying on moderately underspecified residuals ($\Theta_k = \Theta$ when the residual ratio is 2:1) might be advantageous in terms of class enumeration accuracy under some specific conditions, whereas models relying on severely underspecified residuals ($\Theta_k = \Theta$ when the residual ratio is 5:1) systematically resulted in a substantial decrease in class enumeration accuracy. In particular, models relying on underspecified residual structures tended to be more sensitive to the impact of additional design conditions. In contrast, models relying on an overspecified residual structure almost never resulted in a decrease in class enumeration accuracy compared to the exactly specified models. Overall, our results show that relying on models in which both the latent variance–covariance matrix ($\Psi_k$) and the residual structure ($\Theta_k$) are freely estimated almost systematically result in rates of accuracy comparable to those obtained with exactly specified models, across studies and conditions.

Fourth, our results show that the impact of various forms of model misspecifications described thus far is far larger than that of the other design conditions (sample size, mixing ratio, class separation), and generally reveal only limited variations across indicators (IC, LRT). Regarding the accuracy of these indicators, our results generally support previous research showing that the AIC is not trustworthy and should not be used in the class enumeration process (e.g., Li & Hser, 2011; Nylund et al., 2007; Tofghi & Enders, 2007). The results show that each of the four remaining IC (CAIC, BIC, SCAIC, SBIC) outperformed the others within specific design conditions, and thus all appear useful and complementary. Across studies and conditions, the accuracy of these four IC was also much higher than that of the LMR or ALMR and BLRT. Although these results are in line with those obtained in previous studies regarding the poor performance of the LMR or ALMR (e.g., Li & Hser, 2011; Nylund et al., 2007; Tofghi & Enders, 2007), they are surprising regarding the BLRT. Indeed, the BLRT possesses strong statistical foundations and has been found to be a reliable indicator of class enumeration in previous studies (e.g., Li & Hser, 2011; Nylund et al., 2007). However, if one looks at applied GMM studies in which BLRT results are published, this indicator very seldom proves to be useful in practice (e.g., Morin et al., 2011; Prince & Maisto, 2013). Taken together, these observations call for further examination of the BLRT performance across a wider range of conditions.

In conducting this series of studies, our objective was to provide applied users with clear practical guidelines regarding the appropriateness of various forms of restrictions to impose on GMM components as part of the model estimation process to maximize the chances of identifying the true number of latent classes present in the data. In this regard, our results clearly show that, to maximize the accuracy of the class enumeration process in situations where the underlying true population model remains unknown, users should
rely on a GMM specification in which the latent variance–covariance matrix (Ψk), and the residual structure (Θk) are freely estimated in all classes.

However, the situation is not that simple. As noted in the introduction, the free estimation of all GMM parameters within all latent classes often results in inadmissible solutions.1 So far, simulation studies of LCM (e.g., Diallo & Morin, 2015; Diallo et al., 2014; Fan & Fan, 2005) and GMMs (e.g., Liu & Hancock, 2014; Nylund et al., 2007; Tolvanen, 2007) have shown that the frequency of these inadmissible solutions tends to rapidly increase as a function of model complexity, and to decrease as a function of the joint effect of sample size and number of measurement occasions. In this series of studies, rates of inadmissible solutions remained generally low, ranging (across studies and conditions) from 0.05% for the simplest models with no within-class latent variability (Model 1: [Ψk = 0; Θk = Θ] and Model 2: [Ψk = 0; Θk]), to 6.22% for models with a class-invariant latent variance–covariance matrix (Model 3: [Ψk = Ψ; Θk = Θ] and Model 4: [Ψk = Ψ; Θk]), and 12.47% for the models with a freely estimated latent variance–covariance matrix (Model 5: [Ψk; Θk = Θ] and Model 6: [Ψk; Θk]). These results only showed minimal variations as a function of the other design conditions.

A possible explanation for the fact that rates of inadmissible solutions remained low is that we considered models including seven measurement occasions, based on results from prior simulation studies (e.g., Tofigli & Enders, 2007) showing that this factor had only a limited impact on the class enumeration performance of GMMs. Thus, although our results suggest that GMMs should be estimated using a model corresponding to Model 6, this model is also accompanied by a higher rate of inadmissible conditions than the more parsimonious models. These rates of inadmissible solutions are also likely to increase with real data sets, especially when fewer than seven measurement occasions are available. In particular, following from previous simulation studies (Enders & Tofigli, 2008; Li & Hser, 2011; Tofigli & Enders, 2007), we relied on a simplified parameterization of the residual structure as invariant across classes or differing across classes based on a 2:1 or 5:1 ratio. Based on a series of pilot analyses conducted prior to the main simulation studies, we also considered a more extreme between-class residual ratio condition (10:1), as well as situations where the residuals were also specified as differing across time points within a single class, or as differing across both classes and time points. These more extreme conditions were found to result in much higher rates of nonconvergence and inadmissible solutions. As with most simulation studies, choices had to be made to keep this study within reasonable boundaries, leading us to abandon these more extreme conditions. Similarly, this series of studies relied on a diagonal specification of the residual structure Θk, whereas prior studies of alternative types of mixture models showed that violation of this conditional independence assumption (i.e., assuming the orthogonality of the residuals conditional on the model) could also lead to potentially severe biases in terms of class enumeration accuracy (Peugh & Fan, 2013), as well as parameter or standard error estimates (Heggeseth & Jewell, 2013). Future studies should devote attention to the impact of violations of this conditional independence assumption, as well as to the performance of GMMs when the residual structure differs across time points.

Importantly, our results also suggest that rates of inadmissible solutions tended to increase as a function of the number of latent classes included in the estimated model. For example, for Model 5 and Model 6, rates of inadmissible solutions for models including one, two, three, and four classes were respectively 0%, 9.88%, 11.22%, and 28.79%. Similar observations have led some authors to suggest considering inadmissible solutions as evidence in favor of the model with fewer latent classes (e.g., Nylund et al., 2007). However, even exactly specified models including the correct number of latent classes frequently result in inadmissible solutions (e.g., for exactly specified three-class models based on Model 5 or Model 6, 10.81% of the solutions were inadmissible), showing that inadmissible solutions cannot realistically be taken as evidence supporting a model including fewer latent classes. Echoing this conclusion, Liu and Hancock (2014; also see Pastor & Gagné, 2013) noted that in practice, a challenge facing researchers is to balance the costs of parsimony (possibly resulting in misspecifications) versus the costs of complexity (providing a more exact depiction of the data, but also increasing the risk of inadmissible solutions).

Keeping in mind that inadmissible solutions are highly likely, we propose the following sequential strategy to the estimation of GMM:

1. The class enumeration process should be conducted using a model corresponding to Model 6, including a free estimation of the latent variance–covariance matrix (Ψk) and residual structure (Θk) in all classes. There are four possible outcomes to this initial class enumeration process:

---

1 Inadmissible solutions include nonconvergent solutions and improper solutions. Following the standard practice in mixture modeling simulation studies (e.g., Bauer & Curran, 2003; McLachlan & Peel, 2000) a replication was labeled nonconvergent when that replication failed to converge on a solution with any of the sets of the starting values that were used. A convergent solution was labeled improper when it included out-of-bound parameter estimates (i.e., negative variances, correlations greater than 1 or smaller than −1). Because inadmissible solutions are rarely interpreted (Chen, Bollen, Paxton, Curran, & Kirby, 2001) they were excluded from the results reported here.
i. The model converges on a proper solution with well-replicated log-likelihood. In this situation, the M6 specification should be retained.

ii. 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). In this situation, these out-of-bound parameter estimates might simply reflect random sampling variations or the presence of a slight overparameterization in a subset of classes (e.g., Tolvanen, 2007). For example, estimating a quadratic growth model in a class for which a linear growth process is sufficient is likely to result in out-of-bound parameters estimates in \( \Psi_k \) for the components associated with the quadratic slope. In these situations, the model suggested as optimal as part of this class enumeration process can be re-estimated with parameter constraints on the out-of-bound estimates (either constraining them to be zero, or to be higher than zero; see, e.g., Morin, Maïano, Marsh, Nagengast, & Janosz, 2013). If these out-of-bound estimates do indeed simply reflect random sampling variations or a slight overparameterization, these constraints should not change the nature of the estimated latent classes and converge on a proper solution. This solution can then be retained as the final model.

iii. The model converges on a solution with well-replicated log-likelihood, but includes multiple severely out-of-bound parameter estimates for which the solution proposed in the preceding scenario does not work, or results in a degenerate solution including empty classes. In this situation, users should move to the second strategy. However, the result of the class enumeration process (in terms of number of classes) conducted using Model 6 can be retained as a rough guideline to help with model selection.

iv. The model simply fails to converge, or to converge on a solution with well-replicated log-likelihood. In this situation, users should move to Strategy 2.

2. When the Model 6 approach fails, a more constrained model is required. However, the decision of which constraints to implement first is not easy to determine. As a first step, we suggest conducting the class enumeration process using the simplest possible model among those that can be recommended (i.e., excluding Model 1 \( [\Psi_k = 0; \Theta_k = \Theta] \) and Model 2 \( [\Psi_k = 0; \Theta_k] \)).

i. The class enumeration process should be conducted using Model 3 \( [\Psi_k = \Psi; \Theta_k = \Theta] \) as a starting point, which was found to be able to properly differentiate between homogenous and heterogeneous population models. Whenever this process results in the identification of a homogenous (one-class) solution, then a simpler LCM approach should be preferred.

ii. In all other situations, alternative specifications can be explored and compared to Model 3, allowing either \( \Psi_k \) (Model 5) or \( \Theta_k \) (Model 4) to be freely estimated across all classes at a time. Ideally, the final decision should be informed by a comparison of the results obtained using Model 3, Model 4, and Model 5, or those specifications converging on proper solutions (a strategy similar to the one described earlier can also be used when necessary). This comparison can be done using the IC presented earlier to assess the relative fit of each parameterization (based on the comparison of models including the same number of latent classes). Similarly, the optimal solution (in terms of number of latent classes) retained through each of these parameterizations can be contrasted.

iii. The decision of which model to retain should be guided by theory, common sense, and a comparison of results. Substantively, the free estimation of \( \Psi_k \) involves the recognition that the level of within-class variability (or interindividual similarity) can differ across latent classes. In contrast, the free estimation of \( \Theta_k \) involves the recognition that the estimated model does not provide a similarly adequate representation of the repeated measures in each class. Following others, we doubt the realism of assuming equal amounts of within-class variability (e.g., Morin et al., 2011; Peugh & Fan, 2013). In contrast, it appears more reasonable to assume that a statistical model will provide an explanatory power that generalizes to the various subpopulations represented in a specific sample. However, specific research contexts might attribute a greater importance to the free estimation of \( \Theta_k \) than of \( \Psi_k \) (e.g., Morin et al., 2013). If we go back to our results, they suggest that the advantages of using models involving the free estimation of \( \Psi_k \) generally tended to outweigh those of using models involving the free estimation of \( \Theta_k \), except in the most extreme condition where the residuals differ by a ratio of 5:1. For these reasons, our recommendation is to prioritize models corresponding to Model 5. However, observing that Model 4 results in a much improved fit to the data in comparison to Model 5 argues in favor of the Model 4 parameterization.

In some situations, Strategy 2 will also result in inadmissible solutions. When this happens, our recommendation is to abandon the GMM approach and
possibly attempt to collect additional data, or time waves. On the basis of the results presented here, we do not recommend the use of an LCGA approach, or do so only with great caution and awareness of the important risks of overextraction associated with this approach. Alternatively, other approaches can also be investigated, such as unstructured cluster analyses or mixture models (e.g., Liu & Hancock, 2014; Martin & Von Oertzen, 2015), Bayesian models (e.g., Depaoli, 2014; Lu, Zhang, & Lubke, 2011), or time-structured clustering algorithms (e.g., Genolini & Falissard, 2011).

At this stage, it should be kept in mind that our recommendations are based on the results from studies in which the focus was on the performance of alternative GMM specifications to recover the true number of latent classes in the data. Future studies are needed to complement this information with a thorough assessment of the ability of these alternative models (when they include the proper number of latent classes) to recover the true underlying population parameters. Fortunately, our results suggest that these future studies could focus their attention on the population models described in Studies 3 and 4, given that the performance of LCMs is already well documented (e.g., Diallo & Morin, 2015; Diallo et al., 2014) once the class enumeration process has identified a homogenous solution, and that our results suggest that LCGA should generally be avoided. At this stage, we strongly urge researchers to consider the proposed framework only as a rough guideline that might guide, in combination with theory, common sense, and a detailed examination of parameter estimates, the GMM estimation process, rather than as a rigid set of restrictive procedures that need to be followed to the letter. It is our hope that this series of studies will contribute to stimulate further research, which will help to continuously improve on current best practices in the use of GMMs.

REFERENCES


