



Covariate inclusion in factor mixture modeling: Evaluating one-step and three-step approaches under model misspecification and overfitting

Yan Wang¹ · Chunhua Cao² · Eunsook Kim³

Accepted: 24 August 2022 / Published online: 12 September 2022
© The Psychonomic Society, Inc. 2022

Abstract

Factor mixture modeling (FMM) has been increasingly used in behavioral and social sciences to examine unobserved population heterogeneity. Covariates (e.g., gender, race) are often included in FMM to help understand the formation and characterization of latent subgroups or classes. This Monte Carlo simulation study evaluated the performance of one-step and three-step approaches to covariate inclusion across three scenarios, i.e., correct specification (study 1), model misspecification (study 2), and model overfitting (study 3), in terms of direct covariate effects on factors. Results showed that the performance of these two approaches was comparable when class separation was large and the specification of covariate effect was correct. However, one-step FMM had better class enumeration than the three-step approach when class separation was poor, and was more robust to the misspecification or overfitting concerning direct covariate effects. Recommendations regarding covariate inclusion approaches are provided herein depending on class separation and sample size. Large sample size (1000 or more) and the use of sample size-adjusted BIC (saBIC) in class enumeration are recommended.

Keywords Factor mixture model · Covariate · One-step · Three-step · Model misspecification

Introduction

Factor mixture modeling (FMM) has been increasingly used in behavioral and social sciences to examine unobserved population heterogeneity (e.g., Allan et al., 2014; Bernstein et al., 2013; Dimitrov et al., 2015; Elhai et al., 2011). FMM combines common factor model and latent class analysis (LCA) by allowing for the simultaneous presence of a continuous latent variable (factor) and a categorical latent variable (latent class) in the model (Lubke & Muthén, 2005). Distinct latent classes that differ in parameters of the common factor model (e.g., factor mean, factor variance, loadings,

intercepts) would emerge. Given that classes are unobserved, covariates (e.g., gender, race) are often linked to the latent class variable to help understand the formation and characterization of latent classes. Specifically, a significant covariate effect would indicate that the latent class membership can be explained by this covariate. For instance, if gender is a significant covariate, one latent class might be characterized by having a large proportion of females and the other class might be dominated by males.

To evaluate the role of covariates in FMM, there are two decisions that have to be made in specifying covariate effects. The first is about when to include covariates, and two common options are one-step and three-step approaches, both of which concern the covariate effect on the latent class variable. In one-step FMM, such a covariate effect is included in the process of identifying the optimal number of latent classes (i.e., class enumeration). This approach might improve the class enumeration and class assignment, as the incorporation of covariates can increase class separation (Lubke & Muthén, 2007; Wang et al., 2020; Wang et al., 2021). The downside of this approach is that class enumeration and class assignment might change considerably when different covariates are included (Lubke & Muthén, 2005;

✉ Yan Wang
Yan_Wang1@uml.edu

¹ Department of Psychology, University of Massachusetts Lowell, Lowell, MA 01854, USA

² Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, Tuscaloosa, AL 35487, USA

³ Department of Educational and Psychological Studies, University of South Florida, Tampa, FL 33620, USA

Nylund-Gibson & Masyn, 2016). To prevent this change of classification, some researchers have suggested excluding covariates in the latent class enumeration, i.e., specifying an unconditional mixture model, and including covariate effects in the subsequent analyses. Based on a three-step maximum likelihood (ML) procedure proposed by Vermunt (2010), latent class enumeration is conducted with the unconditional model (step 1), and then observations are classified into one of the latent classes based on the most likely class membership (step 2). Step 3 is to regress the class variable on covariates while taking into account the classification error. Note that this approach was referred to as a three-step ML or modal ML procedure in Vermunt (2010) and the three-step approach in this paper.

Whereas the first decision focuses on when to estimate covariate effects on the latent class variable, the second decision is about how to specify direct covariate effects, i.e., whether or not the direct covariate effect on the factor should be estimated, which is of focal interest in this study. Due to the complexity of FMM, covariate specification also becomes complex, and one example is the presence of the covariate effect on the factor in addition to the latent class variable, indicating that within-class variation in the factor scores can be explained by the covariate. For instance, Lubke and Muthén (2005) demonstrated that gender and urban status accounted for some within-class variations in math achievement. Because the population model is unknown in applied research, model misspecification might occur when direct covariate effects are omitted in the analysis; model overfitting is possible when direct covariate effects do not exist in the population but are estimated.

Given that the impact of these two decisions is not well understood in FMM, the overarching goal of the paper is to comprehensively evaluate the performance of one-step and three-step approaches in FMM via Monte Carlo simulations, while taking into account the potential presence of direct covariate effects. Previous methodological studies that compared one-step and three-step approaches have largely focused on LCA and growth mixture modeling (GMM) (e.g., Asparouhov & Muthén, 2014; Cetin-Berber & Leite, 2018; No & Hong, 2018; Park & Yu, 2018; Vermunt, 2010). Among these previous studies on LCA and GMM, inconsistent findings have been observed regarding the performance of one-step and three-step approaches and, more importantly, a comprehensive evaluation of covariate inclusion approaches is still lacking given that misspecification and/or overfitting of covariate effects has not been investigated considering both class enumeration and parameter recovery. In addition, it remains unknown to what extent previous findings on LCA and GMM would be applicable to FMM, given that FMM can be considered as a more complex model with a freely estimated measurement model, whereas a measurement model is absent in LCA and is constrained (loadings

and intercepts fixed to constant values for parameterization) in GMM. In particular, due to the increasing model complexity of FMM, we hypothesize that the one-step approach might outperform the three-step approach in terms of class enumeration due to the contribution of proper covariates to class separation, when FMM is correctly specified (Lubke & Muthén, 2007; Wang et al., 2021). However, it is unclear how class enumeration of these two approaches will be impacted by misspecification and overfitting of FMM—the benefits of covariate inclusion might be offset by the misspecification or overfitting for the one-step approach, whereas the three-step approach does not have the aid of covariates but might be robust to misspecification or overfitting. Moreover, parameter recovery (especially covariate effects) has not been systematically investigated regardless of FMM specification. Such insufficient understanding of covariate inclusion in FMM is in stark contrast to the increasing popularity of FMM and the prevalence of covariate inclusion to help understand and characterize heterogeneity, highlighting the need for a comprehensive evaluation of one-step and three-step approaches under FMM.

In response to this need, this study evaluated the performance of one-step and three-step approaches across different scenarios, i.e., correct specification, model misspecification, and model overfitting, in terms of direct covariate effects on factors. The first study compared the performance of the two approaches when only the covariate effect on the latent class variable is considered (i.e., correct specification). The second study evaluated the performance of these two approaches under model misspecification. That is, direct covariate effects on factors are ignored in one-step and three-step procedures. The third study focused on the scenario of model overfitting when the covariate effects on the factors can be zero in the population but was included in the analytical model. Both class enumeration (i.e., proportion of replications that selected the correct number of classes) and parameter recovery (e.g., bias in covariate effects and factor mean difference) were evaluated across studies to provide a more comprehensive investigation and well-grounded recommendations to practitioners.

Factor mixture modeling

FMM integrates LCA and a common factor model (Lubke & Muthén, 2005). With i denoting an individual and k referring to the class the individual is assigned to ($k = 1, 2, \dots, K$), the common factor model can be expressed as:

$$Y_{ik} = \tau_k + \Lambda_k \eta_{ik} + \varepsilon_{ik}. \quad (1)$$

Y_{ik} , a $J \times 1$ vector of responses with J denoting the number of items, is a function of a $J \times 1$ vector of item intercepts

τ_k , a $J \times R$ matrix of factor loadings Λ_k with R denoting the number of factors, a $R \times 1$ vector of factor scores η_{ik} , and a $J \times 1$ vector of item residuals ϵ_{ik} . Residuals are assumed to be multivariate normally distributed with mean $\mathbf{0}$ and variance-covariance matrix Θ_k (dimension $J \times J$). The subscript k associated with the model parameters indicates that they can vary across latent classes. Factor scores are assumed to be normally distributed, with α_k representing the vector of factor means and Ψ_k the covariance matrix of factors. Thus, the class-specific mean vectors and class-specific variance–covariance matrices can be expressed as:

$$\mu_k = \tau_k + \Lambda_k \alpha_k, \tag{2}$$

$$\Sigma_k = \Lambda_k \Psi_k \Lambda_k' + \Theta_k. \tag{3}$$

Covariates are often included in FMM to explain the latent class membership and help researchers understand the composition or characteristics of latent classes (e.g., Bernstein et al., 2013; Elhai et al., 2011). The probability of belonging to latent class k over a reference class r is estimated through a multinomial regression model with covariates X :

$$\ln \left[\frac{P(C_i = k|X_i)}{P(C_i = r|X_i)} \right] = \nu_k + \Gamma_k X_i, \tag{4}$$

where ν_k and Γ_k represent vectors of intercepts and regression coefficients, respectively.

Additionally, covariates can be included in FMM to explain within-class variations in factor scores (Bauer, 2007; Lubke & Muthén, 2005). When such an effect is present, factor scores can be expressed as:

$$\eta_{ik} = A_k + \Gamma_k^\eta X_{ik} + \zeta_{ik}. \tag{5}$$

In this equation, factor scores are a function of intercepts A_k , effect of covariate effect on factor Γ_k^η , covariates X_{ik} , and residuals ζ_{ik} . Note that although direct covariate effects on items or observed indicators are also possible, they are not considered in this study given that they substantively represent within-class measurement noninvariance in terms of the covariate (De Ayala et al., 2002; Lee & Beretvas, 2014; Lubke & Muthén, 2005; Tay et al., 2011), which is beyond the scope of this study.

Approaches to covariate inclusion

In this section, we review previous simulation studies that evaluated the performance of one-step and three-step approaches in mixture modeling (see Table 1 for a summary table and https://osf.io/amupe/?view_only=f72fb1198c

4947dbab731cebb5c416a3 for a more detailed summary). We first review studies that were conducted in the context of correct specification of covariate effects, that is, only the covariate effect on the latent class variable was in the population, and both one-step and three-step approaches correctly specified the covariate effect. Next, we review simulation studies that evaluated one-step and/or three-step approaches when covariate effects were misspecified or overfitted.

Correct specification of covariate effects

Overall, mixed findings have been reported by simulation studies that evaluated the performance of one-step and three-step approaches using different mixture models. For example, using LCA with predictor or outcome variables, No and Hong (2018) found that the three-step approach produced more stable results in estimating the relationships between the latent class variable and external variables than the one-step approach. With nested data, multilevel FMM using the three-step approach performed well in detecting between-level latent classes among which measurement noninvariance was present (Kim & Wang, 2018). By contrast, other studies found that the inclusion of proper covariates in FMM improved class enumeration and assignment, the coverage of factor mean differences, and measurement invariance testing (e.g., Lubke & Muthén, 2007; Wang et al., 2020; Wang et al., 2021). In the context of GMM with covariate effects on either class membership or growth factors or both, Diallo and Lu (2017) found that the correctly specified one-step approach outperformed the three-step approach in terms of the accuracy of covariate effect and standard error estimates. Note that despite these benefits of covariate inclusion, the one-step approach has been criticized because the inclusion of covariates in the measurement model impacts the formation and interpretation of latent classes (Asparouhov & Muthén, 2014; Bakk et al., 2013; Vermunt, 2010).

Alternatively, some simulation studies found that the relative performance of one-step and three-step procedures depended upon manipulated factors. For instance, the performance of the three-step approach has been shown to be comparable to the one-step approach in sufficiently good class separation in LCA and GMM (Asparouhov & Muthén, 2014; Cetin-Berber & Leite, 2018; Li & Haring, 2017; Park & Yu, 2018). However, Park and Yu (2018) noted superior performance of the one-step procedure in the recovery of the effects of continuous covariates when class separation was poor and/or sample size was small with multilevel LCA. Li and Haring (2017) found that with GMM, when class separation was poor and the covariate effect was weak, the estimation of dichotomous covariate effects was problematic for both approaches but more severely for the three-step procedure. Stegmann and Grimm (2018) also highlighted the importance of covariate effect strength in GMM: covariate inclusion (as in the

Table 1 Summary of previous simulation studies on approaches to covariate inclusion in LCA, GMM, and FMM

Article	Model	Covariate effect in population	Approaches	Covariate effect specification	Simulation outcomes	General recommendation
Asparouhov & Muthén (2014)	LCA/GMM	Latent class variable and item/latent class variable and growth factors	One-step, three-step, adjusted three-step	Correct, misspecification	Recovery of covariate effects	One-step, adjusted three-step
Célin-Berber & Leite (2018)	GMM	Latent class variable	One-step, three-step	Correct	Recovery of various parameters, class assignment accuracy	One-step, three-step
Diallo & Lu (2017)	GMM	Latent class variable, latent class variable and growth factors	One-step, step 1 of three-step	Correct, misspecification	Class assignment accuracy, recovery of covariate effects	One-step
Diallo et al. (2017)	GMM	Latent class variable, latent class variable and growth factors	One-step, step 1 of three-step	Correct, misspecification, overfitting	Class enumeration	Three-step (unconditional model for class enumeration)
Hu et al. (2017)	GMM	Latent class variable and growth factors	One-step, step 1 of three-step	Correct, misspecification	Class enumeration	Depends on class separation and sample size
Kim & Wang (2018)	Multilevel FMM	Latent class variable (at the cluster level)	Three-step	Correct	Detection rates of measurement invariance level, class enumeration, class assignment accuracy, recovery of covariate effects	Three-step
Li & Harring (2017)	GMM	Latent class variable	One-step, three-step	Correct	Recovery of covariate effects	Depends on class separation
Lubke & Muthén (2007)	FMM	Latent class variable	One-step, step 1 of three-step	Correct	Recovery of various parameters, class assignment accuracy, entropy	One-step
No & Hong (2018)	LCA	Latent class variable	One-step, three-step	Correct	Recovery of covariate effects	Three-step
Park & Yu (2018)	Multilevel LCA	Latent class variables (both levels)	One-step, three-step	Correct	Recovery of covariate effects	Depends on class separation and sample size
Stegmann & Grimm (2018)	GMM	Latent class variable	One-step, step 1 of three-step	Correct	Class assignment accuracy, entropy	Three-step
Wang et al. (2020)	FMM	Latent class variable and factor	One-step, step 1 of three-step	Correct, misspecification, overfitting	Class enumeration	One-step
Wang et al. (2021)	FMM	Latent class variable	One-step, step 1 of three-step	Correct	Class enumeration, recovery of factor means	One-step

LCA = latent class analysis; GMM = growth mixture modeling; FMM = factor mixture modeling. Step 1 of three-step refers to the specification of an unconditional model, and the covariate effects are not estimated subsequently

one-step approach) was only beneficial when the association between classes underlying the covariates and growth classes was strong and classes underlying the covariates were at least moderately separated; in all other cases class recovery was negatively affected by covariate inclusion.

Misspecification or overfitting of covariates effects

When modeling covariates in FMM, researchers are likely to misspecify or overfit the covariate effects. As mentioned previously, misspecification and overfitting in this study refer to the direct covariate effects on the factors. Therefore, the literature reviewed in this section is limited to methodological studies that examined the performance of the one-step and/or the three-step that misspecified or overfitted direct covariate effects on factors. Overall, the robustness of the three-step approach to the misspecification of covariate effects has been evidenced in most studies. For instance, Hu et al. (2017) observed reliable performance of the unconditional GMM when direct covariate effects on the growth factors (intercept and slope) were simulated in the population. GMMs with direct covariate effects on growth factors outperformed the unconditional model only when class separation and sample size were small, but correct enumeration rates of the conditional GMMs were still very low. Diallo et al. (2017) examined the impact of partial or total inclusion or exclusion of active or inactive covariates on class enumeration in GMM. Their findings also suggested that class enumeration in GMM should be conducted without covariates.

On the other hand, Asparouhov and Muthén (2014) showed that the three-step approach ignoring direct covariate effects did not perform as well as the one-step procedure or an adjusted three-step approach that included the covariate effects in the first step. Specifically, when direct effects on growth factors in GMM were ignored, severe bias and low coverage in the estimation of covariate effects on the latent class variable occurred, especially when class separation was poor and/or the omitted effects were strong. Relatedly, the impact of ignoring direct covariate effects on model fit in FMM was examined in Wang et al. (2020), who compared the fit of correctly specified and misspecified FMMs along with varying numbers of classes and levels of invariance. They found that misspecified FMMs that ignored the direct covariate effect were rarely selected as the best-fitting model by the Bayesian information criterion (BIC) and sample size-adjusted BIC (saBIC), indicating that such misspecification led to worse fit than the correctly specified FMMs.

The present study

Through the extensive literature review above, it is apparent that previous studies on the performance of one-step and three-step approaches have largely focused on LCA and

GMM. A few methodological studies on FMM have focused on the performance of only one approach (i.e., three-step; Kim & Wang, 2018), or compared the two approaches but examined only class enumeration (Wang et al., 2020) or parameter recovery of factor mean difference under correctly specified FMM (Wang et al., 2021). Building upon these prior studies, the current study aims to comprehensively evaluate the efficacy of one-step and three-step approaches in FMM with regard to class enumeration and the recovery of covariate effects and factor mean difference in three scenarios, i.e., correct specification (study 1), misspecification (study 2), and overfitting (study 3), in terms of direct covariate effects on factors. This comprehensive evaluation of the approaches to covariate inclusion is warranted, given that all three scenarios could possibly occur in applied research as the population model is unknown.

Study 1

Method

Population model

The population model (see Fig. 1a) was a two-class FMM with three factors, with each factor measured by five multivariate normal items. The three factors had a correlation of .25 between them in each class. Measurement invariance held across classes, implying that item intercepts, factor loadings, and residual variances were identical for the two latent classes. The factor loadings of each factor were set at .70, .80, .70, .60, and .80 for the five items, respectively, and the residual variances of the five items were .51, .36, .51, .64, and .36 to obtain unit variance for each item. The factor mean for one class was set at zero, whereas the factor mean of the other class varied as the design factor of effect size (i.e., factor mean difference between classes), which will be explained in the next section. The variances of the factors were set at 1. The covariate was a normally distributed continuous variable with a mean of zero and a variance of 1.

Manipulated factors

The results of previous simulation studies about FMM indicated that the following four design factors had an impact on the performance of FMM in terms of class enumeration and parameter estimates: sample size, effect size, strength of covariate effect, and mixing proportions.

Sample size Previous research showed that sample size was an important factor in class enumeration in mixture

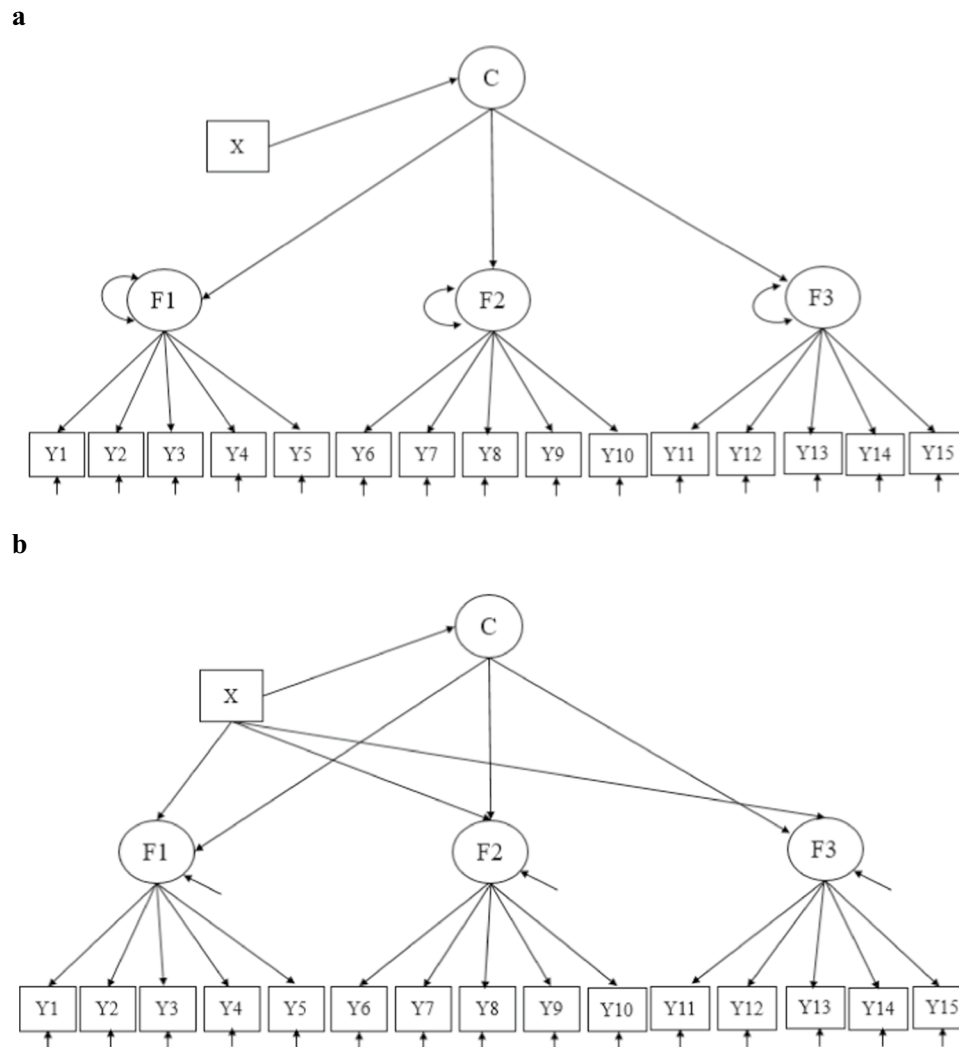


Fig. 1 **a** Population model for studies 1 and 3, **b** Population model for study 2. F1, F2, and F3 are three latent factors each measured by five items, Y1–Y5, Y6–Y10, and Y11–Y15, respectively. C is the latent class variable and X is a covariate

modeling (Li & Hser, 2011; Nylund et al., 2007; Wang et al., 2021). In this study, sample size was manipulated at four levels: 250, 500, 1000, and 2000. These sample sizes represent small, moderate, large, and very large samples in applied research studies of FMM.

Effect size The degree of class separation in FMM represented by the effect size of factor mean differences between the two latent classes also had an impact on the performance of FMM (Lubke & Muthén, 2007). Cohen's *d* measure was used to gauge the effect size of factor mean difference. The factor mean of the reference class was set at zero, and the factor mean of the other class was set at 1, 1.50, and 2 for all three factors, to represent small to large effect size of factor mean difference. These values of effect size accorded with empirical applications of FMM (e.g., Jensen, 2017; Piper et al., 2008; Rice et al., 2014).

Covariate effect In study 1, the covariate had an effect only on the class membership. The covariate effect on the logit of belonging to a specific class over the reference class was manipulated at three levels: 0, 0.50, and 2, corresponding to an odds ratio of 1, 1.65, and 7.39, respectively. These values were consistent with previous research into FMM with a covariate effect (e.g., Wang et al., 2020).

Mixing proportions Previous research in mixture modeling showed that it was easier to recover the true number of latent classes when the proportions of the latent classes were more balanced than dramatically different (e.g., Nylund et al., 2007). The mixing proportions of the two latent classes were set at either balanced (.50/.50) or unbalanced (.75/.25) (e.g., Nylund et al., 2007).

In addition to these conditions in the original design, we included some conditions that had higher factor correlations (.50) and lower factor loadings (.70, .55, .45, .60, and .55

for each of five items per factor) based on reviewers' suggestions, to increase the generalizability of findings. Higher factor correlations and lower loadings were fully crossed with sample sizes (250, 500, 1000, and 2000), effect sizes (1 and 2) and covariate effects (0, 0.50, and 2) for equal proportions. There was a total of 96 conditions ($4 \times 3 \times 3 \times 2$ original conditions and $4 \times 2 \times 3$ additional conditions). Two hundred replications of each condition were generated and analyzed using *Mplus* 8.4 (Muthén & Muthén, 1998–2017). The full simulation codes (across all three studies) can be found here https://osf.io/amupe/?view_only=f72fb1198c4947dbab731cebb5c416a3.

Analytical models

Each data set was analyzed using two approaches, i.e., one-step and three-step FMM (FMM-1S and FMM-3S hereinafter, respectively), correctly specifying the covariate effect on the latent class membership. These two models including one to three latent classes were run and systematically evaluated for all of the designed conditions in these studies. For the one-class model, the identification of FMM was similar to a common factor model where factor loading of the first item was fixed at 1 for each factor and factor means were fixed at zero, which is the default setting of *Mplus*. For two- and three-class models, factor loading of the first item was constrained to 1 for each factor across classes, and factor means of the last class were fixed at zero as the default identification of *Mplus*. In this study, we only allowed for class-specific factor means (except the last class) and imposed the equality constraint on all other parameters (i.e., factor loadings, intercepts, residual variances, factor variances/covariances, covariate mean/variance, and covariate effects) across classes. Thus, the specification of the analytical models was consistent with the population model, and there was no misspecification or overfitting that might contaminate findings of study 1.

The primary simulation outcome was correct class enumeration rate (i.e., the proportion of replications that correctly supported the two-class model). Model selection was based on the following information criteria (ICs)¹: Akaike information criterion (AIC; Akaike, 1974), BIC (Schwarz, 1978), and saBIC (Sclove, 1987). The model with the lowest value of the ICs among the three competing models (one- to three-class) was selected as the best-fitting model. The secondary outcome was parameter recovery including relative bias, type I error rates, and power of (1) covariate effect on latent class membership and (2) effect size, which were investigated for replications that had correct class enumeration.

¹ Additional approaches to class enumeration are available such as the Lo-Mendell-Rubin (LMR) test and adjusted LMR test (Lo et al., 2001), and bootstrap likelihood ratio test (BLRT; McLachlan & Peel, 2000). However, the former two were not examined given that previ-

Results

Figure 2 presents the correct class enumeration rates for FMM-1S and FMM-3S under equal proportions for the original set of conditions. Unequal proportions and the additional set of conditions with higher factor correlations and lower loadings resulted in lower correct enumeration rates for both FMM-1S and FMM-3S, but the relative performance of the two approaches remained the same. Therefore, results for unequal proportions and the post hoc conditions are not presented or discussed here, but can be found in the supplemental tables (see Tables S1–S4).

Overall, the relative performance of FMM-1S and FMM-3S depended upon effect size. When effect size was 1.00, FMM-1S clearly outperformed FMM-3S, with higher correct enumeration rates across ICs. As effect size increased, the discrepancy in correct enumeration rates between the two approaches decreased such that with an effect size of 2.00, the performance of FMM-3S was comparable to that of FMM-1S across ICs. The impact of the covariate effect strength was observed for FMM-1S, i.e., larger correct enumeration rates were observed with larger covariate effects, controlling for effect size. However, the strength of the covariate effect had a negligible impact on the performance of FMM-3S. Larger sample size was associated with higher correct enumeration rates across conditions and models. BIC did not perform as well as saBIC across most conditions, but larger effect size, stronger covariate effect, and larger sample size all helped improve its correct enumeration rates. When correct enumeration rates were low for BIC and saBIC, the one-class model was supported. Overall, AIC outperformed BIC and saBIC when the effect size was small and/or covariate effect was weak. However, the correct enumeration rates of AIC under these conditions were not yet satisfactory (below .70), as AIC tended to over-extract the number of classes.

Because of the comparable performance of FMM-1S and FMM-3S in class enumeration, parameter recovery (see Table 2) was only examined under a large effect size. Overall, the covariate effect was accurately estimated with large sample sizes (i.e., 1000 and 2000), but positive and negative bias occurred with small sample sizes (i.e., 250 and 500) for FMM-1S and FMM-3S, respectively. For both approaches,

Footnote 1 (continued)

ous simulation studies have shown that they do not perform as well as ICs (Henson et al., 2007; Nylund et al., 2007; Tein et al., 2013). BLRT was not considered due to the long execution time. Entropy was also not included due to the unreliable performance that has been documented in the literature (e.g., E. Kim et al., 2016; Tein et al., 2013), which was confirmed by an examination of class enumeration based on entropy for a subset of conditions (interested readers are referred to Tables S3 and S4). In addition, a major limitation with the use of entropy for class enumeration is that the one-class model cannot be compared with models with two or more classes in terms of fit.

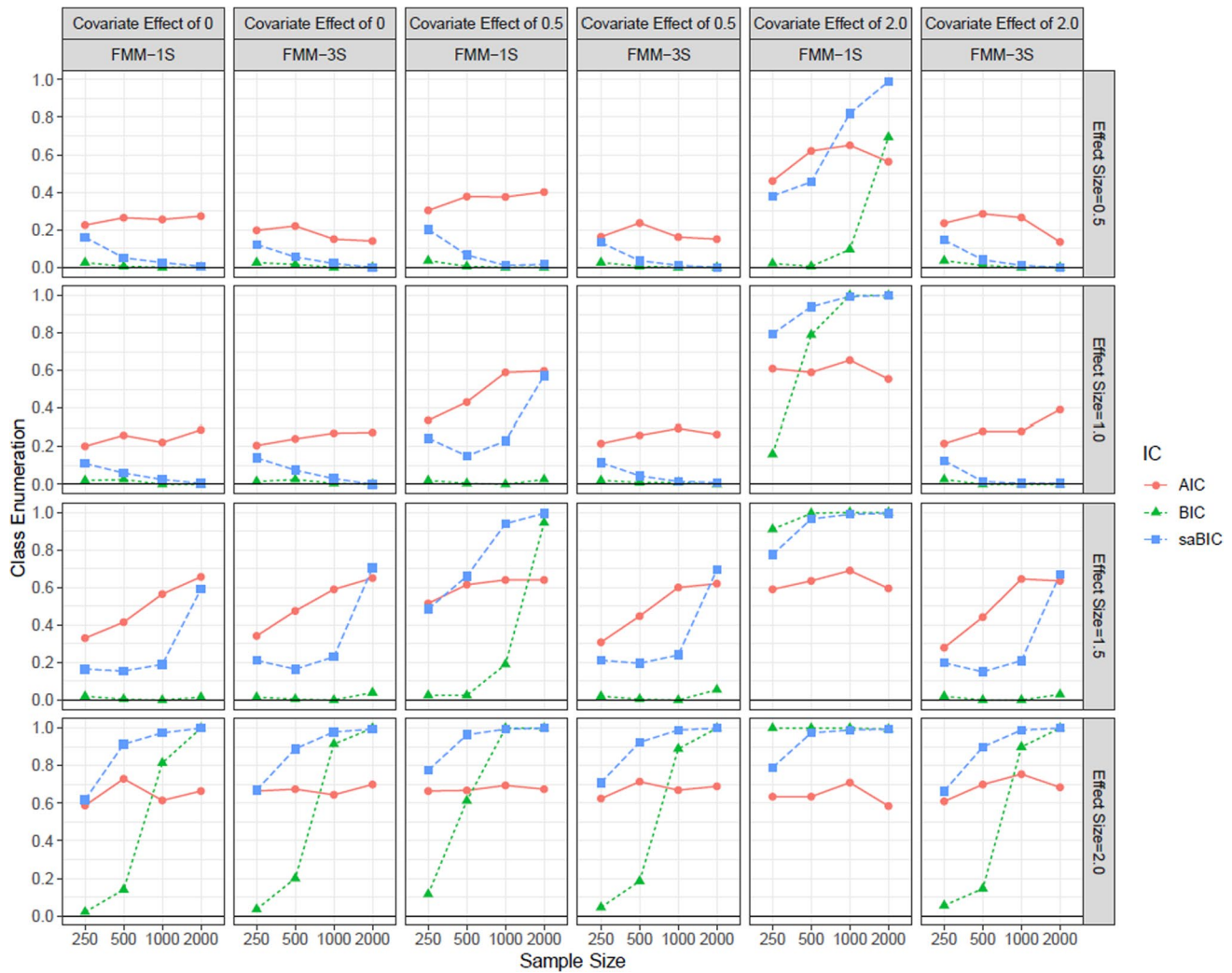


Fig. 2 Correct class enumeration rates of one-step and three-step FMMs (study 1). FMM-1S and FMM-3S refer to one-step and three-step FMM with covariate effect on class, respectively

type I error control was adequate and power in detecting covariate effects remained high across most conditions. Bias and power in detecting factor mean difference were also comparable between the two approaches. *To summarize, when only the covariate effect on the latent class variable was simulated and estimated (i.e., correct specification), the superiority of FMM-1S was evidenced in class enumeration with small effect size.*

Study 2

Method

Built on study 1, the simulation design was modified to address the research question of study 2: how one-step and three-step FMM perform when the covariate effect is misspecified (i.e., ignoring the covariate effect on

factor). To this end, the only modification to the population model is that the covariate had effects on both latent class membership and each of the three factors, as shown in Fig. 1b, whereas the population model in study 1 included covariate effect on the latent class membership only (Fig. 1a). Manipulated factors included effect size (1 and 2), sample size (250, 500, 1000, and 2000), covariate effect on latent class membership (0.50 and 2), and covariate effect on factor (0.20 and 0.60, which were selected to represent a small and large effect). Because the results of study 1 showed trivial differences in the relative performance for the one-step and three-step between equal and unequal proportions or higher and lower factor correlations/loadings, only the equal proportions conditions in the original set of conditions were included here (and in study 3). Study 2 had a total of 32 ($2 \times 4 \times 2 \times 2$) conditions, and 200 replications were generated for each condition.

Table 2 Parameter recovery of one-step and three-step FMMs under large effect size (study 1)

ES	COVC	N	Covariate effect				Effect size			
			FMM-1S		FMM-3S		FMM-1S		FMM-3S	
			Rebias	Type I/Power	Rebias	Type I/Power	Rebias	Power	Rebias	Power
2	0	250	.15	.06	.15	.05	.08	.98	.06	.97
		500	.10	.06	.10	.06	.04	.99	.03	1.00
		1000	.06	.03	.06	.04	.01	1.00	.01	1.00
		2000	.04	.04	.04	.02	.00	1.00	.00	1.00
	.5	250	.19	.88	-.01	.83	.06	.99	.07	.98
		500	.02	.98	-.04	.98	.02	1.00	.03	.99
		1000	.03	1.00	.01	1.00	.01	1.00	.01	1.00
		2000	-.01	1.00	-.02	1.00	.00	1.00	.00	1.00
	2	250	.09	.99	-.12	1.00	.03	1.00	.06	.98
		500	.01	1.00	-.06	1.00	.01	1.00	.03	.99
		1000	.01	1.00	-.02	1.00	.00	1.00	.00	1.00
		2000	.01	1.00	-.02	1.00	.00	1.00	.00	1.00

ES = effect size; COVC = covariate effect on class; Rebias = relative bias; Relative bias with absolute value $\leq .05$ are in bold

Each replication was analyzed using four models: one-step and three-step FMM with covariate effect on the latent class membership only (FMM-1S and FMM-3S, respectively), both of which were misspecified models; one-step FMM with covariate effects on both latent class membership and factor (FMM-CF-1S); and adjusted three-step FMM with covariate effect on factor in the first step (FMM-F-3S). The latter two models were correctly specified and thus could serve as a baseline with which the impact of omitting the covariate effect on factor could be evaluated. The identification and specification of FMM-1S and FMM-3S was the same as in study 1. For FMM-CF-1S and FMM-F-3S, class-invariant covariate effects on factors were specified and the rest of model specification remained the same as FMM-1S and FMM-F-3S, respectively. Note that instead of fixing factor means of the last class at zero for identification purposes, factor intercepts were constrained to zero for FMM-CF-1S and FMM-F-3S due to the paths from the covariate to factors. For class enumeration, AIC was not reported given its unsatisfactory performance in study 1 but can be found in supplemental Table S5. For parameter recovery, the relative bias and statistical power of the covariate effect on factor were also reported when the effect was estimated. The rest of the simulation outcomes remained the same as in study 1.

Results

As shown in Fig. 3, FMM-1S ignored the covariate effect on factors, and the impact of such model misspecification on class enumeration was mixed. Specifically, correct enumeration rates of saBIC were lower than those under FMM-CF-1S across most conditions. BIC was more robust to covariate

effect misspecification than saBIC, and in fact, substantially higher correct enumeration rates were observed for FMM-1S than FMM-CF-1S when the ignored effect on factors was 0.20 (small), except for a few conditions in which effect size was large and covariate effect on class was large. Larger sample size was associated with worse class enumeration, as FMM-1S tended to over-extract the number of classes. Correct enumeration rates were high (above .70) for FMM-CF-1S when class separation was large (i.e., large effect size and/or large covariate effect on class). Larger sample size also contributed to better class enumeration. FMM-3S failed to detect two classes across conditions and the one-class model was supported instead. FMM-F-3S took into account the covariate effect on factor in class enumeration but still failed to detect two classes when effect size was 1. As effect size increased to 2, it outperformed all other models including FMM-CF-1S when covariate effect on class was 0.50 (small). However, when covariate effect on class was 2 (large), it did not perform as well as FMM-CF-1S but performed better than FMM-3S.

Parameter recovery (see Table 3) was investigated for all replications that converged and had admissible solutions. FMM-3S and FMM-F-3S were excluded from this investigation given the unsatisfactory enumeration results. For FMM-CF-1S, covariate effects on the latent class variable and factors were severely overestimated and power remained high, especially for the latter. Effect size was underestimated and, not surprisingly, power was relatively low across conditions. However, for conditions that had high enumeration rates, all examined parameters had minimal bias (around or below 0.05) and power was sufficient (over .85). For FMM-1S, severe overestimation was observed for covariate effect

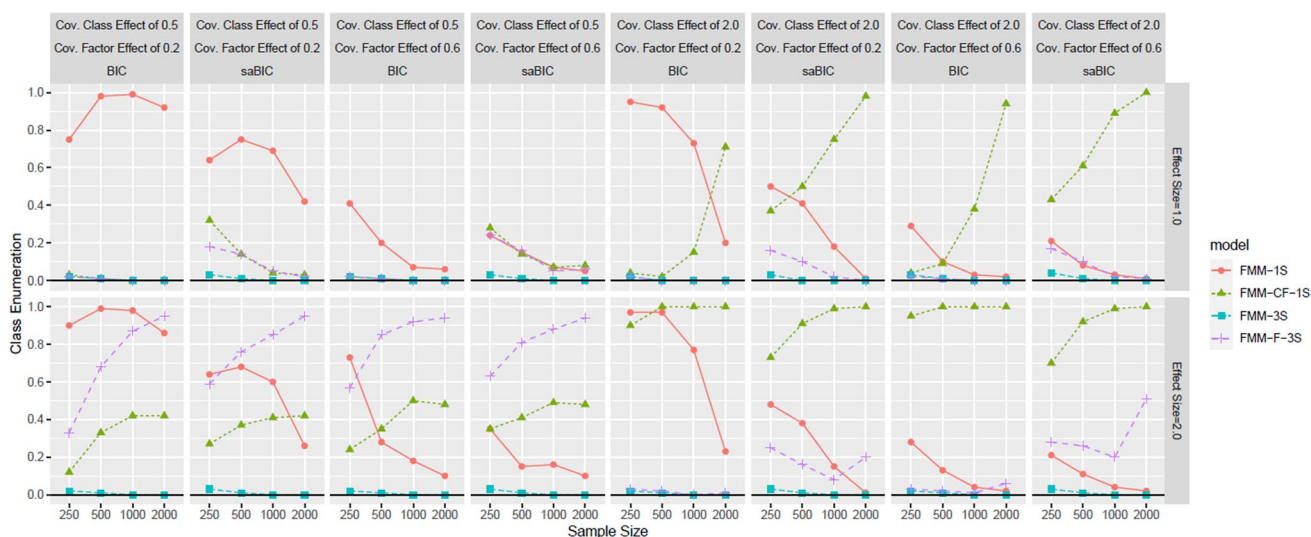


Fig. 3 Correct class enumeration rates of one-step and three-step FMMs (study 2). FMM-1S and FMM-3S refer to one-step and three-step FMM with covariate effect on class, respectively; FMM-CF-1S

is one-step FMM with covariate effects on class and factor; FMM-F-3S is the adjusted three-step FMM with covariate effect on factor in the first step

on class and effect size across most conditions regardless of class enumeration.

Taken together, FMM-3S performed poorly when omitting the covariate effect on factors; FMM-F-3S had improved yet unsatisfactory performance; FMM-1S was more robust to the misspecification than FMM-3S in class enumeration but parameter recovery was poor; and FMM-CF-1S, the correct model, can be recommended with the caveats that it requires large class separation and/or large sample size.

Study 3

Method

Whereas study 2 considered a scenario in which the covariate effect on factors was omitted, study 3 aimed at evaluating the performance of one-step and three-step FMMs when model overfitting occurred. Specifically, this overfitting refers to the estimation of covariate effects on factors when such effects do not exist in the population model. To this end, we reanalyzed a subset of conditions in study 1, including effect size (1 and 2), covariate effect on the latent class variable (0.50 and 2), and sample size (250, 500, 1000, and 2000), with these three factors fully crossed. Each of the 200 replications was analyzed using FMM-CF-1S and FMM-F-3S as in study 2. Both models overfitted the covariate effect because they estimated covariate effects on factors when only the covariate effect on the latent class variable was present in the population model. Study 3 included the

same outcomes as in studies 1 and 2, i.e., class enumeration (AIC reported in the supplemental Table S6) as well as relative bias and type I error rates/power of covariate effects and effect size estimates.

Results

Table 4 presents correct enumeration rates of FMM-CF-1S and FMM-F-3S when both models were overspecified. Note that correct enumeration rates for FMM-1S and FMM-3S in Fig. 1 can serve as a benchmark against which the impact of model overfitting can be evaluated. FMM-F-3S had high correct enumeration rates when the effect size was large and the covariate effect on the latent class variable was small. However, the overall performance of this model was worse than FMM-3S, which correctly specified the covariate effect. In contrast, FMM-CF-1S was more robust to the overfitting of covariate effects. It performed well and comparably to FMM-1S when class separation was large, i.e., large effect size and/or large covariate effect on the latent class variable. Larger sample size was associated with higher correct enumeration rates. Under conditions with smaller separation, FMM-CF-1S tended to under-extract classes.

Parameter recovery was investigated for FMM-CF-1S (see Table 4). FMM-F-3S was excluded from this analysis given its overall poor correct enumeration rates. FMM-CF-1S tended to overestimate covariate effects on class and factors but underestimate effect size, which is aligned with the finding in study 2 when this model was correctly

Table 3 Parameter recovery of one-step FMMs (study 2)

ES	COVC	COVF	N	COVC				ES				COVF	
				FMM-1S		FMM-CF-1S		FMM-1S		FMM-CF-1S		FMM-CF-1S	
				Rebias	Power	Rebias	Power	Rebias	Power	Rebias	Power	Rebias	Power
1	.5	.2	250	4.94	.72	8.79	.60	.12	.89	-.28	.45	.42	.71
			500	2.38	.85	4.05	.63	.09	.97	-.41	.44	.34	.76
			1000	2.15	.97	5.04	.69	.09	.99	-.52	.41	.32	.86
			2000	1.93	1.00	2.22	.57	.07	1.00	-.56	.38	.32	.91
		.6	250	7.00	.92	3.73	.57	.62	.99	-.41	.46	.02	.99
			500	7.34	.95	2.67	.57	.60	.99	-.50	.43	.02	.99
			1000	6.26	1.00	2.87	.61	.58	.99	-.55	.44	.01	.99
			2000	6.07	1.00	3.52	.58	.57	1.00	-.58	.41	.01	.99
	2	.2	250	.70	.95	.66	.65	.36	.99	-.13	.54	.55	.63
			500	.43	.98	.16	.81	.35	1.00	-.16	.64	.30	.65
			1000	.37	1.00	.07	.94	.35	1.00	-.04	.86	.09	.84
			2000	.43	1.00	.03	.99	.34	1.00	.01	.98	.00	.98
		.6	250	1.59	1.00	.73	.76	1.01	1.00	-.18	.59	.08	1.00
			500	1.46	.99	.14	.84	.98	1.00	-.12	.75	.04	.99
			1000	1.38	1.00	.05	.95	.98	1.00	-.02	.94	.01	1.00
			2000	1.31	1.00	.02	1.00	.98	1.00	.01	1.00	.00	1.00
2	.5	.2	250	1.02	.96	2.97	.60	-.01	.98	-.58	.47	.48	.71
			500	1.60	.99	1.70	.71	.02	1.00	-.52	.57	.39	.83
			1000	.61	1.00	1.56	.88	.03	1.00	-.47	.61	.33	.94
			2000	.57	.99	1.33	.82	.03	1.00	-.50	.58	.31	.98
		.6	250	3.84	.92	2.40	.69	-.01	.97	-.55	.54	-.05	.99
			500	3.22	.97	1.70	.73	-.01	1.00	-.53	.56	-.04	1.00
			1000	3.09	1.00	1.25	.89	-.01	1.00	-.42	.67	-.03	1.00
			2000	2.96	1.00	1.40	.82	-.02	1.00	-.47	.60	-.04	.99
	2	.2	250	.24	.99	.11	.98	.14	1.00	-.01	.98	.08	.56
			500	.19	1.00	.00	1.00	.12	1.00	.00	1.00	.02	.82
			1000	.19	1.00	.01	1.00	.12	1.00	.00	1.00	.02	.99
			2000	.18	1.00	.01	1.00	.12	1.00	.00	1.00	-.01	1.00
		.6	250	.67	.99	.10	.99	.40	.99	-.01	.98	.01	1.00
			500	.62	1.00	.00	1.00	.40	1.00	.00	1.00	.00	1.00
			1000	.60	1.00	.01	1.00	.40	1.00	.00	1.00	.00	1.00
			2000	.58	1.00	.01	1.00	.40	1.00	.00	1.00	.00	1.00

ES = effect size; COVC = covariate effect on class; COVF = covariate effect on factor; Rebias = relative bias. Relative bias with absolute value ≤ .05 are in bold

specified. When correct enumeration rates were high, all three parameters reached minimal bias. Of note, severe inflation of type I error rates was observed for covariate effects on factors, unless enumeration rates were high.

To sum up, overspecification of FMM by estimating covariate effects on factors when they were zero in the population would have a negative impact on class enumeration, but more so for the three-step approach. Large class separation and/or large sample size are needed to ensure adequate performance of FMM-CF-1S with regard to class enumeration and parameter recovery.

Discussion

To comprehensively evaluate the performance of one-step and three-step approaches in FMM, we conducted a series of simulation studies to examine class enumeration and parameter recovery under correct specification, misspecification, and overfitting concerning direct covariate effect(s) on factors. Major findings are summarized in Table 5 for each simulation study and discussed here with regard to two decisions in covariate inclusion: when to include covariates (one-step and three-step approaches) and how to specify

Table 4 Results of one-step and three-step FMMs with overfitting of covariate effects (study 3)

ES	COVC	N	FMM-CF-1S								FMM-F-3S	
			Class Enumeration		COVC		COVF		ES		Class Enumeration	
			BIC	saBIC	Rebias	Power	Bias	Type I	Rebias	Power	BIC	saBIC
1	.5	250	.03	.30	8.26	.58	.11	.36	-.24	.45	.02	.20
		500	.01	.12	2.96	.59	.09	.41	-.39	.44	.01	.11
		1000	.00	.03	4.65	.64	.09	.47	-.50	.42	.00	.04
		2000	.00	.03	3.54	.60	.09	.48	-.55	.39	.00	.02
	2	250	.04	.38	1.22	.67	.13	.41	-.12	.55	.02	.17
		500	.01	.48	.18	.79	.08	.30	-.16	.64	.00	.09
		1000	.13	.72	.10	.91	.02	.12	-.05	.86	.00	.02
		2000	.67	.98	.03	.99	.00	.07	.00	.98	.00	.00
2	.5	250	.15	.31	2.79	.62	.14	.39	-.52	.51	.28	.53
		500	.30	.34	1.72	.71	.13	.44	-.53	.56	.63	.77
		1000	.42	.41	2.86	.89	.11	.47	-.47	.62	.82	.81
		2000	.42	.42	1.27	.78	.10	.49	-.50	.58	.96	.96
	2	250	.86	.72	.13	.98	.02	.09	-.02	.97	.03	.23
		500	1.00	.91	.01	1.00	.01	.06	.00	1.00	.01	.15
		1000	1.00	.99	.01	1.00	.00	.05	.00	1.00	.00	.07
		2000	1.00	1.00	.01	1.00	.00	.05	.00	1.00	.01	.18

ES = effect size; COVC = covariate effect on class; BIC = Bayesian information criterion; saBIC = sample size-adjusted BIC; Rebias = relative bias. Correct enumeration rates $\geq .70$ and relative bias with absolute value $\leq .05$ are in bold

covariate effects (whether or not the covariate effect on factor should be estimated).

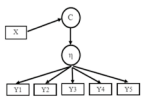
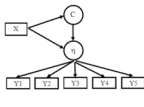
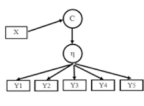
When correctly specified, one-step and three-step perform equally well if (1) a large effect size (i.e., 2) is coupled with sample size of 500 or above, or (2) a moderate effect size (1.50) is coupled with sample size of 2000. Otherwise, the one-step approach outperforms the three-step approach given more accurate class enumeration. The superior performance of the one-step approach is consistent with previous simulation studies on mixture models and can be explained by the benefit of including covariate effects on the latent class variable—enhancing class separation (Lubke & Muthén, 2007; Park & Yu, 2018; Wang et al., 2020; Wang et al., 2021).

Results also showed that the one-step FMM is more robust than the three-step approach to model misspecification or overfitting concerning the direct covariate effect on factor. In this study, the three-step FMM performed poorly, consistently under-extracting classes across conditions, which was contrary to the finding of previous studies under regression mixtures, LCA, and GMM that class enumeration was adequate (Diallo et al., 2017; Hu et al., 2017; M. Kim et al., 2016; Nylund-Gibson & Masyn, 2016). This gap might occur because the data generated in this scenario are quite complex (with two types of covariate effects), and without good covariates that can contribute to class separation, three-step FMM fails to accurately identify the

heterogeneity in the data. Class enumeration was improved for the adjusted three-step FMM that estimated the covariate effect on factor in the first step, but only when large effect size was coupled with small covariate effect on latent class. This approach also had bias across the examined parameters even when class enumeration was accurate, which was also found in Asparouhov and Muthén (2014) in the context of GMM when class separation was comparable between these two studies. However, they found no bias when class separation was greater (i.e., 4 and 6 standard deviations apart in the intercept factor mean or corresponding entropy values of .85 and .95).

When one-step FMM is adopted, the optimal specification of covariate effects depends on class separation and sample size. When class separation was small and sample size was small (i.e., 250 or 500 examined in this study), FMMs with covariate effects on the latent class variable only were shown to perform adequately in terms of class enumeration and parameter recovery. This was true even when the covariate effect was misspecified (covariate effects on the factors were ignored). Thus, it is recommended to avoid overfitting in this scenario. When class separation was small but sample size was large, misspecification became a more concerning issue, as it would lead to severe over-extraction of latent classes, whereas overfitting of covariate effects (estimating covariate effects on both the latent class variable and the factor when the latter did not exist in the population) could still lead to

Table 5 Summary of findings for studies 1, 2, and 3

Population models ^a	Analytical models	Class enumeration	Parameter recovery			Conclusions
			Covariate effect on class	Effect size	Covariate effect on factor	
<i>Study 1: Correct Specification of Covariate Effect</i>						
	FMM-1S	High correct enumeration rates especially with larger class separation and larger sample size	Overestimation with small sample size, adequate type I error control, high power	Bias only when small sample size coupled with small class separation	—	When class separation was large, FMM-1S and FMM-3S performed equally well. When class separation was small, FMM-1S was preferred.
	FMM-3S	Class enumeration comparable to FMM-1S when class separation was large; otherwise, lower correct enumeration rates	Underestimation with small sample size, adequate type I error control, high power	—	—	
<i>Study 2: Misspecification of covariate effect (omitting effect on factor)</i>						
	FMM-CF-1S	High correct enumeration rates especially with larger class separation and larger sample size	Accurate with high enumeration rates, overestimation in other conditions	Accurate with high enumeration rates, underestimation in other conditions	Accurate with large class separation	FMM-3S performed poorly. FMM-F-3S had improved performance. FMM-C-1S performed well when the ignored effect on factor was small. FMM-CF-1S was recommended.
	FMM-1S	High correct enumeration rates when the ignored covariate effect on factor was small	Overestimation	Overestimation	—	
	FMM-3S	Poor enumeration, under-extraction of classes	—	—	—	
	FMM-F-3S	High correct enumeration rates only when large effect size coupled with small covariate effect on class	—	—	—	
<i>Study 3: Overfitting of covariate effect (estimating a true zero effect on factor)</i>						
	FMM-CF-1S	High correct enumeration rates with large class separation, poor enumeration (under-extraction of classes) in other conditions	Accurate with high enumeration rates, overestimation in other conditions	Accurate with high enumeration rates, underestimation in other conditions	Accurate with high enumeration rates, overestimation in other conditions	FMM-CF-1S had better enumeration than FMM-F-3S.
	FMM-F-3S	High correct enumeration rates with large effect size and small covariate effect on class	—	—	—	

FMM-1S and FMM-3S = one-step and three-step FMM with covariate effect on class; FMM-CF-1S = one-step FMM with covariate effects on class and factor; FMM-F-3S = adjusted three-step FMM with covariate effect on factor in the first step. ^aFigures in this table are generic population models that represent the covariate effect on the latent class variable only (for studies 1 and 3), and covariate effects on both the latent class variable and factors (study 2)

adequate class enumeration and parameter recovery when the sample size was 1000 and over. Thus, in this case we would recommend fitting covariate effects on both the latent class variable and the factor.

When large class separation (defined as an effect size of 2.0 combined with a covariate effect of 2.0 on the latent class variable) is expected, FMMs with covariate effects on both the latent class variable and the factor can be fitted. When the covariate effect on the factor was truly present, this model outperformed others in terms of class enumeration and parameter recovery. When the covariate effect on the factor was zero in the population, this overspecified model yielded satisfactory class enumeration and parameter recovery under large class separation. Note that when class separation was not large, the overspecified model tended to under-extract the number of classes, underestimate effect size, and yield high type I error rates for the covariate effect on the factor. This might occur because the between-class differences in factor means are absorbed by the covariate effect on the factor as within-class variations in factor scores (Wang et al., 2020).

We provide a few additional recommendations that might help practitioners conduct FMM analyses. First, saBIC is more reliable than BIC in class enumeration when class separation is small and/or sample size is small (1000 or below), which is aligned with the methodological literature (e.g., E. Kim et al., 2016; Wang et al., 2020). Otherwise, the performance of the two indices is comparable. When class separation is small, AIC can be used in conjunction with saBIC for model selection, but generally speaking, AIC is not recommended due to the tendency to over-extract the number of classes (Cho & Cohen, 2010; Henson et al., 2007; Nylund et al., 2007; Wang et al., 2020). Second, large sample size will benefit class enumeration and parameter recovery. While a sample size of 500 seems to be the minimum for FMM, sample size of 1000 or more is needed in particular when more complex covariate effects are estimated (effects on both class and factor). Interested readers can refer to Wang et al. (2021) for additional guidelines on sample size requirements for FMM. Third, since the inclusion of good predictors of latent class membership would improve class separation and thus class enumeration, it is important to

identify potentially strong predictors of latent class membership, which might be obtained or approximated by consulting substantive theories or relevant literature. Note that we observed minimal contributions of a small covariate effect on class (defined as .50 in this study or odds ratio of 1.65) to class enumeration and parameter recovery when both covariate effects on class and factor were present. Therefore, we would recommend the search for covariates that potentially have a stronger effect than .50. Lastly, applied researchers are strongly advised to investigate the psychometric properties of the scale prior to conducting FMM analyses. Specifically, the additional set of simulation we conducted showed that lower factor loadings and higher factor correlations would worsen the performance of both one-step and three-step approaches. Thus, high factor loadings and reasonable (i.e., not too high) factor correlations will be desirable for the subsequent FMM analyses.

Limitations and conclusions

Although this series of simulation studies provided a relatively comprehensive evaluation of one-step and three-step approaches in FMM, generalization of findings beyond the set of conditions examined in the current study should be exercised with caution. Specifically, this study only considered one covariate, whereas in applied research, multiple covariates might be available and considered to be included in FMM. We expect that the inclusion of good covariates as in the one-step approach would benefit class enumeration regardless of the number of covariates; however, it remains unknown whether the suggested specification of covariate effects (i.e., on both class and factor) would be tenable when multiple covariates are present. The covariate effects are much more complex than the scenarios examined in this study. For instance, when one covariate has an effect on both class and factor, another covariate might only impact latent class membership or have no impact on FMM parameters at all. Considering the complexity of these scenarios with multiple covariate effects, additional research is needed to further investigate the specification of covariate effects, including but not limited to the impact of misspecification or overfitting.

Another important direction for future research is to comprehensively evaluate and examine covariate inclusion approaches in FMM when there are direct covariate effects on items. This study focused on direct covariate effects on factors, but direct covariate effects on items have been discussed in FMM and the general mixture modeling framework (De Ayala et al., 2002; Lee & Beretvas, 2014; Lubke & Muthén, 2005; Masyn, 2017; Tay et al., 2011; Vermunt & Magidson, 2021). In particular, Vermunt and Magidson (2021) proposed a modified three-step approach in LCA to account for direct covariate effects on indicators

which conceptually indicate measurement noninvariance with respect to covariates. In their proposed step-one analysis, covariates with direct effects on indicators should be included as well as the effects of these covariates on the latent class variable. The step-three analysis allows the classification error correction matrix to differ across categories of covariates that have direct effects on indicators. It remains unknown how this proposed approach would perform in FMM relative to the one-step and three-step approaches considered in this study.

Although we recommend the one-step approach over the three-step, we are aware of the criticism that the formation of latent classes and class assignment is model-dependent. When different covariates are included, latent class solutions are subject to change. Therefore, it is important to identify proper covariates that are well grounded in substantive theories so that latent class solutions are meaningful and interpretable.

Despite the limitations, we believe the study provides insightful information for practitioners in conducting FMM. Although the performance of three-step FMM is comparable to that of one-step FMM when class separation is large and model specification is correct, one-step FMM is the preferred approach to covariate inclusion, as the inclusion of good covariates can benefit class enumeration and the model is more robust to misspecification or overfitting with regard to covariate effects. We suggest fitting FMM with covariate effects on the latent class variable and the factor. We also highlight the importance of identifying theoretically grounded proper covariates and obtaining a large sample size (1000 or more), as well as the reliable performance of saBIC in class enumeration.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-022-01964-8>.

Authors' contributions Yan Wang designed the study and worked on the simulation codes. All three authors ran simulations and analyzed data. Yan Wang led the writing of the manuscript, and Chunhua Cao and Eunsook Kim both contributed to the writing.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The data sets generated and/or analyzed during the current study are not publicly available due to the large volume of data in Monte Carlo simulations (i.e., 200 replications per condition), but are available from the corresponding author on reasonable request.

Code availability Example codes for generating and analyzing data are available on the Open Science Framework (https://osf.io/amupe/?view_only=9cfd249a13b4ae5b9031f713acc1ede).

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflicts of interest/Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Allan, N. P., MacPherson, L., Young, K. C., Lejuez, C. W., & Schmidt, N. B. (2014). Examining the latent structure of anxiety sensitivity in adolescents using factor mixture modeling. *Psychological Assessment*, 26(3), 741–751. <https://doi.org/10.1037/a0036744>
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 329–341. <https://doi.org/10.1080/10705511.2014.915181>
- Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias adjusted three-step approaches. *Sociological Methodology*, 43(1), 272–311. <https://doi.org/10.1177/0081175012470644>
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757–786. <https://doi.org/10.1080/00273170701710338>
- Bernstein, A., Stickle, T. R., & Schmidt, N. B. (2013). Factor mixture model of anxiety sensitivity and anxiety psychopathology vulnerability. *Journal of Affective Disorders*, 149(1-3), 406–417. <https://doi.org/10.1016/j.jad.2012.11.024>
- Cetin-Berber, D. D., & Leite, W. L. (2018). A comparison of one-step and three-step approaches for including covariates in the shared parameter growth mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 588–599. <https://doi.org/10.1080/10705511.2018.1428806>
- Cho, S., & Cohen, A. S. (2010). A multilevel mixture IRT model with an application to DIF. *Journal of Educational and Behavioral Statistics*, 35(3), 336–370. <https://doi.org/10.3102/1076998609353111>
- De Ayala, R. J., Kim, S.-H., Stapleton, L. M., & Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing*, 2(3-4), 243–276. <https://doi.org/10.1080/15305058.2002.9669495>
- Diallo, T. M. O., & Lu, H. (2017). On the application of the three-step approach to growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5), 714–732. <https://doi.org/10.1080/10705511.2017.1322516>
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, 22(1), 166–190. <https://doi.org/10.1037/met0000084>
- Dimitrov, D. M., Al-Saud, F. A. A.-M., & Alsadaawi, A. S. (2015). Investigating population heterogeneity and interaction effects of covariates: The case of a large-scale assessment for teacher licensure in Saudi Arabia. *Journal of Psychoeducational Assessment*, 33(7), 674–686. <https://doi.org/10.1177/0734282914562121>
- Elhai, J. D., Naifeh, J. A., Forbes, D., Ractliffe, K. C., & Tamburrino, M. (2011). Heterogeneity in clinical presentations of posttraumatic stress disorder among medical patients: Testing factor structure variation using factor mixture modeling. *Journal of Traumatic Stress*, 24(4), 435–443. <https://doi.org/10.1002/jts.20653>
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(2), 202–226. <https://doi.org/10.1080/10705510709336744>
- Hu, J., Leite, W. L., & Gao, M. (2017). An evaluation of the use of covariates to assist in class enumeration in linear growth mixture modeling. *Behavior Research Methods*, 49(3), 1179–1190. <https://doi.org/10.3758/s13428-016-0778-1>
- Jensen, T. M. (2017). Constellations of dyadic relationship quality in stepfamilies: A factor mixture model. *Journal of Family Psychology*, 31(8), 1051–1062. <https://doi.org/10.1037/fam0000355>
- Kim, M., Vermunt, J., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4), 601–614. <https://doi.org/10.1080/10705511.2016.1158655>
- Kim, E. S., Joo, S.-H., Lee, P., Wang, Y., & Stark, S. (2016). Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 870–887. <https://doi.org/10.1080/10705511.2016.1196108>
- Kim, E. S., & Wang, Y. (2018). Investigating sources of heterogeneity with 3-step multilevel factor mixture modeling: Beyond testing measurement invariance in cross-national studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(2), 165–181. <https://doi.org/10.1080/10705511.2018.1521284>
- Lee, H., & Beretvas, S. N. (2014). Evaluation of two types of differential item functioning in factor mixture models with binary outcomes. *Educational and Psychological Measurement*, 74(5), 831–858. <https://doi.org/10.1177/0013164414526881>
- Li, M., & Harring, J. R. (2017). Investigating approaches to estimating covariate effects in growth mixture modeling: A simulation study. *Educational and Psychological Measurement*, 77(5), 766–791. <https://doi.org/10.1177/0013164416653789>
- Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, 46(2), 266–302. <https://doi.org/10.1080/00273171.2011.556549>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778. <https://doi.org/10.1093/biomet/88.3.767>
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21–39. <https://doi.org/10.1037/1082-989X.10.1.21>
- Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(1), 26–47. <https://doi.org/10.1080/10705510709336735>
- Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2), 180–197. <https://doi.org/10.1080/10705511.2016.1254049>
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: Wiley.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- No, U., & Hong, S. (2018). A comparison of mixture modeling approaches in latent class models with external variables under small samples. *Educational and Psychological Measurement*, 78(6), 925–951. <https://doi.org/10.1177/0013164417726828>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>

- Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 782–797. <https://doi.org/10.1080/10705511.2016.1221313>
- Park, J., & Yu, H.-T. (2018). A comparison of approaches for estimating covariate effects in nonparametric multilevel latent class models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 778–790. <https://doi.org/10.1080/10705511.2018.1448711>
- Piper, M. E., Bolt, D. M., Kim, S.-Y., Japuntich, S. J., Smith, S. S., Niederdeppe, J., Cannon, D. S., & Baker, T. B. (2008). Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives. *Journal of Abnormal Psychology*, 117(4), 747–761. <https://doi.org/10.1037/a0013298>
- Rice, K. G., Richardson, C. M. E., & Tueller, S. (2014). The short form of the Revised Almost Perfect Scale. *Journal of Personality Assessment*, 96(3), 368–379. <https://doi.org/10.1080/00223891.2013.838172>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3), 333–343. <https://doi.org/10.1007/BF02294360>
- Stegmann, G., & Grimm, K. J. (2018). A new perspective on the effects of covariates in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 167–178. <https://doi.org/10.1080/10705511.2017.1318070>
- Tay, L., Newman, D. A., & Vermunt, J. K. (2011). Using mixed-measurement item response theory with covariates (MM-IRT-C) to ascertain observed and unobserved measurement equivalence. *Organizational Research Methods*, 14(1), 147–176. <https://doi.org/10.1177/1094428110366037>
- Tein, J., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 640–657. <https://doi.org/10.1080/10705511.2013.824781>
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469. <https://doi.org/10.1093/pan/mpq025>
- Vermunt, J. K., & Magidson, J. (2021). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(3), 356–364. <https://doi.org/10.1080/10705511.2020.1818084>
- Wang, Y., Hsu, H.-Y., & Kim, E. (2021). Investigating the impact of covariate inclusion on sample size requirements of factor mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*. Advance online publication. <https://doi.org/10.1080/10705511.2021.1910036>
- Wang, Y., Kim, E., Ferron, J. M., Dedrick, R. F., Tan, T. X., & Stark, S. (2020). Testing measurement invariance across unobserved groups: The role of covariates in factor mixture modeling. *Educational and Psychological Measurement*, 81(1), 61–89. <https://doi.org/10.1177/0013164420925122>

Open practices statement Supplemental tables and simulation codes for data generation and analysis are available on the Open Science Framework (https://osf.io/amupe/?view_only=f72fb1198c4947dbab731cebb5c416a3).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.