Covariance Pattern Mixture Models:
Eliminating Random Effects to Improve Convergence and Performance

Daniel McNeish
Arizona State University

&

Jeffrey Harring
University of Maryland

**Funding**

Both authors were supported by a research grant from the Institute for Educational Sciences, 305D190011. Any opinions, findings, or conclusions are those of the authors and do not necessarily reflect the views of the supporting agency.

**Date of Publication**

The peer-reviewed journal version of the paper appeared online first at *Behavior Research Methods* on September 11, 2019. This version of the paper can be found at https://link.springer.com/article/10.3758/s13428-019-01292-4

**Contact Information**

Daniel McNeish, Department of Psychology, PO Box 871104, Tempe, AZ, USA 85287. Email: dmcneish@asu.edu.

## Abstract

Growth mixture models (GMMs) are prevalent for modeling unknown population heterogeneity via distinct latent classes. However, GMMs are riddled with convergence issues, often requiring researchers to atheoretically alter the model with cross-class constraints to obtain convergence. We discuss how within-class random effects in GMMs exacerbate convergence issues even though these random effects rarely help to answer typical research questions. That is, latent classes provide a discretization of continuous random effects, so including additional random effects within latent classes can unnecessarily complicate the model. These random effects are commonly included to properly specify the marginal covariance; however, random effects are inefficient for patterning a covariance matrix, resulting in estimation issues. Such a goal can be achieved more simply covariance pattern models, which we extend to the mixture model context in this paper (covariance pattern mixture models, CPMMs). We provide evidence from theory, simulation, and an empirical example showing that employing CPMMs (even if misspecified) instead of GMMs can circumvent computational difficulties that can plague GMMs without sacrificing the ability to answer the type of questions commonly asked in empirical studies. Results show advantages of CPMMs with respect to improved class enumeration, and less biased class-specific growth trajectories in addition to vastly improved convergence rates. Results also show that constraining covariance parameters across classes to bypass convergence issues with GMMs leads to poor results. An extensive software appendix is included to assist researchers run CPMMs in M*plus*.

# Covariance Pattern Mixture Models: Eliminating Random Effects to Improve Convergence and Performance

In longitudinal data analysis, mixture models are commonplace in the empirical literature where the primary goal is to identify unobserved, latent classes of growth trajectories (Jung & Wickrama, 2007). As a hypothetical example, researchers may follow students' test scores over time and wish to identify which students in the sample are "on-pace" learners, "accelerated" learners or "slow" learners (e.g., Musu-Gillette, Wigfield, Harring & Eccles, 2015). These subgroups are latent and are not identified a priori as observed variables like other independent variables that may be of interest (e.g., gender, SES, treatment condition). Instead, their existence must be inferred from characteristics of the growth patterns themselves.

Common goals of a mixture analysis in this longitudinal context are to identify how many classes exist, to which of these latent classes an individual belongs, and to estimate the growth trajectory of each of the latent classes. These interests parallel how a researcher may want to estimate different growth trajectories for different levels of an observed variable. For instance, a researcher may wish to separately model growth trajectories for students identified as coming from households with high and low SES (i.e., whether SES moderates growth curves). The major difference in the mixture model context is that the classification variable of interest happens to be latent. When adding latent classes to growth models, two methods are common: latent class growth models and growth mixture models.

Latent class growth models (LCGMs; Nagin, 1999; Nagin & Tremblay, 2001; Nagin, 2005) use trajectory groups to approximate a complex, non-normal underlying distribution (Nagin, 2005). To explain, a typical growth model incorporates continuous random effects for each individual which are typically assumed to be normally distributed in order to create a unique subject-specific growth curve for each individual in the data. For continuous outcomes, LCGMs discretize this continuous distribution by estimating a handful of classes, each with a unique mean trajectory. Individuals are

assigned to the class trajectory that most closely represents what their subject-specific growth curve would have been in a standard growth model with continuous random effects. This discretization of the random effects facilitates interpretation because the dimensionality is reduced from the number of individuals (where each individual has a unique growth curve) to a small number of easily interpretable representative trajectories. LCGMs do not allow for subject-specific growth trajectories within classes, meaning that any deviation from the class trajectory is assigned to an error term that is assumed to be independently and identically distributed across time with common variance. Though conceptually appealing, a noted downside of LCGMs is that the number of classes tends to be overextracted, especially at larger sample sizes (i.e., extracted classes are not necessarily substantively different from one another). The specification of the model is not always flexible enough to properly model the marginal covariances among the repeated measures, so covariance misspecification manifests as additional latent classes (e.g., Kreuter & Muthén, 2008).

Growth mixture models (GMMs) represent another method by which to model latent classes with longitudinal data (Muthén, 2001; Muthén & Shedden, 1999). Similar to LCGMs, GMMs estimate latent classes, each with a unique mean growth trajectory. Unlike LCGMs, GMMs also specify random effects of the growth factors *within each class* to allow between-individual and within-individual variability within classes. The benefit of such an approach is that the model more easily accommodates proper specification of the covariance of the repeated measures. As a result, the model is less likely to extract the spurious classes as a consequence of a misspecified covariance structure.

Though effective for accurately capturing the marginal covariance of the repeated measures, GMMs' reliance on continuous within-class random effects can be inefficient. As we discuss in more detail in subsequent sections, GMMs are computationally demanding and as a result, routinely encounter estimation issues. In psychology-adjacent fields that are interested in modeling change over time such as epidemiology and public health, population-averaged models (a.k.a. marginal

models; Liang & Zeger, 1986) are popular alternatives to random effects models for longitudinal data because they can appropriately model the marginal covariance while requiring fewer assumptions and are much less demanding to estimate (Burton, Gurrin, & Sly, 1998; Harring & Blozis, 2016). To date, there has been no coverage in the literature extolling the advantages of and applying population-average models within a mixture model framework. This is precisely the goal of this manuscript.

Specifically, we will use the existing literature, statistical theory, Monte Carlo simulation, and an empirical example to argue that most research questions being addressed with GMMs do not require the within-class random effects. That is, researchers are primarily interested in the discretization of individual growth-curves provided by the latent classes — individual deviation within each latent class is rarely a research focus and the use of random effects primarily serves the secondary function of properly specifying the marginal covariance between repeated measures. A primary aim of this paper is to show that this objective can be satisfied in a less computationally demanding fashion with population-averaged models, which ultimately yields better convergence rates, reduced need for atheoretical model constraints, and better statistical properties of model estimates.

To outline the structure of this manuscript, we first overview the generic latent growth model and demonstrate how it naturally extends to GMMs. We discuss how the random effects approach makes estimation more demanding, which can augment computational difficulties. We then provide evidence from the post-traumatic stress literature – where mixture models frequently appear – to demonstrate that research questions rarely make use of the information provided by within-class random effects featured in GMMs.. Population-averaged models are then overviewed with specific focus on covariance pattern models and the advantages they provide in the context of mixture models for longitudinal data. We provide a Monte Carlo simulation study to highlight how covariance pattern mixture models can address issues that tend to plague applications of GMMs. Specifically,

we explore convergence rate, class trajectory bias, classification accuracy, and class enumeration. We then provide an empirical example to compare and contrast the traditional GMM with our proposed population-averaged approach. A detailed appendix of annotated M*plus* code is also provided to facilitate the use of these models by empirical researchers.

## Overview of LGMs and GMMs

### The Latent Growth Model

First, consider a traditional LGM, which can be thought of as a special case of a GMM with only one class. The general linear latent growth model with *q* time-invariant covariates can be written as a restricted confirmatory factor analysis model with a structured mean vector of the observed variables such that

$$\mathbf{y}_i = \mathbf{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

and

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \,. \tag{2}$$

In Equation 1, $\mathbf{y}_i$ is a $n_i \times 1$ vector of responses where $n_i$ is the number of observations for individual *i*, $\mathbf{\Lambda}_i$ is a $n_i \times q$ matrix of loadings for *q* the number of growth factors where the loadings are commonly, but not always, pre-specified to fit a specific type of growth trajectory,[1] $\boldsymbol{\eta}_i$ is a $q \times 1$ vector of individual-specific growth factor scores for individual *i*, and $\boldsymbol{\varepsilon}_i$ is a $n_i \times 1$ vector of time specific residuals where $\boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \mathbf{\Theta}_i)$, and where $\mathbf{\Theta}_i$ depends on *i* only through its dimension, although this assumption can be relaxed (Davidian & Giltinan, 1995). In Equation 2, the individual-specific growth factor scores are equal to a $q \times 1$ vector of factor means $\boldsymbol{\alpha}$, a $q \times p$ matrix of time-

[1] Note that the $\mathbf{\Lambda}$ matrix has an *i* subscript and by implication would assume that the loading matrix can be person-specific. This is contrary to the typical specification when modeling growth as a multivariate system in the SEM framework, which is less adept at handling time-unstructured data (McNeish & Matta, 2018). However, including the *i* subscript is the most general form of the model because there are methods to handle time unstructured data and the dimension of $\mathbf{\Lambda}$ can be person-specific even with time-structured data in the common context of missing data (Codd & Cudeck, 2014). Methods to handle time-unstructured data in LGM framework have been developed (e.g., Mehta & West, 2000), but, to our knowledge, this issue has not been widely investigated in the context of mixture modeling.

invariant coefficients in $\boldsymbol{\Gamma}$ for *p*, the number of time-invariant covariates, a $p \times 1$ vector of time-invariant covariate values $\mathbf{x}_i$, and a $q \times 1$ vector of random effects, $\zeta_i \sim MVN(\mathbf{0}, \boldsymbol{\Psi})$.

The model-implied mean and covariance structures of the repeated measures are thus,

$$E[\mathbf{y}_i] = \boldsymbol{\mu}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\alpha} + \boldsymbol{\Gamma\kappa}) \tag{3}$$

and

$$\text{var}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma\Phi\Gamma}^{\text{T}} + \boldsymbol{\Psi})\boldsymbol{\Lambda}_i^{\text{T}} + \boldsymbol{\Theta}_i, \tag{4}$$

where $\boldsymbol{\kappa}$ is the vector of covariate means and $\boldsymbol{\Phi}$ is the covariance matrix of the time-invariant covariates.

**Multiple group growth models**. Populations are often heterogeneous and different segments of the population may follow different growth trajectories. If the heterogeneity is the result of known group membership, parameters can be estimated separately for each group with what is referred to as a *multiple group model* (e.g., Muthén & Curran, 1997). Notationally, the vectors and matrices of Equations 1 through 4 would take a *g* subscript (where *g* = 1, …, *G*) to denote to which observed group the parameters belong.

Although the conceptual idea is alluring, multiple group models require that the grouping variable be an observed variable in the data, which is not frequently the case with heterogeneous populations (e.g., it is hard to objectively assign “fast” learner or “slow” learner labels to individuals). Group membership is more often latent and not known a priori (Nylund-Gibson, Grimm, Quirk, & Furlong, 2014), leading to the use of mixture models to sort individuals via latent classes.

**Growth Mixture Model**

GMMs are a generalization of the multiple group framework where group membership is unobserved (Muthén, 2001; Muthén & Shedden, 1999; Nagin, 1999; Verbeke & Lesaffre, 1996). Instead of a known value for group membership, each observation receives a probability of

membership in each of the estimated latent classes. Assuming multivariate normality, the composite density of a vector of continuous outcome variables for the *i*th individual, $\mathbf{y}_i$, can be written as

$$f(\mathbf{y}_i \mid \boldsymbol{\varphi}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \sum_{k=1}^{K} \varphi_k f_k(\mathbf{y}_i \mid \boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}), \tag{5}$$

where *K* is the number of latent classes that the researcher specifies, $f_k$ is the component density for the *k*th class, $\boldsymbol{\mu}_{ik}$ is the model-implied mean vector for the *k*th class, $\boldsymbol{\Sigma}_{ik}$ is the model-implied covariance matrix for the *k*th class, and $\varphi_k$ is the mixing proportion for the *k*th class where $0 \leq \varphi_k \leq 1$ and $\varphi_K = 1 - \sum_{k=1}^{K-1} \varphi_k$. Given class *k*, Equations 1 and 2 can be extended to accommodate the inclusion of latent classes as

$$\mathbf{y}_i = \boldsymbol{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \tag{6}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \mathbf{x}_i + \boldsymbol{\zeta}_i, \tag{7}$$

with the assumption that the residuals and random effects for the *i*th individual follow separate multivariate distributions, $\boldsymbol{\varepsilon}_i \mid k \sim MVN(\mathbf{0}, \boldsymbol{\Theta}_{ik})$, and $\boldsymbol{\zeta}_i \mid k \sim MVN(\mathbf{0}, \boldsymbol{\Psi}_k)$. Following the notation from Equations 3 and 4, the model-implied mean vector and model-implied covariance matrix from Equations 6 and 7 can be written as

$$\boldsymbol{\mu}_i \mid k = \boldsymbol{\Lambda}_i (\boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \boldsymbol{\kappa}) \tag{8}$$

$$\boldsymbol{\Sigma}_i \mid k = \boldsymbol{\Lambda}_i (\boldsymbol{\Gamma}_k \boldsymbol{\Phi} \boldsymbol{\Gamma}_k^{\mathrm{T}} + \boldsymbol{\Psi}_k) \boldsymbol{\Lambda}_i^{\mathrm{T}} + \boldsymbol{\Theta}_{ik}. \tag{9}$$

More conceptually, GMMs add a discrete latent variable with a specific number of categories to a traditional LGM. This discrete latent variable then serves as a moderator for the whole model, allowing parameter estimates to differ for the different categories of the discrete latent variable.

## Within-Class Variation

Though GMMs have the advantage that they summarize the covariances among the repeated measures in a more realistic manner (based on Equation 9) compared to LCGMs, this richer model specification containing (latent) within-class random effects can be difficult to estimate on top of the latent classes. That is, variances of growth factor random effects are difficult to estimate in any growth model, so a model that requires unique growth factor variances and time-specific residual variances within each *class* quickly becomes challenging. Indeed, the frequency of inadmissible or non-converging solutions is notable when the variance parameters are uniquely estimated for each class and the frequency rapidly increases as the complexity of the model increases (Diallo et al., 2016; Liu & Hancock, 2014).

A common method by which to work around estimation issues encountered with estimating unique covariance parameters in each class is to constrain covariance parameters to be equal across classes (i.e., $\boldsymbol{\Theta}_{ik} = \boldsymbol{\Theta}_i$; and $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}$ for all $k$ in Equation 9; see, e.g., Petras & Masyn, 2010). In fact, this constraint is applied by default in M*plus* and must be actively overridden (Diallo, Morin, & Lu, 2016). This simplifies Equation 9 by removing the $k$ subscript on $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, and/or $\boldsymbol{\Theta}$.

More conceptually, if the growth factor variances and covariances are difficult to estimate, applying constraints so that the model features fewer of these parameters will simplify the estimation. The rationale for this decision is often rooted in a desire to reduce the complexity of the estimation rather than for substantive reasons (Bauer & Curran, 2003; Enders & Tofighi, 2008; Gilthorope et al., 2014; Harring & Hodis, 2016, Infurna & Grimm, 2017; Infurna & Luthar, 2016; van de Schoot, Sijbrandij, Winter, Depaoli, & Vermunt, 2017).

Though commonly implemented in empirical settings (Infurna & Grimm, 2017), the approach of constraining variance terms across classes has been widely criticized in the methodological literature. The main reason being that the rationale behind this modeling decision is to aid estimation

rather than because theory posits that each latent class actually has equal variance(s). Bauer and Curran (2003) explicitly question the choice to apply constraints across classes by stating,

> Although [covariance equality constraints] are statistically expedient, we do not regard these equality constraints as optimal from a theoretical standpoint, and in our experience, they are rarely found to be tenable in practice. Indeed, implementing these constraints is in some ways inconsistent with the spirit of the analysis, because one is forcing the majority of the parameter estimates to be the same over classes (permitting only mean differences in the within-class trajectories) (p. 346).

Furthermore, recent methodological studies by Diallo et al. (2016) and Morin et al. (2011) have demonstrated that implementing covariance equality constraints directly impacts class enumeration because the estimation attempts to classify individuals who best mirror the within-class growth characteristics defined by the model. Thus, individuals must necessarily vary around the within-class mean trajectories in equal amounts while holding the amount of variability within individuals across time to be equal as well. This has the effect of adding an additional homogeneity of variance assumption into the model that is questionably tenable and likely false in many applications.

The critical take-home message is that the individuals assigned to each latent class, the number of enumerated classes, and the trajectories of the classes are all impacted by the constraining covariance parameters across classes. However, this choice is frequently based on whether the model converges rather than on criteria related to the theory being tested. With such rampant frequency of convergence issues and the current solution of cross-class constraints known to cause so many issues, a natural question that emerges is whether the complexity of GMMs is necessary to answer questions being asked by researchers or whether a simpler model may suffice. The next section reviews a segment of the psychological literature that frequently uses GMMs to explore whether researchers' questions necessitate a model as complex as GMMs.

## Do Mixture Model Research Questions Require Random Effects?

Despite GMMs often being described as a person-oriented or person-centered modeling approach (Bergman & Magnusson, 1997; Laursen & Hoff, 2006; Muthén & Muthén, 2000), the most common interest in empirical studies applying GMMs pertains to classes, not individuals (Cole & Bauer, 2016; Sterba & Bauer, 2010, 2014). To provide evidence that researchers are employing

GMMs with random effects when the research questions of interest do not require them, we reviewed research questions, modeling practices, and reporting practices in empirical articles using mixture models in PTSD research. We use the literature review of van de Schoot et al. (2018), the results of which are available on the original author's Open Science Framework page. Van de Schoot et al. (2018) conducted a thorough review of all studies employing any type of mixture model within the field of PTSD research up to October 2016; ultimately locating 34 papers from 11,395 initially identified papers that satisfied keywords (full details are available in Appendix A of van de Schoot et al., 2018).

Our interest in these papers deviated from the original authors' interests, so we reviewed each of these 34 studies to identify (a) which type of mixture model was used in the study, (b) if the model constrained covariance parameters across classes, (c) if the growth factor covariance estimates were reported, and (d) if any subject-specific information was reported or required to answer the research questions.

Results of our review found that 18 studies (53%) used GMMs (the other 47% used LCGMs). A surprising minority of these GMM studies reported any information related to the within-class random effects: 83% did not report any covariance parameter estimates. Furthermore, 39% reported applying cross-class covariance constraints to aid convergence while another 44% did not provide enough information to determine if covariance constraints were present or not (i.e., only 17% definitively did not constrain covariance estimates across classes). None of the studies that applied cross-class constraints reported doing so for a theoretical reason (i.e., it is highly probable that constraints were uniformly applied to address convergence issues or as a default software option)

Most importantly and similar to points made in Cole and Bauer (2016), *zero* studies reported or asked any research questions about subject-specific curves. Studies invariably had the same three basic interests: (a) how many classes exist, (b) what do the class trajectories look like, and (c) which covariates predict class membership. Notably, none of these three interests requires within-class

random effects. Although the individual is the central focus of latent growth models, the *latent class* is the central focus of GMMs. From this evidence, it seems that the within-class random effects are not providing answers to substantively motivated questions and do not appear to be a factor that researchers are seriously considering when modeling their data (e.g., perhaps GMMs are used because researchers have been exposed to them rather than for their correspondence with the research question).

This is extremely relevant because researchers regularly encounter rampant convergence issues because of overly complex models, which ultimately leads them to atheoretically constrain any parameters they can across classes with the sole purpose of achieving model convergence. Poignantly, this process is undertaken to obtain quantities (partitioned variance components and subject-specific curves) that are irrelevant to the research questions.

Instead, we argue that a more advantageous modeling strategy is to bypass within-class random effects and adopt a population-average approach, a class of models that specifically focuses on the broader mean trajectory, while accounting for the variances and correlations among the repeated measures. As outlined in detail in the next section, these methods can similarly account for complex covariance structures but do so without relying on random effects.

## Modeling Change without Random Effects

As an extension of latent growth models, GMMs explicitly model between-individual variability and within-individual variability. However, the random effects framework is not necessarily required in order to properly model repeated measures data in all circumstances, especially when subject-specific curves are not needed. Although the random effects framework is omnipresent for growth models in psychology, the subject-specific focus concomitant with these models has largely been forgotten (e.g., Cudeck & Codd, 2012; Liu, Rovine, & Molenaar, 2012; McNeish, Stapleton, & Silverman, 2017; Molenaar, 2004; Molenaar & Campbell, 2009). Cudeck and Codd (2012) aptly summarize the disconnect between the widespread use of random effect models

and their ensuing model interpretation by noting, “the current curious state of practice is to sing the praises of the model as an ideal method for the study of individual change, but then ignore the individuals and resort to an analysis of the mean change profile.” (p. 5).

If the research questions can be sufficiently addressed without needing to inspect subject-specific curves or if the interest is in the mean trajectories within each class while properly accounting for within-class variation, the high computational demands of GMMs are needlessly taken on. In such cases, researchers (perhaps unknowingly) are augmenting the complexity of an already complex model with random effects and making an already difficult estimation problem more difficult, all for the purpose of obtaining information that is not central to the primary goals of a mixture model analysis and whose estimates are often not reported.

**A Different Perspective: Population-Averaged Models**

The population-averaged approach in non-mixture contexts has been written about extensively in the biostatistics literature (e.g., Diggle, Heagerty, Liang, & Zeger, 2002; Vonesh, 2013). In the context of continuous outcomes measured longitudinally, the goal is to obtain the average growth trajectory in the sample (conditional on any relevant covariates like sex or treatment group) while accommodating the covariance that arises due to the dependent nature of the repeated measures *without partitioning the variance or estimating subject-specific random effects* (Fitzmaurice, Laird, & Ware, 2011; Jennrich & Schluchter, 1986, Verbeke & Molenberghs, 2000). Put another way, the goal of population-averaged models is to *describe* the covariance between repeated measures rather than try to *explain* the covariance between repeated measures with random effects, as is the goal in latent growth models.

As an advantage, the estimation of population-averaged models, even in the non-mixture context, is much easier due to the simplified form of the covariance structure. This approach to summarizing the underlying change process has received very little attention in the behavioral sciences, especially in mixture contexts where the appeal of simplified estimation would seem to be

very attractive given widespread convergence issues encountered with GMMs. We review a specific type of population-averaged model — the covariance pattern model — in the next section and compare it to the latent growth model.

**Differentiating Between Covariance Pattern and Latent Growth Models**

Consider the role of the marginal covariance in growth models. Because data within an individual are dependent, the off-diagonal terms representing the covariance between pairs of repeated measures within a person are likely non-zero. For some arbitrary design where time $t = 1, \ldots, T$, the covariance of the raw repeated measures of the outcome $Y$ takes the general form

$$\begin{bmatrix} Var(Y_1) & & & \\ Cov(Y_1, T_2) & Var(Y_2) & & \\ \vdots & \ddots & \ddots & \\ Cov(Y_1, Y_T) & \cdots & Cov(Y_{T-1}, Y_T) & Var(Y_T) \end{bmatrix}.$$

In latent growth models, the marginal covariance structure is computed by partitioning the variance in between-individual ($\boldsymbol{\Psi}$) and within-individual ($\boldsymbol{\Theta}_i$) components with random effects and then combining the between-individual and within-individual covariance matrices together by adhering to distributional assumptions and following standard methods of deriving the model-implied second moment: $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Psi})\boldsymbol{\Lambda}_i^{\mathrm{T}} + \boldsymbol{\Theta}_i$.

In covariance pattern models, the variance is not partitioned and the marginal covariance is directly modeled. That is, a standard single-level regression model is fit to the data but assumptions about the residuals are relaxed. Rather than assuming constant variance and that the residuals are independent (e.g., $\varepsilon \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right)$ as in ordinary least squares), maximum likelihood allows the residuals from the same individual to be related to each another: $\boldsymbol{\varepsilon}_i \sim MVN\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right)$ where $\boldsymbol{\Sigma}$ is a residual covariance structure specified by the researcher. The model is called a “covariance pattern model” because the researcher selects a covariance structure that patterns how the residuals are

related to each other. The elements of the structure that is selected for $\boldsymbol{\Sigma}$ are then directly estimated as parameters with maximum likelihood For instance, $Cov(Y_1, T_2)$ from the above matrix would be directly estimated with a covariance pattern model whereas in a latent growth model, it be an indirect combination of $\boldsymbol{\Psi}$ (the between-individual covariance matrix) and elements of $\boldsymbol{\Theta}_i$ (the within-individual covariance matrix). The word "pattern" is used because a parsimonious structure is typically applied (e.g., all repeated measures that are one occasion apart have the same correlation) rather than uniquely estimating each individual element of the matrix (though simply estimating all unique elements is also possible). The covariance pattern model has fallen out of favor in recent years within psychology as computational advances have removed the computational barrier that once existed for the estimation of random effects models. However, the model continues to receives attention for its generality and flexibility in the quantitative psychology literature (e.g., Azevedo, Fox, & Andrade, 2016; Liu, Rovine, & Molenaar, 2012a; 2012b; Lix & Sajobi, 2010).

Though the mechanism adopted by latent growth models and covariance pattern models differs, each model essentially has the same goal: to provide estimates that reproduce the observed covariances as closely as possible. The next subsection discusses some of the common structures used in covariance pattern models and when different structures may be best applied.

**Common Covariance Pattern Covariance Structures**

Similar to the requirement that researchers select the structure of the within- and between-individual covariance matrices in latent growth models, researchers must similarly select the covariance pattern structure in covariance pattern models. This is often accomplished through an exploration of the repeated measures data taking into account longitudinal design features (e.g., spacing of the measurements and whether the measurement occasions are fixed across subjects).

We discuss four possible structures in the following subsections. Note that these structures are not exclusive to covariance pattern models and are sometimes used to describe the within-individual

covariance matrix in latent growth models (though less complicated structures with all off-diagonal elements constrained to zero remain the most popular; Grimm & Widaman, 2010). The difference in covariance pattern models is that the structure captures *all* residual covariance – not just *within-person* residual covariance as in latent growth models – so structures with non-zero off-diagonal terms are typically required with covariance pattern models.

**Compound Symmetry.** One common patterned structure is a compound symmetric (a.k.a., exchangeable) structure where $\mathbf{\Sigma}_i = \sigma_c \mathbf{J}_{n_i} + \sigma^2 \mathbf{I}_{n_i}$ for $\mathbf{J}_{n_i}$ a matrix of ones of dimension $n_i$ and $\mathbf{I}_{n_i}$ is an identity matrix of dimension, $n_i$. This results in a correlation matrix with equal off-diagonal elements. The variance terms on the diagonal can also be heterogeneous if the variance of the repeated measures changes over time. In a hypothetical case of 5 repeated measures, the compound symmetric correlation matrix would be

$$\begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho & \rho & 1 & & \\ \rho & \rho & \rho & 1 & \\ \rho & \rho & \rho & \rho & 1 \end{bmatrix}.$$

**First-Order Autoregressive.** Another popular structure features residuals that follow a first-order autoregressive process such that the $(j, j')$th element is $\mathbf{\Sigma}_{i(j,j')} = \sigma^2 \rho^{|j-j'|}$. In this first-order autoregressive structure, pairs of repeated measures separated by the one measurement occasion (i.e., that are lagged by one) are correlated equally. Repeated measures with larger lags are correlated to a lesser degree based on an exponential function of the one-lag correlation. For instance, with this structure, repeated measures that are one lag apart (e.g., Time 1 and Time 2) are correlated at some estimated value $\rho$ and repeated measures that are two lags apart (e.g., Time 1 and Time 3) would be correlated at $\rho^2$. The structure increases flexibility without sacrificing parsimony because it allows for correlations to vary across lags (i.e., measurements that are further apart are less related) but does

not require additional parameters to be estimated because each lag is a function of a single estimate, $\rho$. This structure could be embellished to allow for heterogeneous variances across time. In a hypothetical case of 5 repeated measures, the autoregressive correlation matrix would be

$$\begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho^2 & \rho & 1 & & \\ \rho^3 & \rho^2 & \rho & 1 & \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

**Toeplitz.** A structure that is similar to the first-order autoregressive structure that maintains more flexibility is the Toeplitz structure. In a Toeplitz structure, the autocorrelation process for the (*j*, *j′*)th element is $\mathbf{\Sigma}_{i(j,j')} = \sigma_{|j-j'|+1}$. Similar to the first-order autoregressive structure, all measures separated by one measurement occasion are equally correlated. However, for measures lagged by two, the Toeplitz structure estimates a separate correlation rather than simply squaring the lag-one correlation. Each subsequent lag also receives a unique estimate, so the number of off-diagonal parameters to be estimated is equal to the number of measurement occasions minus one. Like the first-order autoregressive structure, this allows the correlation between measures to change as the measures become more distant in time. Unlike the first-order autoregressive structure, a Toeplitz structure does not require that the change in correlation follow a specific function of the lag-one correlation. In a hypothetical case of 5 repeated measures, the Toeplitz correlation matrix would be

$$\begin{bmatrix} 1 & & & & \\ \rho_1 & 1 & & & \\ \rho_2 & \rho_1 & 1 & & \\ \rho_3 & \rho_2 & \rho_1 & 1 & \\ \rho_4 & \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}.$$

**Unstructured.** The most flexible of all within-class covariance structures is one that is completely unstructured where every element is uniquely estimated. In a hypothetical case of 5 repeated measures, the unstructured correlation matrix would be

$$\begin{bmatrix} 1 & & & & \\ \rho_{21} & 1 & & & \\ \rho_{31} & \rho_{32} & 1 & & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 & \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1 \end{bmatrix}.$$

This structure is reminiscent of the covariance structure in MANOVA within a general linear modeling framework (Liu et al., 2012a).

**Selecting a Covariance Structure.** The selection of which type of structure to use in a covariance pattern model can be a challenge to researchers not well-versed in this framework. Many options exist and readers looking for a good summary of possible options may wish to consult the treatment provided in the SAS 9.2 manual under the Repeated Statement section of the PROC MIXED chapter (PROC MIXED is the SAS procedure used to fit covariance pattern models even though they are not technically mixed-effects models). Chapter 7 of Fitzmaurice, Laird, and Ware (2011) is also dedicated to discussing covariance pattern models. As general guidance,

- Compound symmetry tends to be most useful when there are few repeated measures or when repeated measures are spaced very closely together. Compound symmetry in a covariance pattern model produces an identical marginal covariance matrix as a latent growth model with random intercepts but no random slopes.
- First-order autoregressive structures are most useful when there are many repeated measures and the spacing between measurement occasions is equal or nearly equal.
- Toeplitz is best suited for a moderate number of repeated measures but correlations are not expected to decrease exponentially over time.

- Unstructured is typically reserved for very few measurement occasions or when the measurement occasions have an unorthodox structure.

Ultimately, the goal is to strive for parsimony such that the covariance pattern structure reflects the dependency among repeated measures with as few parameters as possible.

**Equivalency of Mean Structures**

Though latent growth models and covariance pattern models differ with respect to the formation of the covariance structure and whether subject-specific curves are available, either model will produce the same mean trajectory with identical interpretations with continuous outcomes. Covariance pattern models estimate the growth trajectory for the typical individual in a sample but does not include random effect to capture individual deviation from the mean trajectory. Therefore, the means structure is written very similarly to the latent growth mean structure in Equation 1 except that the growth factors have no *i* subscript because they do not vary by individual. This can be written as

$$\mathbf{y}_i = \mathbf{\Lambda}_i \boldsymbol{\eta} + \boldsymbol{\varepsilon}_i \tag{10}$$

The mean growth trajectory of the covariance pattern model can be obtained by taking the expectation of $\mathbf{y}_i$:

$$E(\mathbf{y}_i) = E(\mathbf{\Lambda}_i \boldsymbol{\eta}) + E(\boldsymbol{\varepsilon}_i) \tag{11a}$$
$$= \mathbf{\Lambda}_i \boldsymbol{\alpha}, \tag{11b}$$

The mean growth trajectory of the latent growth model can similarly be obtained by taking the expectation of Equation 3, which yields the identical quantity as Equation 11b.

$$E(\mathbf{y}_i) = E(\mathbf{\Lambda}_i \boldsymbol{\eta}_i) + E(\mathbf{\Lambda}_i \boldsymbol{\zeta}_i) + E(\boldsymbol{\varepsilon}_i) \tag{12a}$$
$$= \mathbf{\Lambda}_i \boldsymbol{\alpha}, \tag{12b}$$

given that the random effects and residuals have a zero mean vector (i.e., $E(\boldsymbol{\zeta}_i) = E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$) and are uncorrelated [i.e., $Cov(\boldsymbol{\zeta}_i, \boldsymbol{\varepsilon}_i) = \mathbf{0}$ ].

## Extending Covariance Pattern Models to the Mixture Context

To place the covariance pattern mixture model (CPMM) on the current continuum of mixture models for repeated measures data, CPMMs fall between the LCGM from Nagin (2005) and the GMM from Muthén and Shedden (1999). Like LCGMs, CPMMs acknowledge that the latent classes are already a discretization of the random effects distribution and the discrete classes are the focus of the interpretation, so random effects within classes are not necessarily required. CPMMs address possible issues in LCGMs by expanding the marginal covariance structure so that extra classes are not extracted because of a covariance structure misspecification. Like GMMs, CPMMs fully model all variation by including a patterned marginal covariance structure that reflects between- and within-individual variation. Unlike GMMs, the marginal covariance in CPMMs is directly estimated rather than a combination of partitioned variance components.

If researchers are primarily interested in enumerating classes and interpreting the mean trajectory for each class while satisfactorily summarizing the pattern of variances and covariances among the repeated measures, a CPMM accomplishes these tasks in a more parsimonious and more efficient manner than GMMs. Concurrently, the CPMM has a simpler specification compared to a GMM, which should theoretically make convergence, inadmissible solutions, atheoretical parameter constraints, and other estimation-related issues less frequent. These claims are explicitly assessed and demonstrated via simulation evidence in the next section.

## Simulation Design

### Data Generation

The data generation model is based on the so-called “Cat’s Cradle” pattern that emerges in substance use (Sher, Jackson, & Steinley, 2011) and post-traumatic stress research (Bonanno, 2004). In these research domains, four classes typically emerge: one class that starts at higher values and maintains high values (the “Chronic” class), a second class that starts low and maintains low values (the “Unaffected” class), a third class that starts high but decreases over time (the “Recovery” class),

and a fourth class that starts low and increases over time (the "Delayed Onset" class). In both research domains where such solutions are found, the Chronic and Unaffected classes typically comprise a majority of the data (approximately 65% to 80%; Bonanno, 2004; Sher et al., 2011). Of the remaining data, the Recovery class tends to be about twice as big as the Delayed Onset class.

Figure 1 shows a plot of the trajectories in each of the four classes and Figure 2 shows the general path diagram of the model. Table 1 shows the model equations and covariance structures that were used to generate data from these trajectories within M*plus* Version 8; the population generating model is a GMM.

Table 1
*Data generation equations for population model*

| Class | Model Equation | Covariance Structures |
|---|---|---|
| 1 (Unaffected) | $y_{ij} = \eta_{0i} + \eta_{1i}t + \varepsilon_{ij}$<br>$\eta_{0i} = 15.0 + \zeta_{0i}$<br>$\eta_{1i} = -0.15 + \zeta_{1i}$ | $\boldsymbol{\Psi} = \begin{bmatrix} 40 & \\ 0 & 0.04 \end{bmatrix}$<br>$\boldsymbol{\Theta} = 28\mathbf{I}_5$ |
| 2 (Recovery) | $y_{ij} = \eta_{0i} + \eta_{1i}t + \eta_{2i}t^2 + \varepsilon_{ij}$<br>$\eta_{0i} = 38.0 + \zeta_{0i}$<br>$\eta_{1i} = -0.1 + \zeta_{1i}$<br>$\eta_{2i} = -0.008$ | $\boldsymbol{\Psi} = \begin{bmatrix} 160 & \\ 0 & 0.04 \end{bmatrix}$<br>$\boldsymbol{\Theta} = 72\mathbf{I}_5$ |
| 3 (Chronic) | $y_{ij} = \eta_{0i} + \eta_{1i}t + \eta_{2i}t^2 + \varepsilon_{ij}$<br>$\eta_{0i} = 41.0 + \zeta_{0i}$<br>$\eta_{1i} = 0.12 + \zeta_{1i}$<br>$\eta_{2i} = -0.002$ | $\boldsymbol{\Psi} = \begin{bmatrix} 120 & \\ 0 & 0.04 \end{bmatrix}$<br>$\boldsymbol{\Theta} = 60\mathbf{I}_5$ |
| 4 (Delayed Onset) | $y_{ij} = \eta_{0i} + \eta_{1i}t + \eta_{2i}t^2 + \varepsilon_{ij}$<br>$\eta_{0i} = 18.0 + \zeta_{0i}$<br>$\eta_{1i} = 0.2 + \zeta_{1i}$<br>$\eta_{2i} = 0.015$ | $\boldsymbol{\Psi} = \begin{bmatrix} 72 & \\ 0 & .035 \end{bmatrix}$<br>$\boldsymbol{\Theta} = 60\mathbf{I}_5$ |

*Note*: $t = 0, 1, 10, 18, 26$

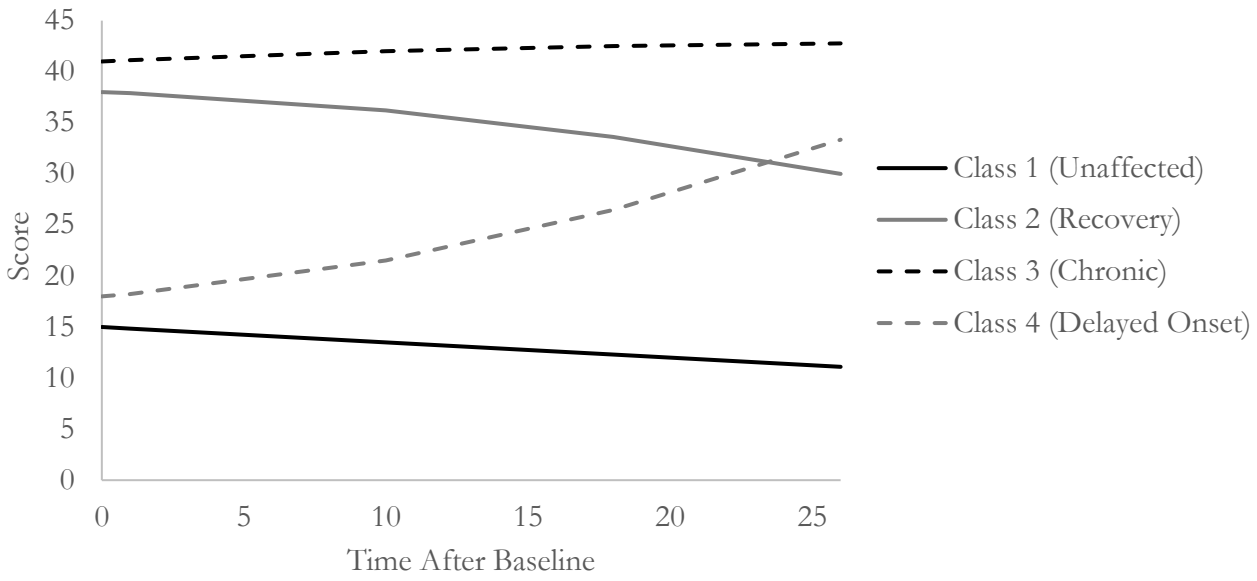


*Figure 1.* Plots of the mean trajectories in each of the 4 simulated classes in Model 2. Based on class proportions in a study by Depaoli, van de Schoot, van Loey, and Sijbrandij (2015), we assigned 63% of the sample to Class 1, 12% to Class 2, 19% to Class 3, and 6% to Class 4.

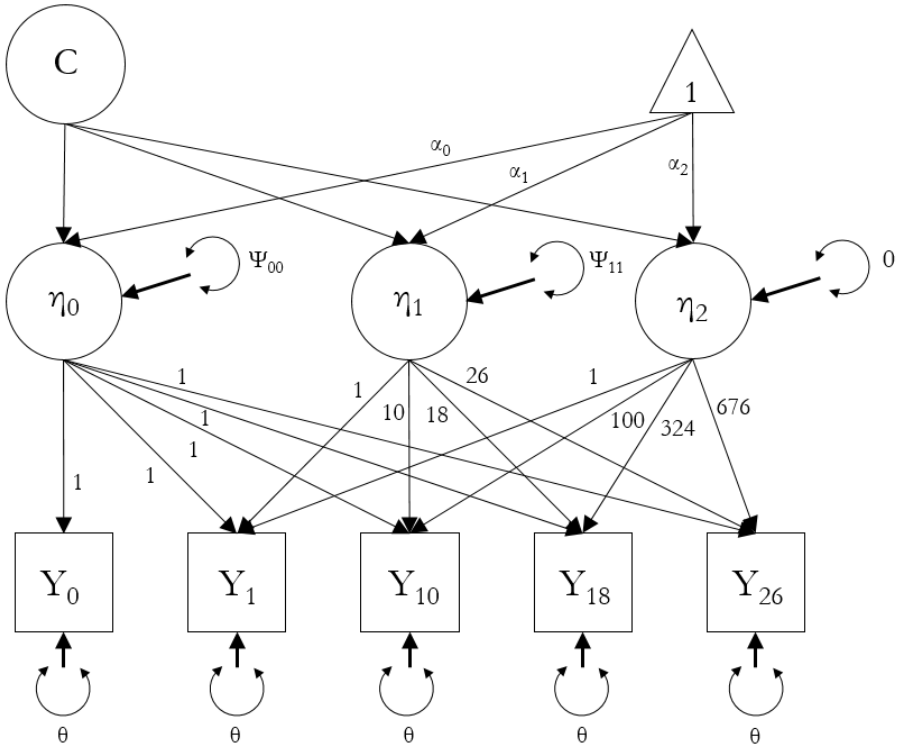


*Figure 2.* Path diagram of data generation model. C is a discrete latent variable represents the different classes. The α parameters represent the latent variable means. The $\eta_0$ latent variable represents the intercept, $\eta_1$ represents the linear slope, and $\eta_2$ represents the quadratic slope. The intercept latent variable varies across people within a class (with variance $\Psi_{00}$) as does the linear slope latent variable (with variance $\Psi_{11}$). The quadratic slope latent variable does not vary across people within a class. The latent variable loadings are fixed because data are time-structured such that all people have the same occasions of observations. The residual variances are constrained to be equal across time.

The growth trajectory in each class has both linear and quadratic components to achieve non-linear trajectories. The linear slope varies across individuals within classes but the quadratic slope variance was constrained to zero in the population. The Unaffected group comprised 63% of the population, the Recovery class 12%, the Chronic class 19%, and the Delayed Onset class 6% in attempt to mirror empirical applications of mixture models where class proportions are disparate. The data feature 5 time-points that represent either months after baseline in the substance use context or weeks in the post-traumatic stress context. The loadings from the linear slope factor to the observed

variable are 0 (baseline), 1, 10, 18, and 26. Loadings from the quadratic slope factor to the observed variables are the square of these loadings. The growth factor variances are rather large relative to the growth factor means for the intercept and linear slope, which was intentional in order to generate data that were not well separated, as is typically encountered in empirical examples.

**Simulation Conditions**

The simulation features sample sizes of 500 and 1500. Our assessment of the van de Schoot et al. (2018) literature review on mixture models in PTSD research resulted in a median sample size of 509, which informed our smaller sample size condition. 1500 was selected to represent a study that is far above average because 1500 corresponded to about the 85$^{th}$ percentile of sample sizes in the van de Schoot et al. (2018) review.

Three different models were fit to the data: a CPMM with a compound symmetric structure with homogeneous variances that was unconstrained across classes (CPMM; 2 covariance parameters per class: 1 residual variance, 1 covariance), a GMM with all variances unconstrained across classes (GMM; 4 covariance parameters per class: 1 residual variance, intercept variance, slope variance, and a covariance between intercepts and slopes), and a GMM with all covariance parameters constrained across classes (GMMC; 4 covariance parameters total: 1 residual variance, intercept variance, slope variance, and a covariance between intercepts and slopes). The residual variance was constrained to be equal within classes across all models to match the data generation process. The path diagram for the GMM is identical to Figure 2. The path diagram for the GMMC is similar to Figure 2 with the exception that the $\Psi$ and $\theta$ parameters are constrained to be equal across all classes. The CPMM path diagram is shown in Figure 3.

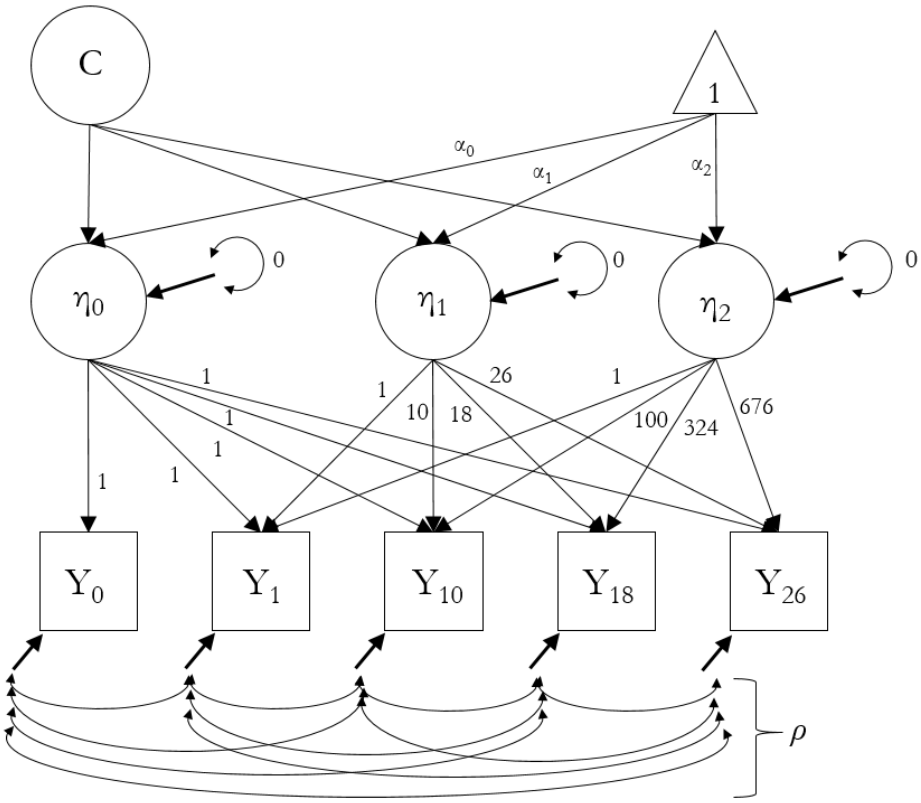


*Figure 3.* Path diagram of Covariance Pattern Mixture Model. Parameters definitions are the same as in Figure 2. The newly added ρ is a residual covariance; with the compound symmetric structure, all residual covariances are constrained to be equal. The residual variances are not shown to avoid overcrowding, but each residual variance is constrained to be θ as in Figure 2. Also, note that the latent variable variances are all constrained to 0, which forces all the residual variance to the observed repeated measures rather than partitioning it into within-person and between-person components. A covariance pattern is then applied directly to the residuals of the repeated measures

As an important note, the 4-Class GMM is identical to the data generation model. Also of note, because the data generation included non-trivial random slopes, the CPMM covariance structure will be moderately misspecified. This was done intentionally. Using a misspecified CPMM will paint a more realistic picture of performance and will avoid artificially inflating the performance of CPMMs by unrealistically modeling the true covariance structure. Instead, the results will reflect the results that would be obtained if someone were to use CPMMs somewhat naively and select an unfavorable covariance structure. So, keep in mind that the results we present compare a *perfect*

GMM with a misspecified CPMM. As a secondary consideration, we hope that this will alleviate potential fears researchers may have about switching model types and not specifying the model perfectly – our results already build in some possible user error that may be encountered if using an unfamiliar modeling framework.

**Simulation Outcomes**

We will follow four outcomes in the simulation: class enumeration, percent of convergent models, relative bias in the class-specific growth trajectories, and classification accuracy. For the last three outcomes (convergence, bias, and classification), the results can be easily automated within M*plus* Version 8 using the MonteCarlo module. These results are based off 500 replications for each sample size condition. In each of these 500 replications, results will be based on the assumption that the correct 4-Class solution has been fit the data. For these three outcomes measures, the population values for the class trajectories were used as starting values for the mean structure in each class. Default M*plus* starting values were used for all covariance structure parameters. To study class enumeration, a separate simulation using a different setup is required, which is outlined next.

**Details of separate class enumeration simulation**. To study class enumeration behavior of each of the three fitted models, we generated 100 datasets for each sample size condition and then fit a 3-Class, 4-Class, and 5-Class model to each generated dataset. Fewer replications were used than for other simulation outcomes because class enumeration requires fitting multiple models per replication. We then compared the sample size adjusted BIC (SA-BIC; Sclove, 1987) across the three different class solutions for each replication, for each model. The solution with the lowest SA-BIC was then selected for each replication. SA-BIC was chosen because Yang (2006) and Tofighi and Enders (2007) have found that it tends to perform better than other information criteria for class enumeration. Tofighi and Enders (2007) in particular note that SA-BIC is the clear choice for enumeration with moderate sample sizes (p. 332), poorly separated classes (p. 333), or when there is a large disparity in the class proportions (p. 334), all of which exist in some or all conditions of our simulation design.

The goal of the simulation was to track the number of times each model type selected the correct 4-Class solution. The class enumeration simulation used M*plus* default starting values rather than using the population values as starting values. Default starting values were used because (a) it was unclear which starting values should be used for the incorrect 3-Class and 5-Class solutions as there would not be any population values for these classes, (b) default starting values are more representative of how classes are enumerated in empirical studies in the initial stages of analyses (i.e., researchers do not have a good idea of the class trajectories before they know how many classes there are), and (c) it may be helpful to present simulation results for how starting values may impact convergence. We also tracked the percentage of non-convergent replications for each of the fitted models. Replications where the best likelihood was not replicated from multiple random sets of random starting values were not treated as non-converging for this simulation.

These models were run by calling M*plus* Version 8 via SAS PROC IML to facilitate aggregating the results because this type of simulation cannot be performed entirely within the M*plus* MonteCarlo module (to our knowledge). We used 100 random starts and 10 final stage optimizations based on recommendations in Liu and Hancock (2014) and Li, Harring, and Macready (2014) so that the full likelihood surface could be explored. All code for the simulation from SAS and M*plus* are included on the first author's Open Science Framework webpage (https://osf.io/yh6kf/).

## Simulation Results: Unknown Number of Classes

### Class Enumeration

The initial step of a mixture model analysis is typically to determine the number of latent classes that are present, so we begin with the enumeration and convergence results first. Table 2 shows the number of replications that converged for each class solution and the number of replications selecting each class solution by model type and sample size. The number of replications selecting each class solution may not add up to 100 because there were some replications in which none of the competing class solutions converged (and therefore none of the competing options were selected). We did include cases where some but not all class solutions converged. For example, if

only the 3-Class solution converged but the 4-Class and 5-Class solutions did not, the replication was recorded as selecting the 3-Class solution. We followed this criterion because it most closely mirrors how we felt the situation would be handled with empirical data.

Table 2
*Number of replications selecting a 3-, 4-, or 5-class solution based on the SA-BIC and the convergence rate by the number of classes*

| | *N* =500 | | | *N* =1500 | | |
|---|---|---|---|---|---|---|
| | CPMM | GMM | GMMC | CPMM | GMM | GMMC |
| % Converged | | | | | | |
| 3-Class | 98 | 38 | 77 | 100 | 66 | 83 |
| 4-Class | 93 | 5 | 67 | 100 | 16 | 72 |
| 5-Class | 88 | 0 | 83 | 97 | 3 | 98 |
| | | | | | | |
| 3-Class Count | 37 | 37 | 11 | 2 | 66 | 0 |
| 4-Class Count | 33 | 2 | 32 | 74 | 2 | 5 |
| 5-Class Count | 30 | --- | 53 | 24 | 0 | 95 |

*Note:* CPMM = Covariance Pattern Mixture Model, GMM = Growth Mixture Model, GMMC = Growth Mixture Model with Covariance Parameters Constrained to Equality across Classes. --- = not applicable because there were no viable replications for this condition. Some columns do not add up to 100 because none of the models converged for some replications.

***N* = 500 condition**. In the *N* = 500 condition, the CPMM narrowly had the highest number of replications in which the true 4-Class solution was selected (33 out of 100). This vastly exceeds the GMM which only selected the 4-Class solution in 2 replications (poor performance was largely driven by convergence issues) but only narrowly eclipses the number of times the GMMC selected the correct 4-Class solution (32 out of 100).

Regarding convergence, using M*plus* default starting values, the convergence issues encountered by the GMM are readily apparent: only 5% of the 4-Class models converged, even though this was the exact model from which data were generated. Constraining all the covariance

structure parameters to be equal across classes is clearly effective for convergence, as convergence rates of the GMMC where in the high 60s to low 80s. Do note that the CPMM convergence was in the high 80s to high 90s without requiring assumptions implied by constraints, however. Constrained variances in the GMMC were not warranted on some parameters (i.e., intercept variance, residual variances). This may explain why the GMMC seemed to prefer the 5-Class solution because the misspecified covariance structure may be emerging as a separate class, especially when considering that BIC-based measures tend to be conservative and underextract the number of classes (e.g., Diallo, Morin, & Lu, 2017; Dziak, Lanza, & Tan, 2014). Despite the improved convergence of the GMMC over the GMM, the CPMM uniformly had the highest convergence rates, especially for the correct 4-Class solution.

***N* = 1500 condition**. In the *N*=1500 condition, the frequency with which the CPMM selected the correct 4-Class solution increased to 74 out of 100 and convergence issues were essentially negligible across conditions. For the GMM, even though the model was identical to the data generation model and the sample size was in the 85$^{th}$ percentile of empirical studies in this area, convergence issues remained immensely problematic with only 16 out of 100 replications converging for the 4-Class solution. As a result, the GMM most often selected the 3-Class solution (66 out of 100 replications) and only selected the correct 4-Class solution in 2 replications. As expected, the GMMC vastly improved convergence rates compared to the GMM. However, the GMMC overwhelmingly favored the 5-Class solution, which is an unconventional finding given the conservative nature of the BIC-based metrics and their tendency to under-extract. The spurious class is likely attributable to the covariance structure misspecification such that additional classes represent assumption violations rather than a substantively interesting group of people (e.g., Bauer & Curran, 2003). Across all conditions, the convergence rate of the CPMM was, at worst, within 1% of the GMMC and CPMM convergence rates exceeded the GMMC rates by a wide margin in other conditions, particularly in the true 4-Class solution condition.

Although the CPMM is not perfect (or even necessarily good in the $N = 500$ condition in an absolute sense), the CPMM gives the best *relative* chance to select the correct number of classes. Part of this improved performance is related to improved convergence rates – in the GMM (and GMMC to a lesser extent), the 4-Class solution could not converge due to the augmented complexity of the model, so there was no chance that the correct solution could be selected. Though the GMMC certainly improves convergence when good starting values are not known a priori, convergence is worse than the CPMM, it requires more assumptions about constraints across classes, and it selects the proper number of classes less often than the CPMM. The behavior as sample size increases is also telling – the GMMC performed much more poorly in the larger sample size, possibly suggesting that the 4-Class solutions in the smaller sample size conditions may be attributable to uncertainty or the conservative nature of BIC-based measures. On the other hand, the CPMM dramatically improved at the larger sample size.

## Simulation Results: Known Number of Classes

Results in the previous section approached the analysis as an empirical study in that the number of classes is unknown a priori. Even though there were discrepancies in the ability of each approach to correctly identify the number of classes, the results presented in this section are fit as if the enumeration yielded the correct 4-Class solution. This was done with an interest in gauging the quality of the model estimates independent of each approach's ability to detect the correct number of classes. These analyses also used the population values from the data generation for the mean structure in each class, as it would be more reasonable that a researcher would have a better idea about the different trajectories that exist in the data once the number of classes was determined.

### Convergence Rates, Population Starting Values

Results previously reported in Table 2 contained information about convergence when the M*plus* default starting values are used. Figure 4 shows the percent of the 500 replications with a 4-Class solution that successfully converged with the population values used as starting values for each

class. As noted previously, convergence is a major obstacle to fitting GMMs in empirical studies –even with good starting values –, which is reflected in Figure 4. Only 26% of GMMs converged when $N$ = 500 and only 33% converged when $N$ = 1500. Good starting values made an improvement over the 4-Class values in Table 2, but the results still are troubling considering that this is true model with the population values for the starting value of each class. The GMMC was again effective for convergence as evidenced by the 68% and 100% convergence rates for GMMC in the $N$ = 500 and $N$ = 1500 conditions, respectively. However, the covariance pattern approach led to the best convergence rates in Figure 4 With 99% and 100% convergence in the $N$ = 500 and $N$ = 1500 conditions, respectively. Coinciding with the argument that CPMMs reduce model complexity, the starting values made the smallest difference in convergence for the CPMMs compared to the GMMs or GMMCs.

Though a helpful starting point, simply achieving convergence and obtaining estimated values in software output is not indicative of improved performance. The next subsection investigates the estimated trajectories of the classes to assess the quality of the estimates that are obtained from each model type once convergence is achieved.

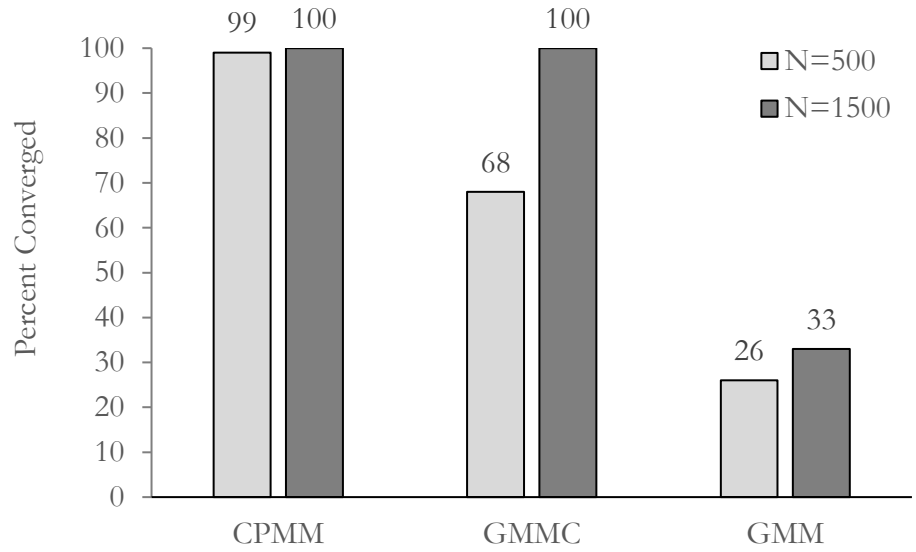

*Figure 4.* Plot of convergence rates from 500 replications of the 4-Class solution by model type and sample size condition. CPMM = Covariance Pattern Mixture Model, GMMC = Growth Mixture Model with Covariance Parameters Constrained to Equality across Classes, GMM = Unconstrained Growth Mixture Model.

**Trajectory Bias**

Figure 5 shows the class trajectories from each model type with a 4-Class solution averaged over replications that converged for the $N = 500$ condition. We only present the $N = 500$ results in-text for brevity due to similarities across conditions, but the bias for the $N = 1500$ condition is available from the supplementary material for interested readers.

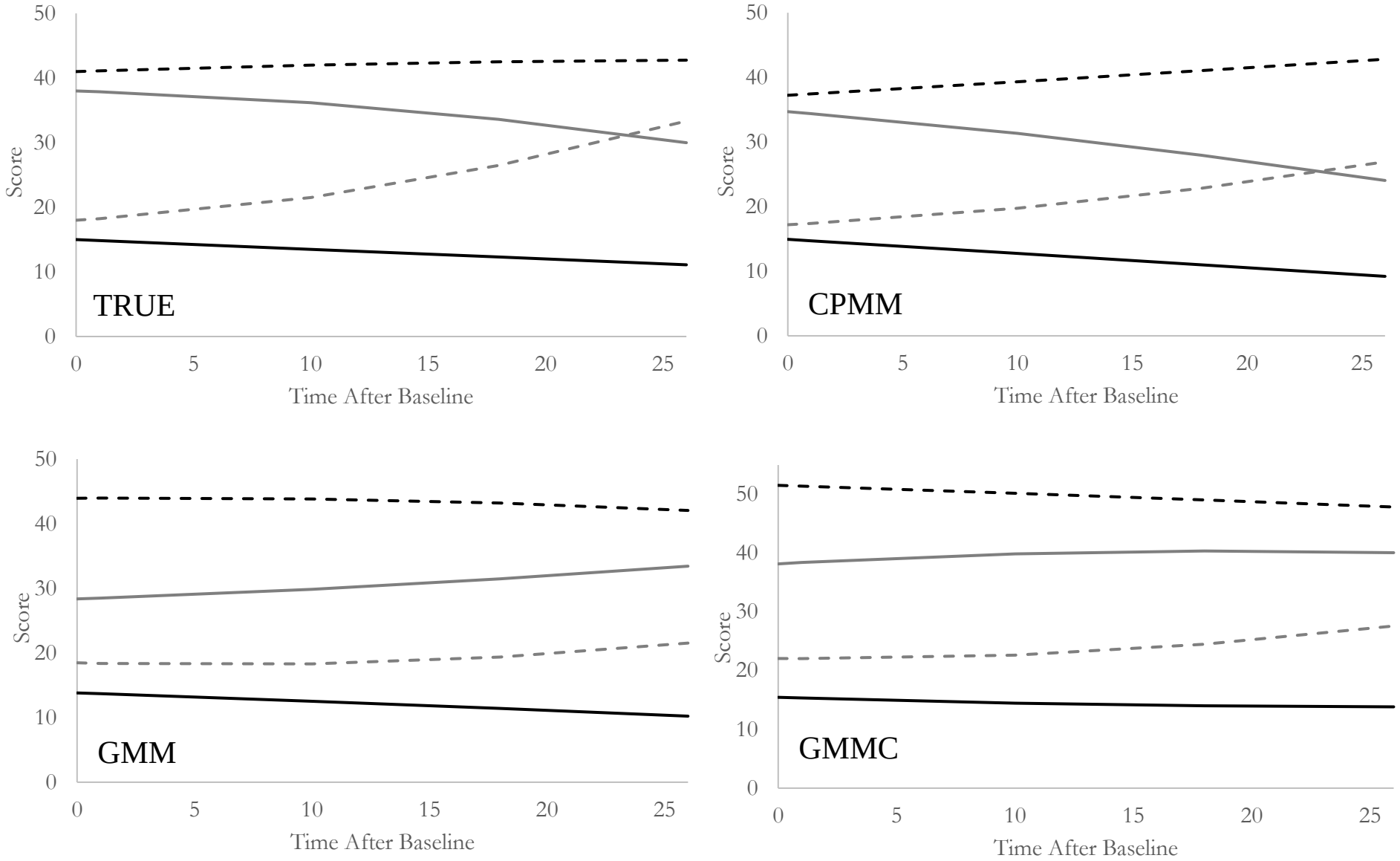


*Figure 5.* Comparison of true population average class trajectories (upper left) to covariance pattern mixture model (CPMM, upper right), unconstrained growth mixture model (GMM, lower left), and constrained growth mixture model (GMMC, lower right) for the $N$ = 500 condition

The generated population trajectories are shown in the upper left as a reference. The CPMM model in the upper right does not perfectly match the population trajectories – the Recovery Class slope in solid grey is noticeably steeper and the Delayed Onset Class slope in dashed grey is noticeably flatter. In fact, the relative bias of these slopes and the Unaffected Class slope in solid black exceed the 10% threshold typically used in simulation studies (Flora & Curran, 2004), though the intercept bias was negligible for all classes. Despite this bias, the four classes from the generating model are still rather clear: the "Cat's Cradle" pattern is quite apparent and the basic substantive interpretation of the classes is discernible.

The GMM and GMMC trajectories shown in the bottom panels, on the other hand, do not reflect the population trajectories very accurately. With both the GMM and the GMMC, the Recovery Class slope in solid grey goes in the wrong direction and increases over time whereas the Delayed Onset Class slope in dashed grey is essentially flat. The relative bias for the slopes in the Recovery and Delayed Onset classes in these model types ranged from -120% to -334%. Of ultimate importance, the class trajectories from either of the GMM or GMMC do not show the "Cat's Cradle" pattern and instead show four nearly horizontal lines. Though the GMMC helped improve convergence as noted in Figure 4, the class trajectories produced by this model have the highest relative bias for all but one parameter (slope of the Unaffected class in solid black; the CPMM has the highest relative bias for this parameter).

So far, the CPMM shows better enumeration, convergence rates, and improved (but not perfect) estimates of class trajectories, even as compared to the true GMM. However, a typical substantive interest of mixture models for longitudinal data is the ability to assign individuals to the proper class. The performance of classification accuracy is covered in the next subsection to address this property of each model type.

**Classification Accuracy**

Table 3 shows the classification percentages for each model type and sample size condition for converged replications for the 4-Class solution. This outcome measure shows the percentage of simulated individuals who were assigned to the appropriate class by the model (this is possible in a simulation because the true class is known). For example, the "96%" value in the CPMM column for $N = 500$ means that 96% of the simulated individuals who were generated to be in the Unaffected Class were assigned to the Unaffected Class by the CPMM. The total number of correctly classified individuals is included at the bottom of the table (because the class sizes are very different, this value is not equal to the unweighted average in each column).

Table 3
*Classification accuracy for simulated data for each sample size and model type showing the percent of simulated people in who were assigned to the correct class.*

| | $N = 500$ | | | $N = 1500$ | | |
|---|---|---|---|---|---|---|
| Class | CPMM | GMM | GMMC | CPMM | GMM | GMMC |
| Unaffected Class | 96% | 96% | 97% | 96% | 97% | 96% |
| Recovery Class | 54% | 71% | 77% | 47% | 39% | 57% |
| Chronic Class | 89% | 92% | 87% | 85% | 80% | 91% |
| Delayed Onset Class | 71% | 55% | 0% | 58% | 41% | 0% |
| Percent Correctly Classified | 89% | 89% | 88% | 87% | 84% | 86% |

*Note:* CPMM = Covariance Pattern Mixture Model, GMM = Growth Mixture Model, GMMC = Growth Mixture Model with Covariance Parameters Constrained to Equality across Classes

In Table 3, the classification accuracy was the lowest for the Recovery and Delayed Onset classes (the two middle classes whose trajectories cross in Figure 1). This makes sense from the data generation because the intercepts of these classes are very similar to the other classes and the slope variances were rather large relative to magnitude of the growth factor means. Additionally, the Recovery and Delayed Onset classes were the smallest classes (12% and 6%, respectively).

In both sample size conditions, the CPMM had the highest overall classification rate but the CPMM was not always the best as classifying individuals for each class. Classification percentages for the Unaffected and Chronic classes differed slightly across model type and sample size, but the percentages were roughly the same. However, the CPMM tended to be worse for the Recovery Class compared to the GMM and GMMC models whereas the GMM and GMMC performed noticeably worse for the Delayed Onset Class. The classification rates for the Delayed Onset class also decreased with sample size for both the CPMM and GMM.

Of particular note is that the GMMC did not accurately assign any individuals from the Delayed Onset Class. GMMC did assign some individuals who were generated to be members of other classes into the Delayed Onset Class so the class was not empty; however, these individuals truthfully belonged to a different class. Nearly all (97% in both sample size conditions) of the individuals generated to be in the Delayed Onset Class were assigned to the Unaffected Class in the GMMC. This somewhat odd finding may be attributable to the large trajectory bias seen in Figure 4. That is, given the values for the intercept and slope variances and the fact that the GMMC class trajectories were essentially four horizontal lines, it is questionable whether the Delayed Onset Class has the same meaning in the GMM or GMMC as it does in the data generation model. Conversely, the CPMM has some difficulty accurately assigned individuals to the small, overlapping classes but the classes at least appear to have the same general meaning as intended in the data generation model.

In the next section, we show how the general findings from the simulation study apply to an empirical dataset.

## Empirical Example

Consider a subset of 405 children and mothers from the National Longitudinal Survey of Youth (NLSY) that can be found in Hox (2010). Each child's Reading Recognition are measured at four different time-points where children were between 6 and 8 years old at baseline. To these data, we fit a CPMM and GMM/GMMC to outline the difference in approaches with a single empirical dataset. Mixture models were estimated in M*plus* 8 with 100 random starts and 10 final stage optimizations. Complete files containing the M*plus* code and results for models used in the example as well as the data are provided on the first author's Open Science Framework webpage. M*plus* code for each of the common covariance structures we discussed previous for CPMMs are provided in the Appendix.

### Determining the Mean Structure

To this data, we first fit an unconditional growth model to the Reading Recognition variable without extracting any latent classes (means at the four time-points are 2.52, 4.08, 5.00, and 5.77). When looking at the empirical means as well as exploratory plots, it seemed plausible that the growth trajectory may be non-linear because the difference between successive time points decreases for larger values of time. When fitting the LGM without multiple latent classes, we fit a linear growth model and a quadratic growth model. The quadratic model resulted in a significant likelihood ratio test ( $\chi^2(1) = 146.99,\ p < .01$ ) and this mean structure was retained.

### Adding Mixture Components

Then we subsequently fit a CPMM and a GMM. The CPMM was fit with a homogeneous Toeplitz structure because the measures are equally spaced and the raw variances at each point are rather close (0.86, 1.17, 1.35, 1.56 for Time 1 through Time 4, respectively). Raw correlations

between time-points were also high (range: .45 to .80) and decreased over time but in a not consistent fashion. The GMM was fit with a homogeneous diagonal residual structure; the quadratic growth factor variance was set to 0 but the intercept and slope variance were estimated and allowed to covary. We did consider heterogeneous variances for both models as well, but the SA-BIC was worse with heterogeneous variances in all instances.

For growth in academic measures like reading, it is typical to find three latent classes generally corresponding to "fast" learners, "on-time" learners, and "slow" learners (e.g., Musu-Gillette et al., 2015). Along with this theory, we compared the 2-Class, 3-Class, and 4-Class models for both the CPMM and the GMM using the SA-BIC and the bootstrapped likelihood ratio tests with 100 replications (BLRT; McLachlan, 1987). Though the BLRT was not included in the simulation because of its heavy computational demand, but its use has been advocated for along with BIC-based measures in previous studies (Nylund, Asparouhov, & Muthén, 2007; Nylund-Gibson & Masyn, 2016).

**Enumerating Classes**

**CPMM enumeration.** With the CPMM, the SA-BIC of the 3-Class solution was smaller than the 2-Class solution (3-Class SA-BIC = 3107 vs. 2-Class SA-BIC = 3154) and the BLRT was significant in the 3-Class model ( $\Delta$-2LL = 69.81, $p_{BLRT} < .01$ ), suggesting that the 3-Class solution fit better. The 4-Class solution best likelihood could not be replicated with different starting values. The issue likely stemmed from the fact that the fourth class was very small (only 2% of the sample which is 10 individuals for this data). Furthermore, the estimates that were output (along with the warning message) had a nearly identical class trajectory to another class, suggesting that the four classes was an over-extraction. We therefore continued with the 3-Class solution for the CPMM.

**GMM enumeration**. When fitting an unconstrained GMM, the 2-Class solution converged without issue (SA-BIC = 3097). When allowing all parameters to be class-specific, the 3-Class GMM

did not converge due to a non-positive definite growth factor covariance matrix. Constraining the growth factor covariance matrices to be equal across classes did not lead to convergence, but constraining the growth factor covariance matrices and residual covariance matrices across classes did allow the model to converge (i.e., a GMMC was required to achieve convergence; 3-Class SA-BIC = 3180). The BLRT was significant for the 3-Class solution ($\Delta$-2LL = 32.93, $p_{BLRT} < .01$), suggesting that 3 classes fit better than 2 classes. As with the CPMM, the 4-Class GMM solution could not converge with all covariances freely estimated across classes. Constraining growth factor covariance matrices did not help, but constraining all covariance parameters did converge (4-Class SA-BIC = 3166). The BLRT was significant for the 4-Class solution ($\Delta$-2LL = 25.62, $p_{BLRT} < .01$), suggesting that the 4-Class solution fit better than the 3-Class solution. Although the 2-Class SA-BIC was lower, the BLRT was significant, so we proceed with the 4-Class solution because both the SA-BIC and BLRT support 4 classes over 3 classes and the BLRT also supported 3 classes over 2 classes. We acknowledge that the 2-Class solution could also be used, depending on which metrics researchers choose to weigh most heavily.

**Class Trajectories and Proportions**

A plot of the class trajectories and the percentage of the sample assigned to each class for the CPMM is shown in the top panel of Figure 6. The classes were well-separated with average latent class probabilities of .877 for Class 1, .806 for Class 2, and .870 for Class 3. The bottom panel of Figure 6 shows the results from the 4-Class GMMC (with all covariance parameters were constrained across classes). The classes were also well-separated with average latent class probabilities of .839 for Class 1, .806 for Class 2, .858 for Class 3, and .851 for Class 4.

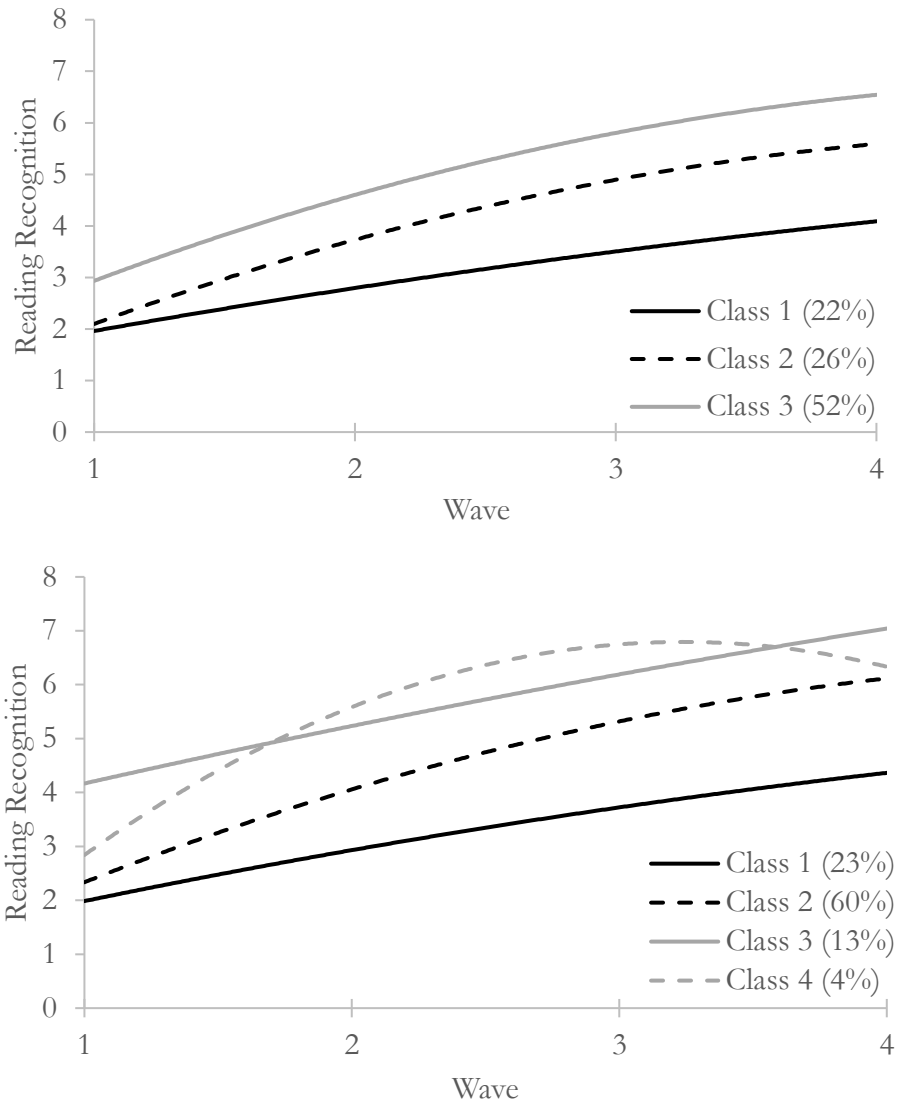


*Figure 6.* Growth trajectories of 3 extracted classes using a covariance pattern mixture model (top panel) compared to growth trajectories of 4 extracted classes using a growth mixture model with all covariance parameters constrained across classes (bottom panel).

When comparing the class trajectories between the CPMM and the GMMC in Figure 6, the GMMC has more classes and assignment of participants to the classes is quite different. Class 1 contains about the same proportion of the sample in the CPMM (22%) and the GMMC (23%). In the GMMC, Class 2 makes up a majority of the sample at 60%. In the CPMM, Class 2 contains only 26% of the sample. A majority of the sample in the CPMM were assigned to Class 3 (52%) whereas Class 3 (13%) and Class 4 (4%) together resulted in only 17% of the sample in the GMMC.

Differences in the empirical example mirror findings from the simulation study. Namely, the more theoretically desirable GMM showed notable convergence problems, necessitating a switch to a GMMC for the sake of achieving convergence. Also as seen in the simulation, the likely misspecification present in the GMMC resulted in additional extracted classes, which reflect the misspecification to the covariance structure rather than a substantively motivated class. In the CPMM, the class-specific marginal covariances (e.g., Class 1 Lag-1 = .076, Class 2 Lag-1 = .205, Class 3 Lag-1 = .520) and class-specific residual variances (Class 1 = .186, Class 2 = .308, Class 3 =.943) were quite different, suggesting that the constraints applied in the GMM were an inappropriate methodological shortcut required to reach convergence. The CPMM was fit and converged without issues, allowing the desired model (without undesirable constraints) to be fit and interpreted as intended.

## Discussion

Though random effects modeling is thoroughly engrained in psychology, there is a growing body of literature questioning its status as the default methods when considering the types of questions being asked. In the context of mixture models for longitudinal data, the random effects can make an already complex estimation process more complex, leading to higher rates of convergence issues and poor statistical properties of estimates, even if the model is the exact true model and sample size is large. Moreover, as found in our review of PTSD mixture model studies, the most concerning aspect of universal random effect usage is that researchers are not using these quantities that are responsible for making the estimation so demanding.

If the goal is to obtain the average growth trajectory in each latent class, population-averaged models are the simplest type of model that is capable of addressing these questions while still accounting for the variances and covariances among the repeated measures in a sensible manner. The CPMM as presented in this paper is one such approach and our results unambiguously support that CPMMs vastly improves convergence, class enumeration, and class trajectories compared to GMMs.

In this way, CPMMs align with recommendations in Wilkinson (1999) which calls for psychological researchers to choose a minimally sufficient analysis (p. 598). The simpler population-averaged approach is equally capable as random effect models for answering research questions asked in the empirical literature. Though GMMs (and the GMMC variation used to combat convergence issues) are currently used to address these questions, the model tends to be overly complex for the intended purpose, which leads to convergence issues, poor performance, and, commonly, the need for questionable cross-class constraints. Though convergence issues can arise for a variety of reasons, a mismatch in model complexity and data quality can be a primary culprit and the oft-noted convergence issues with GMMs are a strong indicator that less complex alternatives are a fruitful avenue to explore.

One concern concomitant with the CPMM approach is that researcher must select the covariance structure and may not choose the optimal structure. However, the results from the current investigation suggest that this may not be overly important. In our simulation studies, we purposefully choose the most misaligned covariance structure given the data generation and even a noticeably misspecified CPMM was clearly superior to a GMM that was identical to the data generation model and that used the population values as starting values as well as the GMMC version of the model often used in practice. In a relative sense, the more adverse issue is that GMMs are too complex to be reasonably applied in most contexts rather than whether the CPMM covariance structure is misspecified.

Furthermore, we want to emphasize that the GMM model-implied covariance structured is not necessarily correct in the empirical data in all cases either. As noted earlier (Equation 9), GMMs pose a specific functional form to the model-implied covariance, which may or may not adequately summarize the variability among the repeated measures. This could possibly lead to some misspecification, especially given that the model-implied covariance similarly requires the researcher to select the appropriate structures of $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$, similar to CPMMs. The most general, parsimonious

form of the covariance structure is $\boldsymbol{\Sigma} = \mathbf{D}^{1/2}\mathbf{P}\mathbf{D}^{1/2}$ where the variance matrix, **D**, can have its own structure uncoupled from the correlational (off-diagonal) structure in **P** (see, Harring & Blozis, 2014). Covariance structures in either CPMMs and GMMs can be seen as restricted versions of this most general structure. CPMMs are not necessarily more misspecified than GMMs – both are at risk of misspecification based on researchers' modeling decisions. CPMMs happened to be (intentionally) more misspecified in our simulation study because the generating model was a GMM, so GMM or GMMC could degrade even further if the covariance structure were not properly specified.

Nonetheless, if researchers continue to be concerned about possible covariance structure misspecification in CPMMs, traditional choices like compound symmetry, Toeplitz, or autoregressive can be sidestepped if they are deemed insufficient. An alternative method would be to inspect the observed covariance or correlation matrix of the repeated measures for guidance about the appropriate structure. If the structure does not appear to follow one of the traditional structures, the flexibility of the structural equation modeling framework under which CPMMs fall makes custom covariance structures easy to specify (Grimm & Widaman, 2010). For example, the observed marginal correlation matrix from the NLSY data has a form that may deviate from traditional structures:

$$\begin{bmatrix} 1 & & & \\ .66 & 1 & & \\ .54 & .78 & 1 & \\ .45 & .76 & .80 & 1 \end{bmatrix}$$

One could customize a marginal structure where the (3,2), (4,2), and (4,3) elements (.78, .76, and ,.80) are captured by a one parameter while the (2,1), (3,1), and (4,1) elements (.66, .54, and .45) are each captured by unique parameters so the marginal correlation structure would be

$$\begin{bmatrix} 1 & & & \\ \rho_1 & 1 & & \\ \rho_3 & \rho_2 & 1 & \\ \rho_4 & \rho_2 & \rho_2 & 1 \end{bmatrix}$$

This does not adhere to one of the traditional forms, but in terms of implementation, it is no more difficult to fit in software as the traditional forms and could help further reduce the risk of misspecification (example M*plus* code for this structure is provided in the Appendix).

**Concluding Remarks for Empirical Researchers**

Random effects models are the default method for longitudinal data analysis in psychology. Though reasons for this preference are defensible in the context of data without latent classes (e.g., Grimm & Stegmann, 2018), the role of the within-class random effects is much reduced in latent class models. The classes provide the primary latent information for the research questions by qualitatively grouping would-be continuous random effects, thereby reducing the dimensionality of the solution. This makes the extra latent information provided by the within-class random effects of little utility in many cases, beyond properly specifying the marginal covariance. Though not incorrect, using random effect strictly to pattern the marginal covariance is inefficient compared to approaches taken by population-averaged models.

In naturally complex models like mixture models for longitudinal data, unnecessarily employing within-class random effects leads to exceedingly high levels of non-convergence and inadmissible solutions. The current method to circumvent these issues is to apply cross-class constraints; however, though this does improve convergence, it also raises additional issues and typically results in highly biased class trajectories and poor class enumeration. Much of the computational complexity of GMMs can be avoided with more theoretically aligned CPMMs, which avoid the within-class random effects but retain the flexibility to properly model the covariance structure.

Model choice in mixture modeling should operate in the same way as any other model: the simplest model that can answer the question at hand should be preferred. We encourage researchers to consider which questions they wish to answer and to think critically about which is the simplest model capable of answering these questions. In our view, GMMs are rarely the answer to this question if CPMMs are simultaneously considered.

## Open Sciences Practices Statement

All simulated data sets, simulation code, and simulation data management code are provided and are publically available on the Open Science Framework. The anonymized link to access this information is https://osf.io/yh6kf/. Data used for the empirical example as well as all software code used to analyze the empirical example data are also accessible via the same OSF link. No aspects of this simulation study were preregistered.

## References


1. Azevedo, C. L., Fox, J. P., & Andrade, D. F. (2016). Bayesian longitudinal item response modeling with restricted covariance pattern structures. *Statistics and Computing*, *26*, 443-460.

2. Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338-363. http://dx.doi.org/10.1037/1082-989X.8.3.338

3. Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods*, *16*, 373-390. http://dx.doi.org/10.1037/a0025813

4. Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and psychopathology*, *9*, 291-319. https://doi.org/10.1017/S095457949700206X

5. Bonanno, G. A. (2004). Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events?. *American Psychologist*, *59*, 20-28. http://dx.doi.org/10.1037/0003-066X.59.1.20

6. Burton, P., Gurrin, L., & Sly, P. (1998). Tutorial in biostatistics. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modeling. *Statistics in Medicine*, *17*, 1261-1291. http://dx.doi.org/10.1002/0470023724.ch1a

7. Codd, C. L., & Cudeck, R. (2014). Nonlinear random-effects mixture models for repeated measures. *Psychometrika*, *79*, 60-83. http://dx.doi.org/10.1007/s11336-013-9358-9

8. Cole, V. T., & Bauer, D. J. (2016). A note on the use of mixture models for individual prediction. *Structural Equation Modeling*, *23*, 615-631. https://doi.org/10.1080/10705511.2016.1168266

9. Cudeck, R., & Codd, C. L. (2012). A template for describing individual differences in longitudinal data with application to the connection between learning and ability. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 3-24). Charlotte, NC: Information Age

10. Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529-569. http://dx.doi.org/10.1207/s15327906mbr3804_5

11. Davidian, M., & Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. New York: Chapman & Hall.

12. Depaoli, S., van de Schoot, R., van Loey, N., & Sijbrandij, M. (2015). Using Bayesian statistics for modeling PTSD through Latent Growth Mixture Modeling: implementation and discussion. *European Journal of Psychotraumatology, 6,* 27516. http://dx.doi.org/10.3402/ejpt.v6.27516

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

14. Diallo, T. M., Morin, A. J., & Lu, H. (2016). Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling, 23,* 507-531. http://dx.doi.org/10.1080/10705511.2016.1169188

15. Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. L. (2002). A*nalysis of longitudinal data* (2nd ed.). New York: Oxford University Press.

16. Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling, 21,* 534-552.

17. Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15,* 75-95. http://dx.doi.org/10.1080/10705510701758281

18. Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011) *Applied longitudinal analysis* (2nd ed.). Philadelphia: Wiley.

19. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466-491.

20. Gilthorpe, M. S., Dahly, D. L., Tu, Y. K., Kubzansky, L. D., & Goodman, E. (2014). Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *Journal of Developmental Origins of Health and Disease, 5,* 197-205.

21. Grimm, K. J., & Stegmann, G. (2018). Modeling change trajectories with count and zero-inflated outcomes: Challenges and recommendations. *Addictive Behaviors,* advance online publication.

22. Grimm, K. J., & Widaman, K. F. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling, 17,* 424-442. http://dx.doi.org/10.1080/10705511.2010.489006

23. Harring, J. R., & Blozis, S. A. (2016). A note on recurring misconceptions when fitting nonlinear mixed models. *Multivariate Behavioral Research*, *51*, 805-817. http://dx.doi.org/10.1080/00273171.2016.1239522

24. Harring, J. R., & Blozis, S. A. (2014). Fitting correlated residual error structures in nonlinear mixed-effects models using SAS PROC NLMIXED. *Behavior research methods*, *46*(2), 372-384.

25. Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, *51*, 354-367. http://dx.doi.org/10.1080/00461520.2016.1207176

26. Hox J. (2010). *Multilevel analyses: Techniques and applications*. Mahwah, NJ: Erlbaum

27. Infurna, F. J., & Grimm, K. J. (2017). The use of growth mixture modeling for studying resilience to major life stressors in adulthood and old age: Lessons for class size and identification and model selection. *Journals of Gerontology Series B, 73*, 148-159. http://dx.doi.org/10.1093/geronb/gbx019

28. Infurna, F. J., & Luthar, S. S. (2016). Resilience to major life stressors is not as common as thought. *Perspectives on Psychological Science*, *11*, 175-194. http://dx.doi.org/10.1177/1745691615621271

29. Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, *42*, 805-820. http://dx.doi.org/10.2307/2530695

30. Jung, T., & Mickrama, K. A. S. (2007). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302–317.

31. Kreuter, F., & Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology, 24*, 1-31. http://dx.doi.org/10.1007/s10940-007-9036-0

32. Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*, 377-389.

33. Li, M., Harring, J. R., & Macready, G. B. (2014). Investigating the feasibility of using Mplus in the estimation of growth mixture models. *Journal of Modern Applied Statistical Methods, 13*, 31. http://dx.doi.org/10.22237/jmasm/1398918600

34. Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13-22. http://dx.doi.org/10.1093/biomet/73.1.13

35. Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement, 74*, 557-584. http://dx.doi.org/10.1177/0013164413519798

36. Liu, S., Rovine, M. J., & Molenaar, P. (2012a). Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods, 17*, 15-30. http://dx.doi.org/10.1037/a0026971

37. Liu, S., Rovine, M. J., & Molenaar, P. C. (2012b). Using fit indexes to select a covariance model for longitudinal data. *Structural Equation Modeling, 19*, 633-650.

38. Lix, L., & Sajobi, T. (2010). Discriminant analysis for repeated measures data: a review. *Frontiers in psychology, 1*, 146.

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

40. McNeish, D. & Matta, T. (2018). Differentiating between mixed effects and latent curve approaches to growth modeling. *Behavior Research Methods, 50*, 1398-1414. http://dx.doi.org/10.3758/s13428-017-0976-5

41. McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*, 114-140. http://dx.doi.org/10.1037/met0000078

42. Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23-43. http://dx.doi.org/10.1037/1082-989X.5.1.23

43. Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*, 201-218. http://dx.doi.org/10.1207/s15366359mea0204_1

44. Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112-117.

45. Morin, A. J., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling, 18*, 613-648. http://dx.doi.org/10.1111/j.1467-8721.2009.01619.x

46. Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation, 21*, 343-370. http://dx.doi.org/10.1080/13803611.2015.1057161

47. Muthén, B.O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: *New opportunities for latent class/latent growth modeling*. In Collins, L. M. & Sayer, A. (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, D.C.: American Psychological Association.

48. Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research, 24*, 882-891. https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

49. Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*, 371-402. http://dx.doi.org/10.1037/1082-989X.2.4.371

50. Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463-469. http://dx.doi.org/10.1111/j.0006-341X.1999.00463.x

51. Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139-157. http://dx.doi.org/10.1037/1082-989X.4.2.139

52. Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods, 6*, 18-34. http://dx.doi.org/10.1037/1082-989X.6.1.18

53. Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

54. 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, 23*, 782-797. http://dx.doi.org/10.1080/10705511.2016.1221313

55. Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling, 21*, 439-454.

56. Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535-569. http://dx.doi.org/10.1080/10705510701575396

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

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

59. Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: cause for concern?. *Journal of Abnormal Psychology, 120,* 322-335. http://dx.doi.org/10.1037/a0021813

60. Stegmann, G., & Grimm, K. J. (2018). A new perspective on the effects of covariates in mixture models. *Structural Equation Modeling, 25,* 167-178. http://dx.doi.org/10.1080/10705511.2017.1318070

61. Sterba, S. K., & Bauer, D. J. (2014). Predictions of individual change recovered with latent class or random coefficient growth models. *Structural Equation Modeling, 21,* 342-360. https://doi.org/10.1080/10705511.2014.915189

62. Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology, 22,* 239-254. https://doi.org/10.1017/S0954579410000015

63. Thurstone, L. L. (1919). The learning curve equation. *Psychological Monographs, 26,* 1-51. http://dx.doi.org/10.1037/h0093187

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

65. van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & Van Loey, N. E. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research, 53,* 267-291. http://dx.doi.org/10.1080/00273171.2017.1412293

66. Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91,* 217-221.

67. Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York:
68. Springer.

69. Vonesh, E. F. (2013). *Generalized linear and nonlinear models for correlated data: Theory and applications using SAS*. New York: Springer-Verlag.

70. Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.

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

72. Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121-130. http://dx.doi.org/10.2307/2531248

73. Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics, 44*, 1049-1060. http://dx.doi.org/10.2307/2531734

# Appendix

*Mplus Code for Fitting Marginal Growth Mixture Models*

The following code shows how to fit marginal models with exchangeable, Autogressive, Toeplitz, and unstructured covariance matrices. All code will assume that the residual variances are heterogeneous and that three classes are being extracted. To provide some context, we will use the Curran data; this data can be downloaded from http://www.joophox.net/mlbook2/MLbook.htm or can be found on the first Author's Open Science Framework page for this project (https://osf.io/yh6kf/)

M*plus* code for marginal growth mixture models is not appreciably different from standard growth mixture models. The primary difference stems from how the covariance matrices are modeled. Other options (e.g., TECH11 for the Lo-Mendell-Rubin test or TECH14 for the BLRT) are not changed in M*plus* when using a marginal growth mixture model compared to a standard growth mixture model.

To specify a marginal model in M*plus*, the "trick" is to not partition the variance so that all the variability is housed in what is normally the residual covariance matrix. This is done by constraining the growth factor variances to 0. The "residual variances" (which now represent the marginal covariance structure) are then structured as desired.

## Exchangeable

*The exchangeable structure models all covariances to be equal, regardless of how far apart the measures are in time. It is also referred to as a compound symmetric or equi-correlated structure. A marginal model with an exchangeable structure is equivalent to a random intercepts model. Though parsimonious in that all covariances are modeled with a single parameter, it can be too simplistic for some types of data where repeated measures are less related when they are further apart in time.*

```
data:
  FILE IS \\Client\C$\Users\Dan\Desktop\curnl.csv;
   VARIABLE:
        Names are Anti1-anti4 read1-read4 sex momage kidage homecog homeemo;
           Usevariables are read1-read4;
  CLASSES = c(3);
     Missing are .

  ANALYSIS:

  TYPE = MIXTURE;
  STARTS = 50 10;
  STITERATIONS = 100;
  ITERATIONS = 1000;
  SDITERATIONS = 250;
  MITERATIONS = 500;
  MCONVERGENCE = 1E-5;

  MODEL:
```

```
%overall%

I S Q| read1@0 read2@1 read3@2 read4@3;

%c#1%
I S Q| read1@0 read2@1 read3@2 read4@3;
read1 with read2 (a);
read1 with read3 (a);
read1 with read4 (a);
read2 with read3 (a);
read2 with read4 (a);
read3 with read4 (a);
read1-read4 ;
I@0; S@0; Q@0;

%c#2%

I S Q| read1@0 read2@1 read3@2 read4@3;
read1 with read2 (b);
read1 with read3 (b);
read1 with read4 (b);
read2 with read3 (b);
read2 with read4 (b);
read3 with read4 (b);
read1-read4 ;
I@0; S@0; Q@0;

%c#3%

I S Q| read1@0 read2@1 read3@2 read4@3;
read1 with read2 (c);
read1 with read3 (c);
read1 with read4 (c);
read2 with read3 (c);
read2 with read4 (c);
read3 with read4 (c);
read1-read4 ;
I@0; S@0; Q@0;
```

**Autoregressive**

*The autoregressive structure allows for observations to be less related as the observations grow further apart. To maintain parsimony, the autoregressive structure only estimates a single covariance parameter ρ. The covariance at Lag-1 are equal to ρ, Lag-2 are equal to $\rho^2$, Lag-3 are equal to $\rho^3$, etc. Though this seems a little tricky to program in Mplus, it is rather straightforward with a MODEL CONSTRAINT command.*

```
data:
  FILE IS \\Client\C$\Users\Dan\Desktop\curnl.csv;
   VARIABLE:
        Names are Anti1-anti4 read1-read4 sex momage kidage homecog homeemo;
           Usevariables are read1- read4;
  CLASSES = c(3);
     Missing are .

  ANALYSIS:

  TYPE = MIXTURE;
  STARTS = 50 10;
  STITERATIONS = 100;
  ITERATIONS = 1000;
  SDITERATIONS = 250;
  MITERATIONS = 500;
  MCONVERGENCE = 1E-5;

  MODEL:

  %overall%

  I S Q| read1@0 read2@1 read3@2 read4@3;

  %c#1%
  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 (a1);
  read1 with read3 (a2);
  read1 with read4 (a3);
  read2 with read3 (a1);
  read2 with read4 (a2);
  read3 with read4 (a1);
  read1-read4 ;
  I@0; S@0; Q@0;

  %c#2%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 (b1);
  read1 with read3 (b2);
  read1 with read4 (b3);
  read2 with read3 (b1);
  read2 with read4 (b2);
  read3 with read4 (b1);
  read1-read4 ;
  I@0; S@0; Q@0;
```

```
  %c#3%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 (c1);
  read1 with read3 (c2);
  read1 with read4 (c3);
  read2 with read3 (c1);
  read2 with read4 (c2);
  read3 with read4 (c1);
  read1-read4 ;
  I@0; S@0; Q@0;

Model Constraint:
new(rho1);
new(rho2);
new(rho3);

a1= rho1;
a2= rho1*rho1;
a3= rho1*rho1*rho1;

b1= rho2;
b2= rho2*rho2;
b3= rho2*rho2*rho2;

c1= rho3;
c2= rho3*rho3;
c3= rho3*rho3*rho3;
```

**Toeplitz**

*The main idea of a Toeplitz structure is that the covariance at each equivalent lag is equal but that unequal lags are freely estimated. So, covariances for time-points that are 1 lag apart are constrained to be the same; the same is true for covariances of time-points that are 2 lags apart. However, unlike an autoregressive structure, the Lag-1 covariance is completely unrelated to the Lage-2 covariance (e.g., the Lag-2 covariance is not the square of the Lag-1 covariance as with the autoregressive structure).*

```
data:
  FILE IS \\Client\C$\Users\Dan\Desktop\curnl.csv;
   VARIABLE:
        Names are Anti1-anti4 read1-read4 sex momage kidage homecog homeemo;
           Usevariables are read1-read4;
  CLASSES = c(3);
     Missing are .

  ANALYSIS:

  TYPE = MIXTURE;
  STARTS = 50 10;
  STITERATIONS = 100;
  ITERATIONS = 1000;
  SDITERATIONS = 250;
  MITERATIONS = 500;
  MCONVERGENCE = 1E-5;

  MODEL:

  %overall%

  I S Q| read1@0 read2@1 read3@2 read4@3;

  %c#1%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 (a1);
  read1 with read3 (a2);
  read1 with read4 (a3);
  read2 with read3 (a1);
  read2 with read4 (a2);
  read3 with read4 (a1);
  read1-read4 ;
  I@0; S@0; Q@0;

  %c#2%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 (b1);
  read1 with read3 (b2);
  read1 with read4 (b3);
  read2 with read3 (b1);
  read2 with read4 (b2);
  read3 with read4 (b1);
  read1-read4 ;
```

```
I@0; S@0; Q@0;



%c#3%

I S Q| read1@0 read2@1 read3@2 read4@3;
read1 with read2 (c1);
read1 with read3 (c2);
read1 with read4 (c3);
read2 with read3 (c1);
read2 with read4 (c2);
read3 with read4 (c1);
read1-read4 ;
I@0; S@0; Q@0;
```

**Unstructured**

*An unstructured matrix is the most flexible but the least parsimonious. Every element of the covariance matrix is uniquely estimated, so if there are many repeated measures or many latent classes, this can quickly add up to a large number of parameters. Reaching convergence with an unstructured matrix can be difficult and often requires a very large sample. If the model converges, it can sometimes provide a good initial idea for alternative structures that may be more parsimonious (e.g., if the estimates look approximately equal down the diagonals, perhaps an autoregressive or Toeplitz structure may provide similar fit with fewer parameters).*

```
data:
  FILE IS \\Client\C$\Users\Dan\Desktop\curnl.csv;
   VARIABLE:
        Names are Anti1-anti4 read1-read4 sex momage kidage homecog homeemo;
           Usevariables are read1-read4;
  CLASSES = c(3);
     Missing are .

  ANALYSIS:

  TYPE = MIXTURE;
  STARTS = 50 10;
  STITERATIONS = 100;
  ITERATIONS = 1000;
  SDITERATIONS = 250;
  MITERATIONS = 500;
  MCONVERGENCE = 1E-5;

  MODEL:

  %overall%

  I S Q| read1@0 read2@1 read3@2 read4@3;

  %c#1%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 ;
  read1 with read3 ;
  read1 with read4 ;
  read2 with read3 ;
  read2 with read4 ;
  read3 with read4 ;
  read1-read4 ;
  I@0; S@0; Q@0;

  %c#2%

  I S Q| read1@0 read2@1 read3@2 read4@3;
  read1 with read2 ;
  read1 with read3 ;
  read1 with read4 ;
  read2 with read3 ;
  read2 with read4 ;
  read3 with read4 ;
```

```
read1-read4 ;
I@0; S@0; Q@0;

%c#3%

I S Q| read1@0 read2@1 read3@2 read4@3;
read1 with read2 ;
read1 with read3 ;
read1 with read4 ;
read2 with read3 ;
read2 with read4 ;
read3 with read4 ;
read1-read4 ;
I@0; S@0; Q@0;
```