



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Bayesian approach to piecewise growth mixture modeling: Issues and applications in school psychology

Ihnwhi Heo^{*}, Sarah Depaoli, Fan Jia, Haiyan Liu

Department of Psychological Sciences, University of California, Merced, United States

ARTICLE INFO

Guest Action Editor: Garret Hall
Editor: Craig A. Albers

Keywords:

Bayesian mixture model
Class enumeration
Class separation
Mathematics achievement
Piecewise growth mixture model
Prior sensitivity

ABSTRACT

Bayesian piecewise growth mixture models (PGMMs) are a powerful statistical tool based on the Bayesian framework for modeling nonlinear, phasic developmental trajectories of heterogeneous subpopulations over time. Although Bayesian PGMMs can benefit school psychology research, their empirical applications within the field remain limited. This article introduces Bayesian PGMMs, addresses three key methodological considerations (i.e., class separation, class enumeration, and prior sensitivity), and provides practical guidance for their implementation. By analyzing a dataset from the Early Childhood Longitudinal Study-Kindergarten Cohort, we illustrate the application of Bayesian PGMMs to model piecewise growth trajectories of mathematics achievement across latent classes. We underscore the importance of considering both statistical criteria and substantive theories when making decisions in analytic procedures. Additionally, we discuss the importance of transparent reporting of the results and provide caveats for researchers in the field to promote the wide usage of Bayesian PGMMs.

1. Introduction

Bayesian piecewise growth mixture modeling combines the advantages of Bayesian methods with a statistical framework to identify distinct subgroups within a population and understand how various factors influence these subgroups at different stages over time. This approach can enhance our understanding of developmental trajectories, particularly when analyzing groups (e.g., students, teachers) segmented into distinct phases or “pieces”. In this article, we introduce Bayesian piecewise growth mixture models (PGMMs) for school psychologists. Unlike general mixture modeling, PGMMs have the potential to improve the quality of interventions and outcomes being examined by describing and understanding nonlinear development across the segmentation.

We believe this approach has relevance to the types of questions being addressed within the field of school psychology. Our main goal is to make these methods more accessible to a broad audience by describing the concepts and applications of Bayesian PGMMs. As a preview of the topics covered, we begin with a brief overview of the core elements that comprise this article, including (a) mixture modeling, (b) estimation issues, and (c) the relevance of Bayesian PGMMs in school psychology. Then we move into sections outlining more details for these methods and topics. As we work through these various elements, it is important to note that [Table 1](#) provides a guide for defining some of the technical jargon that is necessary when introducing these methods.

^{*} Corresponding author at: Department of Psychological Sciences, University of California, Merced, 5200 N. Lake Road, Merced, CA 95343, United States.

E-mail address: ihnwhi.heo@gmail.com (I. Heo).

<https://doi.org/10.1016/j.jsp.2024.101366>

Received 11 August 2023; Received in revised form 30 August 2024; Accepted 31 August 2024

Available online 27 September 2024

0022-4405/© 2024 The Authors. Published by Elsevier Ltd on behalf of Society for the Study of School Psychology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1.1. Why mixture modeling in school psychology?

A central research interest of school psychologists is to gain insights into the complex psychological and behavioral processes of individuals in school settings. Traditionally, it is assumed that these individuals come from a single population, meaning all individuals share the same properties and the influence of hypothesized covariates is applied uniformly. However, empirical studies suggest that individuals in school environments can be categorized into qualitatively different subpopulations (e.g., Byrd & Andrews, 2016; Coyle et al., 2021; Herman et al., 2018; Ingram et al., 2019; Ren et al., 2021). These subpopulations may be substantively important to decipher, but they may also be unknown to the researcher.

To address this, mixture modeling is a statistical technique that can account for this heterogeneity within a population and identify hidden subgroups (McLachlan & Peel, 2000). For school psychologists, detecting differences across these latent subpopulations—or classes—can enhance their understanding of different mechanisms and intervention effects on social, cognitive, health, and achievement-related outcomes in schools.

Mixture modeling operates on the principle that unobserved heterogeneity exists in the data, meaning there are underlying groups within the population that we cannot directly observe. Researchers must determine these unknown class memberships based on observed patterns (e.g., Does an individual *A* belong to Group 1 or Group 2 given the observed data patterns?). The observed data patterns, either through the outcome of interest or other covariates, are clues about how individuals may group together to form unobserved classes. Put simply, mixture modeling suggests that different covariates may influence different groups of students or teachers in different ways. The challenge then becomes identifying the correct set of classes so that group differences and similarities can be accurately interpreted. The question of how many unobserved groups exist and to which group each individual should be assigned is closely tied to the accuracy of the model estimation technique. In other words, can we accurately determine the number of unobserved groups and correctly assign each individual to their respective groups? In the next section, we introduce the Bayesian estimation framework, a method that is particularly effective for addressing these issues.

Table 1
Glossary of technical terms based on Depaoli (2021).

Terms	Conceptual Definition
Burn-in samples	The initial samples to form the Markov chain are often highly dependent on starting values and may not precisely approximate the target posterior distribution. These initial samples are referred to as the burn-in samples and they are discarded from the estimated posterior. The burn-in samples are also referred to as the warm-up samples.
Class enumeration	The ability to properly identify the number of latent classes in a sample. In other words, the ability to identify the number of populations that the sample data represents. Information criteria (i.e., DIC, WAIC, and LOOIC) are used for class enumeration.
Class proportions	The estimated size of the latent classes. Class proportions should add up to 1.
Class separation	How similar or dissimilar the latent classes are to one another. Highly separated classes in a piecewise growth mixture model would show very different piecewise growth trajectories and/or knot locations. Poorly separated classes would show similar growth trajectories and/or knot locations. The more similar classes are, the more difficult it is to properly identify them without the use of prior information.
Convergence	The post-burn-in phase of the Markov chain must be assessed for convergence through diagnostics. Convergence indicates a stable mean and variance of the chain throughout the entire section representing the posterior distribution.
Diffuse priors	Priors that reflect complete uncertainty about population parameters. For instance, a normal prior distribution that is diffuse has a very large variance hyperparameter that may make this prior nearly flat.
Growth factors	The latent variables (i.e., factors) in the piecewise growth mixture model that govern the specific growth trajectory being estimated. These factors include the intercept (i.e., starting point of the trajectory) and the respective slopes (i.e., first linear rate of change for the trajectory).
Gibbs sampler	This is one type of Markov chain Monte Carlo algorithms used for sampling values from an unknown probability distribution.
Growth trajectory	The rate of change, or growth, that is tested in the piecewise growth mixture model. Multiple growth trajectories are obtained within the piecewise growth mixture model to uncover the overall rate of change across individuals from different latent classes.
Hyperparameter	The parameters that form a prior distribution are referred to as hyperparameters. For instance, the normal distribution is defined by a mean and a variance, and these represent the hyperparameters. Hyperparameters can be manipulated to create more or less informed priors.
Informativeness	The degree of prior informativeness captures how much (un)certainity there is in the prior. Informativeness can be captured along a continuum from complete uncertainty to relative certainty.
Informative priors	Priors that reflect a high degree of certainty surrounding a population parameter value.
Latent class	An unobserved group of individuals modeled through a mixture model.
Likelihood	A conditional probability of the data given the unknown model parameters (composing the model).
Markov chain Monte Carlo	An iterative process implementing sampling methods used to construct an estimated posterior through an indirect process involving simulations.
Posterior distribution	A distribution that captures prior knowledge and observed data (via the likelihood).
Prior distribution	A probability distribution that represents beliefs about model parameters. Priors are typically constructed before examining the sample data.
Prior sensitivity analysis	An analysis that allows the researcher to examine the impact of different prior settings on final model estimates to determine how robust or varied results are when priors are modified.
Weakly informative priors	Priors that contain some information about the population parameter while remaining less certain than informative priors.

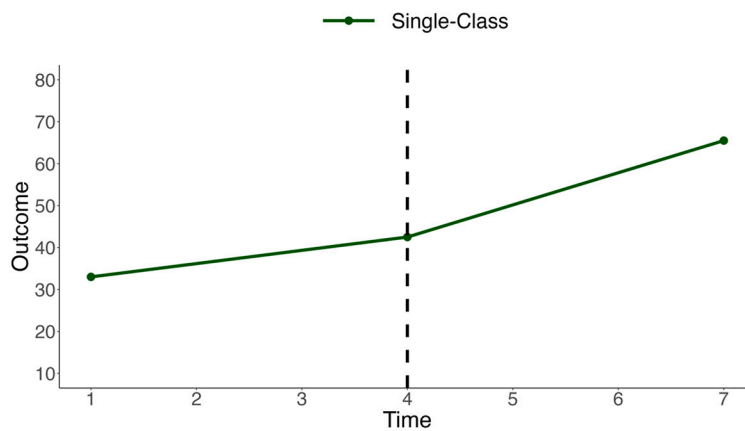
1.2. Why Bayesian methods?

The frequentist estimation framework, such as maximum likelihood via expectation maximization, has traditionally been the dominant method for estimating mixture models (Jedidi et al., 1997; Muthén & Shedden, 1999; Tueller & Lubke, 2010). However, this approach can sometimes struggle with accurately identifying the number of latent classes, as well as class membership (e.g., Which particular latent class an individual should belong to?). The estimation issues within the frequentist estimation framework are especially prevalent for complex statistical models (e.g., models with mixture components or a multilevel structure) or when smaller sample sizes are present. When these complexities are present, the Bayesian estimation framework offers a promising alternative that has been shown to provide more accurate estimation results (Depaoli, 2021; Frühwirth-Schnatter, 2010; Kaplan & Depaoli, 2012; Lee, 2007).

Bayesian estimation methods have similarities to the frequentist methods that most researchers are familiar with, but there is an important difference that alters the implementation and interpretation of results. That difference is through the incorporation of *prior distributions* (or *priors*). Priors are used to incorporate knowledge about each model parameter into the estimation process. That knowledge can come from a variety of places, including previous research findings, expert opinions, or any other relevant information that informs the likely values of model parameters. The purpose is to introduce some “extra information” about each model parameter through the prior; this is notably different from traditional frequentist methods, which do not incorporate any information about model parameters into the estimation process.

Prior distributions are formulated from information coming from a variety of places as noted and combined with information from the observed data to estimate model parameters. In the case of mixture models, Bayesian priors are extremely helpful when determining the number of latent classes, as well as properly assigning individuals to those classes. Furthermore, Bayesian estimation is particularly helpful in cases of small sample sizes or when latent classes have overlap or similarities that make them harder to

(a) Growth Trajectories of a Non-Mixture Piecewise Growth Curve Model



(b) Growth Trajectories of a 2-Class PGMM.

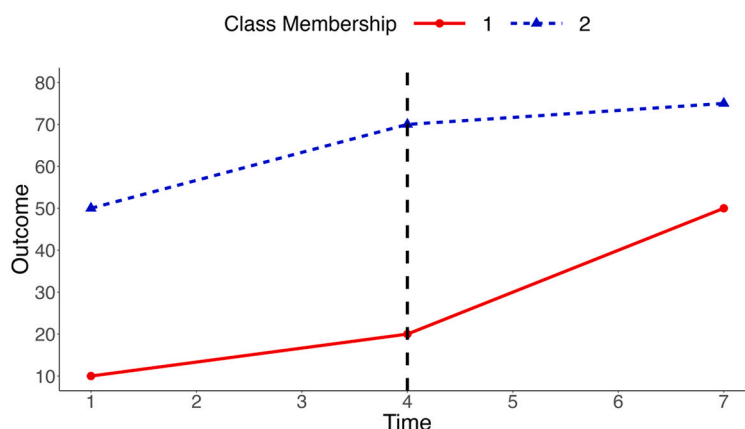


Fig. 1. A hypothetical example of piecewise growth trajectories.

distinguish from one another (van de Schoot et al., 2015, 2018; Zhang et al., 2007).

1.3. Practical relevance of Bayesian PGMMs in school psychology

Analyzing data with latent classes has been applied to various forms of data measured from a single time point to multiple time points. PGMM is a popular method for investigating complex growth over time. In the analysis of longitudinal data, researchers often observe distinct, phasic developmental periods rather than a smooth and uninterrupted growth trajectory. In the non-mixture implementation of this model (i.e., the piecewise growth curve model), two growth segments are joined together by a changepoint (or *knot*). The knot represents a point in time where the growth rate changes, sometimes abruptly. Fig. 1a illustrates an example of this pattern. The figure shows a piecewise growth trajectory of a single class (i.e., non-mixture) shown by the green solid line where the growth rate changes at Time 4, indicated by the vertical black dashed line. The trajectory before and after the knot shows distinct linear segments.

Adding a mixture element to this model allows for multiple groups, each represented by different growth trajectories. These multiple groups can have the same knot location (i.e., groups that change at the same time point) or different locations (i.e., groups that change at different time points). In addition, there can be a single knot (i.e., one time point where a change in growth occurs) or multiple knots (i.e., multiple time points where there is a shift in the growth pattern or rate) depending on the complexity of the outcome variable being tracked over time. An example of a PGMM with two latent classes can be found in Fig. 1b. Note that there is a clear and distinct growth pattern for the two groups. Even though they each have a changepoint at Time 4, they display different growth patterns (i.e., intercepts and slopes) before and after the knot at Time 4.

There are many examples of research questions in school psychology designed to advance our understanding of individuals (e.g., students, teachers) via PGMMs. As a hypothetical example, assume that a researcher wants to implement a new intervention for enhancing students' math scores in the classroom. The researcher may opt to collect data at multiple time points leading up to the intervention and then also for multiple time points after the intervention. This longitudinal design would allow the researcher to track math scores before the intervention and assess the impact of the intervention by examining scores collected afterward. The intervention point, located at the knot, may produce an abrupt change in math scores, thus producing two segments of the growth model (i.e., pre- and post-intervention). The piecewise growth curve model (i.e., non-mixture) can be used to capture math ability over time in this context, with the knot placed at the time point where the intervention was implemented (e.g., akin to Time 4 in Fig. 1a). If the researcher suspected multiple latent groups existed (e.g., the piecewise growth trajectories should be different between high-achieving and low-achieving groups of students), then the PGMM could be implemented. In that case, the researcher would estimate a series of models to determine the optimal number of latent subgroups that exist (e.g., akin to the two-class structure in Fig. 1b). Each subgroup can represent different growth patterns as well as capture different effects from the intervention.

1.4. Organization of the remaining sections

The above sections provided the basic context as follows. Mixture models can help disentangle substantively different yet unobserved subgroups from one another. However, frequentist methods may struggle to identify the correct subgroups, especially when models are complex or sample sizes are small. As a result, researchers can turn to Bayesian estimation, which can improve the estimation of mixture modeling using "extra information" about the model parameters, defined through *priors*. A notable and relevant type of mixture model to school psychology inquiries is the PGMM, which can track phasic growth over time for different unobserved subgroups. The interesting aspect of the PGMM is that it can incorporate abrupt changes in growth over time through a *knot*. With these basic ideas in mind, we now transition to sections detailing more of the technical aspects of this modeling approach, Bayesian estimation, and important methodological considerations in the implementation of PGMMs. Then we present an empirical application of Bayesian PGMMs to longitudinal data. We conclude with a summary of important findings, discuss additional issues, suggest future research directions, and provide practical advice for researchers in the field of school psychology looking to implement these methods. Furthermore, we provide Appendices that offer technical details beyond the scope of the text.

2. Piecewise growth mixture modeling

PGMM is a methodological option to adequately capture the varying growth trajectories across different developmental periods and latent classes. This approach allows researchers to break down longitudinal trajectories into separate distinct phases and analyze them within subpopulations. As such, PGMMs offer flexibility to capture complex trajectory patterns with multiple growth phases across different latent classes. In a PGMM, respective phases are linked to the previous or next phase via knots. Locations of knots can be predetermined (Depaoli et al., 2023; Heo et al., 2024) or estimated freely (Kohli, Hughes, et al., 2015; Lock et al., 2018). The presence of knots allows researchers to implement interventions and understand or construct substantive theories surrounding piecewise growth trajectories. Depending on the specific research questions, PGMMs can be tailored to accommodate varying degrees of complexity. To illustrate these methods, the present article focuses on a PGMM with two linear segments connected by a single knot that is freely estimated due to its simplicity and general applicability (e.g., Kohli, Hughes, et al., 2015; Kohli & Sullivan, 2019).

2.1. Visualization of trajectories

Before we introduce the functional form of PGMMs, it is important to have a clear visualization of how the model can be

represented. PGMMs can be displayed through plots of growth trajectories. For this purpose, we revisit Fig. 1b that contains a hypothetical example of a two-class PGMM with a knot located midway through the trajectories. Imagine this scenario as a school psychology study where the separate growth trajectories and knot locations across the two classes are the result of applying a Bayesian PGMM to describe the piecewise growth before and after an intervention. This intervention could be the implementation of a new math course or a change of teachers. Here, the outcome variable is students' math examination scores.

In this example, two latent classes exist, with Class 1 in Fig. 1b depicted by the red solid line and Class 2 represented by the blue dotted line. The knot is located at time point 4, as indicated by the vertical black dashed line. Class 1 shows a steeper growth trajectory after the knot location. Conversely, Class 2 demonstrates a trajectory that becomes less steep after the knot location.

Substantively, students from Class 2 can be defined as a high-achieving group whose performance was high before the intervention and continued to increase after the intervention, albeit at a decelerated learning rate. Conversely, students from Class 1 can be defined as a low-achieving group who had low math achievement before the intervention, but whose math scores increased dramatically with acceleration after the intervention. This suggests that the intervention was particularly effective for the low-achieving group. As such, we can observe a clear distinction between these two classes in terms of their growth trajectories.

2.2. Model definition

Next, we explain how PGMMs are defined. We strive to keep the technical details of PGMMs brief and succinct, yet researchers may need more explanation about the model formulas. To this end, we provide illustrations of additional technical details behind PGMMs in Appendix A.

The mixture modeling approach, including PGMMs, assumes that observed data are generated from a mixture distribution. Mixture models like PGMMs can be viewed as combining several subpopulations, each with its own distinct characteristics. In this framework, there are C latent classes, and some parameters are allowed to vary across these classes. Researchers need to estimate the number of latent classes (C) and then determine class-specific parameters, such as class-specific intercepts and slopes. For example, in a population with two latent classes, researchers would estimate parameters for each class separately, capturing unique growth trajectories and changes over time.

A PGMM can be defined from a structural equation modeling perspective (Grimm et al., 2016). To formulate a PGMM using the structural equation modeling framework, we consider a measurement model part and a structural model part. The measurement model links the observed data to growth factors, which can be formulated as

$$y_{ic} = \Lambda_y \eta_{ic} + \epsilon_{ic}, \tag{1}$$

where y_{ic} is a vector of repeatedly measured dependent variables for an individual i in class c , Λ_y refers to a $T \times m$ matrix of factor loadings where T is the number of time points and m is the number of growth factors (i.e., latent intercept and latent slope), η_{ic} is an $m \times 1$ vector of growth factors, and ϵ_{ic} is a vector of T residuals that cannot be accounted for by the trajectory in class c . In any formulation of PGMMs, specifying the Λ_y matrix is important because piecewise growth trajectories are expressed. Usually, the first column consists of 1s because it is associated with the intercept that does not change over time. The second and the remaining subsequent columns indicate patterns of piecewise trajectories hypothesized by a researcher.

Suppose we have measured observed variables across seven equidistant time points and will fit a PGMM that consists of two linear segments joined at a knot at the fourth time point, as shown in Fig. 1b. For linear slope factors, the first time point is typically set to be 0. As illustrated in Fig. 1b, the knot at Time 4 is included in the model, with a slope estimated prior to Time 4 and another slope estimated after Time 4. The pre-knot slope captures the rate of change in measurements leading up to the knot location, whereas the post-knot slope captures the rate of change in measurements following the knot location. This approach allows the model to separately account for changes before and after the knot.¹

The structural model forms relationships between growth factors and can be written as

$$\eta_{ic} = \alpha_c + \zeta_{ic}, \tag{2}$$

¹ For those interested in the specific matrix formulation in this example, the Λ_y matrix used to define the factor loadings can be represented as follows:

$$\Lambda_y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \end{bmatrix}.$$

In this matrix, the second column defines the loadings of the first slope factor that captures the rate of change in measurements leading up to the knot location. The third column defines the loadings of the second slope factor that captures the rate of change in measurements following the knot location. The pre-knot slope does not impact measurements after the knot location, and vice versa. For a generalized expression of this matrix, researchers are encouraged to consult Grimm et al. (2016).

where α_c is a vector of growth factor means, and ζ_{ic} is a vector of residuals that express interindividual variation from α_c . The complete formulation of PGMMs under the structural equation modeling framework entails the combination of the measurement model (Eq. (1)) and the structural model (Eq. (2)). For more technical details about this complete form of PGMMs and the model-implied mean and covariance structure, we refer readers to Appendix A.

3. Bayesian inference in piecewise growth mixture modeling

The Bayesian framework for statistical inference differs from the conventional frequentist framework in that prior information is updated and then weighted by observed data to form posterior knowledge, which acts as the output of results. Utilizing the Bayesian estimation framework for PGMMs enables researchers to manage computational challenges associated with the nonlinear nature and large parameter space of PGMMs (Kohli, Hughes, et al., 2015).

A central tenet that underlies this updating process is the Bayes theorem:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \tag{3}$$

where θ and \mathbf{y} respectively represent all model parameters and the observed sample data. Eq. (3) states that the prior distribution, $p(\theta)$, that represents prior knowledge about parameters, is updated to the posterior distribution, $p(\theta|\mathbf{y})$, that represents our updated knowledge via taking into account likelihood, $p(\mathbf{y}|\theta)$, that corresponds to information obtained from the sample data. The term in the denominator of the right side of the equation, $p(\mathbf{y})$, refers to the marginal likelihood that indicates the evidence and serves as the normalizing constant in Bayesian estimation; hence, the formula in Eq. (3) usually reduces to $p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta)$, meaning the posterior distribution is proportional to the multiplication between the likelihood and prior distribution.

3.1. Prior specification for PGMMs

In Bayesian analysis, the role of prior distributions is crucial as they influence the final estimates of the model parameters. Depending on the desired level of informativeness, prior distributions can be classified as either diffuse, weakly informative, or informative. Researchers leverage different degrees of informativeness by adjusting the parameters of the prior distributions that we normally refer to as hyperparameters. For example, the normal prior distribution is defined through a mean and a variance that act as the distribution's hyperparameters. *Diffuse priors* provide little information about parameter values, largely relying on observed data for the updating process. Diffuse priors are often represented by a uniform distribution or a normal distribution with a very large variance. *Weakly informative priors* provide partial knowledge about parameter values. They are usually represented as normal distributions with a smaller variance term than diffuse priors. Weakly informative priors are used when researchers still hope to incorporate knowledge about parameters but do not want the prior distributions to have a substantial impact on the estimation process. *Informative priors* reflect specific information about parameters and can be purposefully used by researchers. An example of informative priors is a normal distribution with a smaller variance term than diffuse or weakly informative priors.

To implement Bayesian inference in PGMMs, it is necessary to define prior distributions for all unknown parameters. For linear-linear PGMMs with one unknown knot location, the priors should be specified for five key parameters, including (a) mixture class proportions, (b) latent means (i.e., means of growth factors), (c) variances of error terms associated with repeatedly measured variables, (d) latent covariance matrix (i.e., covariance of growth factors), and (e) knot location. Table 2 presents the notation, description, and prior specification for these parameters. In addition, Appendix B details the hyperparameters of these prior distributions. An important aspect to consider is that researchers can set these priors either identically or differently across different latent classes.

Table 2
Parameters and corresponding prior distributions in bayesian piecewise growth mixture models.

Parameter	Notation	Description	Prior Specification
Class proportion	π	The size of different latent classes	Dirichlet distribution: $\mathcal{D}(d_1, \dots, d_c)$
Factor mean	α_{mc}	The mean of each growth factor (e.g., the intercept and the respective slopes)	Normal distribution: $\mathcal{N}(\mu_{\alpha_{mc}}, \sigma_{\alpha_{mc}}^2)$
Variance of error terms	θ_{ϵ_t}	The variance of error terms associated with observed variables (e.g., repeatedly measured outcome variables)	Inverse gamma distribution: $\mathcal{IG}(a_{\theta_{\epsilon_t}}, b_{\theta_{\epsilon_t}})$
Factor covariance matrix	Ψ_{η}	The variances and covariances between growth factors	Inverse Wishart distribution: $\mathcal{IW}(\Psi, \nu)$
Knot location	γ	The time at which transitions arise between separate developmental phases	Uniform distribution: $\mathcal{U}(\min(t), \max(t))$ Truncated normal distribution: $\mathcal{N}(\mu_{\gamma}, \sigma_{\gamma}^2) \cdot \mathcal{I}(\min(t), \max(t))$

Note. Details of notation for parameters and prior specifications are provided in Appendix B.

3.2. Posterior computation

The final step in the Bayesian estimation of PGMMs is to obtain the posterior distribution. However, deriving the analytical form of the posterior distribution via the Bayes theorem in Eq. (3) is often challenging in practice due to the integrations over the high-dimensional parameter space, which is also the case for PGMMs. An alternative and practical approach is to approximate the posterior distribution using random samples from the posterior distribution, where the Markov chain Monte Carlo (MCMC) estimation algorithms such as the Gibbs sampler (Geman & Geman, 1984) are often used.

The Gibbs sampler generates a sequence of samples from the joint posterior distribution by iteratively drawing samples from the conditional posterior distributions of each parameter given the current values of the other parameters (Geman & Geman, 1984). The convergence of the Gibbs sampler is essential to obtain reliable and valid posterior estimates. After applying the MCMC algorithm, samples from the initial iterations, often referred to as “burn-in” samples, may not be representative of the target posterior distribution. These samples are discarded from the sequence of samples to avoid any potential bias. The remaining samples are then used for further analysis. Next, assessing the convergence of the obtained posterior samples is an essential step in the MCMC process. A common diagnosis tool for this purpose is the statistic \hat{R} , also known as the potential scale reduction factor. The \hat{R} statistic measures the potential reduction in the scale of the parameter estimates when multiple chains are used in the MCMC simulation. It compares the within-chain variability to the between-chain variability, and values close to 1 indicate convergence. Additionally, trace plots are a useful visual measure to assess convergence (Gelman, Carlin, et al., 2014). By using these diagnostics tools, researchers can ensure that the MCMC chains have adequately explored the posterior distribution and have reached convergence. Proper convergence assessment is crucial to produce reliable and unbiased posterior estimates and to gain confidence in the validity of the Bayesian analysis results.

After confirming the evidence of convergence, the posterior distributions can be summarized by calculating key statistics such as the posterior mean, posterior standard deviation, and credible intervals. These posterior summaries offer an understanding of the estimated parameter values and their uncertainty within a Bayesian framework.

4. Methodological considerations of Bayesian PGMMs

Having established the theoretical foundations and motivation underlying the Bayesian approach to PGMMs, we now focus on key methodological considerations in their implementation. Specifically, we introduce the topics of class separation, class enumeration, and prior sensitivity. These topics have significant implications for the accurate estimation of latent classes and the interpretation of substantive findings. We explore how these methodological considerations relate to the context of PGMMs.

4.1. Class separation

The first consideration is class separation, which pertains to how each latent class is similar to or distinct from one another in the population (Depaoli, 2021; Henson et al., 2007). Researchers typically define class separation based on disparate parameter estimates and substantive interpretations (Depaoli, 2021). Well-separated latent classes indicate that there is a clear distinction between individuals from different subpopulations. Conversely, poor class separation leads to ambiguous distinctions across classes, making the latent groups more difficult to distinguish from one another.

Regarding PGMMs, this concept refers to the level of overlap observed among separate growth trajectories or placement of knots associated with different latent classes. Consider the hypothetical example in Fig. 1b. We observe that their initial outcome levels (i.e., students' math examination scores) differ by 40 units. After this initial time point, there is no further overlap between the two classes (i.e., high-achieving group and low-achieving group) in their respective growth trajectories. In addition, their growth trends exhibit distinct patterns. As such, we can claim that these latent classes do not overlap, and hence, are dissimilar to each other in terms of the growth trajectories. We point out that the two classes in Fig. 1b share the same knot location. Although knot locations can be predetermined due to study design (e.g., intervention), they can be estimated from the data. In this regard, the dissimilarity between latent classes can also be linked to the placement of knots. If there are two classes where Classes 1 and 2 have respective knot locations estimated at the second and fifth time points, these two classes are considered dissimilar in terms of knot placement.

When latent classes exhibit higher degrees of class separation, little to no overlap is expected for individual trajectories or knot locations across classes, resulting in distinct parameter estimates for each latent class. Conversely, poor class separation is defined through substantial overlap between classes, making class-specific parameters difficult to distinguish from one another. For instance, let's say a school psychology researcher applies a Bayesian PGMM to a group of students with ADHD to identify latent classes and describe their executive functioning abilities each semester. The researcher especially aims to differentiate between attentive-ADHD and hyperactive-ADHD by fitting a two-class model based on substantive theory. In this case, it is likely that the parameter estimates for these two classes are very close, and accordingly, the growth trajectories of students with attentive-ADHD and hyperactive-ADHD are similar, reflecting poor class separation. Therefore, distinguishing between the classes and understanding the unique needs of each group becomes challenging.

Class separation can be determined in a variety of ways. Visually, separation is easiest to see if intercepts or slopes are quite different across the classes. In that case, separation would be in terms of the latent factor means. However, there can also be separation in the structural part of the model, for example, in the covariance matrix for the latent growth factors. Regardless of how separation is defined, under conditions of poor class separation the risk of mislabeling individual cases to the “wrong” class increases and it becomes more difficult to substantively define latent classes as being different when they overlap so heavily with one another. In addition,

convergence issues may arise, particularly in cases of poor class separation when sample sizes are small in one or more of the classes (Tofighi & Enders, 2008; Tueller & Lubke, 2010). To address these challenges, larger sample sizes are generally recommended as they improve class identification, facilitate obtaining reliable parameter estimates, and achieve better convergence (Depaoli, 2013; Tolvanen, 2007; Tueller & Lubke, 2010; van de Schoot et al., 2015).

When obtaining a larger sample size is challenging, Bayesian estimation can aid in recovering accurate parameter estimates. Methodological literature has shown that even with class sizes as low as 20, Bayesian methods can still effectively recover parameters (e.g., Zhang et al., 2007). Research on linear and quadratic growth mixture models has demonstrated that using informative priors can assist in accurately recovering latent classes and parameter estimates, particularly when latent classes are poorly separated (Depaoli, 2013). However, if information about parameters is limited and the use of informative priors is not feasible, relying on diffuse prior specifications yields poor recovery (Depaoli, 2013).

Importantly, the sample size challenge extends beyond absolute sample sizes; relative sample sizes across latent classes are also an important factor. For instance, if there is a minority latent class that is relatively smaller than the other classes, the recovery of this minority latent class becomes difficult (Depaoli, 2013; Tofighi & Enders, 2008). The incorporation of prior distributions is helpful in the context of a minority latent class, but prior impact is increasingly important in this case (Depaoli, 2013; Depaoli et al., 2017). These considerations emphasize the need to address class separation in relation to prior sensitivity, which is discussed later. Furthermore, given the complexity of PGMs, the construction of generalized sample size requirements for accurate modeling and estimation is vexingly compounded. There are a variety of aspects that affect whether such a result can be generalized. We discuss this topic in more detail in the Discussion.

4.2. Class enumeration

Another important consideration is class enumeration, which refers to properly determining the number of latent classes in the sample data (Depaoli, 2021; Muthén, 2004; Nylund et al., 2007). Class enumeration is closely tied to class separation. When latent classes are clearly separated, researchers can easily define them as distinct classes and determine the number of classes. However, poorly separated classes are more challenging to define accurately, making it difficult to identify the number of latent classes. Recall our ADHD example in the class separation above where a researcher applies a two-class PGM to a group of students with ADHD based on substantive theory. If the parameter estimates for these two classes are very close in describing the piecewise growth trajectories of executive functioning abilities, this reflects poor class separation. Assuming there are two latent classes (i.e., attentive-ADHD and hyperactive-ADHD) at the population level, it is challenging to accurately determine that there are indeed two distinct classes. Therefore, the low class separation makes it difficult to determine the true number of latent classes and complicates the class enumeration.

The decision regarding the number of classes can be understood as a model evaluation process using information criteria. The model with the true number of latent classes according to the population is assumed to best represent the data from a set of competing solutions (McLachlan & Peel, 2000). For example, a researcher may attempt to find a final model solution by estimating two-class, three-class, and four-class solutions. Then, substantive considerations together with statistical criteria can aid the researcher in determining the final model to interpret.

Examples of information criteria are the Bayesian information criterion (BIC; Schwarz, 1978), deviance information criterion (DIC; Spiegelhalter et al., 2002), Watanabe-Akaike information criterion (WAIC; Watanabe, 2010), and leave-one-out cross-validation information criterion (LOOIC; Gelman, Carlin, et al., 2014; Vehtari et al., 2017), which should be used for model selection along with substantive considerations. These indices are briefly illustrated in the following paragraph.

The BIC approximates the Bayes factor (Kass & Raftery, 1995) and is popular for model selection in both frequentist and Bayesian applications. The BIC evaluates the likelihood as a measure of model fit and penalizes the model complexity that is dependent on both sample size and the number of parameters. The formulation of the BIC relies on maximum likelihood estimates of model parameters and does not consider the posterior distribution. For this reason, the BIC cannot be considered purely Bayesian (Gelman, Hwang, & Vehtari, 2014). A measure that aligns more closely with Bayesian principles is the DIC that is formulated based on Bayesian deviance as a measure of model fit. Compared to the BIC, the DIC can prevent over-penalizing the model complexity because the DIC can only include the effective number of parameters. However, the DIC is criticized because it assumes the normality of the posterior distribution (Spiegelhalter et al., 2002). In addition, some have argued that the DIC is partially Bayesian (e.g., Liang & Luo, 2020) because it evaluates a model when parameter estimates are fixed to posterior means and hence does not consider the entire posterior distribution. The WAIC and LOOIC are information criteria that are considered fully Bayesian because both indices consider the entire posterior distribution. The WAIC assesses the predictive performance of a model by evaluating the likelihood using posterior estimates. Conversely, the LOOIC is based on the idea of cross-validation within the Bayesian framework to examine how well a model can predict the observed data. Specifically, the LOOIC evaluates the performance of a model to predict a data point using data that exclude that certain data point. When deciding the number of latent classes to retain, models with the lowest values of the information criteria are considered optimal and hence favored.

In the Bayesian PGM literature, the BIC is the only index that has been used for class enumeration purposes in applied research. For instance, Kohli, Hughes, et al. (2015) employed the BIC to evaluate one-, two-, and three-class solutions and ultimately selected the two-class solution based on the lowest BIC value and substantive theories. Further adoption of other Bayesian model selection indices such as the DIC, WAIC, and LOOIC in guiding model selection for PGMs holds promise for enhancing the process of class enumeration in substantive research.

The methodological performance of the DIC, WAIC, and LOOIC, which are truer “Bayesian indices” given their specification based

on the posterior distribution (Gelman, Hwang, & Vehtari, 2014), has not been extensively studied within the context of PGMMs. Instead, we found two studies that could be of relevance. A related study by Depaoli et al. (2023) focused on a PGMM with a single latent class, referred to as a piecewise growth curve model (i.e., not a mixture). They used the BIC and DIC as model selection indices to detect misspecifications in knot locations and growth trajectories. The findings suggested that the DIC outperformed the BIC, particularly when knot locations were ignored. In addition, a minimum sample size of 300 was necessary for both indices to reliably detect model misspecification.

Kim et al. (2021) examined the performance of the DIC, WAIC, and LOOIC in the context of linear growth mixture models, where the assumption of normality for outcome variables was violated, and the model was estimated based on conditional medians. For a normal mixture, results indicated the DIC, WAIC, and LOOIC performed comparably well in detecting the correct number of latent classes. However, in the presence of outliers, the WAIC and LOOIC more accurately detected the number of latent classes compared to the DIC.

4.3. Prior sensitivity

The last consideration surrounds the specification of prior distributions. The nontrivial role of prior distributions, particularly in linear growth mixture models, has been extensively explored in previous studies (e.g., Depaoli, 2013; Depaoli et al., 2017) that highlighted the need for thoughtful consideration when eliciting and incorporating prior knowledge into mixture modeling. Prior distributions can be advantageous in recovering latent classes and obtaining true model parameters (Depaoli, 2013). In addition, incorporating prior information into the estimation can mitigate convergence issues when dealing with small total sample sizes or latent classes that contain relatively few cases compared to larger classes (Tofighi & Enders, 2008; Tueller & Lubke, 2010).

Importantly, there are caveats to consider when implementing priors. The choice of prior specifications can significantly impact the results of mixture models (Depaoli, 2013; Depaoli et al., 2017). Depaoli (2013) found that specifying diffuse priors yielded inaccurate parameter estimates, whereas accurately specifying informative priors successfully recovered the true class mixture proportions. Therefore, understanding the role of priors in addressing substantive theories is crucial to ensuring the robustness and stability of the obtained results. It is strongly suggested and recommended that researchers conduct analyses under different prior settings and compare the results to evaluate the impact of these specifications on their conclusions. As discussed by van de Schoot et al. (2018), researchers should openly discuss the potential instability of results arising from different prior settings and carefully reassess the alignment between proposed Bayesian mixture models and substantive theories.

In the context of PGMMs, prior distributions can aid in improving the accuracy of describing piecewise growth trajectories, identifying disparate estimates across classes, and determining the number of classes. The inclusion of additional parameters such as knot locations or separate growth means (Depaoli et al., 2023; Kohli, Hughes, et al., 2015) adds complexity to prior specification. Priors for growth factor means and knot locations have received relatively less attention in both methodological and applied research within the context of PGMMs. Although previous studies such as Depaoli et al. (2023) and Heo et al. (2024) have explored priors for growth factor means within the framework of a piecewise growth curve model (i.e., a PGMM with a single-class solution) for model evaluation purposes, minimal impact of priors on slope trajectories was found. Priors for knot locations, which indicate the timing of changes in separate developmental trajectories, have primarily been utilized as diffuse priors (Kohli, Hughes, et al., 2015). The exploration of the effects of different prior settings on class proportions in Bayesian PGMMs remains understudied.

To illustrate the importance of prior sensitivity in an applied context, consider our ADHD example. Depending on how the priors for mixture proportions are specified, the resulting number of classes can vary. For instance, depending on the informativeness of priors (e.g., diffuse or informative priors), the model might suggest two classes (e.g., attentive-ADHD, hyperactive-ADHD), or it is possible that a third class could emerge, indicating a more nuanced differentiation within the ADHD population. Priors can also be specified for separate growth trajectories (e.g., latent means) across classes. These priors can influence the final model estimates and potentially alter the determination of the number of classes. Although this example is purely for showcasing the sensitivity of PGMMs to prior specifications, we reiterate the need for careful consideration and justification of priors in real research settings.

5. Empirical application

To address methodological concerns and demonstrate the practical application of Bayesian PGMMs, we conducted an analysis using data from the Early Childhood Longitudinal Studies-Kindergarten Cohort of 1998 (ECLS-K) study. The ECLS-K study was a comprehensive investigation that examined various aspects of children's early school experiences. Our specific focus was on exploring the trajectory of students' mathematics achievement from kindergarten to eighth grade. Not only is math achievement an important educational outcome, but it is associated with factors such as family environment and cognitive development, making it a good construct for the current illustration (Garon-Carrier et al., 2018; Siegler et al., 2012).

Notably, previous research has suggested that mathematics achievement follows a piecewise functional growth pattern characterized by two distinct linear segments (Kohli, Sullivan, et al., 2015). This indicates that there may be qualitatively distinct subpopulations within this population, with different growth trajectories. By employing a Bayesian PGMM, we can identify these latent classes, describe piecewise trajectories under each latent class, and further examine, for example, whether factors such as family environment influence these groups differently. In this context, we employed a Bayesian PGMM to investigate latent classes within the domain of mathematical achievement, aiming to contribute to educational research. By identifying subgroups within the student population, we can deeply investigate diverse learning profiles and specific needs.

5.1. Data and measures

The dataset consisted of nationally representative data from the United States collected from children who began kindergarten in the 1998–1999 academic year. Data were repeatedly collected at seven time points, including fall and spring of kindergarten, fall and spring of first grade, spring of third grade, spring of fifth grade, and spring of eighth grade. The original dataset included 21,407 participants from diverse socioeconomic, racial, and ethnic backgrounds. For our analysis, we randomly selected a subsample of 500 children who had complete responses for mathematics item response theory (IRT) scores. The sample size of 500 was chosen because it was the medium sample size that can be typically encountered in literature on piecewise modeling (e.g., Kohli et al., 2016). For this chosen sample of 500 children, we ensured there were no missing values.² The math IRT scores served as our target observed variable, representing a measure of mathematics achievement. Demographic information of our analytic sample was as follows: 53.4% of our sample were female; 82.6% attended public schools; 66.2% were White, 10.2% were black, 13.8% were Hispanic, 4.2% were Asian, and the remaining 5.6% represented other races or ethnicities.

5.2. Analytic procedures

The analysis of our sample of 500 children involved considering three different mixture class solutions within the PGMM framework, including one-class, two-class, and three-class models. We coded the time variable as 0, 0.5, 1, 1.5, 3.5, 5.5, and 8.5, corresponding to the fall and spring of kindergarten, fall and spring of first grade, spring of third grade, spring of fifth grade, and spring of eighth grade, respectively. We did not predetermine knot locations; instead, we allowed them to be freely estimated.

Our analysis consisted of two main steps. In the first step, we fitted three PGMMs (i.e., one-class, two-class, and three-class solutions) using common diffuse prior settings, as described in Merkle and Rosseel (2018) and listed in Appendix B. For the latent class proportions, we used a Dirichlet prior with hyperparameters of 1 *s* as diffuse priors. We interpreted the results, evaluated class separation and enumeration, and selected the final model based on both statistical criteria and substantive interpretations. We proceeded with analyses using the diffuse prior settings to mirror scenarios commonly encountered by researchers who lack prior knowledge. When prior knowledge is available, researchers can create prior distributions based on that information and implement either informative or weakly informative priors by adjusting the level of informativeness accordingly. An instrumental task is to examine the robustness of the results under different prior specifications via a prior sensitivity analysis, which we detail next.

In the second step, we conducted a prior sensitivity analysis by specifying informative and weakly informative priors to account for the impact of different prior specifications on the results. A detailed presentation of the prior distributions is provided in Table 3. For the one-class or two-class PGMMs, we referred to a previous study in school psychology (Kohli & Sullivan, 2019) to derive our informative and weakly informative priors. In the case of a three-class PGMM, where prior knowledge was lacking, we empirically derived prior distributions through a pilot analysis. Specifically, we randomly selected half of the dataset and fitted a three-class PGMM using the maximum likelihood estimation framework using *Mplus* (L. K. Muthén & Muthén, 1998–2017). Subsequently, the obtained parameter estimates were taken into account to specify priors for the three-class PGMM. The parameters we considered for the prior sensitivity analysis were the latent class mixture proportions and latent growth factor means (which form the estimated growth trajectories).³

For the latent class proportions of the two-class PGMM, we specified informative priors based on estimated class proportions from Kohli and Sullivan (2019). As for weakly informative priors, the hyperparameter values were derived from informative and diffuse prior settings. For instance, the informative prior for the latent class proportions was $\mathcal{D}(0.24, 0.76)$, which assumed 24% and 76% of assignments to Class 1 and Class 2, respectively. The diffuse prior was $\mathcal{D}(1, 1)$, which assumed 50% to 50% of assignments to each latent class. Then, we took the average of the proportions for each class from the informative and diffuse priors to compute the weakly informative prior: $\mathcal{D}\left(\frac{0.24+0.50}{2}, \frac{0.76+0.50}{2}\right) = \mathcal{D}(0.37, 0.63)$, indicating 37% and 63% of assignments to each latent class. The same procedure was applied to determine informative and weakly informative priors for the three-class PGMM. Using the estimated class proportions from pilot analysis, the informative prior was $\mathcal{D}(0.21, 0.29, 0.50)$, which assumed 21%, 29%, and 50% of assignments to Class 1, Class 2, and Class 3, respectively. The diffuse prior was $\mathcal{D}(1, 1, 1)$, which assumed that 33%, 33%, and 33% of assignments to each latent class. Then, the weakly informative prior distribution was calculated by taking the average such that $\mathcal{D}\left(\frac{0.21+0.33}{2}, \frac{0.29+0.33}{2}, \frac{0.50+0.33}{2}\right) = \mathcal{D}(0.27, 0.31, 0.42)$, which indicated 27%, 31%, and 42% of assignments to each latent class.

For the growth factor means of the two-class PGMM, the centers of the normal prior distributions were set as the maximum likelihood estimates from Kohli and Sullivan (2019). For the three-class PGMM, we used the maximum likelihood estimates obtained from pilot analysis. The variance hyperparameters for both the two-class and three-class PGMMs were fixed at 0.05 for informative

² We opted to select complete cases (as opposed to those with missing data present) for demonstration purposes. That decision allowed us to focus more technical efforts on the Bayesian estimation elements, as well as model result interpretation for PGMMs. However, we note here that PGMMs and the Bayesian framework can handle missing data situations. In particular, the Bayesian estimation framework treats missing data via data augmentation (Tanner & Wong, 1987). In that case, the missing data are “filled in” throughout the Bayesian estimation process (Enders, 2022). For more information on this topic, we refer the reader to Daniels and Hogan (2008) and Enders (2022).

³ Rather than disentangling the impact of each prior distribution, we simultaneously manipulated priors on both the growth factor means and the latent class proportions. We thus examined the magnitude of change in the estimates of these parameters.

Table 3
Informative and weakly informative prior specifications for one-, two-, and three-class PGMMs.

	Informative Priors			Weakly Informative Priors		
	One-Class	Two-Class	Three-Class	One-Class	Two-Class	Three-Class
Class Proportion		$\mathcal{D}(0.24, 0.76)$	$\mathcal{D}(0.21, 0.29, 0.50)$		$\mathcal{D}(0.37, 0.63)$	$\mathcal{D}(0.27, 0.31, 0.42)$
Class 1						
I, Mean	$\mathcal{N}(27.08, 0.05)$	$\mathcal{N}(21.83, 0.05)$	$\mathcal{N}(22.59, 0.05)$	$\mathcal{N}(27.08, 0.1)$	$\mathcal{N}(21.83, 0.1)$	$\mathcal{N}(22.59, 0.1)$
S1, Mean	$\mathcal{N}(22.46, 0.05)$	$\mathcal{N}(17.14, 0.05)$	$\mathcal{N}(22.32, 0.05)$	$\mathcal{N}(22.46, 0.1)$	$\mathcal{N}(17.14, 0.1)$	$\mathcal{N}(22.32, 0.1)$
S2, Mean	$\mathcal{N}(5.64, 0.05)$	$\mathcal{N}(7.07, 0.05)$	$\mathcal{N}(8.70, 0.05)$	$\mathcal{N}(5.64, 0.1)$	$\mathcal{N}(7.07, 0.1)$	$\mathcal{N}(8.70, 0.1)$
Class 2						
I, Mean		$\mathcal{N}(29.16, 0.05)$	$\mathcal{N}(36.80, 0.05)$		$\mathcal{N}(29.16, 0.1)$	$\mathcal{N}(36.80, 0.1)$
S1, Mean		$\mathcal{N}(23.92, 0.05)$	$\mathcal{N}(24.76, 0.05)$		$\mathcal{N}(23.92, 0.1)$	$\mathcal{N}(24.76, 0.1)$
S2, Mean		$\mathcal{N}(5.42, 0.05)$	$\mathcal{N}(4.90, 0.05)$		$\mathcal{N}(5.42, 0.1)$	$\mathcal{N}(4.90, 0.1)$
Class 3						
I, Mean			$\mathcal{N}(21.56, 0.05)$			$\mathcal{N}(21.56, 0.1)$
S1, Mean			$\mathcal{N}(20.26, 0.05)$			$\mathcal{N}(20.26, 0.1)$
S2, Mean			$\mathcal{N}(4.90, 0.05)$			$\mathcal{N}(4.90, 0.1)$

Note. I = Intercept; S1 = Slope 1 (slope of the first phase); S2 = Slope 2 (slope of the second phase).

priors and 0.1 for weakly informative priors, based on Depaoli et al. (2020).

In all Bayesian analyses, we used the R jagsUI package, which wrapped the rjags package and integrated the JAGS software into the R environment (Kellner, 2021; Plummer, 2016; Plummer et al., 2003). This integration provided a flexible and efficient platform for conducting Bayesian inference in our proposed PGMMs. We implemented the Gibbs sampler with a single chain to obtain samples from the posterior distribution.⁴ Specifically, for the one- and two-class PGMMs, we performed 70,000 iterations with a burn-in period of 10,000 samples. For the three-class PGMM, we increased the number of iterations to 112,000, with a burn-in period of 52,000 samples. The choices behind the number of iterations and length of burn-in was determined after conducting several preliminary tests and visually inspecting convergence using trace plots. To assess the convergence of the chains, we monitored the \hat{R} statistics, which is a numerical measure commonly used for chain convergence evaluation (Vehtari et al., 2021).⁵ By considering these measures and adjusting the lengths of the iterations and burn-in samples, we made sure that our chains adequately converged. To facilitate the application of Bayesian PGMM methods, we have made the R, JAGS, and Mplus code available in the Open Science Framework (OSF) materials.⁶

6. Results

6.1. Convergence

We first fitted three PGMMs under the diffuse prior settings. According to the results, all parameters had \hat{R} statistics lower than 1.1, indicating adequate evidence of convergence. An additional two models (i.e., PGMMs under informative and weakly informative prior settings) fitted for prior sensitivity analysis also indicated evidence of convergence according to the \hat{R} statistic.

6.2. Model estimation results

We present the results of the one-class PGMM using diffuse prior distributions in Table 4, which includes posterior means, posterior standard deviations, and 95% credible intervals for the estimated parameters. The average mathematics IRT score at kindergarten (i.e., the intercept mean) was 27.361. The growth trajectory of mathematics achievement from kindergarten to eighth grade was divided into two phases. In the first phase (i.e., S1: the slope of the first phase), the mathematics IRT scores increased rapidly by an average of 21.739 per unit of time until the knot location at 4.231. In the second phase after the knot location (i.e., S2: the slope of the second phase), the increase in scores slowed to an average of 5.740 per unit of time. The estimated knot location of 4.231 indicates that the transition from the first phase to the second phase of development occurred between the fall and spring of the fourth grade. The

⁴ We used a single chain to avoid label switching, which refers to the arbitrary rearrangement of latent class order (Jasra et al., 2005; Stephens, 2000). However, label switching can even occur when sampling from a single chain switches back and forth from different latent classes. Therefore, we additionally inspected trace plots of latent class proportions and implemented the equivalence classes representatives algorithm (Papastamoulis, 2016). For alternative approaches to handle label switching, researchers may refer to Depaoli (2021), Tueller et al. (2011), and Yao and Lindsay (2009).

⁵ We implemented the single-chain computation of the \hat{R} statistics following Mplus because it aligned with our goals for avoiding between-chain label switching. After discarding the burn-in samples, we split the remaining samples in half and treated these two halves as two separate chains for computing the \hat{R} statistics. For more on how Mplus implements the \hat{R} statistics in single-chain situations and associated issues, see Asparouhov and Muthén (2010), Depaoli (2021), and Vats and Knudson (2021).

⁶ <https://osf.io/udrb8/>.

Table 4
Parameter estimates for a one-class PGMM.

	Diffuse Priors		
	<i>M</i>	<i>SD</i>	95% CI
I, Mean	27.361	0.499	[26.379, 28.337]
S1, Mean	21.739	0.220	[21.310, 22.169]
S2, Mean	5.740	0.191	[5.369, 6.115]
Knot, Mean	4.231	0.039	[4.153, 4.307]
I, Variance	100.979	7.919	[86.348, 117.411]
S1, Variance	16.479	1.425	[13.845, 19.432]
S2, Variance	5.425	0.880	[3.777, 7.228]
COV(I, S1)	20.892	2.448	[16.197, 25.800]
COV(I, S2)	-16.682	1.886	[-20.516, -13.127]
COV(S1, S2)	-4.677	0.876	[-6.474, -3.029]
Error Variance	7.624	0.127	[7.382, 7.880]

Note. *M* = Posterior Mean; *SD* = Posterior Standard Deviation; CI = Credible Interval; I = Intercept; S1 = Slope 1 (slope of the first phase); S2 = Slope 2 (slope of the second phase); Knot = Knot Location; COV = Covariance.

variances of the intercept, the first slope, and the second slope were 100.979, 16.479, and 5.425, respectively. These results suggest there was large variability in the initial math IRT scores compared to the first and second slopes.

Table 5 presents a summary of posterior inference for the two-class PGMM under diffuse priors, highlighting the distinct trajectories of mathematics achievement for students in two latent classes. According to the estimated class proportions, about 58.2% and 41.8% of students were classified to Class 1 and Class 2, respectively. Students belonging to Class 2 demonstrated higher initial mathematics IRT scores, characterized by an intercept mean of 34.227, whereas students in Class 1 had an intercept mean of 22.515. Moreover, students in Class 2 demonstrated a greater increase in mathematics IRT scores before the knot location, with an average increase of 24.909 per additional year, compared to an increase of 19.321 for students in Class 1. After the knot location, growth rates slowed down in both classes. Students in Class 2 showed an average increase of 5.189, whereas students in Class 1 had an increase of 5.913, which suggested that students in Class 1 had a higher rate of increase in mathematics achievement during the second developmental phase as compared to students in Class 2. The knot location for Class 2 was estimated to be at 4.038, indicating that a transition happened right after the fall of the fourth grade. In contrast, students in Class 1 experienced a change of rate a semester later at 4.530, corresponding to a time point right after the spring of the fourth grade. These results indicate that students in Class 2 showed a higher level of mathematics achievement upon entering kindergarten and experienced greater rates of increase during the first developmental phase, compared to

Table 5
Parameter estimates for a two-class PGMM.

	Diffuse Priors		
	<i>M</i>	<i>SD</i>	95% CI
Class 1			
I, Mean	22.515	0.512	[21.500, 23.510]
S1, Mean	19.321	0.450	[18.429, 20.195]
S2, Mean	5.913	0.289	[5.333, 6.470]
Knot, Mean	4.530	0.080	[4.377, 4.693]
I, Variance	27.059	4.715	[18.487, 36.945]
S1, Variance	15.043	2.027	[11.441, 19.414]
S2, Variance	8.542	1.455	[5.918, 11.657]
COV(I, S1)	9.457	2.118	[5.370, 13.739]
COV(I, S2)	-4.888	1.756	[-8.394, -1.496]
COV(S1, S2)	-4.280	1.224	[-6.794, -1.988]
Error Variance	6.489	0.186	[6.128, 6.857]
Class Proportion	0.582	0.047	[0.487, 0.670]
Class 2			
I, Mean	34.227	1.233	[31.889, 36.734]
S1, Mean	24.909	0.388	[24.144, 25.667]
S2, Mean	5.189	0.385	[4.442, 5.956]
Knot, Mean	4.038	0.066	[3.909, 4.167]
I, Variance	116.674	16.373	[87.818, 151.701]
S1, Variance	5.291	1.671	[2.453, 8.985]
S2, Variance	3.273	1.174	[1.440, 6.047]
COV(I, S1)	0.905	4.269	[-7.541, 9.382]
COV(I, S2)	-16.268	3.652	[-24.053, -9.682]
COV(S1, S2)	-1.987	1.185	[-4.674, -0.067]
Error Variance	8.523	0.246	[8.065, 9.033]
Class Proportion	0.418	0.047	[0.330, 0.513]

Note. *M* = Posterior Mean; *SD* = Posterior Standard Deviation; CI = Credible Interval; I = Intercept; S1 = Slope 1 (slope of the first phase); S2 = Slope 2 (slope of the second phase); Knot = Knot Location; COV = Covariance.

students in Class 1. The variances of the intercept were 27.059 for Class 1 and 116.674 for Class 2, indicating a much larger variability of the initial mathematics scores for students in Class 2. However, the variances of the first slope were 15.043 for Class 1 and 5.291 for Class 2, and those of the second slope were 8.542 for Class 1 and 3.273 for Class 2. These values indicate larger variability in both slopes for students in Class 1 as compared to Class 2.

Parameter estimates for the three-class PGMM under diffuse prior specifications are displayed in Table 6. About 47.0%, 42.4%, and 10.6% of students, respectively, were classified to Class 1, Class 2, and Class 3. Among these classes, students in Class 2 showed the highest initial level of mathematics scores, with the intercept mean of 34.071, compared to students in Class 1 (23.814) and Class 3 (16.317). In addition, students in Class 2 demonstrated a substantial increase during the first phase, with an average increase of 24.497 per additional year, compared to students in Class 1 (19.996) and Class 3 (17.729). In the second phase, students in Class 1 exhibited the highest increase of 6.178 per year on average, compared to students in Class 2 (5.229) and Class 3 (−0.511). For Class 1 and Class 2, a similar pattern of mathematics achievement development was observed characterized by a substantial increase in the first phase followed by a slower rate of change in the second phase. For Class 3, however, mathematics achievement decreased in the second development phase. The estimated knot location for students in Class 2 was the fastest at 4.100, which corresponds to a transition from the fall to the spring of the fourth grade. Students in Class 1 had a changepoint at 4.238, which is again between the fall and spring of the fourth grade but a bit delayed compared to Class 2. Students in Class 3 had the slowest knot location at 6.214, being between the fall and spring of the sixth grade. Of note, students in Class 2 demonstrated the most prominent mathematics achievement, with the highest initial level and the highest rate of increase during the first phase. Between the three classes, students in Class 2 had the largest variability at the initial level (122.454), followed by those in Class 1 (11.261) and Class 3 (8.867). The variability at the first slope was the greatest for students in Class 1 (19.594), followed by those in Class 2 (5.626) and those in Class 3 (5.270). The variability at the second slope was the greatest in Class 1 (8.022), followed by those in Class 3 (6.998) and in Class 2 (4.090).

Table 6
Parameter estimates for a three-class PGMM.

	Diffuse Priors		
	<i>M</i>	<i>SD</i>	95% CI
Class 1			
I, Mean	23.814	0.530	[22.782, 24.853]
S1, Mean	19.996	0.561	[18.953, 21.153]
S2, Mean	6.178	0.344	[5.489, 6.840]
Knot, Mean	4.238	0.088	[4.060, 4.409]
I, Variance	11.261	4.082	[4.756, 20.978]
S1, Variance	19.594	2.961	[14.446, 26.016]
S2, Variance	8.022	1.580	[5.228, 11.423]
COV(I, S1)	12.346	2.419	[7.695, 16.833]
COV(I, S2)	−4.853	1.785	[−8.335, −1.335]
COV(S1, S2)	−5.243	1.504	[−8.442, −2.529]
Error Variance	6.426	0.231	[5.976, 6.883]
Class Proportion	0.470	0.048	[0.374, 0.560]
Class 2			
I, Mean	34.071	1.195	[31.834, 36.499]
S1, Mean	24.497	0.399	[23.711, 25.277]
S2, Mean	5.229	0.404	[4.432, 6.035]
Knot, Mean	4.100	0.067	[3.967, 4.230]
I, Variance	122.454	16.696	[93.045, 158.379]
S1, Variance	5.626	1.771	[2.501, 9.421]
S2, Variance	4.090	1.200	[2.054, 6.744]
COV(I, S1)	2.696	4.227	[−5.687, 10.944]
COV(I, S2)	−18.903	3.761	[−26.935, −12.165]
COV(S1, S2)	−2.581	1.182	[−5.105, −0.497]
Error Variance	8.592	0.256	[8.124, 9.130]
Class Proportion	0.424	0.044	[0.341, 0.512]
Class 3			
I, Mean	16.317	1.013	[14.384, 18.518]
S1, Mean	17.729	0.581	[16.545, 18.767]
S2, Mean	−0.511	3.966	[−8.073, 5.381]
Knot, Mean	6.214	0.473	[5.397, 6.932]
I, Variance	8.867	6.072	[1.188, 24.106]
S1, Variance	5.270	2.041	[2.207, 10.294]
S2, Variance	6.998	8.122	[0.271, 29.653]
COV(I, S1)	5.119	2.369	[1.176, 10.475]
COV(I, S2)	1.042	4.466	[−8.264, 10.594]
COV(S1, S2)	−0.415	2.742	[−7.241, 3.648]
Error Variance	5.895	0.391	[5.154, 6.706]
Class Proportion	0.106	0.029	[0.062, 0.166]

Note. *M* = Posterior Mean; *SD* = Posterior Standard Deviation; CI = Credible Interval; I = Intercept; S1 = Slope 1 (slope of the first phase); S2 = Slope 2 (slope of the second phase); Knot = Knot Location; COV = Covariance.

6.3. Class separation

We evaluated class separation based on trajectory plots and differences in parameter estimates between latent classes. In Fig. 2, we present the individual and estimated trajectories for both the two- and three-class PGMMs. The upper two plots (i.e., Plots 2a and 2b) illustrate the two-class PGMM, with the solid line representing Class 1 and the dotted line representing Class 2. The lower two plots (i.e., Plots 2c and 2d) correspond to the three-class PGMM, with the dotted line still denoting Class 2, whereas the solid line and the dashed line represent Class 1 and Class 3, respectively. Plots on the left (i.e., Plots 2a and 2c) display individual trajectories, whereas those on the right (i.e., Plots 2b and 2d) describe the estimated mean trajectories.

The analysis of the two-class solution demonstrates a substantial distinction between the two latent classes. In Plots 2a and 2b, we observed a higher intercept and a steeper growth rate in the first slope in Class 2, whereas Class 1 showed a lower starting point and a less steep growth rate in the first slope. During the second phase after the knot location, the trajectories of these two classes displayed minimal overlap. This distinction substantively tells that Class 2 can be described as a high-achieving group whereas Class 1 can be represented as a low-achieving group. A high degree of class separation was corroborated by statistical elements in Table 5 such that, for instance, estimates of the intercept and first slope were higher for Class 2 compared to Class 1. The two-class solution effectively captures the distinct nature of the two latent classes.

For the three-class solution, we encountered a different scenario. According to Plots 2c and 2d, Class 2 was clearly separated from both Class 1 and Class 3. However, the distinction between Class 1 and Class 3 was less pronounced, making it difficult to discern their growth trajectories. In addition, these two classes overlapped around the sixth and seventh grades according to the estimated mean trajectories in Plot 2d. This observation suggests that Class 1 and Class 3 may be duplicates of one another and hence may have originated from the same population. These findings were supported by the parameter estimates in Table 6. For instance, Class 2 displayed higher estimates for the intercept and the first slope compared to the other two classes. Consequently, substantively justifying three latent classes is more challenging.

6.4. Class enumeration

Determining the “optimal” number of latent classes involves consideration of both statistical criteria and substantive interpretation. In our analyses, we assessed the fit of three PGMMs using information criteria such as DIC, WAIC, and LOOIC. For each of the one-class, two-class, and three-class PGMMs, the DIC values were 27,461.47, 26,907.42, and 26,980.58, respectively; the WAIC values were 24,935.13, 24,686.54, and 24,607.19, respectively; and the LOOIC values were 25,535.13, 25,247.44, and 25,174.84, respectively. Based on these results, the three-class PGMM was preferred by the WAIC and LOOIC, whereas the two-class PGMM was preferred by the DIC. Although two of the statistical criteria favored the choice of the three-class solution, it is important to consider the substantive interpretation of the latent classes. Upon revisiting the class separation issue, Class 1 and Class 3 were less pronounced from one another as compared to the more intuitively distinct two-class solution for the PGMM.

Based on class separation, class enumeration, and substantive interpretations, we selected the two-class solution for our subsequent prior sensitivity analysis due to its clearer and more interpretable distinction between the latent classes. In practice, researchers may arrive at different conclusions, which is perfectly appropriate given the subjective nature of latent class modeling.

Upon selecting and justifying the final model solution, it is important to assess the robustness of results under different sets of prior distributions. We conducted a prior sensitivity analysis under different prior specifications to further examine the influence of the prior settings specified. In particular, we examined whether parameter estimates of the two-class solution became disparate or not under weakly-informative and informative prior settings.

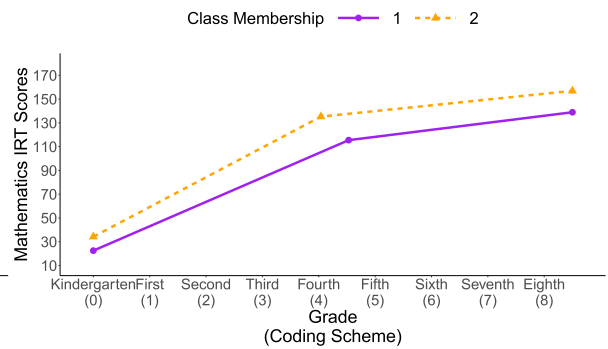
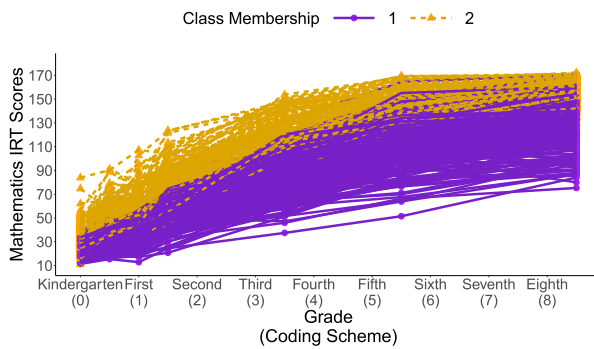
6.5. Prior sensitivity

Table 7 displays the results of the prior sensitivity analysis for the two-class PGMM. We focused on the estimates of the intercept, first slope, second slope, and class proportion because these were the parameters that we modified priors for. Compared to the results under the diffuse prior settings in Table 5, when weakly informative priors were used, the initial level of mathematics achievement decreased for both classes. For example, the mean of the intercept for Class 2 decreased from 34.227 to 29.561. Similar decreases in estimates were observed for the first slopes of both classes. The second slopes, however, increased from 5.913 to 6.552 for Class 1 and from 5.189 to 5.489 for Class 2. Despite these observed differences, the substantive interpretation of the latent classes remained intact such that Class 2 still represented the high-achieving group and Class 1 represented the low-achieving group. A more drastic change under weakly informative prior settings was the class proportion. Previously, more than half of the students (58.2%) were assigned to Class 1, which was defined as the low-achieving group. However, under the weakly informative prior settings, approximately 40.9% and 59.1% of the students were assigned to Class 1 and Class 2, respectively. This shift is noticeable as it substantively suggested that less than half of the students (40.9%) were assigned to the low-achieving group. When informative priors were specified, the estimates of the intercept, first slope, and second slope were comparable to those under the weakly informative prior settings. However, for the latent class proportion, an even smaller proportion of students belonged to Class 1 with about 39.3% of students, as compared to about 60.7% of students who belonged to Class 2. These findings indicate that as the degree of informativeness increased, the latent class proportion shifted a lot in a way that less than 40% of students were classified as the low-achieving group. Finally, for growth factor means and latent class proportions, the posterior standard deviations decreased and the width of the 95% credible intervals became narrower as the informativeness of the prior distributions increased.

It is crucial to acknowledge that different prior settings can indeed influence the estimates of the model, including knot locations in

(a) Individual Trajectories for Two Classes

(b) Estimated Trajectories for Two Classes



(c) Individual Trajectories for Three Classes

(d) Estimated Trajectories for Three Classes

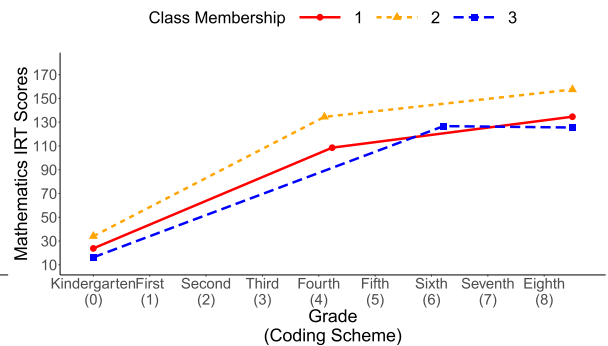
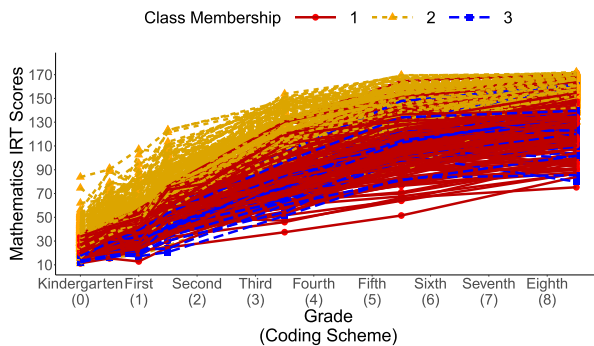


Fig. 2. Individual and estimated trajectory plots of two- and three-class PGMMs.

PGMMs. In our analysis, we observed that the estimates of knot location were affected by the prior specification. Initially, under a set of diffuse priors, the knot location for Class 1 was estimated to be 4.530, indicating a transition right after the spring of the fourth grade. Under a set of weakly informative and informative priors, this estimate changed to 4.705 and 4.700, respectively. Although both values still pointed to a transition from the spring of the fourth year to the fall of the fifth year, we observed a delay compared to the estimate under diffuse prior settings. This highlights that it is imperative for researchers to thoroughly examine the potential shifts in all parameters and reassess the substantive interpretations and underlying theories accordingly.

7. Discussion

Mixture modeling is a useful statistical framework for identifying hidden subgroups and accounting for population heterogeneity, which has implications for research and practice in school psychology. The Bayesian approach to mixture modeling has proven its usefulness in various applied settings and the increasing applications of Bayesian mixture modeling are to be expected. In this article, we focused on the practical relevance of Bayesian PGMMs for describing complex longitudinal phenomena in school psychology research. Bayesian PGMMs can be applied to understand phasic developmental trajectories across qualitatively different, heterogeneous latent classes. However, methodological considerations associated with its application from the perspective of applied researchers have remained limited. This article has aimed to introduce Bayesian PGMMs and address three important methodological considerations in their implementation, including class separation, class enumeration, and prior sensitivity. We outlined the analytic procedures using a real-data example and provided demonstrations on addressing these methodological challenges.

We used the ECLS-K dataset to examine the piecewise growth trajectories of mathematics achievement from kindergarten to eighth grade. The results from the one-class, two-class, and three-class PGMMs under diffuse priors consistently indicated separate and distinct developmental trajectories before and after the knot location. These findings are similar to patterns found in previous studies (Kohli, Hughes, et al., 2015; Kohli & Sullivan, 2019), suggesting that piecewise trajectories seem a reasonable approach to understanding mathematics achievement. For class separation, the two-class PGMM provided meaningful substantive interpretations based on distinct growth trajectories and divergent parameter estimates, supporting high class separation. During the process of class enumeration, the selection of the three-class PGMM was favored by the WAIC and LOOIC, whereas the DIC favored the two-class

Table 7
Results of prior sensitivity analysis.

	Weakly Informative Priors			Informative Priors		
	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI
Class 1						
I, Mean	21.814	0.278	[21.263, 22.356]	21.805	0.208	[21.399, 22.215]
S1, Mean	17.441	0.280	[16.890, 17.984]	17.290	0.209	[16.884, 17.702]
S2, Mean	6.552	0.241	[6.078, 7.025]	6.746	0.193	[6.371, 7.124]
Knot, Mean	4.705	0.115	[4.488, 4.937]	4.700	0.113	[4.480, 4.927]
I, Variance	22.683	5.184	[12.926, 33.419]	21.094	5.834	[10.052, 32.733]
S1, Variance	12.760	1.992	[9.248, 17.032]	12.573	2.036	[9.013, 16.969]
S2, Variance	11.369	2.168	[7.492, 16.018]	11.650	2.227	[7.713, 16.421]
COV(I, S1)	8.076	2.476	[3.393, 13.145]	8.479	2.705	[3.445, 13.964]
COV(I, S2)	-3.037	2.270	[-7.540, 1.383]	-3.049	2.443	[-7.977, 1.624]
COV(S1, S2)	-3.777	1.481	[-6.836, -1.049]	-3.661	1.570	[-6.937, -0.753]
Error Variance	6.214	0.215	[5.806, 6.652]	6.222	0.237	[5.779, 6.714]
Class Proportion	0.409	0.035	[0.342, 0.479]	0.393	0.033	[0.329, 0.458]
Class 2						
I, Mean	29.561	0.294	[28.984, 30.136]	29.366	0.215	[28.946, 29.787]
S1, Mean	24.241	0.227	[23.802, 24.688]	24.116	0.185	[23.755, 24.481]
S2, Mean	5.489	0.218	[5.058, 5.908]	5.474	0.172	[5.138, 5.814]
Knot, Mean	4.105	0.049	[4.007, 4.201]	4.119	0.045	[4.027, 4.204]
I, Variance	117.930	12.983	[94.822, 145.495]	118.004	12.634	[95.474, 144.903]
S1, Variance	6.340	1.443	[3.733, 9.345]	6.514	1.353	[4.069, 9.350]
S2, Variance	2.934	0.787	[1.569, 4.671]	2.893	0.711	[1.645, 4.438]
COV(I, S1)	5.457	2.956	[-0.462, 11.203]	6.010	2.790	[0.449, 11.437]
COV(I, S2)	-16.553	2.738	[-22.232, -11.538]	-16.530	2.539	[-21.807, -11.790]
COV(S1, S2)	-2.092	0.891	[-4.001, -0.492]	-2.117	0.808	[-3.816, -0.674]
Error Variance	8.169	0.180	[7.827, 8.532]	8.144	0.174	[7.815, 8.495]
Class Proportion	0.591	0.035	[0.521, 0.658]	0.607	0.033	[0.542, 0.671]

Note. *M* = Posterior Mean; *SD* = Posterior Standard Deviation; CI = Credible Interval; I = Intercept; S1 = Slope 1 (slope of the first phase); S2 = Slope 2 (slope of the second phase); COV = Covariance.

PGMM. The selection of the two-class PGMM by the DIC aligns with a previous study by Kohli, Hughes, et al. (2015), where the BIC favored the two-class PGMM over one- or three-class PGMMs. Importantly, such discrepancies in the results from different information criteria showcase that researchers may encounter conflicting results when relying solely on statistical criteria. Therefore, we recommend that researchers consider multiple aspects, including class separation, substantive interpretation, and previous research in their decision-making process. In our illustration, we adopted the two-class PGMM as the most appropriate choice based on a variety of facets described above.

To assess the robustness of our results under the two-class solution, we conducted a prior sensitivity analysis and examined the impact of weakly informative and informative priors on parameter estimation. Although the estimates of growth means showed some changes compared to the results obtained under diffuse prior settings, the most notable change occurred in the latent class proportions. Under informative and weakly informative prior settings, over half of the students were assigned to Class 2, which represents the high-achieving group, whereas under diffuse prior settings, less than half of the students were assigned to the high-achieving group. The evident impact of the Dirichlet prior on latent class proportions was thus found in our study, which echoes the importance of prior distributions on latent class proportions examined by Depaoli (2013) and Depaoli et al. (2017). In addition, priors can impact multiple parameters. For instance, we observed shifts in other parameter estimates such as the knot location. These findings highlight the importance of considering different prior specifications in Bayesian PGMMs and their implications for the interpretation of results. Overall, our analysis provided a comprehensive overview of the implementation of Bayesian PGMMs from the perspective of applied researchers, demonstrating the steps involved in class separation, class enumeration, prior sensitivity analysis, and interpretation of results.

We emphasize that any decisions made throughout Bayesian mixture modeling offer considerable flexibility, providing researchers with ample room to tailor their own analytic procedures and make informed decisions. For instance, in terms of class separation, researchers may opt to collapse latent classes, as we did, or choose to retain separate classes based on substantive theories or expert knowledge (Depaoli, 2021). Similarly, when it comes to class enumeration, researchers may place more emphasis on statistical criteria or substantive theories. It is even possible that different information criteria may favor different class solutions. In our example, the two-class PGMM was preferred by the DIC, whereas the WAIC and LOOIC favored the three-class PGMM. It is thus important to review all the available statistical criteria and consider substantive meanings of latent classes. After class separation and enumeration have been assessed, checking the robustness of their findings based on the available information becomes a primary concern (Depaoli et al., 2017). In practice, eliciting and specifying prior distributions in a proper and accurate way may not always be feasible and information may be limited (van de Schoot et al., 2018). In this regard, prior distributions inherently contain some level of inaccuracy that can impact parameter estimation and, consequently, research conclusions. If researchers encounter fluctuating results under different prior specifications, this suggests that the data may not align well with the underlying theories, and this aspect should be thoroughly discussed as an important substantive conclusion. If different prior settings have minimal impact on the final estimates, this indicates

the underlying theory is strong and reinforces confidence in the stability of the model results.

Recognizing the considerable latitude for statistical decision-making in the context of Bayesian PGMMs, we underscore the paramount importance of method implementation and reporting of the results across all analytic procedures in a transparent manner (Depaoli, 2021; Depaoli & van de Schoot, 2017). To achieve this, researchers should plan their statistical analyses prior to data collection and analysis. In addition, it is crucial to contextualize any meaningful findings in relation to substantive theories. By considering both the statistical results and the substantive theories, researchers can engage in comprehensive discussions that lead to well-rounded research conclusions. It is through this harmonious integration of statistical and substantive perspectives that a deeper understanding of the phenomena under investigation can be achieved. Therefore, researchers must strive to effectively communicate their findings and foster a balanced discourse that encapsulates both statistical rigor and theoretical relevance.

7.1. Future methodological work

Given the complexity that PGMMs can take on, it is particularly difficult to construct generalized sample size requirements for accurate modeling and estimation purposes. For example, the number of latent classes, relative size of those classes, and the separation of those classes in relation to one another are all aspects that interplay and create a complex picture of sample size requirements in conventional GMMs (Depaoli, 2013). When the GMM is extended by including knots, then aspects such as segment length, overall length of the trajectory, and the number and placement of knots can also play important roles in affecting base-level sample size requirements. Because of these complexities, and the importance of determining sample size bounds for applied inquiries, the methodological literature should work to establish an online power analysis tool where researchers can enter specifications (e.g., number of classes, number and placement of knots, approximate size of classes) and a brief simulation would be run to determine lower-bound sample size levels that produce unbiased and efficient estimates. Such a tool does not yet exist, but it would be of great benefit to the empirical PGMM literature.

7.2. Concluding remarks

In conclusion, we hope that this article has provided practical guidance for applied researchers and practitioners interested in implementing Bayesian PGMMs. By addressing key methodological considerations such as class separation, class enumeration, and prior sensitivity, we encourage researchers to confidently and thoroughly implement Bayesian PGMMs in their research endeavors. We believe that the adoption of this modeling approach can greatly benefit the school psychology field by uncovering hidden subgroups, addressing population heterogeneity, and investigating complex nonlinear phenomena. This can ultimately lead to a deeper understanding of student development and inform targeted interventions. We encourage researchers to embrace Bayesian PGMMs as a powerful tool for enhancing the quality of educational support in school settings.

CRediT authorship contribution statement

Ihnwhi Heo: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarah Depaoli:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Conceptualization. **Fan Jia:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Conceptualization. **Haiyan Liu:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Conceptualization.

Declaration of competing interest

We have no conflicts of interest to disclose.

Appendix A. Model definition

This section provides additional information about the model definition for PGMMs. The mixture modeling approach, including PGMMs, assumes that observed data are generated from a mixture distribution. Put differently, mixture models such as PGMMs can be expressed as a combination of probability densities of each latent class. For a latent class c , we can represent the mixture density, $f(\mathbf{y}_i|\mathbf{\Omega})$, as follows:

$$f(\mathbf{y}_i|\mathbf{\Omega}) = \sum_c^C \pi_c f_c(\mathbf{y}_i|\theta_c). \quad (\text{A1})$$

In Eq. (A1), \mathbf{y}_i is a vector of repeatedly measured observed variables for an individual i across T time points, $\mathbf{\Omega}$ refers to a vector of parameters being estimated, and π_c is the unknown mixture proportion for a class c where $c = 1, 2, \dots, C$, $f_c(\cdot)$ are the densities across all the C classes assumed to follow a multivariate normal distribution. Because π_c refers to mixture class proportions of a latent class c , these should add up to 1. For instance, if the true population consists of two classes where 30% and 70% of the population belong to Class 1 and Class 2, respectively, $\pi_1 = 0.3$, $\pi_2 = 0.7$, and $\pi_1 + \pi_2 = 1.0$. Note that any elements with subscript c denote those elements can vary across the latent classes. If necessary, certain parameters can be fixed across classes by researchers.

It is assumed that the densities across all the C latent classes are multivariate normally distributed. In Eq. (A1), the multivariate

normality of the data in each latent class can be expressed such that $y | c \sim \mathcal{MN}(\mu_c, \Sigma_c)$ with μ_c and Σ_c being respectively referred to as a mean vector and a covariance matrix for a class c . In addition, we have introduced that Ω is a vector that contains parameters being estimated. The Ω vector can be extended such that $\Omega = (\pi, \Theta)'$, where π contains latent class proportions for all the C latent classes and Θ contains model parameters θ_c for each latent class c such that $\Theta = (\theta_1, \theta_2, \dots, \theta_c)$.

The formulation of a PGMM under the structural equation modeling framework, which combines the measurement model and the structural model, leads to the following reduced form:

$$y_{ic} = \Lambda_y(\alpha_c + \xi_{ic}) + \varepsilon_{ic}. \tag{A2}$$

The model-implied mean and covariance structure of the Eq. (A2) are as follows:

$$\begin{aligned} \mu_c(\theta) &= \Lambda_y \alpha_c, \\ \Sigma_c(\theta) &= \Lambda_y \Psi_\eta \Lambda_y' + \Theta_{ec}. \end{aligned} \tag{A3}$$

In Eq. (A3), Ψ_η indicates the covariance matrix of growth factors, and Θ_{ec} refers to the covariance matrix of the error terms of repeatedly measured observed variables. The growth factor covariance matrix, Ψ_η , can be treated as heterogeneous across latent classes by the addition of the subscript c .

Appendix B. Prior distributions

Appendix B is devoted to illustrating the prior specifications for model parameters in PGMMs. The first parameter to consider is the mixture class proportions, π . The most common prior distribution used is the Dirichlet prior

$$\pi \sim \mathcal{D}(d_1, \dots, d_C), \tag{B1}$$

where the hyperparameters d_1, \dots, d_C describe the proportion of cases in the C latent classes, thus they control how the latent classes are balanced in the prior. For example, if a researcher assumes that sample data consist of equal proportions of participants from Classes 1, 2, and 3, the research may specify the prior distribution as $\mathcal{D}(1, 1, 1)$. However, if there is a reason to believe that the sample dataset consists of 20%, 30%, and 50% of participants from Classes 1, 2, and 3, the corresponding prior distribution can be specified as $\mathcal{D}(0.2, 0.3, 0.5)$.

The second set of parameters to consider is the growth factor means. These parameters are typically assumed to be normally distributed, such that

$$\alpha_{mc} \sim \mathcal{N}(\mu_{\alpha_{mc}}, \sigma_{\alpha_{mc}}^2), \tag{B2}$$

where α_{mc} indicates the mean of growth factor $m = 1, 2, \dots, F$ of class c , $\mu_{\alpha_{mc}}$ indicates the center of this prior distribution, and $\sigma_{\alpha_{mc}}^2$ indicates the informativeness of this prior distribution via the distribution's variance.

Third, we define the prior distributions for the variance of the error terms in Θ_{ec} (see Eq. (A3)). A typical way to put priors on this matrix is to map an inverse gamma distribution to an individual element of the matrix Θ_{ec} . Let θ_{e_t} denote the variance of an individual element associated with the measurement at time point t such that $t = 1, 2, \dots, T$. We can specify prior distributions in the following way

$$\theta_{e_t} \sim \mathcal{IG}(a_{\theta_{e_t}}, b_{\theta_{e_t}}), \tag{B3}$$

where $a_{\theta_{e_t}}$ and $b_{\theta_{e_t}}$ represent the shape and scale for the inverse gamma prior distribution. We omitted the subscript c to enhance visibility, but different prior specifications are possible across latent classes if desired.

The fourth parameter is the factor covariance matrix Ψ_η . Commonly, the inverse Wishart prior distribution is used such that

$$\Psi_\eta \sim \mathcal{IW}(\Psi, \nu), \tag{B4}$$

where Ψ is a positive definite matrix of size m , and ν is the degrees of freedom that control the informativeness of the inverse Wishart prior distribution. This prior distribution can also vary across latent classes by adding a subscript c .

The fifth, and final, parameter is the knot location, γ . We present two possible forms of prior distributions. One method is to implement the uniform distribution to indicate a diffuse prior distribution over the knot location:

$$\gamma \sim \mathcal{U}(\min(t), \max(t)), \tag{B5}$$

where $\min(t)$ and $\max(t)$ represent the initial and last time points. This uniform prior distribution indicates that all values for γ are equally likely, which makes the uniform prior setting diffuse. The other method is to implement a truncated normal distribution to leverage the informativeness with respect to knot location:

$$\gamma \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2) \mathcal{T}(\min(t), \max(t)), \tag{B6}$$

where μ_γ and σ_γ^2 respectively define the location and the informativeness (variance) of this prior distribution, and $\mathcal{T}(\min(t), \max(t))$ represents that this normal prior distribution is truncated ranging from $\min(t)$ to $\max(t)$. It is natural to assume that knot locations are different across latent classes. Researchers can thus specify different prior distributions for knot locations across the latent class.

The diffuse prior distributions that were implemented in the initial analysis were, in part, pulled from Merkle and Rosseel (2018) as follows:

$$\begin{aligned} \pi &\sim \mathcal{D}(1, 1) \text{ for a two-class solution,} \\ \pi &\sim \mathcal{D}(1, 1, 1) \text{ for a three-class solution,} \\ \alpha_{mc} &\sim \mathcal{N}(0, 10^2), \\ \theta_{\epsilon_i} &\sim \mathcal{IG}(1, .5), \\ \Psi_{\eta_j} &\sim \mathcal{IW}(I, 3), \\ \gamma &\sim \mathcal{U}(0, 8.5), \end{aligned} \tag{B7}$$

where I is the identity matrix. Readers can consult the OSF materials for the code associated with the prior specifications.⁷

References

- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Muthén & Muthén.
- Byrd, C. M., & Andrews, D. J. C. (2016). Variations in students' perceived reasons for, sources of, and forms of in-school discrimination: A latent class analysis. *Journal of School Psychology, 57*, 1–14. <https://doi.org/10.1016/j.jsp.2016.05.001>.
- Coyle, S., Cipra, A., & Rueger, S. Y. (2021). Bullying types and roles in early adolescence: Latent classes of perpetrators and victims. *Journal of School Psychology, 89*, 51–71. <https://doi.org/10.1016/j.jsp.2021.09.003>.
- Daniels, M. J., & Hogan, J. W. (2008). *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. Chapman & Hall/CRC.
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*(2), 186–219. <https://doi.org/10.1037/a0031609>.
- Depaoli, S. (2021). *Bayesian structural equation modeling*. Guilford Press.
- Depaoli, S., Jia, F., & Heo, I. (2023). Detecting model misspecification in Bayesian piecewise growth models. *Structural Equation Modeling: A Multidisciplinary Journal, 30*(4), 574–591. <https://doi.org/10.1080/10705511.2022.2144865>.
- Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The wambs-checklist. *Psychological Methods, 22*(2), 240–261. <https://doi.org/10.1037/met0000065>.
- Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in Psychology, 11*, Article 608045. <https://doi.org/10.3389/fpsyg.2020.608045>.
- Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian statistics to model uncertainty in mixture models: A sensitivity analysis of priors. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(2), 198–215. <https://doi.org/10.1080/10705511.2016.1250640>.
- Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). Guilford Press.
- Frühwirth-Schnatter, S. (2010). *Finite mixture and markov switching models*. Springer.
- Garon-Carrier, G., Boivin, M., Lemelin, J.-P., Kovas, Y., Parent, S., Séguin, J. R., Vitaro, F., Tremblay, R. E., & Dionne, G. (2018). Early developmental trajectories of number knowledge and math achievement from 4 to 10 years: Low-persistent profile and early-life predictors. *Journal of School Psychology, 68*, 84–98. <https://doi.org/10.1016/j.jsp.2018.02.004>.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*(6), 997–1016. <https://doi.org/10.1007/s11222-013-9416-2>.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(2), 202–226. <https://doi.org/10.1080/10705510709336744>.
- Heo, I., Jia, F., & Depaoli, S. (2024). Performance of model fit and selection indices for Bayesian piecewise growth modeling with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 31*(3), 455–476. <https://doi.org/10.1080/10705511.2023.2264514>.
- Herman, K. C., Cohen, D., Reinke, W. M., Ostrander, R., Burrell, L., McFarlane, E., & Duggan, A. K. (2018). Using latent profile and transition analyses to understand patterns of informant ratings of child depressive symptoms. *Journal of School Psychology, 69*, 84–99. <https://doi.org/10.1016/j.jsp.2018.05.004>.
- Ingram, K. M., Davis, J. P., Espelage, D. L., Hatchel, T., Merrin, G. J., Valido, A., & Torgal, C. (2019). Longitudinal associations between features of toxic masculinity and bystander willingness to intervene in bullying among middle school boys. *Journal of School Psychology, 77*, 139–151. <https://doi.org/10.1016/j.jsp.2019.10.007>.
- Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science, 20*, 50–67. <https://doi.org/10.1214/088342305000000016>.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*(1), 39–59. <https://doi.org/10.1287/mksc.16.1.39>.
- Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). The Guilford Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>.
- Kellner, K. (2021). jagsUI: A wrapper around 'rjags' to streamline 'jags' analyses. <https://CRAN.R-project.org/package=jagsUI>.
- Kim, S., Tong, X., & Ke, Z. (2021). Exploring class enumeration in Bayesian growth mixture modeling based on conditional medians. *Frontiers in Education, 6*, Article 624149. <https://doi.org/10.3389/educ.2021.624149>.

⁷ <https://osf.io/udrb8/>.

- Kohli, N., Harring, J. R., & Zopluoglu, C. (2016). A finite mixture of nonlinear random coefficient models for continuous repeated measures data. *Psychometrika*, 81, 851–880. <https://doi.org/10.1007/s11336-015-9462-0>.
- Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear–linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods*, 20(2), 259–275.
- Kohli, N., & Sullivan, A. L. (2019). Linear-linear piecewise growth mixture models with unknown random knots: A primer for school psychology. *Journal of School Psychology*, 73, 89–100. <https://doi.org/10.1016/j.jsp.2019.03.004>.
- Kohli, N., Sullivan, A. L., Sadeh, S., & Zopluoglu, C. (2015). Longitudinal mathematics development of students with learning disabilities and students without disabilities: A comparison of linear, quadratic, and piecewise linear mixed effects models. *Journal of School Psychology*, 53(2), 105–120. <https://doi.org/10.1016/j.jsp.2014.12.002>.
- Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons.
- Liang, X., & Luo, Y. (2020). A comprehensive comparison of model selection methods for testing factorial invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(3), 380–395. <https://doi.org/10.1080/10705511.2019.1649983>.
- Lock, E. F., Kohli, N., & Bose, M. (2018). Detecting multiple random changepoints in Bayesian piecewise growth mixture models. *Psychometrika*, 83(3), 733–750. <https://doi.org/10.1007/s11336-017-9594-5>.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons.
- Merkle, E., & Rosseel, Y. (2018). Blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 85(4), 1–30. <https://doi.org/10.18637/jss.v085.i04>.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Sage.
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2), 463–469. <https://doi.org/10.1111/j.0006-341X.1999.00463.x>.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>.
- Papastamoulis, P. (2016). Label Switching: An R package for dealing with the label switching problem in MCMC outputs. *Journal of Statistical Software*, 69(1), 1–24. <https://doi.org/10.18637/jss.v069.c01>.
- Plummer, M. (2016). rjags: Bayesian graphical models using MCMC. <https://cran.r-project.org/web/packages/rjags/index.html>.
- Plummer, M., et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In , 124(125.10). *Proceedings of the 3rd international workshop on distributed statistical computing* (pp. 1–10).
- Ren, L., Tong, X., Xu, W., Wu, Z., Zhou, X., & Hu, B. Y. (2021). Distinct patterns of organized activity participation and their associations with school readiness among Chinese preschoolers. *Journal of School Psychology*, 86, 100–119. <https://doi.org/10.1016/j.jsp.2021.03.007>.
- van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, 6(1), Article 25216. <https://doi.org/10.3402/ejpt.v6.25216>.
- 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(2), 267–291. <https://doi.org/10.1080/00273171.2017.1412293>.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, 23(7), 691–697. <https://doi.org/10.1177/0956797612440101>.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64(4), 583–639. <https://doi.org/10.1111/1467-9868.00353>.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 62(4), 795–809. <https://doi.org/10.1111/1467-9868.00265>.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540. <https://doi.org/10.1080/01621459.1987.10478458>.
- Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. In G. Handcock, & K. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Information Age.
- Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study*. Doctoral dissertation. University of Jyväskylä.
- Tueller, S., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(1), 110–131. <https://doi.org/10.1080/10705511.2011.534695>.
- Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 165–192. <https://doi.org/10.1080/10705511003659318>.
- Vats, D., & Knudson, C. (2021). Revisiting the Gelman–Rubin diagnostic. *Statistical Science*, 36(4), 518–529. <https://doi.org/10.1214/20-STS812>.
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- Yao, W., & Lindsay, B. G. (2009). Bayesian mixture labeling by highest posterior density. *Journal of the American Statistical Association*, 104(486), 758–767. <https://doi.org/10.1198/jasa.2009.0237>.
- Zhang, Z., Hamagami, F., Lijuan Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31(4), 374–383. <https://doi.org/10.1177/0165025407077764>.