PSYCHOMETRIKA
DOI: 10.1007/s11336-015-9462-0

A FINITE MIXTURE OF NONLINEAR RANDOM COEFFICIENT MODELS FOR CONTINUOUS REPEATED MEASURES DATA

NIDHI KOHLI

UNIVERSITY OF MINNESOTA

JEFFREY R. HARRING

UNIVERSITY OF MARYLAND

CENGIZ ZOPLUOGLU

UNIVERSITY OF MIAMI

Nonlinear random coefficient models (NRCMs) for continuous longitudinal data are often used for examining individual behaviors that display nonlinear patterns of development (or growth) over time in measured variables. As an extension of this model, this study considers the finite mixture of NRCMs that combine features of NRCMs with the idea of finite mixture (or latent class) models. The efficacy of this model is that it allows the integration of intrinsically nonlinear functions where the data come from a mixture of two or more unobserved subpopulations, thus allowing the simultaneous investigation of *intra*-individual (within-person) variability, *inter*-individual (between-person) variability, and subpopulation heterogeneity. Effectiveness of this model to work under real data analytic conditions was examined by executing a Monte Carlo simulation study. The simulation study was carried out using an R routine specifically developed for the purpose of this study. The R routine used maximum likelihood with the expectation–maximization algorithm. The design of the study mimicked the output obtained from running a two-class mixture model on task completion data.

Key words: estimation, finite mixture models, nonlinear, random coefficient models.

1. Introduction

More and more researchers and practitioners in the field of education and more generally, the social and behavioral sciences, are interested in investigating patterns of change or development over a particular span of time or other condition. The random coefficient model RCM (Raudenbush & Bryk, 2002; Verbeke & Molenberghs, 2000) is one of several statistical methods that have been developed to analyze continuous repeated measures data from longitudinal studies. The RCM refers to a model in which all coefficients of a function describing the change process vary across individuals. When only some coefficients are random while others are fixed across individuals, a more flexible version known as the mixed-effects model (Davidian & Giltinan, 1995; Laird & Ware, 1982) can be specified. Primary goals of analyses that employ RCMs are to represent change in average performance of the outcome over time as well as specific behavior of individuals who may very well change in distinct ways (Cudeck, 1996). Advantageously, RCMs also permit the disaggregation of the covariation among the repeated measures into level-1 (i.e., within-subjects fluctuations) and level-2 (i.e., between-subjects heterogeneity) covariance structures, which can be customized to align with theoretical expectations or to incorporate interesting features of the

Correspondence should be made to Nidhi Kohli, Quantitative Methods in Education Program, Department of Educational Psychology, University of Minnesota, 161 Education Sciences Bldg., 56 East River Road, Minneapolis, MN 55455, USA. Email: nkohli@umn.edu

Published online: 30 April 2015

data (see, e.g., Fitzmaurice, Laird, & Ware, 2011; Singer & Willett, 2003). Greater detail on these structures will be forthcoming. Frequently, static covariates are introduced at a later stage in an analysis to account for variability in the growth parameters defining individual change. This is an important data analytic endeavor and often one of scientific significance; however, the three basic components of the RCM (i.e., the mean vector of regression coefficients, the level-1 within-subjects covariance structure, and level-2 between-subjects covariance structure) must be properly specified for drawing valid inferences on regression coefficients of the model, parameters that practitioners are often most interested in (Cudeck & Harring, 2007).

RCMs, including both linear and nonlinear functions, can be viewed as a generalization of conventional multiple regression models. At its core, linear RCMs are used when individual growth trajectories show straight-line patterns of change, or more generally, when the random effects of the model enter the within-individual function in a linear fashion. The model characterizes the change process by a function common to all subjects, but whose parameterization is allowed to vary among individual subjects, thereby allowing for between-subject variability. When the repeated measures show a curvilinear trend, researchers, by default, frequently turn to higher-order polynomials to model nonlinearity in the data. In fact, within the observed range of the data, increasing the order of the polynomial model will frequently result in an accurate approximation to the true, often nonlinear, time-response relation (Davidian & Giltinan, 1995; Pinheiro & Bates, 2000). Describing systematic change between the response and covariates using polynomial models is convenient and always available. Modeling real-world phenomena under these general suppositions, particularly when more flexible and scientifically relevant alternatives exist, seems quite unsatisfying, however. In contrast to polynomials, RCMs can be expanded to include intrinsically nonlinear functions (i.e., the response function has at least one parameter with corresponding random effect that enters the function nonlinearly). These nonlinear functions can frequently be tailored so that parameters have a natural, physical interpretation which corresponds to substantive characteristics of the research situation. In general, nonlinear models provide greater flexibility than their linear counterparts, requiring fewer parameters for the same quality of fit (Vonesh & Carter, 1992), and often portray a more realistic representation of the developmental or change process (Pinheiro & Bates, 2000).

It is often assumed in RCMs that individuals on whom the repeated measures are observed are drawn from the same population, and thus, share the same functional form of the continuous response variable. However, there are situations where the repeated measures data come from a mixture of two or more unobserved subpopulations (i.e., latent classes). In a situation like this where unobserved heterogeneity in data is suspected, RCMs can be extended to include a categorical latent class variable to analyze mixtures of repeated measures data. In other words, the idea of finite mixture models can be combined with RCMs to analyze and estimate respective growth trajectories (i.e., growth parameters) corresponding to each of the subpopulations.

This leads to the primary objective of the current study where the aim is to extend RCMs with intrinsically nonlinear functions to a framework that integrates features of finite mixture models. For the purpose of this study, we will focus on one particular nonlinear function, that is, a two-phase piecewise function where the location and variance of knot (change point), a nonlinear parameter, will be estimated, along with other model parameters. Efficacy of this modeling framework will be assessed under real data analytic conditions. To elaborate, we first show the utility of the finite mixtures of nonlinear random coefficient models (NRCMs) in the substantive areas of psychology and education using a real data example. We then design a Monte Carlo simulation study that mimics the output obtained from the analysis of real data example. That is, we evaluate the performance of the model with respect to model convergence and accuracy of estimated model parameters. The estimation procedure for this study was carried out by writing a self-coded routine, fitPMM.R (Zopluoglu, Harring, & Kohli, 2014), in the R program version 3.0.1 (R Core Team, 2013) that employs the expectation–maximization (EM) algorithm within a maximum likelihood (ML) estimation framework. Other statistical software programs do have the capacity to

fit finite mixtures of nonlinear random coefficient models. For example, Codd and Cudeck (2014) demonstrated how SAS PROC NLMIXED could be used to fit these types of nonlinear mixture models with marginal maximum likelihood via Gaussian quadrature using the user-defined likelihood feature in the module. However, unlike SAS PROC NLMIXED, the EM algorithm used in the self-coded R routine for the Monte Carlo simulation is not as susceptible to issues of non-convergence incurred by poor starting values—especially for the variance–covariance model parameters.

The remainder of this paper is organized as follows. NRCMs, finite mixture models, finite mixtures of NRCMs, and model estimation are briefly reviewed and discussed. The significance of the finite mixtures of NRCMs in the areas of psychology and education is illustrated using a real data example. In a subsequent method section, the design of the simulation study that follows along the lines of the output obtained from the real data example analysis is described. In a final section, the results from the simulation study are presented.

2. Nonlinear Random Coefficient Model (NRCM)

The NRCM (see, e.g., Cudeck & Harring, 2007; Davidian & Giltinan, 1995; Vonesh, 2012) is appropriate for situations where the change pattern of measured variables occurs more quickly in some periods than in others. In addition, these models are suitable when the functional form of individual trajectories follows the same curvilinear form of population (average) trajectory, yet individual differences with regard to the time-response function are quite evident (Harring, 2012). The NRCM can also accommodate imbalanced longitudinal designs (subjects measured at different time points, or measurement occasions) and data missing at random (Cudeck, 1996). The general form of NRCM for the response y_{ij} for the ith individual at the jth observation is specified as

$$y_{ij} = f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \ldots, \beta_{pi}) + e_{ij} \quad i = 1, \ldots, M;\ \ j = 1, \ldots, n_i,$$

where f denotes an intrinsically nonlinear function, t_{ij} denotes predictor time, $\mathbf{z}_i$ is a vector of individual attributes, $\beta_{1i}, \ldots, \beta_{pi}$ denote the p subject-specific regression parameters, and e_{ij} denotes time-specific residual error. We change the notation slightly and write the function f above using the formulation described by Crowder and Hand (1990, sect. 9.4):

$$\mathbf{y}_i = \mathbf{X}_i(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i + \mathbf{e}_i. \tag{1}$$

This version of the model differs from the typical hierarchical specification defined elsewhere (see, e.g., Davidian & Giltinan, 1995); however, a salient advantage is the straightforward demarcation between parameters that enter the function linearly and those that do not. In addition, this representation facilitates the estimation algorithm that is outlined herein. The $n_i \times p$ design matrix, $\mathbf{X}_i(\boldsymbol{\gamma}_i)$, a function of nonlinear parameter vector $\boldsymbol{\gamma}_i$, incorporates covariates that account for change in $\mathbf{y}_i$ including measurements of time, $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})'$ recorded as elapsed time from the beginning of the study or an individual's age at the jth occasion. Individual regression coefficients are of two types: p values, $\boldsymbol{\beta}_i = (\beta_{1i}, \ldots, \beta_{pi})'$, that enter linearly, and q values, $\boldsymbol{\gamma}_i = (\gamma_{1i}, \ldots, \gamma_{qi})'$, that enter the function in a nonlinear fashion. Vector $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})'$ comprised of time-specific residual errors representing the lack of fit between data for the ith subject and the model accounting for the responses. The distribution of residuals is assumed to be multivariate normal with zero mean vector and covariance structure, $\boldsymbol{\Theta}_i(\boldsymbol{\beta}_i,\ \boldsymbol{\xi})$, which may depend on parameter vector, $\boldsymbol{\xi}$, and possibly an individual's response function characterized by his or her

regression coefficients. Many times, when coupled with the random effects, a simple structure such as a mutually uncorrelated, heterogeneous variance structure $\boldsymbol{\Theta}_i(\boldsymbol{\beta}_i,\ \boldsymbol{\xi}) = \boldsymbol{\Theta}_i(\boldsymbol{\xi}) = \sigma_j^2\mathbf{I}_{n_i}$ for $j = 1,\ldots,n_i$ is adopted where $\boldsymbol{\xi} = (\sigma_1^2,\ldots,\sigma_{n_i}^2)'$. However, other structures that take into account serial correlation can be incorporated (Jennrich & Schluchter, 1986), while a more general specification of the variance components of this structure that depends on an individual's regression coefficients is possible (see, e.g., Davidian & Giltinan, 1995; 2003). From this point forward without loss of generality, we suppress the $\boldsymbol{\beta}_i$ notation and assume that the within-individual covariance structure, $\boldsymbol{\Theta}_i(\boldsymbol{\xi})$, may vary across individuals in terms of its dimension but not otherwise.

At the population level, the individual regression coefficients can be specified simply as the sum of fixed and random effects, although static covariates could certainly be added to each of these models.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i \text{ and } \boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \mathbf{g}_i. \tag{2}$$

The joint distribution of the random effects is assumed to be multivariate normal,

$$\begin{pmatrix}\mathbf{b}_i\\ \mathbf{g}_i\end{pmatrix} \sim N(\mathbf{0},\ \boldsymbol{\Phi}),$$

where

$$\boldsymbol{\Phi} = \begin{pmatrix}\boldsymbol{\Phi}_{bb} & \boldsymbol{\Phi}_{bg}\\ \boldsymbol{\Phi}_{gb} & \boldsymbol{\Phi}_{gg}\end{pmatrix}, \tag{3}$$

and where the order of submatrices $\boldsymbol{\Phi}_{bb}$, $\boldsymbol{\Phi}_{gg}$, and $\boldsymbol{\Phi}_{gb}$ are $(p\ \times\ p)$, $(q\ \times\ q)$, and $(q\ \times\ p)$, respectively. While $\boldsymbol{\Phi}$ can be specified to distinguish random-effects distributions by groups and other types of data nuances, a typical formulation is to let $\boldsymbol{\Phi}$ be unstructured with each element freely estimated.

Unlike linear RCMs which are estimated using ML, the application of ML estimation procedure for the analysis of NRCMs is not seen as frequently (Cudeck, 1996). This is because the joint likelihood function for parameters of the NRCM requires a multidimensional integration over the random effects to obtain the marginal distribution of the response where inferences are typically made. This multidimensional integral does not have a closed form, and thus cannot be calculated explicitly. du Toit and Cudeck (2009) presented a marginal ML estimation method for the estimation of NLME models where parameters of the time-response function could be decoupled into those that enter the function linearly and those that enter into a nonlinear fashion. The import of their article was to show how the marginal distribution of the data required integration of dimension equal to the number of nonlinear parameters with random effects—a significant savings in computing time for approximate estimation methods that perform the integration directly. In their article, the application of marginal ML estimation method was carried out by writing the marginal distribution of $\mathbf{y}_i$ as a conditional distribution of $\mathbf{y}_i$ given the nonlinear random effects $\mathbf{g}_i$. Briefly, for a particular value of the nonlinear random effects, $\mathbf{g}_i = \tilde{\mathbf{g}}$, the conditional distribution of $\boldsymbol{\beta}_i|\mathbf{g}_i = \tilde{\mathbf{g}}$ is

$$(\boldsymbol{\beta}_i|\mathbf{g}_i = \tilde{\mathbf{g}}) \sim N(\boldsymbol{\mu}_b,\ \boldsymbol{\Phi}_b),$$

and given the distribution expressed in Equation 3, the moments of the conditional distribution are

$$\boldsymbol{\mu}_b = \boldsymbol{\beta} + \boldsymbol{\Phi}_{bg}\boldsymbol{\Phi}_{gg}^{-1}\tilde{\mathbf{g}} \quad \boldsymbol{\Phi}_b = \boldsymbol{\Phi}_{bb} - \boldsymbol{\Phi}_{bg}\boldsymbol{\Phi}_{gg}^{-1}\boldsymbol{\Phi}_{gb}.$$

The implication is that $\mathbf{y}_i$ conditional on $\mathbf{g}_i = \tilde{\mathbf{g}}$ can be specified as

$$\mathbf{y}_i|\tilde{\mathbf{g}} = \mathbf{X}_i(\boldsymbol{\gamma} + \tilde{\mathbf{g}})\boldsymbol{\beta}_i + \mathbf{e}_i,$$

which has a normal distribution with mean vector and covariance matrix,

$$\begin{aligned}\boldsymbol{\mu}_y &= E[\mathbf{y}_i|\tilde{\mathbf{g}}] = \mathbf{X}_i(\boldsymbol{\gamma} + \tilde{\mathbf{g}})\boldsymbol{\mu}_b \\ \boldsymbol{\Sigma}_y &= \text{cov}[\mathbf{y}_i|\tilde{\mathbf{g}}] = \mathbf{X}_i(\boldsymbol{\gamma} + \tilde{\mathbf{g}})\boldsymbol{\Phi}_b\mathbf{X}_i(\boldsymbol{\gamma} + \tilde{\mathbf{g}})'.\end{aligned}$$

du Toit and Cudeck (2009) demonstrated that how the marginal distribution of $\mathbf{y}_i$ could be defined as

$$h(\mathbf{y}_i) = \int_g p_{y|g}(\mathbf{y}_i|\mathbf{g}_i)p_g(\mathbf{g}_i)d\mathbf{g}_i, \tag{4}$$

where $p_{y|g}(\mathbf{y}_i|\mathbf{g}_i)$ is the conditional distribution of $\mathbf{y}_i$ given $\mathbf{g}_i$ defined as

$$p_{y|g}(\mathbf{y}_i|\mathbf{g}_i) = (2\pi)^{-\frac{n_i}{2}}\left|\boldsymbol{\Sigma}_y\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_y)'\boldsymbol{\Sigma}_y^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_y)\right],$$

and where the nonlinear random effect has distribution $\mathbf{g}_i \sim N(\boldsymbol{\mu}_g,\ \boldsymbol{\Sigma}_g)$ with density

$$p_g(\mathbf{g}_i) = (2\pi)^{-\frac{q}{2}}\left|\boldsymbol{\Sigma}_g\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{g}_i - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mathbf{g}_i - \boldsymbol{\mu}_g)\right].$$

If $\mathbf{g}_i$ is a scalar, then $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$ will be as well; otherwise, the mean vector and covariance matrix will be of dimension $q \times 1$ and $q \times q$, respectively. In either case, the marginal distribution in Equation 4 can be computed using Gaussian Hermite quadrature although the difficulty in estimation increases with the dimension of $\mathbf{g}_i$. For many nonlinear functions used in the social and behavioral sciences including the linear piecewise model, the reduction in computational complexity is significant because the integration decreases to a single dimension (Harring, 2009).

3. Finite Mixture Models

The area of finite mixture models is sufficiently large and well established as a viable method useful in a number of disciplines for explaining relations in observed data. There are a number of general excellent resources including Everitt and Hand (1981) and McLachlan and Peel (2000). Medical research such as Boos and Brownie (1991) and Luo, Boos, and Tamura (2004) applied mixtures in a regression based framework, and Wedel and DeSarbo (1995) broadly described application of mixtures for generalized linear models (GLM). Others, such as Pavlic, Brand, and Cumming (2001) and Xu and Hedeker (2001), considered mixtures in two group experimental cross-sectional and longitudinal designs, respectively, with applied research and simulation studies, while Arminger, Stein, and Wittenberg (1999) developed finite mixtures of covariance structure models. The natures of these applications and empirical studies are varied; some simulations illustrated the viability of a mixture model representation, others specified elaborate models representative of a particular data set, while others focused on a posterior probability aspect of individual group membership.

Recently, mixture models have been extended in a variety of interesting ways (see, Hancock & Samuelsen, 2008; for a convenient summary of some of these elaborations). Pertinent

to the current discussion, mixture models have been integrated in methods and models for analyzing longitudinal data to address analytic situations in which intra-individual correlation and subpopulation heterogeneity exist simultaneously. These include mixtures of linear mixed-effects models (Verbeke & Lesaffre, 1996), growth mixture models (Muthén, 2001, 2002, 2004; Muthén & Shedden, 1999), and generalized mixed-effects mixture models (Hall & Wang, 2005).

Finite mixture models are frequently employed in situations in which latent heterogeneity in the population is suspected. To be specific, it may be reasonable to assume that a sample of observations arises from a composite of underlying subpopulations in which class membership for each observation is unobserved and where the relative proportions of the subpopulations are unknown (Everitt & Hand, 1981). Other applications of mixture models involve simply attempting to approximate complex distributions that exhibit multi-modality, over-dispersion, excessive skew, or other interesting characteristics (Titterington, Smith, & Makov, 1985). In the first type of application, the form of the densities in each of these subpopulations is often specified in advance and typically of the same parametric family with parameters that are allowed to differ across components or latent classes (Aitkin & Rubin, 1985; McLachlan & Peel, 2000). The primary inferential goals are (*i*) to decompose the sample into its mixture components and (*ii*) to estimate the mixture probabilities and the unknown parameters of each component density.

A finite mixture model is a probability model that combines the probability densities across all the subpopulations underlying the data. According to McLachlan and Peel (2000), the general form of a finite mixture model is given as

$$\mathbf{A}(\mathbf{y}) = \pi_1 \mathbf{A}_1(\mathbf{y}, \boldsymbol{\theta}_1) + \cdots + \pi_K \mathbf{A}_K(\mathbf{y}, \boldsymbol{\theta}_K), \tag{5}$$

where $\mathbf{A}(\mathbf{y})$ is the composite density function for all $k = 1, \ldots, K$ number of components. A single density $\mathbf{A}_k(\mathbf{y}, \boldsymbol{\theta}_k)$ is referred as the component density which depends on component specific parameter vector, $\boldsymbol{\theta}_k$. It is typically assumed that the distribution underlying the subpopulations (mixture components) has the same density, although this restriction can be relaxed (Gruen & Leisch, Gruen and Leisch 2007). The parameters $\pi_1, \pi_2, \ldots, \pi_K$ are called the mixing proportions. It is assumed that the mixing proportions π_k are non-negative quantities that sum to one. That is,

$$\sum_{k=1}^{K} \pi_k = 1 \text{ where } 0 < \pi_k < 1.$$

To make this general definition more tangible, consider a finite mixture of simple linear regression models. For a two-class regression mixture and integrating the equality constraint for the mixing proportions, Equation 5 would be defined as

$$\mathbf{A}(\mathbf{y}) = \pi_1 \mathbf{A}_1(\mathbf{y}, \boldsymbol{\theta}_1) + (1 - \pi_1)\mathbf{A}_2(\mathbf{y}, \boldsymbol{\theta}_2),$$

where the two-component densities for sample size n are

$$\mathbf{A}_1(\mathbf{y}, \boldsymbol{\theta}_1) = \left(2\pi\sigma_1^2\right)^{-n/2} \exp\left\{ -(\mathbf{y} - \beta_{01} - \beta_{11}\mathbf{x})^2 / 2\sigma_1^2 \right\},$$

and

$$\mathbf{A}_2(\mathbf{y}, \boldsymbol{\theta}_2) = \left(2\pi\sigma_2^2\right)^{-n/2} \exp\left\{ -(\mathbf{y} - \beta_{02} - \beta_{12}\mathbf{x})^2 / 2\sigma_2^2 \right\},$$

respectfully. There are a total of six parameters defined by the regression mixture with each class having three parameters: regression intercept β_{0k} and slope β_{1k} and the residual error, σ_k^2, for k= 1, 2. The last parameter to be estimated is the mixing proportion, π_1.

For the general mixture model in Equation 5, $\boldsymbol{\theta}_k$ are unknown parameter values that must be estimated from the observed data. This is also the case with the number of components. That is, it is often not known that how many components exist a priori, and thus, the optimal number of components is inferred from the observed data. A standard method of estimating these models is ML estimation via the EM algorithm.

Finite mixture models are a type of a model-based clustering tool that can help to identify more than one unobserved population with the intent to infer qualitatively distinct classes of individuals in the population. When modeling population heterogeneity using finite mixture models, it is typically assumed that data came from a mixture of two or more distributions from the same parametric family with parameters that are allowed to differ across components (Frühwirth-Schnatter, 2006; McLachlan & Peel, 2000).

4. Nonlinear Random Coefficient Mixture Model (NRMM)

The NRMM is an amalgam of the NRCM and finite mixture model. In other words, the NRMM model represents a statistical method that incorporates latent classes (components) into the NRCM framework. This particular model can be effectively applied in situations where the continuous observed response data in vector $\mathbf{y}$ are a mixture of two or more unobserved subpopulation distributions. The incorporation of categorical latent classes allows the regression parameters (growth parameters) of the model to vary across classes, thus allowing for the estimation of class-specific mean parameter vector, and variance and covariance parameters matrices (Harring, 2012). In order to formulate a NRMM, consider that observed response data come from K subpopulations ($k = 1, \ldots, K$), with a latent categorical variable $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$ indicating class membership for individual i in latent class k. Note that c_{ik} is a dichotomous latent indicator that takes the value of 1 if the ith individual belongs to the kth latent class, and 0 otherwise. The NRMM for repeated measures for the ith individual from the population has the following mixture formulation:

$$\mathbf{y}_i = \sum_{k=1}^{K} \pi_k \mathbf{X}_i(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i + \mathbf{e}_i, \tag{6}$$

where $\mathbf{X}_i(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i$ is defined identically to that in Equation 1. Of note are $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$, the $p \times 1$ and $q \times 1$ vectors of individual-specific coefficients, respectively, that govern the functional form of the response for individual i. Codd and Cudeck (2014) partitioned both fixed effects and random effects into those that are specific to a particular class and those that are common across classes. This elaboration allows for some coefficients in $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ to be stochastic, while others are fixed with mean vector and covariance matrix permitted to be class-specific if the data analytic situation warrants.

In our formulation, given class k, the random coefficients in $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ have a multivariate normal distribution with mean vector and corresponding covariance matrix that are both class-specific

$$\begin{pmatrix} \boldsymbol{\beta}_i \\ \boldsymbol{\gamma}_i \end{pmatrix} \Bigg| \; k \sim N\left(\begin{bmatrix} \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{bmatrix}, \; \boldsymbol{\Phi}_k \right), \tag{7}$$

where $\boldsymbol{\Phi}$ was defined in the expression in Equation 3. Within each class, the error terms are also assumed to be normally distributed:

$$\mathbf{e}_i \mid k \sim N(\mathbf{0}, \; \boldsymbol{\Theta}_{ik}(\boldsymbol{\xi})), \tag{8}$$

where the $n_i \times n_i$ covariance matrix $\boldsymbol{\Theta}_{ik}(\boldsymbol{\xi})$ can depend on class-specific parameter vector, $\boldsymbol{\xi}$. For example, if serial correlation was suspected above and beyond what the random effects could account for and measurements were taken at equally spaced time points, the residuals may satisfy a first-order autoregressive process so that the pq-th element of $\boldsymbol{\Theta}_{ik}(\boldsymbol{\xi})$ is given by $[\boldsymbol{\Theta}_{ik}(\boldsymbol{\xi})]_{pq} = \sigma^2\rho^{|p-q|}$ where $\boldsymbol{\xi} = (\sigma^2,\ \rho)'$. Note that the subscript k in Equation 6 and the distributional assumptions at the individual and population levels of the model conveyed in Equations 7 and 8 permit a separate NRCM for each latent class k, thus allowing for between- and within-individuals heterogeneity with the caveat that both $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_{ik}(\boldsymbol{\xi})$ may be constrained to be the same across classes if suggested by the problem at hand.

The NRMM defined in Equations 6–8 is quite general and can be fit to a myriad of nonlinear functions with at least one coefficient that enters the function nonlinearly. For example, an exponential function of the form

$$f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \beta_{2i}, \gamma_i) = \beta_{2i} - (\beta_{2i} - \beta_{1i})\exp\{-\gamma_i t_{ij}\}$$

fits into $\mathbf{X}_i(\gamma_i)\boldsymbol{\beta}_i$ in Equation 6 by defining the vector-valued function

$$\mathbf{X}_i(\gamma_i) = \left[\exp\{-\gamma_i\mathbf{t}\}\ 1 - \exp\{-\gamma_i\mathbf{t}\}\right],$$

and letting $\boldsymbol{\beta}_i = (\beta_{1i},\ \beta_{2i})'$. As a second example, consider the situation where there is differential response to treatment. Suppose that some individuals show improvement initially then at some undetermined amount of time, begin to decline, while other individuals demonstrate continued improvement, but whose positive effect dissipates over time. Assume that the first group follows a quadratic function (see Cudeck & Toit, 2002), while a nonlinear function, such as a variant of the Michaelis–Menten model, would be appropriate to summarize the second group. These two functions are defined as

$$\left[f(t_{ij},\ \mathbf{z}_i, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i)\right]_j = \begin{cases} \beta_{2i} - (\beta_{2i} - \beta_{1i})\left(\frac{t_{ij}}{\gamma_{1i}} - 1\right)^2, & \text{group} = 1, \\ \beta_{4i} + \frac{(\beta_{4i} - \beta_{3i})t_{ij}}{\gamma_{2i} + t_{ij}}, & \text{group} = 2. \end{cases}$$

Although a bit more complicated, this situation can still be handled by the setup in Equation 6 by defining $\mathbf{X}_i(\gamma_i)$ by groups

$$\left[\mathbf{X}_i(\gamma_i)\right]_j = \left[\left(\frac{t_{ij}}{\gamma_{1i}} - 1\right)^2\ 1 - \left(\frac{t_{ij}}{\gamma_{1i}} - 1\right)^2\ 0\ 0\right], \qquad \left[\mathbf{X}_i(\gamma_i)\right]_j = \left[0\ 0\ \frac{-t_{ij}}{\gamma_{2i} + t_{ij}}\ 1 + \frac{t_{ij}}{\gamma_{2i} + t_{ij}}\right]$$

Group 1 Group 2

and defining $\boldsymbol{\beta}_i = (\beta_{1i},\ \beta_{2i}, \beta_{3i}, \beta_{4i})'$. Note that in the first example, there is a single nonlinear coefficient, while in the second example there are two nonlinear coefficients—one for each nonlinear function.

Different kinds of single, dynamic intrinsically nonlinear functions (e.g., exponential) can be used to adequately characterize a nonlinear growth pattern over a span of time. However, in situations or applications where the overall growth process exhibit discrete phases of development in observed repeated measurements, assuming a single, uninterrupted nonlinear functional form may not be unrealistic. For example, the data in Figure 1 are the number of words out of 15 possible that were recalled by a sample of college students over ten thirty-second trials of a single-session

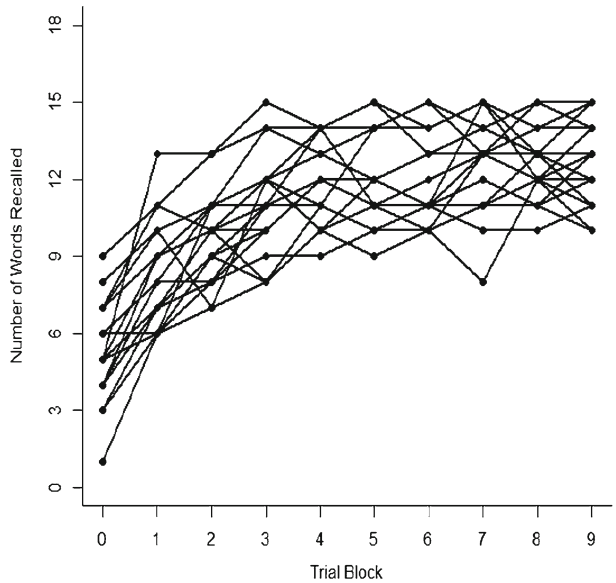

FIGURE 1.
Random subsample ($n = 20$) of repeated response word recall scores from a 30-s learning task.

experiment (Smith & Klebe, 1997). The trials were recorded from 0 to 9. Of the entire sample ($n = 103$), a random subset of 20 individuals is included in the figure. The number of words recalled on the initial trial varied between 3 and 10, whereas at the final trial the range was from 11 to 15. Most participants showed rapid and linear improvement up to trial 3, after which there was modest individual change in the number recalled for the duration of the study.

In the change process portrayed in Figure 1, the trajectory characterizing the rate of change could be composed of a phase of rapid constant improvement that gives way to a later phase of slower constant improvement. When this kind of process is observed, a piecewise function (a kind of nonlinear function) is appropriate for characterizing the distinct phases of development. While many interesting nonlinear functions could be espoused and fit to data with these features including smooth, uninterrupted functions like an exponential and logistic, in this study, we will focus on a two-phase piecewise function because it provides the opportunity to understand when the underlying process changes from one distinct phase of development to another. It is just this type of information that makes spline functions, in general, such popular alternatives to functions following a smooth trajectory. An interesting parameter of any piecewise function is the knot or changepoint. The knot is the value of the predictor, in this case time, at which the response function shifts from one phase to another (Cudeck & Klebe, 2002). The location of knot can be known a priori or can be estimated. For the NRCM like that described in Equation 1, a linear–linear piecewise function would be specified as follows:

$$f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \ldots, \beta_{4i}, \gamma_i) = \mathbf{X}_i(\gamma_i)\boldsymbol{\beta}_i = \begin{cases} \beta_{1i} + \beta_{2i} t_{ij} & t_{ij} \leq \gamma_i \\ \beta_{3i} + \beta_{4i} t_{ij} & t_{ij} > \gamma_i \end{cases}. \tag{9}$$

In Equation 9, β_{1i} and β_{2i} represent the individual intercept and slope regression coefficients of the first phase, respectively; β_{3i} and β_{4i} represent the individual intercept and slope regression coefficients of the second phase, respectively; and γ_i represents the knot. In many applications, the two linear segments join at the knot (i.e., $\beta_{1i} + \beta_{2i}\gamma_i = \beta_{3i} + \beta_{4i}\gamma_i$—the situation we consider

throughout the remainder of this article), which makes one of the regression coefficients redundant, and thus given this constraint, one parameter can be eliminated although the choice as to which coefficient to eliminate is completely arbitrary. Thus, applying this restriction to eliminate β_{3i} from Equation 9 results in the following model:

$$f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \ldots, \beta_{4i}, \gamma_i) = \mathbf{X}_i(\gamma_i)\boldsymbol{\beta}_i = \begin{cases} \beta_{1i} + \beta_{2i} t_{ij} & t_{ij} \leq \gamma_i \\ \beta_{1i} + \beta_{2i}\gamma_i + \beta_{4i}(t_{ij} - \gamma_i) & t_{ij} > \gamma_i \end{cases}. \tag{10}$$

A more compact expression that fits into Equation 1 and emphasizes that the two segments comprise a single function uses the truncation operator,

$$f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \ldots, \beta_{4i}, \gamma_i) = \mathbf{X}_i(\gamma_i)\boldsymbol{\beta}_i = \beta_{1i} + \beta_{2i} t_{ij} + (\beta_{2i} - \beta_{4i})(\gamma_i - t_{ij})_+, \tag{11}$$

where the last term in Equation 11 is defined as zero or the positive value of the argument, depending on its sign

$$v_+ = \begin{cases} v, & v \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

The function in Equation 11 contains three individual-specific linear regression coefficients (i.e., $\beta_{1i}, \beta_{2i}, \beta_{4i}$), and one individual-specific nonlinear coefficient, γ_i. The function on the right-hand side of Equation 11 fits into the design matrix, $\mathbf{X}_i(\boldsymbol{\gamma}_i)$, using the following vector-valued coding scheme

$$\mathbf{X}_i(\gamma + \mathbf{g}_i) = \mathbf{X}_i(\gamma_i) = [\,\mathbf{1}_{n_i}\ \mathbf{t}_i + ([\gamma_i \cdot \mathbf{1}_{n_i}] - \mathbf{t}_i)_+ - ([\gamma_i \cdot \mathbf{1}_{n_i}] - \mathbf{t}_i)_+].$$

Transitioning to mixtures, we assume conditional independence, and for simplicity, we assume $K = 2$ latent classes with the constraint, $\pi_2 = 1 - \pi_1$. Individuals in each latent class are hypothesized to have a linear–linear piecewise function of the following form:

$$\mathbf{y}_i|k = \mathbf{X}_i(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i + \mathbf{e}_i,$$

where for a linear–linear function that joins at the knot

$$\left[\mathbf{X}_{ik}(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i\right]_j = \beta_{1i} + \beta_{2i} t_{ij} + (\beta_{2i} - \beta_{4i})(\gamma_i - t_{ij})_+, \tag{12}$$

where $[\mathbf{u}]_j$ denotes the jth element of vector $\mathbf{u}$. Furthermore, given class k, individuals' regression coefficients in Equation 11 are posited to be the sum of fixed parameters and random effects

$$\beta_{1i}|k = \beta_{1k} + b_{1i}|k \quad \beta_{2i}|k = \beta_{2k} + b_{2i}|k \quad \beta_{4i}|k = \beta_{4k} + b_{4i}|k \quad \gamma_i|k = \gamma_k + g_i|k,$$

where $\beta_{1k}, \beta_{2k}, \beta_{4k},$ and γ_k are the fixed effects for class k. Inference for the two-phase piecewise NRMM is implemented on the marginal distribution of $\mathbf{y}_i$ that is obtained by writing it out as a conditional distribution of $\mathbf{y}_i$ given the single nonlinear random effect, g_i, and latent class k, which inturn reduces the integration to a single dimension.

The NRMM model can be estimated via maximum likelihood. The joint density for $\mathbf{y}_i$ can expressed as

$$\begin{aligned} h(\mathbf{y}_i) &= \int p(\mathbf{y}_i|\mathrm{g}_i)p(\mathrm{g}_i)dg_i \\ &= \int \sum_{k=1}^{K} \pi_k p^k(\mathbf{y}_i|\mathrm{g}_i)p^k(g_i)dg_i \\ &= \sum_{k=1}^{K} \pi_k \int p^k(\mathbf{y}_i|\mathrm{g}_i)p^k(\mathrm{g}_i)dg_i \\ &= \sum_{k=1}^{K} \pi_k h_{ik}(\mathbf{y}_i), \end{aligned} \tag{13}$$

where $p^k(\mathbf{y}_i|\mathrm{g}_i)$ is the conditional distribution of $\mathbf{y}$ given g_i for class k, while $p^k(\mathrm{g}_i)$ is the marginal distribution of g_i for class k. These are both assumed multivariate normal densities. In this form, the marginal density of $\mathbf{y}_i$ comes from a mixture of densities with class probabilities $\pi_1, \ldots, \pi_k$ and class-specific densities $h_{ik}(\mathbf{y}_i)$. The likelihood function can be defined as $L(\boldsymbol{\theta}) = \prod_{i=1}^{M} h(\mathbf{y}_i|\boldsymbol{\theta})$, from which the loglikelihood can be expressed as

$$\begin{aligned} l(\boldsymbol{\theta}) &= \ln\left(\prod_{i=1}^{M} h(\mathbf{y}_i|\boldsymbol{\theta})\right) \\ &= \ln\left(\prod_{i=1}^{M}\sum_{k=1}^{K} \pi_k h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)\right) \\ &= \sum_{i=1}^{M} \ln\left(\sum_{k=1}^{K} \pi_k h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)\right). \end{aligned}$$

To be concrete, for $K = 2$, the loglikelihood would be specified using the probability constraints as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{M} \ln\left([\pi_1 h_{i1}(\mathbf{y}_i|\boldsymbol{\theta}_k)] + [(1-\pi_1)h_{i2}(\mathbf{y}_i|\boldsymbol{\theta}_k)]\right).$$

4.1. Model Estimation

The two-phase piecewise NRMM is a very complicated model that cannot be easily fit with standard commercial software. Recently, Codd and Cudeck (2014) demonstrated that how a nonlinear mixture model could be fit through direct maximization of the loglikelihood function using the programming capabilities in SAS PROC MIXED. While this method shows promise in fitting NRMMs, it is susceptible to poor starting values and has not incorporated other aspects of growth mixture modeling (i.e., multiple sets of random starting values, spanning the parameter space for plausible starting parameter values) that are now accepted practice when estimating these kinds of models. A self-coded R routine (Zopluoglu et al., 2014) in the R program version 3.0.1 (R Core Team, R Core Team 2013) was used to estimate this model using ML estimation via the EM algorithm. The R routine followed the estimation methods presented in du Toit and

Cudeck (2009) and Harring (2012). Following Harring (2012), the complete data loglikelihood function for the response data vector $\mathbf{y}$ and the vector of latent categorical variable $\mathbf{c}$ with K classes can be written as

$$\ln L(\boldsymbol{\theta}|\mathbf{y},\mathbf{c}) = \sum_{i=1}^{M} \ln L_i(\boldsymbol{\theta}|\mathbf{y}_i,\mathbf{c}_i) = \sum_{i=1}^{M}\sum_{k=1}^{K} c_{ik}\left\{\ln(\pi_k) + \ln[h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)]\right\}, \tag{14}$$

where π_k is the *prior* probability of belonging to the kth class; h_{ik} denotes the marginal distribution of $\mathbf{y}_i$ conditional on latent class k, and model parameters $\boldsymbol{\theta}_k$, where $\boldsymbol{\theta}_k = (\beta_{1k}, \beta_{2k}, \beta_{3k}, \gamma_k, vech(\boldsymbol{\Phi}_k)', \sigma_k^2)$, where the $vech(\cdot)$ operator creates a column vector of a symmetric matrix by stacking the lower diagonal elements below one another. Note that for the model in Equation 14 the covariance matrix of random effects $\boldsymbol{\Phi}_k$ has 10 unique elements. Including the class probabilities $(\pi_1, \ldots, \pi_{K-1})$, the total number of model parameters to be estimated is $16K-1$.

The *complete* data loglikelihood equation above is substituted by its conditional expectation function because the vector $\mathbf{c}$ is unobserved. To elaborate, in the E-step of the EM iterative procedure, the posterior probabilities of belonging to a latent class with respect to $\boldsymbol{\theta}$ at iteration υ are computed. Let $\boldsymbol{\theta}^{\upsilon}$ be the value of $\boldsymbol{\theta}$ at iteration υ, then the conditional *expected* loglikelihood value at the E-step is specified as

$$E\left[\ln L(\boldsymbol{\theta}|\mathbf{y},\mathbf{c})|\mathbf{y},\boldsymbol{\theta}^{\upsilon}\right] = \sum_{i=1}^{M}\sum_{k=1}^{K} E\left[c_{ik}|\mathbf{y},\boldsymbol{\theta}^{\upsilon}\right]\{\ln(\pi_k) + \ln[h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)]\}, \tag{15}$$

where $E[c_{ik}|\mathbf{y},\boldsymbol{\theta}^{\upsilon}]$ is the posterior probability of the ith subject who belongs to the kth class and is equal to

$$E\left[c_{ik}|\mathbf{y},\boldsymbol{\theta}^{\upsilon}\right] = \frac{\pi_k^{\upsilon} h_{ik}\left(\mathbf{y}_i|\boldsymbol{\theta}^{\upsilon}\right)}{\sum_{k=1}^{K} \pi_k^{\upsilon} h_{ik}\left(\mathbf{y}_i|\boldsymbol{\theta}^{\upsilon}\right)} = \pi_{ik}\left(\boldsymbol{\theta}^{\upsilon}\right). \tag{16}$$

In the M-step of the EM iterative procedure, the conditional expected loglikelihood function in Equation 15 is maximized with respect to $\boldsymbol{\theta}$ to obtain the updated parameter vector, $\boldsymbol{\theta}^{\upsilon+1}$. The conditional expected loglikelihood function to be maximized at iteration $\upsilon$is

$$\sum_{i=1}^{M}\sum_{k=1}^{K} \pi_{ik}\left(\boldsymbol{\theta}^{\upsilon}\right)\ln(\pi_k) + \sum_{i=1}^{M}\sum_{k=1}^{K} \pi_{ik}\left(\boldsymbol{\theta}^{\upsilon}\right)\ln[h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)]. \tag{17}$$

The class probabilities are updated as a result of the maximization of the first part of the function in Equation 17 in the following way:

$$\pi_k^{\upsilon+1} = \frac{1}{M}\sum_{i=1}^{M} \pi_{ik}\left(\boldsymbol{\theta}^{\upsilon}\right).$$

The second part of Equation 17

$$\sum_{i=1}^{M}\sum_{k=1}^{K} \pi_{ik}\left(\boldsymbol{\theta}^{\upsilon}\right)\ln[h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)]$$

is a loglikelihood function weighted by the posterior probabilities of likely class membership at iteration υ. Component density, $h_{ik}(\mathbf{y}_i|\boldsymbol{\theta}_k)$, is obtained through approximating the integral in Equation 13 using Gaussian Hermite quadrature previously outlined.

5. Purpose of the Study

It has been demonstrated previously in the literature how to estimate a piecewise model with unknown random knot(s) for a single population (see, e.g., Harring, Cudeck, & du Toit, 2006; Kohli & Harring , 2013; Wang & McArdle , 2008), and for a mixture of subpopulations (see, e.g., Harring, 2012; Kohli, Harring, & Hancock, 2013). However, no study up to this point has investigated how reliable (stable) the NRMM with a two-phase piecewise function is with respect to the model convergence and recovering of true model parameters under manipulated conditions. For example, Wang and McArdle (2008) in their simulation study compared the fit of a single-class piecewise mixed-effects model with unknown knot locations using a Bayesian approach and the ML approach. Harring (2012) presented an estimation scheme for fitting nonlinear mixed-effects mixture models, but the study did not show how well the estimation procedure performed under manipulated conditions, and thus there is an absence of knowledge with regard to the quality of the proposed method with respect to model convergence and /or accuracy of the parameter estimates. This leads to the main motivation for the current study and that is, to evaluate the performance of ML approach via the EM algorithm (in terms of model convergence and estimation of parameters) presented in du Toit and Cudeck (2009) and Harring (2012) for estimating NRMM with a two-phase piecewise function under several manipulated conditions described in the later section. But before we describe the design of the simulation study, we first want to illustrate the importance of this kind of model in the substantive areas of psychology and education. The analysis from the real data example will be the basis for the population values used to generate the data sets for the Monte Carlo simulation study. Thus, our goal is to show the efficacy of this model in real data analytic situation.

6. Performance on a Procedural Task: Real Data Example

The methodological work in this study was motivated by data obtained from a learning study[1] in which verbal skill acquisition was assessed from the performance on a procedural task. For each task, study participants were required to learn a set of declarative rules for assessing characteristics of visual stimuli presented in series. Tasks were organized and given together in blocks, with order of administration varied within blocks to lessen ordering effects. There were 24 trial blocks for each task, and each trial block is the mean of 16 individual trials. Thus, each task actually consisted of $24 \times 16 = 384$ basic trials. Score accuracy in addition to response times was recorded. Data for a restricted sample of 214 individuals whose average accuracy score across trial blocks was 85% or better on the task are considered here. The reasoning behind restricting the sample to only those individuals with higher accuracy was to reduce any effect of an accuracy/speed trade-off in response time (Blozis, 2004). To reduce the effects of extreme scores, the median time to respond for the procedural task was computed as an aggregate for each set of 32 trials in 12 separate blocks. As participants learned this task as is shown in the spaghetti plot in Figure 2, there appears to be a rapid decrease in reaction time followed by a period of more gradual and nearly linear improvement. Consequently, the model in Equation 11, a two-phase linear–linear structure with first-order continuity, was deemed appropriate.

[1] The data were obtained from Robert Cudeck who acquired them from Scott Chaiken of the Armstrong Laboratory, Brooks Air Force Base. A more detailed explanation of the origins of the data can be found in Cudeck (1996).

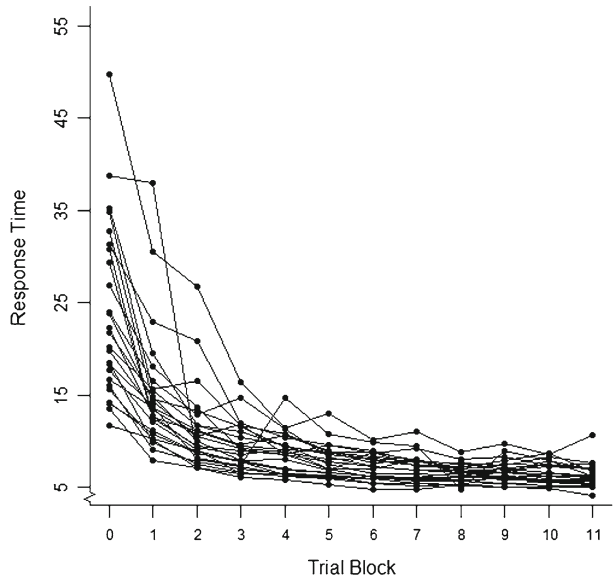

FIGURE 2.
Random 10 % subsample of repeated response time scores on the verbal procedural learning task.

The R routine fitPMM.R (Cudeck & Harring, 2014) was used to fit a series of linear–linear piecewise models with increasingly greater number of latent classes. Prior to fitting the piecewise mixture models, a single-class analysis was executed to inform starting values of parameters, an appropriate residual covariance structure as well as to ascertain whether or not all regression parameters of the model should be stochastic.

7. One-Class Analysis

In developing and eventually deciding on a model, the objective is to find a version that performs well in terms of data-model fit and that produces information about the process being studied that is different from the measurements actually obtained (Cudeck & Harring, Cudeck and Harring 2007). It is the model parameters that give new information. As was pointed out by Cudeck and Harring (2010), the model is formulated so that parameters embody features of individual change that can be utilized to address questions that are difficult to answer without the model. The data displayed in Figure 2 suggest a rapid decline in response from a high starting value down to an individual plateau and is not expected to rebound. While a number of mathematical functions (e.g., exponential, logistic) have been suggested to summarize data exhibiting these change pattern characteristics (see, e.g., Grimm & Ram, 2009; Harring et al., 2006), for illustrative purposes, the linear–linear two-phase model described in Equation 11 will be used and is reproduced here,

$$f(t_{ij},\ \mathbf{z}_i, \beta_{1i}, \ldots, \beta_{4i}, \gamma_i) = \mathbf{X}_i(\gamma_i)\boldsymbol{\beta}_i = \beta_{1i} + \beta_{2i}t_{ij} + (\beta_{2i} - \beta_{4i})(\gamma_i - t_{ij})_+,$$

with the twelve trials numbered $\mathbf{t} = (0,\ 1,\ \ldots,\ 11)'$ for all i, where $j = 1, \ldots, 12$. In this particular data set, there were no missing data. In this form, at $t_j = 0$, the function reduces to $f = \beta_{1i}$ and is considered the initial response time score at the beginning of the study; β_{2i} is the constant linear rate (slope) of the first phase. The coefficient, β_{4i} denotes the slope of the

second linear phase and γ_i represents the change point—the time when the outcome changes its functional form for individual i.

In this example, β_1, β_2, β_4, and γ are the fixed population parameters and b_{1i}, b_{2i}, b_{4i}, and g_i are the random effects unique to individual i. The random effects are assumed to be normally distributed with zero means and covariance matrix, $\mathbf{\Phi}$. Fitting the model in Equation 11 with an unstructured level-2 covariance matrix $\mathbf{\Phi}$ and independent, constant covariance structure for level 1 (i.e., $\mathbf{\Theta}_i(\boldsymbol{\xi}) = \sigma^2\mathbf{I}_{n_i}$) revealed that all parameters, regression coefficients, and level-1 and level-2 covariance parameters were statistically significant ($p < 0.05$). ML parameter estimates for the NRCM analysis are

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 19.90 \\ -4.42 \\ -0.25 \\ 2.65 \end{pmatrix} \qquad \hat{\sigma}_e^2 = 2.27 \qquad \hat{\mathbf{\Phi}} = \begin{pmatrix} 130.51 & & & \\ -38.86 & 16.08 & & \\ -1.58 & 0.23 & 0.05 & \\ 1.80 & 0.57 & -0.06 & 0.39 \end{pmatrix}.$$

Mean response time scores for the verbal task were a little less than 20, $\hat{\beta}_1 = 19.90$. Response time declined at a constant rate of approximately 4.5 seconds per trial, $\hat{\beta}_2 = -4.42$, and leveled off to a much slower linear rate $\hat{\beta}_4 = -0.25$ at approximately midway between trial 2 and trial 3, $\hat{\gamma} = 2.65$. Following the notation outlined in Seber and Wild (1989), the fitted function for the typical subject is given using the truncation operator (+) by

$$\hat{\mu}_j = 19.90 - 4.42t_j + (-0.25 + 4.42)(t_j - 2.65)_+.$$

Individual differences in the four aspects of response time relation for the verbal task can be investigated by examination of the variance components of $\mathbf{\Phi}$. Both their size and statistical significance suggest that individual differences in these change characteristics do indeed exist. The covariance between change characteristics can be assessed by the off-diagonal elements. However, to facilitate interpretation of these parameters, $\mathbf{\Phi}$ can be rescaled so that off-diagonal elements represent correlations:

$$\hat{\mathbf{\Phi}} = \hat{\mathbf{D}}^{1/2}\hat{\mathbf{P}}\hat{\mathbf{D}}^{1/2} = \begin{pmatrix} 11.42 & & & \\ & 4.01 & & \\ & & 0.22 & \\ & & & 0.62 \end{pmatrix} \begin{pmatrix} 1 & & & \\ -0.85 & 1 & & \\ -0.62 & 0.26 & 1 & \\ 0.25 & 0.23 & -0.43 & 1 \end{pmatrix} \begin{pmatrix} 11.42 & & & \\ & 4.01 & & \\ & & 0.22 & \\ & & & 0.62 \end{pmatrix},$$

where $\mathbf{D}$ is a diagonal matrix of variances and $\mathbf{P}$ is a symmetric matrix of correlations. It appears that initial response times and linear slope in the first phase are strongly negatively correlated, $\hat{\rho}_{21} = -.85$. This would indicate that those individuals who began the experiment with slower response times declined less quickly than those who began with response times that were faster. The change point was weakly positively related to both intercepts and slopes of the first phase but modestly negatively correlated with the slope of the second phase $\hat{\rho}_{43} = -.43$, the latter indicating that individuals whose change points occurred at later trials had second-phase slopes that were less steep.

TABLE 1.
BIC values for increasing number of components for the NLMM with a linear–linear piecewise function.

Number of components	Number of parameters	BIC
1	15	10892.8
2	19	10678.2
3	23	10698.6
4	27	N/A

Level-1 and Level-2 covariance structures were invariant across classes.
Note that the 4-class solution did not converge, and thus BIC was unable to be computed.

8. Class Enumeration

A generally accepted modeling framework is to perform a conventional NRCM analysis to get a sense of the extent to which individual regression coefficients vary as well as to assess a reasonable covariance structure for the repeated measures. The final parameter estimates for the covariance structures determined from the "one-class" analysis can be used as starting values for multiple class analyses. The parameters for the mean structure can be obtained by running a series of latent class growth models (Nagin, 2005), which suppresses the between-individual variability within a class, yet permits within-individual variation to be estimated.

Fit of models with distinct number of classes were evaluated using an information criterion as an index of model fit such as BIC. The Bayesian Information Criteria (BIC, Schwarz, 1978), a preferred information criteria for mixture models (Nylund, Asparouhov, & Muthén, 2007), penalizes over-parameterized models and takes into account sample size, and is given as

$$\text{BIC} = -2\ln L + p\ln(N),$$

where p is the number of free parameters and N is the sample size. A table of these values for increasing numbers of components can be found in Table 1, with the model corresponding to the lowest BIC value preferred. The one-class BIC value is for the standard NRCM model. Clearly, BIC values improved by more than one class and two classes seem optimal in the mixture modeling context. Not surprisingly because of the relatively modest sample size, the four-class solution did not converge. In addition, a preliminary investigation of whether the random effects covariance matrix as well as the residual variance should be held constant across classes revealed a model supporting the invariance of both structures (i.e., $\mathbf{\Phi}_1 = \cdots = \mathbf{\Phi}_k = \mathbf{\Phi}$; $\sigma^2_{jk} = \sigma^2\ \forall\ j$ and k).

The parameter estimates, standard errors, and class proportions for the two-class solution are summarized in Tables 2 and 3.

The solid lines shown in Figure 3 are the curves corresponding to the fitted functions of the typical values for an individual in each class superimposed on a random sample of subjects who were classified into that class.

The two-component mixture model clearly subdivided the participants into two groups. The first class is characterized by faster initial task response times whose slope in the first phase is steeper leading to a change point transitioning into the second phase at an earlier trial ($\hat{\gamma} = 2.4$). Study participants whose response times were initially slower declined at a less steep rate in phase 1 and consequently transitioned into the second phase at a later trial ($\hat{\gamma} = 4.3$).

A measure of misclassification, entropy, of the 2-component mixture was computed as 0.89, and the average posterior probabilities from the two-class model were

TABLE 2.
Maximum likelihood estimates of the regression coefficients for the two-class linear–linear Piecewise Model with class-invariant Level-1 and Level-2 covariance structures ($N = 214$).

Parameter	Class 1	Class 2
	Estimate (SE)	Estimate (SE)
π	0.19	0.81
β_1	17.9 (1.03)	25.7 (2.10)
β_2	−4.1 (0.45)	−3.9 (0.56)
β_4	−0.18 (0.06)	−0.25 (0.08)
γ	2.4 (0.06)	4.4 (0.10)

Gaussian–Hermite quadrature was used with $Q = 15$ quadrature points. Standard errors are in parentheses.

TABLE 3.
Maximum likelihood estimates of the elements of the class-invariant Level-1 and Level-2 covariance structures for the two-class linear–linear Piecewise Model ($N = 214$).

Parameter	Estimate (SE)
$\hat{\varphi}_{11}$	55.86 (7.47)
$\hat{\varphi}_{21}$	−18.63 (2.30)
$\hat{\varphi}_{22}$	13.75 (1.40)
$\hat{\varphi}_{31}$	−1.57 (0.40)
$\hat{\varphi}_{32}$	0.13 (0.09)
$\hat{\varphi}_{33}$	0.07 (0.02)
$\hat{\varphi}_{41}$	0.80 (0.29)
$\hat{\varphi}_{42}$	0.97 (0.14)
$\hat{\varphi}_{43}$	−0.08 (0.02)
$\hat{\varphi}_{44}$	0.29 (0.02)
$\hat{\sigma}_e^2$	1.76 (0.28)

Gaussian–Hermite quadrature was used with $Q = 15$ quadrature points. Standard errors are in parentheses.

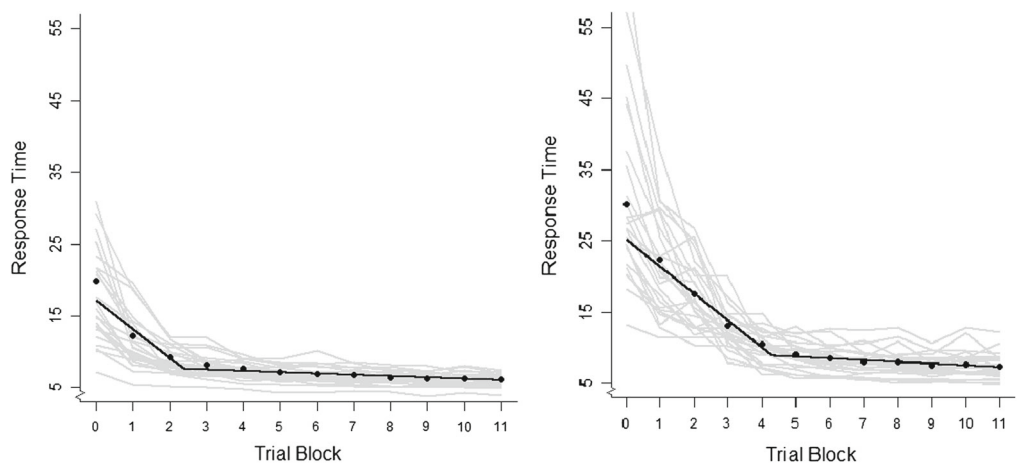

FIGURE 3.
Fitted linear–linear piecewise models for a typical individual for 2-component mixture superimposed on a random sample of 25 subjects whose class membership was determined by their highest posterior probability. The *left panel* is class 1 whose class proportion is $\hat{\pi}_1 = 0.19$. The *right panel* is class 2 whose class proportion is $\hat{\pi}_2 = 0.81$.

	Class 1 ($\hat{\pi}_1 = 0.19$)	Class 2 ($\hat{\pi}_2 = 0.81$)
Class 1	0.974	0.026
Class 2	0.054	0.946

providing additional evidence that the two-class mixture has classified subjects well.

With the findings from the real data analysis in mind, we designed our simulation study to demonstrate the efficacy of the NRMM with a two-phase piecewise function.

9. Design of Simulation Study

A NRMM with a two-phase piecewise function as shown in Equation 10 was used to generate repeated measures data for this study. The R routine fitPMM.R (Zopluoglu et al., 2014) was used to generate the complete and balanced data sets given the known model parameters under various manipulated conditions of sample size, number of time points, class mixing proportion, and class separation (i.e., the distance between two latent classes measured by Mahalanobis distance (MD^2). The detailed rationale for choosing the conditions is described later in this section. In general, the conditions that we did not expect to impact the model convergence and estimation of model parameters, specifically the estimation of the location and variance around the knot, were fixed across replications, and the conditions that we assumed may have an effect on the estimation of model parameters were manipulated. Additionally, the population values for the parameters used to generate the data sets were the same as the parameter estimates obtained from the real data analysis described above (see Tables 2 and 3).

The conditions that were fixed for the data generation routine were number of latent classes ($K = 2$), the mean parameters (regression coefficients) for class 1, the variance–covariance matrix of the random effects, and the residual variance (see Table 4 for fixed conditions and the parameter values for the fixed conditions). The choice of two latent classes was influenced by the real data example. As described in the example above, the two-class model was the best fitting model for the learning task data. Thus, we decided to fit the 2-class NRMM because our intent was to mimic the real data setup. The decision to keep the variance–covariance matrix among the random effects equal across both the classes for simulation purposes was based on the suggestion made by Muthén (2001) that mixture models with large differences in the factor variances and covariances between classes are particularly sensitive to local maxima.

10. Manipulated Conditions

Several conditions including total sample size, number of time points, class mixing proportion, and class separation were manipulated for the data generation routine (see Table 4 for manipulated parameter values). The rationale for choosing these conditions is described in the following section.

The two levels of the sample size condition were small-sized sample ($N = 200$) and medium-sized sample ($N = 500$). We chose these two levels because we expect the estimation of model parameters for small sample size to be much harder with respect to model convergence and accuracy of the model parameters than for the medium sample size. The two levels of class mixing

[2] $\text{MD} = \Delta = \sqrt{(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)' \boldsymbol{\Phi}^{-1} (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)}$, where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ denote the mean vector of fixed / population effects or class 1 and class 2, respectively, and $\boldsymbol{\Phi}$ denotes the variance—covariance matrix of random effects.

TABLE 4.
Fixed and manipulated parameter values.

	Class 1	Class 2
Fixed conditions		
Parameters		
Mean intercept (β_{11})	17.9	
Mean slope 1 (β_{21})	−4.1	
Mean slope 2 (β_{41})	−0.18	
Knot location (γ)	2.4	
Variance–covariance matrix (same for class 1 and 2, respectively		
$\begin{bmatrix} \text{var}(\beta_{11}) = \varphi_{11} = 55.86 & & & \\ -18.63 & \text{var}(\beta_{22}) = \varphi_{22} = 13.75 & & \\ -1.57 & 0.13 & \text{var}(\beta_{44}) = \varphi_{33} = 0.07 & \\ 0.80 & 0.97 & -0.08 & \text{var}(\gamma) = \varphi_{44} = 0.29 \end{bmatrix}$		
Residual variance (σ^2)	1.76	1.76
Manipulated conditions		
1. Sample size		
$N = 200$		
$N = 500$		
2. Number of time points		
t = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)		
t = (0, 1, 2, 3, 4, 5, 6, 7)		
3. Class mixing proportion		
$\pi = (0.50, 0.50)$		
$\pi = (0.20, 0.80)$		
4. Class separation (measured by MD)		
MD = 3		
Mean intercept (β_{11})	21.000	
Mean slope 1 (β_{21})	−3.200	
Mean slope 2 (β_{41})	−0.225	
Knot location (γ)	3.500	
MD = 6		
Mean intercept (β_{11})	25.780	
Mean slope 1 (β_{21})	−3.800	
Mean slope 2 (β_{41})	−0.235	
Knot location (γ)	4.400	

proportion, π_k, chosen, expressed as percentages, were (50/50 and 20/80) based on the simulation study of Nylund et al. (2007). Based on their findings, we expect poor model convergence and estimation of parameters when the class mixing proportion is 20/80, especially when the sample size is $N = 200$ (i.e., $n_1 = 40, n_2 = 160$), as compared to mixing proportion of 50/50 (i.e., $n_1 = 100, n_2 = 100$). The condition of the number of time points had two levels: 8 and 12 equally spaced time points. We chose these two levels because in piecewise growth models it matters whether there are enough time points (i.e., enough information) before and after the knot occurrence as it has an impact on the estimation of model parameters. To elaborate, when there are too few time points before the location of knot, there is not enough information to estimate the model parameters, specifically the slope parameter of the first phase. In contrast, when the knot occurs in the middle, we expect the model parameter estimates to be relatively more precise as compared to the estimates obtained when the knot occurs early.

Lastly, the class separation condition had two levels: MD = 3 and 6. We chose these two levels because we expect that the greater the latent distance between the two classes, the more precise the estimated model parameters would be, and that there will be less convergence issues. In order to create the levels of the condition of MD, we held the mean vector of the fixed regression parameters in Class 1 as $[\beta_{11}, \beta_{21}, \beta_{41}, \gamma_1] = [17.9, -4.1, -0.18, 2.4]$. Note that the population values of the fixed regression parameters for class 1 are same as the estimated parameter values for class 1 in Table 2 from real data analysis. We then manipulated the mean vector of the fixed regression parameters for Class 2 ($\beta_{11}, \beta_{21}, \beta_{41}, \gamma_1\gamma_2$). That is, the mean vector of the fixed regression parameters in Class 2 is $[\beta_{12}, \beta_{22}, \beta_{42}, \gamma_2] = [21, -3.2, -0.225, 3.5]$ to obtain MD = 3, and $[\beta_{12}, \beta_{22}, \beta_{42}, \gamma_2] = [25.78, -3.8, -0.24, 4.4]$ to obtain MD = 6. Note that the population values of the fixed regression parameters for class 2 when MD = 6 are same as the estimated parameter values for class 2 in Table 2 from real data analysis.[3]

The combination of manipulated conditions ($2 \times 2 \times 2 \times 2$) resulted in a Monte Carlo simulation with 16 cells or conditions. For each condition, 100 replications were generated. The simulation study was carried out using a computer with an Intel Xeon Quad-Core E3-1270V2 processor (3.5 GHz) that has a capacity of 32 GB RAM. The average time to run one replication was about 85 minutes for the conditions with the sample size of 200, and about 289 min for the conditions with the sample size of 500.

11. Model Fitting

The parameters of the two-class NRMM with a two-phase piecewise function were estimated using an R routine (Zopluoglu et al., 2014), built upon the Equations 10 to 17 described above in the model estimation section. The variance–covariance matrix of random effects and error variance across classes were fixed in the model fitting, so the fitted model was the same as the model used to generate data. There were 19 parameters in total for the 2-class model fitted to each replication.

In order to avoid local convergence and increase in the likelihood of reaching the global maxima, a number of start value sets were utilized. The starting values were generated by random perturbations of population parameters for the 19 parameters in the model[4]. For every parameter, a lower and upper boundary was created by subtracting and adding 5% of the population value. Then, a random number from a uniform distribution with these boundaries was drawn as a start value for the parameter. This process was repeated for every parameter to generate a vector of start values for 19 parameters. The number of starting value sets for the initial stage iterations was 75. The initial stage iterations were stopped when either the number of iterations exceeded 15 or the difference in the loglikelihood values between two successive iterations was smaller than 0.01. For the final stage iterations, the best starting value set obtained from the initial stage iterations was re-estimated until the convergence criterion was met. That is, the final stage iterations were stopped when either the number of iterations exceeded 1000 or the difference in the loglikelihood values between two successive iterations was smaller than 0.0001. Furthermore, for both initial and final stage iterations, 20 quadrature points were used to approximate the integral described in Equation 13 when maximizing the conditional expected loglikelihood function. This number was chosen as a compromise between accuracy of the integration and the time need to run a single replication within the simulation design.

[3] Results from a preliminary simulation study in which MD values were much small (i.e., less distributional separation) can be found at http://www.cehd.umn.edu/edpsych/people/Faculty/Kohli.html. These results showed poor convergence to a proper solution and large bias of model parameters—particularly the variance–covariance components of the linear–linear NRMM.

[4] We ran a separate simulation study where we empirically generated the starting values. The results from that study can be found on the following webpage: http://www.cehd.umn.edu/edpsych/people/Faculty/Kohli.html

One issue that often occurs when executing simulation studies of mixture models is the problem of label switching. A post-hoc class assignment-based algorithm proposed by Tueller, Drotar, and Lubke (2011) was implemented in R to address this issue in the current study. That is, after the final stage estimation was completed and the model parameters were estimated, the *column maxima switched label detection algorithm* described in Tueller et al. (2011) was used to check whether or not the labels were switched. When label switching was detected, final parameter estimates were repositioned accordingly before computing the outcome measures.

12. Outcome Measures

Two kinds of outcome measures commonly used in simulation studies to evaluate the recovery of the model parameters, bias and root mean squared error (RMSE), were used in the current study. The bias is equal to the average difference between the estimated parameter and the corresponding true value. It was computed using the formula

$$\text{Bias} = \frac{\sum_{r=1}^{R} (\hat{\theta}_{qr} - \theta_{\text{True}})}{R},$$

where $\hat{\theta}_{qr}$ is the estimated value for the qth parameter in the rth replication, θ_{TRUE} is the corresponding population value, and R is the number of converged replications. Note that bias was computed for only successfully converged replications in each cell.

The RMSE is equal to the square root of the averaged squared difference between the estimated parameter and the corresponding true value, and is used as an indicator of estimation error. It was computed using the formula

$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^{R} (\hat{\theta}_{qr} - \theta_{\text{True}})^2}{R}}.$$

Previous literature on simulation studies (see, e.g., Harwell, Stone, Hsu, & Kirisci, 1996; Harwell,1997) recommended deriving inferential analysis from the results of the simulation studies. Thus, to evaluate the impact of manipulated factors on convergence rate, bias, and RMSE, a four-way analysis of variance (ANOVA) with [2 (sample size) × 2 (class mixing proportion × 2 (number of time points) × 2 (class separation) was performed. The main effects of the manipulated factors and the two-way interaction terms were reported and interpreted only when *both* statistical significance and practical significance, as measured by eta squared ($\eta^2 \geq 0.10$), were observed.

13. Results from Simulation Study

13.1. Model Convergence Rate

The model convergence rate across all the 16 conditions is reported in Table 5.

The rate of converged replications for the NRMM with a two-phase piecewise function was in the range of 58% to 98%. From a cursory review, the lower bound of this range seems small; however, upon closer inspection of the conditions that reported small convergence rates, one can

TABLE 5.
Model convergence rate across all conditions.

Sample size	Number of time points	Mixing proportion	MD	Convergence rate
200	8	(.2, .8)	3	0.63
200	8	(.2, .8)	6	0.58
200	8	(.5, .5)	3	0.60
200	8	(.5, .5)	6	0.63
200	12	(.2, .8)	3	0.79
200	12	(.2, .8)	6	0.85
200	12	(.5, .5)	3	0.84
200	12	(.5, .5)	6	0.88
500	8	(.2, .8)	3	0.70
500	8	(.2, .8)	6	0.72
500	8	(.5, .5)	3	0.60
500	8	(.5, .5)	6	0.81
500	12	(.2, .8)	3	0.88
500	12	(.2, .8)	6	0.91
500	12	(.5, .5)	3	0.95
500	12	(.5, .5)	6	0.98

comprehend the reasons behind it. That is, low convergence rates were observed for conditions that had a combination of small sample size ($N = 200$) and the number of time points, $\mathbf{t}_j = 8$. The finding makes sense as we expected model convergence rates to be unfavorably affected by the levels of the stated conditions, respectively.

Additionally, a four-way ANOVA was conducted on the percentage of properly converged replications to investigate if model convergence appeared to be a function of the manipulated conditions. Note that the dependent variable (model convergence rate) was transformed using an arcsine transformation (units are expressed in radians) (Sokal & Rohlf, 1995) before running the four-way ANOVA. This was done so that the distribution of percent or proportions data is closer to normal distribution. The eta squares from the F-test of between-subject effects on model convergence rate are reported in Table 6. The only main effects that satisfied both the statistical and the practical significance criteria were sample size and number of time points. None of the interaction terms met the statistical and the practical significance criteria. This finding is consistent with the discussion in the above paragraph, that is, small sample size ($N = 200$) coupled with a modest number of time points, $\mathbf{t}_j = 8$ attributed to poor convergence rate.

14. Parameter Bias

The average bias across all the 16 conditions is reported in Table 7. The size of the absolute average bias appears to be within the range of 0.001 to 1.385. Overall, the average bias across all the conditions is small for all the model parameters with respect to the corresponding true values for the parameters (see Table 8). For instance, the average bias for the variance of the intercept, φ_{11}, is -1.385 and the population value of $\varphi_{11} = 55.86$, which is about 2.5% of the population value. The variance of the intercept is generally hard to estimate precisely, so the recovery of this parameter appears decent. Likewise, the recovery of other model parameters was good.

A four-way ANOVA was performed on the average parameter bias to see if there was any systematic relation between average parameter bias and the manipulated conditions. The eta

TABLE 6.
Eta squares from between-subjects ANOVA on model convergence rates.

	$\eta^2(SS_{Effect}/SS_{Correctedtotal})$
N	0.146**
NT	0.710***
MP	0.026
MD	0.035
N:NT	0.007
N:MP	0.004
N:MD	0.010
NT:MP	0.020
NT:MD	0.000
MP:MD	0.012

*$p < .05$, ** $p < .01$, *** $p < .001$.
N Sample Size, *NT* Number of time points, *MP* mixing proportion, *MD* mahalanobis distance.

squares from the F-test of between-subject effects on average parameter bias are reported in Table 9. The only factors that stood out from the rest were the manipulated conditions of class separation and class mixing proportion, and the interaction term between these two conditions. That is, when the latent distance between the two classes was small (i.e., MD = 3), and the class mixing proportion was 20/80, the average parameter bias increased as compared to when the latent distance between the two classes was large (i.e., MD = 6), and the class mixing proportion was 50/50. This finding is consistent with what we expected and what other simulation studies examining linear latent growth mixture models have found (Liu & Hancock, 2014).

15. RMSE

RMSE is reported in Table 10 across all the 16 conditions. The size of the average RMSE is within the range of 0.04 to 6.45. The average RMSE across all the conditions in general is small for all the model parameters when compared to the corresponding true values for the parameters (see Table 11). A four-way ANOVA was performed to quantify average RMSE as a function of the manipulated conditions. The eta squares from the F-test of between-subject effects on average RMSE are reported in Table 12. As seen in Table 12, most of the two-way interaction terms were not significant, neither statistically nor practically. The factors that had the most effect on average RMSE were sample size, class mixing proportion, and class separation. As expected, small sample size ($N = 200$), class mixing proportion of 20/80, and small class separation (i.e., MD = 3) produced larger RMSE for most parameter estimates.

Overall, we conclude based on the findings that ML via EM estimation of NRMM with two-phase piecewise function recovers the true parameters well among converged solutions.

16. Discussion

Nonlinear random coefficient models have, for years, been implemented to summarize non-constant individual change. Like their linear counterparts, the efficacy of the NRCM model lies in the model's flexibility to permit individual-specific functions to differ from the mean function over the population of subjects, yet characterizes both the typical individual and individual patterns

TABLE 7.
Average parameter bias across all converged replications within each condition.

N	NT	MP	MD	Class 1				Class 2				Common variance–covariance matrix									
				β_1	β_2	β_4	γ	β_1	β_2	β_4	γ	φ_{11}	φ_{22}	φ_{33}	φ_{44}	φ_{12}	φ_{13}	φ_{14}	φ_{23}	φ_{24}	φ_{34}	σ_ε^2
200	8	(.2,.8)	3	−0.13	0.26	0.02	0.12	0.57	−0.16	−0.01	0.05	−3.36	−0.48	0.00	0.01	0.88	0.14	−0.17	0.02	0.04	0.02	0.00
200	8	(.2,.8)	6	0.49	−0.07	−0.03	0.00	0.11	−0.05	0.00	−0.01	0.97	−0.03	0.07	0.01	−0.04	−0.03	0.11	0.04	−0.01	0.02	0.01
200	8	(.5,.5)	3	−0.02	−0.10	0.04	0.03	0.02	0.00	−0.01	−0.02	−3.62	−0.62	0.00	0.04	0.96	0.20	0.07	−0.04	−0.03	0.01	−0.03
200	8	(.5,.5)	6	−0.10	−0.08	0.03	−0.03	0.02	−0.07	0.01	−0.02	−1.07	0.06	0.04	0.04	0.58	0.01	0.14	0.05	0.05	0.02	−0.02
200	12	(.2,.8)	3	0.05	0.31	0.00	0.13	0.37	−0.19	0.00	0.02	−2.61	−0.22	−0.01	0.03	0.64	0.11	−0.01	−0.01	0.07	0.00	0.00
200	12	(.2,.8)	6	0.08	−0.03	0.00	−0.01	0.00	0.00	0.00	0.00	−0.77	−0.56	0.00	0.01	0.97	−0.02	0.15	−0.01	−0.03	0.00	0.00
200	12	(.5,.5)	3	0.24	−0.18	0.01	0.02	−0.15	0.06	−0.01	−0.05	−4.00	−0.81	0.00	0.05	1.69	0.13	0.14	−0.03	0.01	0.00	0.00
200	12	(.5,.5)	6	0.06	0.01	−0.01	0.01	0.03	−0.05	0.00	0.00	0.34	−0.09	0.00	0.00	0.02	−0.02	0.02	0.01	−0.01	0.00	0.00
500	8	(.2,.8)	3	−0.07	0.32	0.01	0.09	0.36	−0.14	−0.01	0.01	−2.05	0.06	0.00	0.01	0.28	0.05	−0.14	−0.01	0.07	0.00	−0.01
500	8	(.2,.8)	6	−0.06	−0.05	0.01	−0.02	0.01	−0.02	0.01	0.00	−0.41	−0.02	0.01	0.00	0.10	0.03	0.01	0.00	−0.01	0.01	−0.02
500	8	(.5,.5)	3	0.14	0.00	0.00	0.04	0.12	−0.07	0.00	−0.01	−1.52	−0.32	0.00	0.03	0.68	0.06	0.04	−0.02	0.03	0.00	−0.01
500	8	(.5,.5)	6	−0.04	−0.03	0.00	−0.01	0.04	−0.05	0.01	−0.01	−0.23	−0.09	0.00	−0.01	0.10	0.01	−0.01	−0.01	−0.01	0.00	0.00
500	12	(.2,.8)	3	−0.28	0.18	0.01	0.06	0.26	−0.11	0.00	0.01	−1.83	−0.29	0.00	0.00	0.52	0.03	−0.14	0.00	0.03	0.01	0.00
500	12	(.2,.8)	6	0.02	−0.01	0.00	0.01	0.06	0.01	0.00	0.00	0.11	−0.04	0.00	0.00	−0.06	−0.01	−0.01	0.01	−0.01	0.00	0.00
500	12	(.5,.5)	3	−0.22	0.04	0.01	0.00	0.29	−0.10	−0.01	−0.01	−1.80	−0.10	0.00	0.02	0.37	0.06	−0.08	0.00	0.05	0.00	0.00
500	12	(.5,.5)	6	0.03	−0.01	0.00	0.00	0.00	0.03	0.00	0.01	−0.28	−0.04	0.00	−0.01	−0.08	0.00	−0.03	0.01	−0.02	0.00	0.00

N Sample size, *NT* number of time points, *MP* mixing proportion, *MD* Mahalanobis distance.

TABLE 8.
Average parameter bias across all conditions.

Parameter	Class 1	Class 2
β_1	0.011	0.132
β_2	0.036	−0.056
β_4	0.006	−0.002
γ	0.027	−0.002
φ_{11}	−1.385	
φ_{22}	−0.225	
φ_{33}	0.007	
φ_{44}	0.014	
φ_{12}	0.475	
φ_{13}	0.047	
φ_{14}	0.006	
φ_{23}	0.001	
φ_{24}	0.014	
φ_{34}	0.006	
σ^2_ε	−0.005	

TABLE 9.
Eta squares from between-subjects ANOVA for parameter bias.

		N	NT	MP	MD	N*NT	N*MP	N*MD	NT*MP	NT*MD	MP*MD
Class 1	β_1	0.164	0.006	0.000	0.076	0.046	0.048	0.000	0.030	0.000	0.167
	β_2	0.020	0.000	0.295**	0.218**	0.003	0.026	0.006	0.001	0.019	0.346**
	β_4	0.000	0.070	0.061	0.167	0.049	0.205	0.027	0.049	0.005	0.021
	γ	0.018	0.001	0.181**	0.518***	0.005	0.019	0.029	0.000	0.030	0.154**
Class 2	β_1	0.003	0.017	0.228*	0.293*	0.034	0.100	0.009	0.017	0.005	0.171*
	β_2	0.000	0.029	0.151	0.202	0.000	0.057	0.068	0.004	0.020	0.278*
	β_4	0.006	0.042	0.008	0.522**	0.004	0.016	0.040	0.077	0.108	0.027
	γ	0.024	0.007	0.363**	0.018	0.029	0.101	0.003	0.006	0.102	0.241*
	φ_{11}	0.071	0.000	0.009	0.720**	0.000	0.017	0.085	0.002	0.000	0.000
	φ_{22}	0.214	0.030	0.011	0.233	0.017	0.000	0.064	0.019	0.022	0.117
	φ_{33}	0.084	0.219*	0.013	0.184*	0.098	0.004	0.133	0.007	0.102	0.023
	φ_{44}	0.309*	0.026	0.132	0.189	0.001	0.025	0.006	0.015	0.047	0.083
	σ^2_ε	0.003	0.239	0.054	0.080	0.001	0.260	0.010	0.029	0.012	0.017

* $p < .05$, ** $p < .01$, *** $p < .001$.
N Sample size, *NT* number of time points, *MP* mixing proportion, *MD* Mahalanobis distance.

with a response function which has a common nonlinear form. Furthermore, individual subjects need not be measured at the same measurement occasions nor do subjects need to be measured the same number of occasions. When the missing at random assumption is tenable, plausible missing values of the covariates can be accommodated. Lastly, more realistic patterns for the residual structure can be specified when there is empirical evidence to support a more complex structure (Blozis & Cudeck, 1999).

Data from a learning experiment aimed at investigating skill acquisition were used as an example to illustrate the significance of the utility of the linear–linear piecewise growth mixture model in the field of psychology and education. For this sample, a two-component mixture was retained where the first latent class ($\hat{\pi} = 0.19$) was characterized by higher initial task response

TABLE 10.
Average RMSE across all converged replications within each condition.

N	NT	MP	MD	Class 1				Class 2				Common variance–covariance matrix									
				β_1	β_2	β_4	γ	β_1	β_2	β_4	γ	φ_{11}	φ_{22}	φ_{33}	φ_{44}	φ_{12}	φ_{13}	φ_{14}	φ_{23}	φ_{24}	φ_{34}	σ_ε^2
200	8	(.2,.8)	3	2.28	1.47	0.12	0.29	1.37	0.61	0.07	0.11	8.40	1.53	0.03	0.07	2.89	0.38	0.74	0.12	0.31	0.05	0.09
200	8	(.2,.8)	6	2.03	1.29	0.30	0.33	0.86	0.43	0.07	0.10	8.60	1.66	0.31	0.08	3.35	0.81	0.71	0.24	0.25	0.07	0.12
200	8	(.5,.5)	3	2.20	0.96	0.11	0.17	1.87	0.92	0.10	0.16	8.48	2.08	0.02	0.10	2.57	0.45	0.88	0.20	0.48	0.05	0.09
200	8	(.5,.5)	6	1.20	0.71	0.14	0.18	1.30	0.57	0.11	0.13	7.15	2.42	0.15	0.20	2.62	0.52	0.98	0.53	0.46	0.13	0.09
200	12	(.2,.8)	3	2.73	1.35	0.13	0.32	1.47	0.68	0.05	0.10	7.46	1.71	0.01	0.08	2.88	0.26	0.69	0.09	0.37	0.03	0.06
200	12	(.2,.8)	6	1.20	0.69	0.06	0.15	0.70	0.32	0.03	0.08	5.79	1.51	0.01	0.06	2.64	0.18	0.52	0.09	0.20	0.02	0.07
200	12	(.5,.5)	3	2.22	1.15	0.09	0.18	1.92	0.87	0.09	0.15	7.70	2.21	0.02	0.12	3.72	0.31	0.89	0.14	0.42	0.04	0.06
200	12	(.5,.5)	6	0.79	0.40	0.03	0.11	0.77	0.37	0.04	0.10	7.27	1.59	0.01	0.06	2.81	0.23	0.48	0.09	0.26	0.02	0.06
500	8	(.2,.8)	3	1.76	0.92	0.09	0.20	0.97	0.40	0.05	0.09	6.10	1.05	0.01	0.06	1.90	0.24	0.43	0.10	0.22	0.03	0.05
500	8	(.2,.8)	6	0.84	0.48	0.06	0.12	0.46	0.20	0.04	0.05	3.90	0.89	0.02	0.03	1.56	0.22	0.25	0.12	0.13	0.02	0.06
500	8	(.5,.5)	3	1.26	0.68	0.06	0.12	1.64	0.72	0.08	0.11	6.15	1.39	0.02	0.08	2.21	0.25	0.73	0.13	0.33	0.04	0.05
500	8	(.5,.5)	6	0.53	0.28	0.03	0.08	0.56	0.23	0.06	0.07	4.15	0.81	0.01	0.04	1.68	0.18	0.30	0.09	0.17	0.02	0.05
500	12	(.2,.8)	3	1.52	1.00	0.07	0.24	0.86	0.42	0.03	0.07	5.70	1.12	0.01	0.06	2.13	0.17	0.44	0.06	0.22	0.02	0.04
500	12	(.2,.8)	6	0.88	0.44	0.04	0.10	0.39	0.21	0.02	0.05	3.74	0.94	0.01	0.04	1.56	0.14	0.26	0.06	0.13	0.01	0.04
500	12	(.5,.5)	3	1.32	0.63	0.05	0.13	1.20	0.52	0.05	0.12	5.47	1.09	0.01	0.08	1.96	0.18	0.59	0.06	0.28	0.02	0.04
500	12	(.5,.5)	6	0.52	0.31	0.02	0.07	0.51	0.26	0.02	0.06	3.90	0.90	0.01	0.04	1.53	0.14	0.31	0.06	0.14	0.01	0.04

N Sample size, *NT* number of time points, *MP* mixing proportion, *MD* Mahalanobis distance.

TABLE 11.
Average RMSE across all conditions.

Parameter	Class 1	Class 2
β_1	1.46	1.05
β_2	0.80	0.48
β_4	0.09	0.06
γ	0.17	0.10
φ_{11}	6.25	
φ_{22}	1.43	
φ_{33}	0.04	
φ_{44}	0.07	
φ_{12}	2.38	
φ_{13}	0.29	
φ_{14}	0.57	
φ_{23}	0.14	
φ_{24}	0.27	
φ_{34}	0.04	
σ_ε^2	0.06	

TABLE 12.
Eta squares from between-subjects ANOVA for RMSE.

		N	NT	MP	MD	N*NT	N*MP	N*MD	NT*MP	NT*MD	MP*MD
Class 1	β_1	0.326**	0.008	0.092*	0.480**	0.004	0.002	0.011	0.000	0.020	0.004
	β_2	0.308**	0.019	0.184**	0.369**	0.022	0.012	0.000	0.009	0.031	0.000
	β_4	0.293*	0.173	0.091	0.002	0.083	0.014	0.034	0.012	0.110	0.016
	γ	0.286**	0.022	0.300**	0.162*	0.031	0.025	0.010	0.003	0.086*	0.017
Class 2	β_1	0.223**	0.024	0.121**	0.556***	0.000	0.001	0.001	0.009	0.003	0.025
	β_2	0.262**	0.016	0.114**	0.518***	0.002	0.003	0.004	0.012	0.001	0.032
	β_4	0.240**	0.388**	0.142**	0.092*	0.013	0.019	0.000	0.008	0.045	0.010
	γ	0.386***	0.032	0.230**	0.254**	0.009	0.011	0.006	0.000	0.004	0.029
	φ_{11}	0.680***	0.050	0.000	0.173**	0.012	0.000	0.029	0.011	0.000	0.000
	φ_{22}	0.710***	0.010	0.072	0.036	0.005	0.048	0.009	0.020	0.014	0.007
	φ_{33}	0.142	0.163	0.018	0.109	0.118	0.016	0.105	0.019	0.119	0.020
	φ_{44}	0.325*	0.034	0.150	0.034	0.036	0.052	0.056	0.025	0.070	0.000
	σ_ε^2	0.568***	0.280***	0.017	0.012	0.057**	0.005	0.012	0.001	0.022	0.015*

* $p < .05$, ** $p < .01$, *** $p < .001$.
N Sample size, *NT* number of time points, *MP* mixing proportion, *MD* Mahalanobis distance.

times and steeper decline in the first phase. Importantly, this latent group transitioned from the first to the second linear phase at a later trial than the second latent group ($\hat{\gamma}_1 = 4.3$ and 2.4) possibly indicating individual attributes which demonstrate an inability to learn the task as quickly as individuals from the second latent class. It is just this type of finding that makes mixture modeling worthwhile. This example did not include static covariates as part of the model, and thus further investigation to uncover individual determinants that differentiate both growth characteristics and change points would be warranted.

To show the effectiveness of NRCM model in real data analytic situation, we executed a Monte Carlo simulation study using a self-coded R routine. We reported the results of the study undertaken to gain insight into the feasibility of estimating a NLMM with a linear–linear piecewise model

under a non-adaptive Gaussian quadrature estimation algorithm. Although it is not appropriate to draw general conclusions based on a single simulation study, the results suggest some interesting features that may be worthy of further investigation. First, it appeared that the non-adaptive Gaussian quadrature method recovered regression parameters of the linear–linear model in terms of bias and precision in terms of RMSE, although the method did show some difficulty in achieving convergence only under more adverse levels of manipulated conditions. Of the conditions that were manipulated in the study, bias and RMSE were most negatively impacted by small sample size, small class separation, and class mixing proportion of 20/80. Convergence rates from other studies which used an approximation to the multidimensional integration when examining nonlinear mixed- effects models (Hartford & Davidian, 2000) were also reported to be modest. When incorporating another layer of complexity, such as a finite mixture model, into the system, it should come as no surprise that convergence rates would be adversely affected. In all cases, the starting values were generated by random perturbations of true parameter values. As is familiar to analysts using these methods, it is sometimes the case that convergence can never be achieved. This may be the result of poor initial values, practical lack of identifiability with the particular data available from a replication, or other unknown factors. To address the first issue, in practice, it is recommended to try several sets of starting values (Hipp & Bauer, 2006). This was taken into account in the design. The reported results on convergence may be pessimistic with respect to possible convergence issue in real applications and must be interpreted with this caution in mind. However, given the iterative nature of data analysis in practice—issues associated with non-convergence may be ameliorated by a principled data analytic plan that includes many exploratory analyses along the way (Harring, 2012).

Overall, longitudinal mixture models are useful and flexible models in areas of educational and developmental research. In education, for example, researchers and practitioners are often focused on changes in student academic progress, attitude, and motivation. They are interested in studying the effectiveness of a set of treatments or an intervention on students' academic progress where the population of students is composed of two or more latent groups (Kohli et al., 2013). The advantage of employing the NLMM with a piecewise function is that it permits the investigation of change patterns that exhibit distinct phases where researchers specify a functional form in each phase to adequately capture the overall change process within each latent class. While the current study limited the piecewise function to account for two linear regimes, many other variants on this theme are possible (Cudeck & Harring, 2010). For optimal flexibility, the change point, the time when a subject transitions from one phase to the next, was allowed to be individual-specific. This elaboration allows individuals to transition differentially, which in some contexts such as learning a task is a sensible modeling strategy.

References

Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society Series D (Methodological)*, *47*, 67–75.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean and covariance structure models. *Psychometrika*, *64*, 475–494.

Blozis, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures data. *Psychological Methods*, *9*, 334–353.

Blozis, S. A., & Cudeck, R. (1999). Linear mixed-effects models with latent variable covariates. *Journal of Educational and Behavioral Statistics*, *24*, 245–270.

Boos, D., & Brownie, C. (1991). Mixture models for continuous data in dose-response studies when some animals are unaffected by treatment. *Biometrics*, *47*, 1489–1504.

Codd, C. L., & Cudeck, R. (2014). Nonlinear random-effects mixture models for repeated measures. *Psychometrika*, *79*, 60–83.

Crowder, M., & Hand, D. (1990). *Analysis of repeated measures*. London: Chapman and Hall.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*, 371–403.

Cudeck, R., & du Toit, S. H. C. (2002). A nonlinear form of quadratic regression with interpretable parameters. *Multivariate Behavioral Research*, *37*, 501–519.

Cudeck, R., & Harring, J. R. (2007). The analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, *58*, 615–637.

Cudeck, R., & Harring, J. R. (2010). Developing a random coefficient model for nonlinear repeated measures data. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue*. New York: Routledge.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*, 41–63.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.

du Toit, S. H. C., & Cudeck, R. (2009). Estimation of the nonlinear random coefficient model when some random effects are separable. *Psychometrika*, *74*, 65–82.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.

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

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. New York: Springer.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in M plus and SAS. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 676–701.

Gruen, B., & Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis*, *51*, 5247–5252.

Hall, D. B., & Wang, L. (2005). Two-component mixtures of generalized linear mixed effects models for cluster correlated data. *Statistical Modelling*, *5*, 21–37.

Hancock, G. R., & Samuelsen, K. M. (Eds.). (2008). *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing Inc.

Harring, J. R. (2009). A nonlinear mixed effects model for latent variables. *Journal of Educational and Behavioral Statistics*, *34*, 293–318.

Harring, J. R. (2012). Finite mixtures of nonlinear mixed effects models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 159–192). Charlotte, NC: Information Age Publishing Inc.

Harring, J. R., Cudeck, R., & du Toit, S. H. C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research*, *41*, 579–596.

Hartford, A., & Davidian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, *34*, 139–164.

Harwell, M. R. (1997). Analyzing the results of Monte Carlo studies in item response theory. *Educational and Psychological Measurement*, *57*, 266–278.

Harwell, M. R., Stone, C. A., Hsu, T., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, *20*, 101–125.

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

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, *42*, 805–820.

Kohli, N., & Harring, J. R. (2013). Modeling growth in latent variables using a piecewise function. *Multivariate Behavioral Research*, *48*, 370–397.

Kohli, N., Harring, J. R., & Hancock, G. R. (2013). Piecewise linear–linear latent growth mixture models with unknown knots. *Educational and Psychological Measurement*, *73*, 935–955.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, *74*, 557–584.

Luo, X., Boos, D. D., & Tamura, R. N. (2004). Score tests for dose effect in the presence of non-responders. *Statistics in Medicine*, *23*, 3581–3591.

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

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81–117.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

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

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

Nylund, K. L., Asparouhov, T., & Muthén, B. (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*, 535–569.

Pavlic, M., Brand, R. J., & Cumming, S. R. (2001). Estimating probability of nonresponse to treatment using mixture distributions. *Statistics in Medicine*, *20*, 1739–1753.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks: Sage Publications.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Smith, J. L., & Klebe, K. J. (1997, June). *Design issues in learning experiments: A comparison between repeated measures ANOVA and a latent growth curve*. Paper presented at the Annual Meeting of the Psychometric Society, Gatlinburg, TN.
Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practice of statistics in biological research* (3rd ed.). New York: W.H. Freeman.
Titterington, D., Smith, A., & Makov, U. (1985). *Statistical analysis of finite mixture distributions*. Chichester: Wiley.
Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling*, *18*, 110–131.
Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association*, *91*, 217–221.
Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York, NY: Springer-Verlag.
Vonesh, E. F. (2012). *Generalized linear and nonlinear models for correlated data: Theory and applications using SAS*. Cary, NC: SAS Institute.
Vonesh, E. F., & Carter, R. L. (1992). Mixed effects nonlinear regression for unbalanced repeated measures. *Biometrics*, *48*, 1–17.
Wang, L., & McArdle, J. J. (2008). A Simulation study comparison of bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 52–74.
Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, *12*, 21–55.
Xu, W., & Hedeker, D. (2001). A random-effects mixture model for classifying treatment response in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, *11*, 253–273.
Zopluoglu, C., Harring, J. R., & Kohli, N. (2014). FitPMM: An R routine to fit finite mixtures of piecewise mixed-effects models with unknown random knots. *Applied Psychological Measurement*, *38*, 583–584.

Manuscript Received: 16 FEB 2014