ABSTRACT

Title of Document:	A SYSTEMATIC INVESTIGATION OF WITHIN-SUBJECT AND BETWEEN-SUBJECT COVARIANCE STRUCTURES IN GROWTH MIXTURE MODELS
	Junhui Liu, Doctor of Philosophy, 2012
Directed By:	Dr. Jeffrey R. Harring, Department of Human Development and Quantitative Methodology

The current study investigated how between-subject and within-subject variance-covariance structures affected the detection of a finite mixture of unobserved subpopulations and parameter recovery of growth mixture models in the context of linear mixed-effects models. A simulation study was conducted to evaluate the impact of variance-covariance structure difference, mean separation, mixture proportion and sample size on parameter estimates from growth mixture models. Data were generated based on 2-class growth mixture model framework and estimated by 1-, 2-, and 3-class growth mixture models using Mplus. Bias, precision and efficiency of parameter estimates were assessed as well as the model enumeration accuracy and classification quality.

Results suggested that sample size and data overlap were key factors influencing the convergence rates and possibilities of local maxima in the estimation of GMM models. BIC outperformed ABIC and LMR in identifying the correct

number of latent classes. Model enumeration using BIC could be improved by increasing sample size and/or decreasing overall data overlap, and the latter had more impact. Relative bias of parameters was smaller when subpopulation data were more separated. Both the magnitude of mean and variance-covariance separation and variance-covariance differences impacted parameter recovery. Across all conditions, parameter recovery was better for intercept and slope estimates than variance and covariances estimates. Entropy values were as high as the acceptable standards suggested by previous studies for any of the conditions even when data were very well-separated. Class membership assignment was more accurate when mean growth trajectories were more different among subpopulations and mixing proportions were more balanced.

A SYSTEMATIC INVESTIGATION OF WITHIN-SUBJECT AND BETWEEN-SUBJECT COVARIANCE STRUCTURES IN GROWTH MIXTURE MODELS

By

Junhui Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Jeffrey R. Harring, Chair
Professor Gregory R. Hancock
Professor Hong Jiao
Professor George B. Macready
Professor Wayne H. Slater

Dedication

To my dearest parents,

Yaoxin Liu and Shujuan Li,

whose love and support have made this all possible.

谨以此论文献给我的父亲刘耀新和母亲李淑娟。

Acknowledgements

Thank you, Dr. Harring, for being an outstanding academic and dissertation advisor. I am very fortunate to have had your guidance throughout the dissertation process. You were a great influence on the way I conducted and wrote about my research. Without your insights, instruction and encouragement I would not have been able to complete this program. You helped me to develop not only as a graduate student but also as a professional in this field.

I would also like to thank all my committee members, Dr. Hancock, Dr. Macready, Dr. Jiao, Dr. Gottfredson, and Dr. Slater for their valuable input on my dissertation. I am very grateful for Dr. Macready's support and encouragement throughout my study. I would also like to thank my fellow students who have aided in the timely completion of this research. In particular, I would like to thank Youngmi Cho, Dr. Min Liu, Dr. Ru Lu, Dr. Feifei Li, Xiaoshu Zhu, Huili Liu, Yong Luo and Weitian An for your friendship and encouragement throughout the years.

I would also like to thank Dr. Mislevy for providing the opportunity to work for the Cisco project which was an excellent learning experience. I really appreciated your instruction and support. I would like to extend my gratitude to Dr. Daisy Rutstein for the wonderful experience working with you.

I would like to thank Educational Testing Service (ETS) for the graduate fellow position which supported me through completion of the latter part of the dissertation. I am very grateful to my colleagues at ETS, especially Dr. Jinghua Liu and Dr. Hyeonjoo Oh, for their encouragement and support on my dissertation.

Most of all, I would like to thank my parents for their unconditional love and endless support. You are my rock!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

A primary goal of social and behavioral scientists interested in investigating how human behavior changes or develops is to make inferences on features underlying profiles of continuous repeated measures data for a targeted population (Cudeck, 1996). Of particular interest is to study how responses for individuals change over time and to investigate those attributes that may account for individual differences in change characteristics. A distinguishing feature of longitudinal data is that the repeated observations on the same individual are not independent (i.e., repeated measures within the same subject are correlated). Furthermore, the variance of the repeated measurements may not always be constant across multiple time points. Thus, statistical methods, like multiple regression using ordinary least squares estimation, which assumes independent observations and conditional homogeneity of variance, should not be used to estimate model parameters.

Historically, statistical methods such as repeated measures ANOVA (RMA), repeated measures MANOVA (RMM), auto-regressive and cross-lagged multiple regression as well as methods based on calculated quantities or derived values that summarize the repeated measures (e.g., area under the curve) have been the primary methods utilized for analyzing longitudinal data (see, e.g., Collins & Sayer, 2001; Gottman, 1995). Choosing an appropriate analytic method often depends on two primary considerations. First, the analytic method must provide direct evidence that tentatively supports or refutes the research hypotheses posited by the investigator. Hypotheses leading to the use of these more conventional analytic methods tend to focus on aggregate group differences failing to address questions regarding the nature

and determinants of change at the individual level. Secondly, characteristics of the longitudinal design, the data themselves, and the underlying assumptions often dictate which method can be applied in a given situation. Many of these analytic methods suffer from unrealistic assumptions that may limit their usefulness in real world situations. For example, technical assumptions such as sphericity underlying RMA are rarely met in practice in the social sciences (see, e.g., Howell, 2007). Other limitations of traditional methods for longitudinal analyses include their inability to handle missing data or unbalanced designs. As longitudinal data are often collected with long follow up periods, missing data are often inevitable. Sometimes the proportion of missing data can be substantial. Missingness in longitudinal data is usually a result of dropout, mortality, characteristics of the protocol and/or other subtle events that may occur across the study period. Unbalanced designs occur when not all participants are measured at the same time points. For example, it may be known beforehand that the participants will enter the study at different ages and the timing of the waves of measurement will depend on uncontrollable participant factors (e.g., vacation time, forgetfulness). In this scenario, the times that study participants are measured could be entirely unique. Traditional methods like RMA and RMM, which are often viewed as being less flexible in terms of design considerations, would drop cases with missing values (e.g., listwise deletion) at any time point and do not accommodate unbalanced designs.

In part, an increase in computing power brought by new technology in the 1980s made it possible to apply more sophisticated, modern methods to studying change or development. A myriad of statistical models and methods were proposed

and developed to investigate longitudinal change in a wide variety of behavior including human cognition development, crops growth, and so on. One such model, the linear mixed-effects (LME) model (Laird & Ware, 1982), is grounded in the philosophical and mechanistic underpinnings of regression. Unlike its more conventional counterparts, LME models are flexible to handle both data that are missing and observations that are gathered from an unbalanced design. Under the assumption that the mechanism underlying the missingness is missing at random (MAR, Little & Rubin, 1987; Schafer & Graham, 2002), the mixed-effects modeling framework provides a platform for implementing appropriate procedures for drawing valid inferences of model parameters without forcing the researcher to omit cases thereby losing potentially valuable information (Enders, 2010).

As the name suggests, a linear mixed-effects model contains both fixed and random effects (the model will explained in more detail in Chapter 2). Random effects models are often linked to the general analysis of variance models. For example, in a one-way between-subjects ANOVA model, "*effects*", defined as the differences between the group means and the grand mean, are commonly treated as fixed, yet unknown, finite constants. These effects can also be thought of as being randomly selected from an infinite population of effects, and assumed to be independently and identically distributed with mean zero with a certain variance. The LME model may be viewed as a generalization of a variance component regression analysis model. When the number of groups is small and the number of observations per group is large, the group-specific coefficients are treated as fixed as in the regular ANOVA model. When the number of groups is large but the number of observations

per cluster is relatively small, a certain number of groups can be randomly selected and the group-specific coefficients are treated as random (Demidenko, 2004).

In the context of longitudinal data analysis, the fixed effects are parameters that describe population growth characteristics, providing a summary of how a response variable changes systematically as a function of time or other condition. The unobserved heterogeneity of growth among subjects is represented through the random effects. The random effects essentially allow individual subjects to have their own functional form, and thus their own trajectories, but whose functional parameterizations are distinct from the population average trajectory.

Introducing random effects in a longitudinal model also has the advantage of explicitly acknowledging that variability in the repeated measures can be partitioned into at least two components: variability that occurs between subjects and variability occurring within subjects. The variance-covariance structure of the random effects describes between-subject variability in the growth characteristics implied by the functional form of the model. The variance-covariance structure of the individual-level residuals represents a measure of misfit between individuals' data and their own fitted function. Interestingly, if the data permit it, the within-subjects covariance structure can be partitioned further to account for measurement error that is separate from serial correlation induced by within-subject fluctuations accompanying the responses of individual over time (Fitzmaurice, Laird, & Ware, 2011).

In sum, the LME model allows for individual functions to differ from the mean function over the population of subjects, yet characterizes both population and individual patterns as members of a single response function. Different sources of

variability arising from the repeated measures can be acknowledged and explicitly modeled. These important facets of change are thought to summarize growth for a single population. Yet, in some instances this assumption is too restrictive or untenable.

1.1 Population Heterogeneity

In a standard LME model, time-specific within-subject errors and an individual's coefficients (random effects) are often assumed to follow a normal distribution and are indeed subject-specific. These assumptions imply that the data are sampled from a single population with common a mean and variance-covariance structure. In some situations, there exist subpopulations that may differ in one or more population parameters. Sometimes the subpopulations are known by the researcher and thus can be modeled by adding a covariate in the model (e.g., adding a dummy variable indicating subject's gender) or proceeding with a multiple group analysis (Singer & Willett, 2003). In other cases, subpopulations have not been identified by researchers a priori even though theories or previous studies may suggest differences in growth parameters among them. Graphs in Figure 1 are hypothetical examples to demonstrate subgroup differences in growth trajectories. The graph on the left shows the individual trajectories of all people from a target population which is hard to recognize whether there exist subgroups with different growth characteristics. The graph on the right uses different colors to illustrate how two identified subgroups in this particular population distinguish themselves by their growth trends. Without any attention on possible subgroup growth differences, the conventional mixed-effects model may fail to provide accurate estimates for any of

the subgroups since it does not take account of the subpopulation level heterogeneity (Jedidi, Jagpal, & DeSarbo, 1997; Muthén, 1989). Areas of research such as biology, genetics, psychology, social- and cognitive-development regularly encounter situations in which theories support distinct developmental trajectories within unknown subpopulations. For example, Rescorla, Mirak and Singh (2000) studied the development of children's vocabulary and found that two groups of "late-talker" children showed dramatic vocabulary spurts at different ages. The delay in vocabulary acquisition of one group of children had direct clinical implications for diagnosing language delay among children in general.

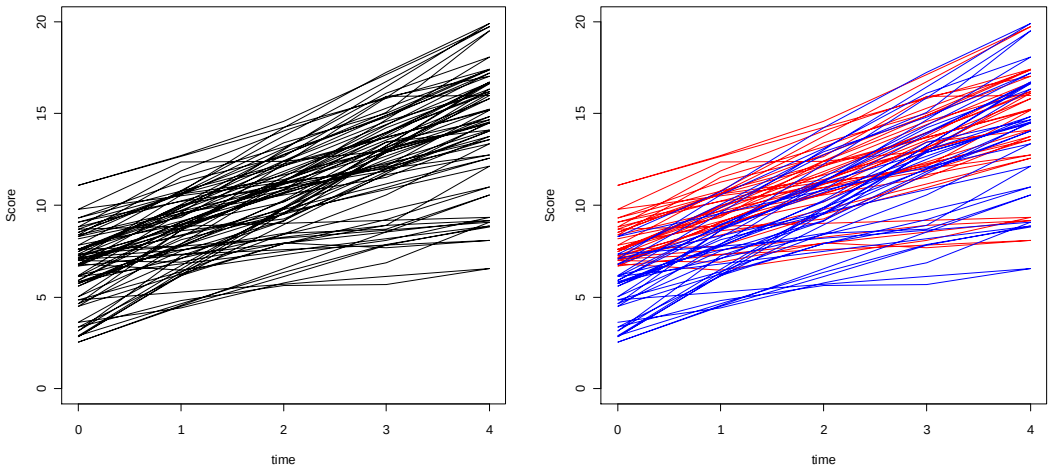

Figure 1. An example of hidden subpopulation heterogeneity in growth.

In response to the demand of modeling population heterogeneity in longitudinal profiles, LME models, and more broadly growth models, have successfully incorporated finite mixture models into this framework (Muthén & Shedden, 1999; Verbeke & Lesaffre, 1996; Verbeke & Molenberghs, 2000). Finite mixture models have been used to depict a variety of phenomena in numerous fields including

biology, physics, economics, psychology and social sciences. One of the earliest studies in mixture modeling was conducted by Karl Pearson over 100 years ago. In his classic paper, Pearson (1894) investigated subspecies among crabs and obtained estimates for a normal mixture distribution using a moment-based approach. In longitudinal analyses, a finite mixture model can be specified in situations where a single parametric family is inadequate to provide a satisfactory description of change characteristics or variability in observed repeated measures data. A finite mixture model relaxes the assumption of a single population and allows parameters to vary across different subpopulations (Muthén, 2004). In sum, a finite mixture of growth models has become a powerful tool to detect heterogeneous growth trajectories of unobserved population subgroups. After group membership identification, further analysis on its relation with possible covariates can be carried out.

1.2 The Current Study

Researchers in the field of growth modeling are sometimes interested in investigating the existence of subpopulations with distinctive growth trajectory characteristics, a model-based post-hoc classification of subjects, or both. The growth characteristics refer to both parameters that describe the functional form of the trajectories as well as variance and covariance components summarizing the patterns of variability of the repeated measures. Investigation of a simple linear growth model, for example, might hypothesize subpopulation differences in intercept and slope parameters. In addition, variability in the repeated measures modeled through the random effects and time-specific residuals may also differ by latent subpopulations. Studies on growth mixture modeling have extensively investigated

issues about parameter recovery of mean structure components; model fit indices, and classification accuracy (Muthén & Shedden, 1999; Nylund, Asparouhov & Muthén, 2007; Tofighi & Enders, 2008; Tolvanen, 2007; Wang & Bordner, 2007). Real data analyses mainly focus on discovering the differences in the mean structure, in other words, the subpopulation intercepts and slopes for the linear model (Colder et al., 2002; Odgers et al., 2007; Verbeke & Lesaffre, 1996) but much less attention has been paid to the variability of the random effects and residuals. Researchers have recognized that class separation among clusters can affect the recovery of parameters and classification accuracy, but none of them have systematically investigated how patterns of variability in the repeated measurements can affect class separation, which in turn impacts the ability of the model to generate estimates. The major objective of this study is to focus on the roles the between-subject and within-subject variance-covariance structures play in detecting a finite mixture of unobserved groups and parameter recovery in the context of LME models as a tool for modeling growth.

Chapter 2: Literature Review

This chapter introduces the linear mixed-effects model and its extension to growth mixture models. As mentioned in Chapter 1, the linear mixed-effects model emerged from regular linear regression models. The beginning of this chapter will briefly talk about regular regression models and the reason why random effects should be added for repeated measures design. Finite mixture distributions will be discussed along with an introduction of measures of distances among component distributions. The growth mixture model which is an extension of the linear mixed-effects model through adding mixture components is explained followed by an illustration of the estimation and applications of the model.

2.1 Regression Models

Modern statistical methods of handling longitudinal data have a strong foundation in regression. Before introducing the linear mixed-effects model for repeated measures data, a brief discussion of the standard linear regression model is warranted. Consider the following general linear model,

$$y_i = \mathbf{X}_i'\boldsymbol{\beta} + e_i \tag{1}$$

where y_i is the response or dependent variable for *i*th subject, $\mathbf{X}_i' = (x_{1i}, x_{2i}, ..., x_{pi})$ is a $1 \times p$ vector whose elements are values on a set of independent variables or predictors, and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ is a $p \times 1$ vector of regression coefficients. In the linear regression model, it is presumed that all individuals have the same population regression coefficients $\boldsymbol{\beta}$ which are often referred to as *fixed parameters* (Kutner, Nachtsheim, Neter, & Li, 2005). The regression model in Equation 1 demonstrates that *y* is characterized by a linear combination of the predictors. The uncertainty in

the relation is modeled through the error term e_i which is generally assumed to be normally and independently distributed with mean zero and common variance, σ^2, and uncorrelated with the predictors in the model. On the right side of Equation 1, e_i is the only random term in the regression model that is allowed to vary among different individuals. Since the error or residual term is randomly distributed among individuals, it is often referred to as "random error."

An ordinary regression analysis assumes that the observations are independent from each other. This assumption is violated when the data are clustered – as they are when the same individuals are measured repeatedly over time. In studies of agriculture, behavioral science and education, clustered data are common. For instance, in the study of crop yield, several individual plants may be planted within the same plot. In this way plants are nested within plot. Other examples of sampling designs that induce a certain correlation among the data include sampling siblings within the same family or students within the same school. Longitudinal data is a special case of clustered data where the clusters are composed of repeated measurements on the same individual across different occasions. Observations within a cluster are not independent and the correlations between multiple observations of a single subject should be accounted for in the analysis.

2.2 Sources of Variability in Repeated Measures Data

Three different sources of variability are often identified to have an impact on correlation among repeated measures: between-subject heterogeneity, within-individual variation and measurement error (Fitzmaurice, Laird, & Ware, 2011). Between-subject heterogeneity reflects the natural variation in individuals' propensity

to respond. Individuals may have different response trajectories over time. For example, in a linear growth analysis, individuals have different intercepts and regression slopes. Within-individual variation can be conceptualized as misspecification of different individuals' response trajectory over time which will induce correlation among repeated measures data. Random measurement error is the last source of variability in longitudinal data. In educational and psychological studies, it is often that measurement instruments or procedures are imprecise, which cause within-subject variation. Reliability is the consistency, or reproducibility, of an instrument to measure certain characteristics of subjects. Scores gathered repeatedly from instruments with low reliability have attenuated correlations among the data. Within-individual variation and measurement error are conceptually two distinct sources of within-subject variation. However, they are rarely modeled separately in longitudinal studies (Fitzmaurice et al., 2011). Instead, they are often combined into a single error term. Figure 2 shows how these three sources of variability are represented in longitudinal data.

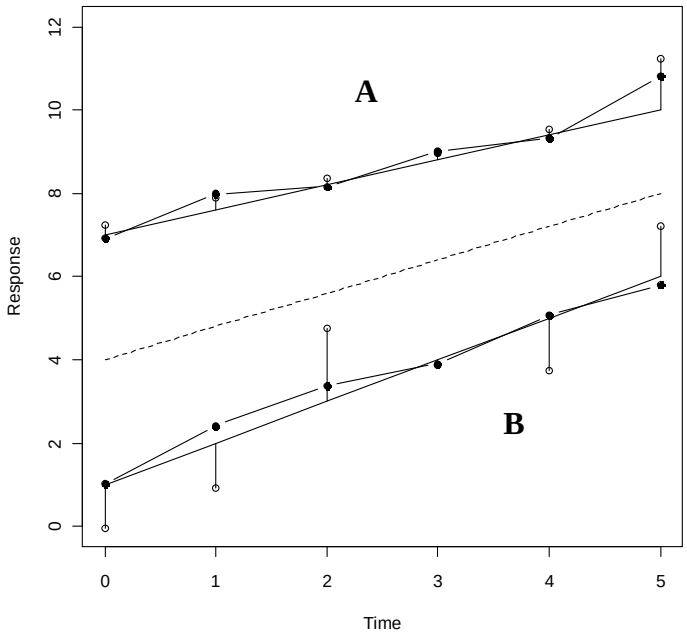

Figure 2. Three sources of variability represented in longitudinal data.

Figure 2 above shows the growth trend of two individuals, A and B, at six measurement occasions. The dotted line is the population growth trend while the straight lines are the individual trajectory for A and B. Separation of the true response profiles (straight lines) for subjects A and B represent heterogeneity (or between-subject variation) in individuals. The black dots are the repeated measures with no measurement error while the open circles denote the observed repeated measures with measurement error. The amount of measurement error resulting from using a particular instrument will largely impact the degree of correlation among repeated measures.

The correlated error structure makes repeated measures data not applicable for regular regression analysis. Conventional approaches to handling repeated measures

data include univariate analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), covariance pattern models, transition models and mixed-effects models. Mixed-effects models have some advantages over these other, more traditional alternatives statistical methods, when analyzing longitudinal data. According to Blozis and Cudeck (1999), this family of models allow (*i*) both population and individual patterns of change to be characterized with a common mathematical function yet whose parameterizations are different; (*ii*) subjects to be measured at unique occasions of time or condition; (*iii*) the number of measurement occasions to be different; (*iv*) specification of more realistic residual covariance structures ; and (*v*) missing data when the missing data are missing at random or other can be handled in a straightforward manner.

To elaborate on this latter point, mixed-effects models are ideal candidates for longitudinal analyses because they can accommodate both unbalanced designs and missing data which are often encountered in practice. Thus, occasions which each individual are measured do not have to be equally spaced, and in fact, can be a completely unique sequence. In longitudinal studies, missing data are almost inevitable since, for many non-experimental protocols, there is greater chance for participants to miss one or multiple observations. Of course, missingness can occur for a variety of reasons including dropout, attrition, or some other unforeseen circumstance. When there exits missing observations, the data are unbalanced over time and not all individuals have the same measurement occasions. Sometimes the unbalanced data in longitudinal studies is planned by the researchers to reduce the time span or cut the cost of the study. The *cohort sequential design* (Duncan, Duncan,

& Stryker, 2006) is a good example of planned missingness while the *rotating panel* (Laird, 1988) design is an example of a planned unbalanced design for longitudinal studies.

2.3 Linear Mixed Effects Models

The linear mixed-effects (LME) model, first mentioned as a two-stage random effects model by Laird and Ware (1982), evolved from the conventional multiple linear regression model with the inclusion of additional random terms for some or all of the fixed regression coefficients. Using vector and matrix notation, the classical linear mixed-effects model for a typical individual selected from the population can be expressed as,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \tag{2}$$

where $\mathbf{y}_i = (y_1, \ldots, y_{n_i})'$ is an $n_i \times 1$ vector of responses for the *i*th individual, $i = 1, \ldots, m$, $\boldsymbol{\beta}$ represents a $p \times 1$ vector of fixed effects, $\mathbf{X}_i$ is a design matrix for the fixed effects specific to the *i*th individual, $\mathbf{b}_i$ is a $q \times 1$ vector of random effects, $\mathbf{Z}_i$ is an $n_i \times q$ design matrix for the random effects, and $\mathbf{e}_i$ is an $n_i \times 1$ vector of regression errors, which is often assumed to normally and independently distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}_i$: $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. In this model, $\mathbf{b}_i$ represents the individual difference in growth, i.e., between-subject variation, while $\mathbf{R}_i$ represents the within-subject variability of data including within-subject variation and measurement error. Conditional on the random effects, $\mathbf{b}_i$, Equation 2 implies

$$\mathrm{E}(\mathbf{y}_i \mid \mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i \qquad \mathrm{cov}(\mathbf{y}_i \mid \mathbf{b}_i) = \mathbf{R}_i.$$

In practice, in the second stage of linear mixed effects models, the $q \times 1$ vector of random effects, $\mathbf{b}_i$, is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and $q \times q$ variance-covariance matrix $\mathbf{D}$, independent of each other and of the $\mathbf{e}_i$. That is,

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

$$\text{cov}(\mathbf{b}_i, \mathbf{b}_{i'}) = \mathbf{0} \qquad \text{cov}(\mathbf{b}_i, \mathbf{e}_i') = \mathbf{0} \qquad \text{cov}(\mathbf{e}_i, \mathbf{e}_{i'}) = \mathbf{0} \quad \text{for } i \neq i'.$$

Given the covariance assumptions above, let $f(\mathbf{y}_i \mid \mathbf{b}_i)$ and $f(\mathbf{b}_i)$ be assumed multivariate normal density functions. The marginal density function of $\mathbf{y}_i$ is then given by

$$f(\mathbf{y}_i) = \int f(\mathbf{y}_i \mid \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i,$$

which can be specified in a closed form by carrying out the integration of the joint density function over $\mathbf{b}_i$. Under these assumptions, the marginal mean and covariance for $\mathbf{y}_i$ is

$$\begin{aligned} \text{E}(\mathbf{y}_i) &= \text{E}\left\{\text{E}(\mathbf{y}_i \mid \mathbf{b}_i)\right\} = \mathbf{X}_i\boldsymbol{\beta} \\ \text{cov}(\mathbf{y}_i) &= \text{E}\left\{\text{cov}(\mathbf{y}_i \mid \mathbf{b}_i)\right\} + \text{cov}\left\{\text{E}(\mathbf{y}_i \mid \mathbf{b}_i)\right\} \\ &= \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' \\ &= \boldsymbol{\Sigma}_i. \end{aligned}$$

As can be seen from the previous individual and marginal mean structures that the random effects quantify the extent to which the regression parameters for the *i*th subject depart from the population regression coefficients. As the random effects have a mean of zero, as shown through matrix integration in Harring (2005), $\mathbf{y}_i$ is an

independent multivariate normally distributed vector with mean $\mathbf{X}_i\boldsymbol{\beta}$ and variance-covariance structure, $\boldsymbol{\Sigma}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$.

2.3.1 Estimation of Model Parameters. Inferences are generally made on the marginal distribution via maximum likelihood estimation. Let $\boldsymbol{\xi}$ be a row vector of the unique elements in $\mathbf{R}_i$, then $\boldsymbol{\theta} = \{\boldsymbol{\beta}', \boldsymbol{\xi}, vech(\mathbf{D})'\}'$, where the $vech(\cdot)$ operator creates a column vector of a symmetric matrix by stacking the diagonal and lower diagonal elements below one another. The resulting contribution of individual i to the marginal loglikelihood can then be written as:

$$\begin{aligned}\ln L_i(\boldsymbol{\theta}) &= \ln\left[\prod_{i=1}^{m}\left\{(2\pi)^{-\frac{n_i}{2}}\left|\boldsymbol{\Sigma}_i\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)'\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right)\right\}\right] \\ &= -\frac{n_i}{2}\ln(2\pi) - \frac{1}{2}\left|\boldsymbol{\Sigma}_i\right| - \frac{1}{2}\sum_{i=1}^{m}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)'\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right).\end{aligned}$$

Estimation can be carried out in a number of ways including gradient-based methods (Demidenko, 2004; Lindstrom & Bates, 1988), the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), or restricted maximum likelihood (Harville, 1977; Laird & Ware, 1982).

2.3.2 Example—LME model of Linear Change. To make the general formulation in the previous section more concrete, consider a basic linear mixed-effects model for straightline change with random intercept and slope. For the model expressed in Equation 2, the design matrix for β and b_i are identical:

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix},$$

where t_{ij} is the subject-specific measurement occasions, $j = 1, \ldots, n_i$. The response score for the *i*th subject at the *j*th time point can be described as:

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}. \quad (3)$$

In Equation 3, each individual *i* has a specific intercept β_{0i} and regression slope, β_{1i}. As a basic convention, the individual regression coefficients β_{0i} and β_{1i} can be decomposed into the sum of fixed and random effects, $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$, where β_0 and β_1 are the population intercept and slope, respectively; and b_{0i} and b_{1i} are deviations of the *i*th individual's intercept and slope from the population parameters. In the majority of cases, the number of columns in $\mathbf{Z}_i$ is a subset of columns in $\mathbf{X}_i$. This allows some regression parameters to be fixed across subjects while others can vary randomly. Furthermore, permitting $\mathbf{Z}_i$ and $\mathbf{X}_i$ to be unique allows potentially different static, individual covariates (i.e., gender, treatment condition) to be incorporated to explain why intercepts and slopes vary among individuals. For example, if gender (*Gender*) is added in the model as a person level covariate, the response model would be specified as

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_0 Gender_i + \gamma_1 Gender_i t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij},$$

where γ_0 and γ_1 are the effects of gender on the intercept and linear growth rate. Suppose for person *i* in the non-reference gender group (coded as 1), the design matrix, $\mathbf{Z}_i$ for the random effects will not change but the design matrix for fixed effects becomes

$$\mathbf{X}_i = \begin{pmatrix} 1 & t_{i1} & 1 & t_{i1} \\ 1 & t_{i2} & 1 & t_{i2} \\ 1 & t_{i3} & 1 & t_{i3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & 1 & t_{in_i} \end{pmatrix}.$$

Recall, that the three sources of variation and covariation among the repeated measures can be modeled via the LME model, within-subject variation, between-subject variation and measurement error. An important feature of longitudinal data is that the repeated measures at different occasions are correlated. For regular repeated measures model without random effects, different intra-individual error structures, such as an autoregressive structure, can be specified to account for the serial correlation among the repeated measures. In LME models, the marginal covariance of response vector $\mathbf{y}_i$ has two components, $\mathbf{D}$ and $\mathbf{R}_i$. In general, $\text{cov}(\mathbf{y}_i)$ has non-zero off diagonal elements capturing the correlation among repeated measures and is decomposed into $\mathbf{D}$ and $\mathbf{R}_i$ where $\mathbf{D}$ accounts for the between-subject variation which induces the correlations among repeated measures of $\mathbf{y}_i$ and $\mathbf{R}_i$ is the within-subject variation. In fact, because the random effects usually account for a large amount of covariance among the repeated measures, there is not a great deal of covariance left among individual errors (Fitzmaurice et al., 2011). Therefore in practice, it is common to adopt a simple structure for the error variance-covariance matrix like, $\sigma^2\mathbf{I}_{n_i}$, where $\mathbf{I}$ is an identity matrix of dimension n_i. This simplified error structure was coined the *conditional-independence model* by Laird and Ware (1982) which indicates that the n_i responses on individual i are independent, conditional on $\mathbf{b}_i$ and $\boldsymbol{\beta}$. In other words, the correlation among the repeated observations on the same

individuals is accounted solely by the correlation of random effects. Then the marginal variance-covariance of $\mathbf{y}_i$ could then be defined as $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i}$. To be specific, for the LME model with random intercepts and random slopes, the variance of response of subject *i* at time *j* can expressed as

$$\text{var}(y_{ij}) = \text{var}(b_{0i}) + 2t_{ij}\,\text{cov}(b_{0i}, b_{1i}) + t_{ij}^2\,\text{var}(b_{1i}) + \sigma^2,$$

and similarly the covariance of y_{ij} and y_{ik} is

$$\text{cov}(y_{ij}, y_{ik}) = \text{var}(b_{0i}) + (t_{ij} + t_{ik})\,\text{cov}(b_{0i}, b_{1i}) + t_{ij}t_{ik}\,\text{var}(b_{1i}).$$

The above variance-covariance structure of y_{ij} suggests that no assumption of homogeneity over time is necessary for the mixed-effects model since this structure allows the variances and covariances to vary as a function of time. Thus, the variances of the repeated measures are already complicated functions of time, which implies that the within-subject component may very well be a simple structure.

The proposed model explicated in Section 2.3 assumes that the subjects come from a single population and the random effects are sampled from a normal distribution. However, the distribution of random effects does not necessarily need be multivariate normal. For example, Pinheiro et al. (2001) demonstrated how the random effects could be modeled with a multivariate *t*-distribution with known or unknown degrees of freedom to obtain more robust and reliable estimates from data with outliers. Oberg and Davidian (2000) proposed using a transformation of response and predictors to achieve approximate within-subject normality. Instead of using the standard logarithmic transformation blindly, their model transformed both responses and regression predictors by a parametric function estimated from the data.

Arellano-Valle et al. (2005) adopted a skew-normal distribution for the random effects and the within-subject errors in mixed-effects models to address non-normality. Another method that has been suggested to account for non-normality in the random effects distribution is to assume a finite mixture distribution. Muthén and Asparouhov (2009) demonstrated how to use mixture modeling with latent classes to represent non-normality of random effects. They referred to their model as a non-parametric representation of random effects, an approach that discretized the random effects distribution into a finite mixture distribution where the latent class means and class probabilities are points and weights of the component distributions.

The above mentioned models for non-normal random effects distributions still assume all individuals come from a single population and that a single growth trajectory can adequately depict the entire population growth characteristics. Yet, existing theories and studies in many fields have suggested different subgroups have different growth trajectories. For example, a large amount of literature in human development have shown people progress differently in a variety of disciplines, such as alcohol usage, cognition, and language acquisition to name just a few (Chassin, Pitts, & Prost , 2002; Connell & Frye, 2006; Nagin & Tremblay, 2001; Rescorla, Mirak, & Singh, 2000). The presence of non-normal random effects distributions can indicate the existence of such sub-populations as well. The growth model can then be combined with latent class analysis or mixture model to capture the unobserved subgroup heterogeneity within a larger population. Verbeke and Lesaffre (1996, 1997) extended the LME model by applying a more flexible distributional assumption on the random effects. In these papers and the book chapter in Verbeke and

Molenberghs (2000, Chpt. 12), the authors referred to this more flexible random effects modeling as the heterogeneity model, which assumes the random effects are sampled from a mixture of normal distributions. The heterogeneity model assumes subgroups in the population with distinct growth trajectories and within each subgroup the random effects form a component of the mixture distribution with specific mean and/or variance-covariance structure. In this case it would be useful to classify people into different subgroups and identify their unique growth trajectories, which will be the focal point of this study. As a point of comparison, if the between-subject variance and covariance estimates within each class are restricted to zero, then the model can be conceptualized as a latent class growth model (Nagin, 1999; Nagin & Land, 1993). For the latent class growth model, all individual growth trajectories within a class are assumed to be homogeneous which greatly improves model convergence in computation. Thus, it can serve as a pre-process for conducting growth mixture modeling.

2.3.3 LME Model and LGC Model Equivalency. As was shown by Muthén and Asparouhov (2009), the LME model defined in Equation 2 is statistically equivalent to the latent growth curve (LGC) model (Bollen & Curran, 2006; Preacher et al., 2008) as implemented in Mplus (Muthén & Muthén, 1999-2010). Consider a linear latent growth process with continuous outcome y, the model can be written as

$$y_{ij} = \eta_{0i} + \eta_{1i} a_{ij} + \varepsilon_{ij}, \tag{4}$$

where a_{ij} indicates the time measurement for subject *i* at occasion *j*, η_{0i} and η_{1i} are the subject-specific intercept and slope, respectively for subject *i*, and ε_{ij} are the time

specific unique factors. An individual's growth characteristics η_{0i} and η_{1i} can be further expressed as a function of a population intercept α_0 and slope α_1 and random residuals ζ_{0i} and ζ_{1i} with mean zero and certain variability. The decomposition can be expressed in the following equations.

$$\eta_{0i} = \alpha_0 + \zeta_{0i} \tag{5}$$

$$\eta_{1i} = \alpha_1 + \zeta_{1i}. \tag{6}$$

In a multilevel modeling framework (Singer & Willett, 2003), Equation 4 represents the level-1, or subject-specific model, while Equations 5 and 6 represent the level-2, or population models. To make the equivalency more explicit, express the LME model in Equation 2 in the language of the LGC model by defining

$$\mathbf{\Lambda}_i = \begin{pmatrix} 1 & a_{i1} \\ 1 & a_{i2} \\ 1 & a_{i3} \\ \vdots & \vdots \\ 1 & a_{in_i} \end{pmatrix},$$

then

$$\mathbf{X}_i = \mathbf{Z}_i = \mathbf{\Lambda}_i,$$

$$\boldsymbol{\beta} = (\alpha_0, \alpha_1)',$$

$$\mathbf{b}_i = (\zeta_{0i}, \zeta_{1i})',$$

$$\mathbf{e}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})'.$$

The LGC model can then be expressed in matrix notation as

$$\mathbf{Y}_i = \mathbf{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i.$$

For the basic model examined here, any difference between the LME model and LGC model is primarily philosophical and not algebraic. The LME model allows for more complex (i.e. dependent) data structures by separating the covariance structures among lower and higher levels of data, whereas LGC models aggregate covariance to a single level structure. Nevertheless, within a two-level growth model, these two models provide identical solutions. Curran (2003) demonstrated the isomorphism between LME and LGC models analytically and empirically. He concluded that estimation of any two-level LME with level-1 and level-2 predictors is equivalent to a similarly specified LGC model. For unbalanced data LGC models should be estimated using full information ML to achieve identical estimates with LME models.

Later in this chapter, the LME model will be extended to finite mixtures and the extension will be equivalent to the finite mixture version of the LGC model. The statistical connection between LME models and LGC models makes it convenient to analyze LME models using SEM software, like M*plus*, which is designed for analyzing LGC models but has the additional flexibility to incorporate finite mixture models.

2.4 Mixture Distributions

In the past decade, finite mixture models have received more attention than ever from broad fields in biology, psychology and the social sciences. A variety of newer statistical techniques has been created based on finite mixture distributions such as latent class analysis, cluster analysis, discriminant analysis and pattern recognition. Mixture models are able to model complex distributions “through an

appropriate choice of its components to represent accurately the local areas of support of the true distribution" (p. 2) (McLachlan & Peel, 2000). It is also useful to adopt a mixture distribution in modeling situations intended to detect potential heterogeneity in the population (Everitt & Hand, 1981; McLachlan & Peel, 2000). In this study, finite mixture distributions will be integrated in the LME modeling framework to investigate different growth profiles among unobserved subpopulations. Because of its algebraic equivalency with latent growth mixture model which is a combination of mixture model and LGC model, this model will be called a growth mixture model (GMM) for the remainder of the paper.

2.4.1 General Formulation. A mixture distribution is a probability distribution which can be expressed as a combination of two or more conditional density functions. The underlying assumption of a mixture distribution is that the random variables are conditionally independent given another random vector. If the random vector is a discrete variable, i.e., the number of conditional density functions is finite, the compound distribution is a finite mixture distribution (Everitt & Hand, 1981). For example, the population distribution of students' weight can be expressed as an infinite superposition of weight density conditional on height or a finite composition of weight density conditional on gender. The present study will focus exclusively on the finite mixture distributions. Let $\mathbf{y}_{1,}\ldots,\mathbf{y}_n$ denote p dimensional random vectors from a random sample of size n. First, let any vector belonging to $\mathbf{y}_1\ldots\mathbf{y}_n$ be a continuous random vector with a probability density function. If y is any multivariate mixture distribution containing K number of density functions

conditional on variable *x* from a multinomial distribution with *K* categories, the density $f(\mathbf{y})$ can be written in the form

$$f(\mathbf{y}) = \sum_{k=1}^{K} p(x_k) f(\mathbf{y} \mid x_k),$$

where $p(x_k)$ is the marginal distribution of variable **x** which is often named π_k in the literature of mixture distributions. The conditional distribution, $f(\mathbf{y} \mid x_k)$, is often written as $f_k(\mathbf{y})$ which is the density of random variable *Y* given group membership *k* and is often called the component densities of the mixture distribution. Thus the density function of a *K*-component mixture distribution can be expressed in the following form as well,

$$f(\mathbf{y}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}),$$

where $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$. The values $\pi_1, ..., \pi_K$ have been referred to in the literature as the mixing proportions or weights (McLachlan & Peel, 2000). The component density, $f_k(\mathbf{y})$, can be any type of distribution but in practice are regularly assumed to come from the same parametric family, like the exponential family.

When the distribution of variable *x* is known, we can use the equations above directly to express the mixture distribution. For example, people are often interested in how subjects of different gender would respond differently to certain treatments or follow distinct growth trends. Nevertheless, in many real analytic situations, data for *x* is unavailable or latent and the overall mixture distribution is the only known quantity. In these cases, it is impossible to observe the underlying variable which splits the observations into groups. Thus the parameters in each conditional

distribution and the mixing proportions or weights become parameters that need to be estimated from the observed data.

Substantial work has been done to study the mathematical and statistical properties of mixture distributions. Many studies were conducted under the circumstance that the existence of mixture distributions and the number and functional forms of component densities were already known. For these applications, theorists have devised many methods for jointly estimating the parameters of mixture distributions and the mixing proportions. The methods range from Pearson's (1894) method of moments, maximum likelihood estimation (McLachlan & Krishnan, 2008; Rao, 1973), a fully Bayesian approach (Diebolt & Roberts, 1994) and informal graphical techniques (Fowlkes, 1979). Within maximum likelihood estimation, the mixture problem is often tackled by the EM (Expectation-Maximization) algorithm and formulated as an incomplete-data problem (McLachlan & Peel, 2000). In reality, the number and functional form of the component densities are often unknown to the researcher. Sometimes it is uncertain whether the data come from a mixture distribution at all. For instance, Bauer and Curran (2003a, 2003b) suggested using mixture models with great caution to distinguish between a single component LGC model with corresponding nonnormal random effects distribution and a true mixture distribution. Their study results showed that the current procedures proposed for model checking of the mixture status as the data may not always effectively differentiate between these two conditions. In the ideal situation, theory would dictate whether or not a finite mixture is plausible or suggested. In the context of an exploration of the data, it is crucial to test for the presence of a mixture distribution,

and if the data support the more sophisticated modeling scenario, how should one proceed to discover the true number of component densities as well as their real function forms. The bootstrap likelihood ratio test and information criteria as AIC and BIC have been commonly used for choosing the number of components for a mixture density.

The focus of the present study is on finite mixture models with normal components. In practice it is common for researchers to assume the mixture distribution is a composite of multivariate normal components. Under many circumstances, a mixture model is built on the basis of non-normal features in the data which are presumed to result from existence of underlying, latent subgroups in the population. The mixture distribution with normal components can be generally defined as

$$f(\mathbf{y}) = \sum_{k=1}^{K} \pi_k \phi_k(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where $\phi_k(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate normal density which is characterized by component mean vector, $\boldsymbol{\mu}_k$ and component covariance matrix, $\boldsymbol{\Sigma}_k$. The multivariate normal mixture is the basis for growth mixture modeling with the noted exception that the mean vector and the variance-covariance matrix for the latter are structured to adhere to the growth process and its attributes. The combination of the linear mixed-effects model with a finite mixture model is defined as a growth mixture model which will be introduced in the following section.

2.5 Growth Mixture Models

Linear mixed effects models are frequently used for longitudinal data analysis. The random effects define the between-subject variance-covariance structure while the regression errors define the within-subjects variance-covariance structure. In general, both random effects and residual errors are assumed to be normally distributed. This assumption is often taken for granted and applied with little thought as to the consequences of violating this assumption. This is largely due to the lack of tools to verify this assumption. In standard linear models, residuals can be plotted against predicted values to check the assumption of normality, constant variance and outliers. These techniques can be applied to linear mixed-effects models for residual diagnostics as well. However, diagnostics for mixed-effects models are more difficult to perform and interpret, due to the presence of random effects and different covariance structures. The predicted random effects values are not eligible for normality assessment since their distribution may not reflect the true distribution of random effects (Verbeke & Molenberghs, 2000; West, Welch, & Gałecki, 2007). When the focus is on finding a population growth trajectory, some important factors that may explain the heterogeneity among individuals may be omitted. For example, studies about human height development commonly use gender, race and other demographic variables to explain why people grow differentially. If the variables that would affect the growth trajectory are well-known, it would be easy to include predictors or covariates in linear mixed effects model to explain group differences. In many research situations, information about sub- populations is unknown to researchers. Treating multiple growth trajectories as a single trajectory for whole

population may result in inconsistent research findings. As Wang and Bonder (2007) pointed out, the reason that previous studies about retirees' psychological well-being found different change trajectories might be that there exists multiple patterns of retirees' psychological well-being changes corresponding to unobserved subpopulations.

Arguably, modeling this type of categorical or class information would help sharpen an understanding of the repeated measures if it were known. That is, understanding differences in gender would be helpful in explaining observed differences in growth of adolescents over time. In the event that classes are unknown, the existence of genuinely different growth patterns in the sample manifested through the individual trajectories themselves may still be suspected. An important relatively recent development in the research on these methods is the extension to latent classes. Unknown classes arise when genuinely distinctive clusters of change exist, but are embedded within individuals' growth patterns. Growth mixture models, which incorporate heterogeneity in the random effects, appear to be a sensible approach in uncovering these latent classes (Muthén & Muthén, 2000; Nagin, 1999; Verbeke & Lesaffre, 1996).

A combination of mixture distributions and linear mixed-effects models is not a new idea in statistics. Verbeke and Lesaffre (1996) have already investigated how to detect a mixture in the distribution of random effects in linear mixed effects model. They did not use the term "growth mixture model" in their paper but referred to their model as "heterogeneity model". However, in Verbeke and Lesaffre's study, only the means of random effects were assumed to vary among component distributions but

not the covariance between random effects. The present study would extend this model to a more general form to account for more possibilities for heterogeneity of individual growth.

2.5.1 Growth Mixture Model Specification. The standard linear mixed-effects model has already been explained in Section 2.2. In this section, a growth mixture model based on the linear mixed-effects model will be introduced. If there exist several sub-populations which have different growth trajectories, the differences among sub-populations can manifest in different places, fixed parameters that describe the mean growth trajectory, the random effects distribution, and residual distribution. A most relaxed formulation of growth mixture model in the linear mixed-effects framework would be

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{7}$$

$$\mathbf{e}_i \sim \sum_{k=1}^{K} \pi_k N(\mathbf{0}, \mathbf{R}_k) \qquad \mathbf{b}_i \sim \sum_{k=1}^{K} \pi_k N(\mathbf{0}, \mathbf{D}_k)$$

$$\sum_{k=1}^{K} \pi_k = 1$$

Equation 7 implies that $\mathrm{E}(\mathbf{y}_i \mid \mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i$. The marginal mean and covariance for $\mathbf{y}_i$ is

$$\begin{aligned}
\mathrm{E}(\mathbf{y}_i) &= \mathrm{E}\left\{\mathrm{E}(\mathbf{y}_i \mid \mathbf{b}_i)\right\} = \mathbf{X}_i\boldsymbol{\beta}_k \\
Cov(\mathbf{y}_i) &= \mathrm{E}\left\{Cov(\mathbf{y}_i \mid \mathbf{b}_i)\right\} + Cov\left\{\mathrm{E}(\mathbf{y}_i \mid \mathbf{b}_i)\right\} \\
&= \mathbf{R}_{ik} + \mathbf{Z}_i\mathbf{D}_k\mathbf{Z}_i' \\
&= \boldsymbol{\Sigma}_{ik}
\end{aligned}$$

Therefore the marginal distribution of $\mathbf{y}_i$ is $\sum_{k=1}^{K} \pi_k N(\mathbf{X}_i\boldsymbol{\beta}_k, \boldsymbol{\Sigma}_{ik})$.

The above unrestricted model can impose estimation difficulties since the likelihood function is unbounded (details forthcoming). In a GMM framework, this model is unidentified. Naturally, researchers make constraints on parameters to make

the model identifiable and to smooth the estimation process. In practice it happens that some parameters in the model may not vary among subgroups. Sometimes subpopulations differ in terms of their mean intercept or slope for a linear model; sometimes they differ only in correlation of the intercept and slope. An important step in conducting a growth mixture analysis is to specify the proper growth mixture model. In this section several possible scenarios where sub-populations show different growth patterns will be introduced and a growth mixture model corresponding to the particular scenario will be specified.

Case 1. Mean growth trajectories vary among sub-populations

The first situation specifies different growth trajectories for each class but assumes the variance-covariance of random effects and residuals remain the same for all sub-populations. This assumption is commonly adopted by many studies in practice within an interest in investigating sub-population heterogeneity of longitudinal data (Colder et al., 2002; Duncan et al., 2006; Verbeke & Lesaffre, 1996; Wang & Bodner, 2007). Some of these studies make this assumption to make the model identifiable in a latent growth model structure. Some also dictate that they have less interest in within-class heterogeneity than the patterns of mean change. Figure 2 shows an example scenario for this case. The graph on the left uses red and black colors to show different subgroup growth profiles, while the graph on the right shows the bivariate distribution of random effects for intercept and slope corresponding to the data in the left graph. The growth mixture model for this scenario can be specified as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}) \text{ and } \mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$$

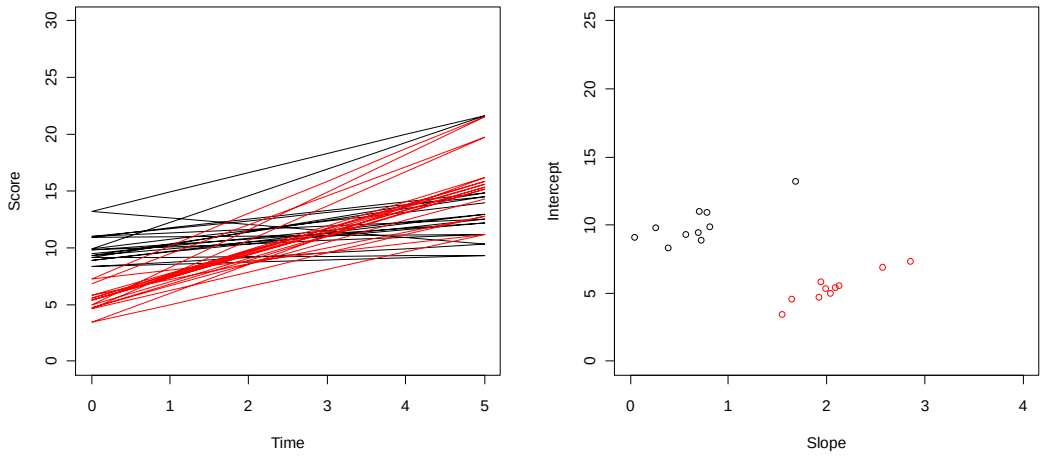

Figure 3. Growth trajectories and intercept-slope distribution of Case 1.

The graph above depicts a popular scenario in a developmental study where some subjects start at lower levels on the outcome but grow faster than those who start at higher levels, and both group reach similar level in the end. Even though the two subgroups start at different levels and grow at different constant rates, the relation between starting point and growth rate remains the same, so does the variability of data.

Case 2. Variance-covariance of intercept and slope vary among sub-populations

Even though the first case scenario is popularly applied in practice, the strong assumption of component-invariant random effects variance-covariance structure makes it unrealistic for many real life phenomena. The assumption is usually applied for convenience or to avoid technical difficulties (estimating the model), yet researchers seldom explore whether this assumption actually holds. In fact, heterogeneous variance-covariance structures among subgroups are likely to be

present in real life applications (Connell & Frye, 2006; deRoon-Cassini et al., 2010; McCullough et al., 2005; Muthén et al., 2000; Muthén et al., 2002; Paririla et al., 2005; Ram & Grimm, 2009). For example, it is reasonable to expect that the slopes vary more for sub-populations with moderate-decreasing and high decreasing levels of depressive symptoms than those at low and high-persistent levels (Stoolmiller, Kim & Capaldi, 2005). Another possibility is that the covariance between intercept and slope can vary across subgroups. Figure 3 demonstrates an example of growth trajectories with these characteristics. The graph on the left shows two subgroups of growth trajectories with different intercepts and slopes; while the graph on the right shows the bivariate distribution of random effects for intercepts and slopes. The corresponding mixture model can be specified as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$

$$\mathbf{b}_i \sim \sum_{k=1}^{g} \pi_k N(\mathbf{0}, \mathbf{D}_k) \text{ and } \mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$$

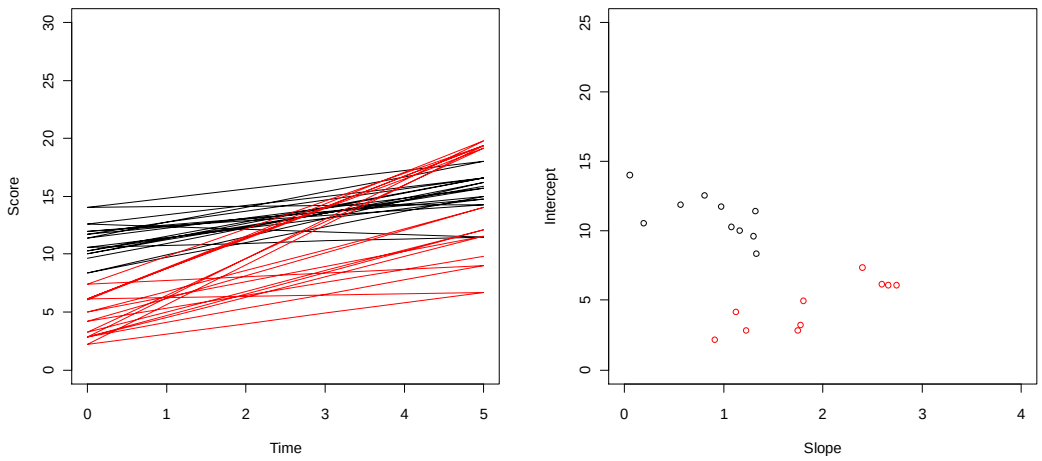

Figure 4. Growth trajectories and intercept-slope distribution of Case 2.

It is clear that the two subgroups illustrated in Figure 4 differ not only in terms of their intercepts and slopes but also with the relation between intercepts and slopes. In the graph on the left, the slopes and intercepts of the red colored group are positively correlated while those of the black group show negative correlation.

Case 3. Error variances vary among sub-populations

The third source of subgroup differences is the within-subject error variances. As elaborated in Chapter 1, within-subject variation comes from within-individual variation or measurement error. Even though they are rarely modeled distinctively in longitudinal studies, some researchers still found significant model improvement by modeling component variant error variances (McCullough et al., 2005; Segawa et al., 2005). Assuming component-specific error variances, the model becomes the ultimate unrestricted model as shown in the beginning of this section. Figure 5 is a scenario based on the model represented in Case 2 with the errors at level-1 coming from a mixture distribution added to the data. The graph on the left of Figure 5 is the mixture distribution of errors with the same zero mean and different variance components. It is clear that the larger error variances of red group definitely increased the data variances of this sub-population. Thus the within-class variation comes from either random parameter variation or within-subject variation. Yet little study has been conducted to examine how these two types of variances and provide correct variance estimates can influence the estimates of GMMs.

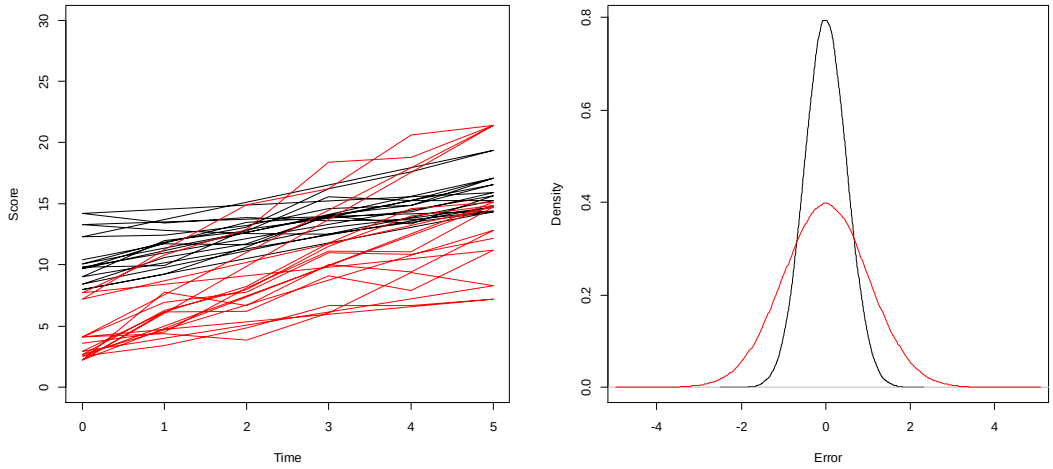

Figure 5. Growth trajectories and intercept-slope distribution of Case 3.

2.5.2 Estimation of Growth Mixture Models. The estimation of growth mixture models are usually implemented via maximum likelihood estimation using the Expectation-Maximization (EM) algorithm. The EM algorithm introduced by Dempster, Laird and Rubin (1977) is a class of optimizers tailored to estimate model parameters via maximum likelihood that can be formulated as a missing data problem. Each iteration of the algorithm consists of two steps, an expectation (or E) step and a maximization (or M) step. The philosophy behind the EM algorithm is to introduce an intermediate, latent variable z whose distribution depends on the unknown parameters and when the loglikelihood is expressed in terms of the distributions of the latent variable, it becomes easier to maximize. In the mixture context, the latent variable is defined as $z_{ik} = 1$ if $\mathbf{y}_i$ is sampled from the *k*th component of the mixture distribution. The prior probability of an individual to

belong to component *k* is $P(z_{ik} = 1) = \pi_k$. The likelihood function corresponding to Equation 7 can be expressed as

$$L(\boldsymbol{\theta} \mid \mathbf{y}) = \prod_{i=1}^{m} \left\{ \sum_{k=1}^{K} \pi_k f_{ik}(\mathbf{y}_i \mid \boldsymbol{\gamma}_k) \right\}, \tag{8}$$

where $\mathbf{y}' = (\mathbf{y}_1', \ldots \mathbf{y}_m')$ is a vector of all observed data and $\boldsymbol{\theta}$ contains all parameters in the marginal model including component probabilities $\boldsymbol{\pi}' = (\pi_1, \ldots \pi_K)$ and $\boldsymbol{\gamma}_k$ which represents all unique parameters in $\boldsymbol{\beta}_k$, $\mathbf{D}_k$, and $\mathbf{R}_k$.

Rewriting the likelihood function for observed data **y** and for the latent variable **z**, the corresponding loglikelihood function is formulated as

$$l(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}) = \sum_{i=1}^{m} \sum_{k=1}^{K} z_{ik} \left\{ \ln \pi_k + \ln f_{ik}(\mathbf{y}_i \mid \boldsymbol{\gamma}_k) \right\}. \tag{9}$$

The above loglikelihood function is composed of two independent parts: the weighted *K* density function $f_k(\mathbf{y}_i \mid \boldsymbol{\gamma}_k)$ and the weighted class proportions.

Compared to the loglikelihood function corresponding to Equation 8, the loglikelihood in Equation 9 is easier to maximize. When maximizing the loglikelihood using the EM algorithm, the latent variable **z** is considered missing. In the E-step the expected values of the probability for the *i*th individual to belong to the *k*th component of the mixture should be calculated for each *i* and *k*. Based on the current parameter estimates $\boldsymbol{\theta}^t$ and $\pi_k^{\,t}$, the posterior probability is given by

$$\pi_{ik}^{(\boldsymbol{\theta}^t)} = E(z_{ik} \mid \mathbf{y}_i, \boldsymbol{\theta}^t) = P(z_{ik} = 1 \mid \mathbf{y}_i, \boldsymbol{\theta}^t)$$

$$= \left. \frac{\pi_k f_{ik}(y_i \mid \gamma_k)}{\sum_{k=1}^{K} \pi_k f_{ik}(y_i \mid \gamma_k)} \right|_{\hat{\pi}^t, \hat{\gamma}^t}$$

The conditional expectation of the loglikelihood in the E-step, $E\left[l(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}) \mid \mathbf{y}_i, \boldsymbol{\theta}^t \right]$, is given by

$$E\left[l(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}) \mid \mathbf{y}_i, \boldsymbol{\theta}^t \right] = \sum_{i=1}^{m} \sum_{k=1}^{K} \pi_{ik}^{(\boldsymbol{\theta}^t)} \left[\ln \pi_k + \ln f_{ik}(\mathbf{y}_i \mid \boldsymbol{\gamma}_k) \right]. \quad (10)$$

In the M-step, the conditional expectation is maximized to get updated estimate $\boldsymbol{\theta}^{t+1}$. Since the two parts of the loglikelihood given by Equation 10 are independent, maximization of these two parts can be carried out separately. The maximization of the first part of the loglikelihood can be done analytically by setting all first-order derivatively to be zero and then solve to get

$$\pi_k^{t+1} = \frac{1}{m} \sum_{i=1}^{m} \pi_{ik}^{(\boldsymbol{\theta}^t)}$$

The second part of the loglikelihood in Equation 10 cannot be maximized analytically but require a numerical maximization procedure such as Newton Raphson. The necessary first- and second-derivatives for the Newton Raphson algorithm within maximum likelihood estimation and restricted maximum likelihood estimation can be found in Lindstrom and Bates (1988, 1994). Once all parameters in $\boldsymbol{\theta}$ in the model have been estimated, the random effects can be calculated using empirical Bayes estimates. The posterior density of random effects $\mathbf{b}_i$ is given by

$$f_i = (\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_{ik}^{\boldsymbol{\theta}} f_{ik}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\gamma}),$$

where $f_{ik}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta})$ is the posterior density function of $\mathbf{b}_i$ given $z_{ik} = 1$. Since the posterior distribution of $\mathbf{b}_i$ is a mixture of different component distributions, the posterior mean of $\mathbf{b}_i$ is

$$\hat{b}_i = \sum_{k=1}^{K} \pi_{ik}^{\boldsymbol{\theta}} E(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\gamma}, z_{ik} = 1).$$

Based on the formula presented by Lindley and Smith (1972), the expected value of $\mathbf{b}_i$ can be calculated by

$$E(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\gamma}, z_{ik} = 1) = \mathbf{D}_k \mathbf{Z}_i' \mathbf{W}_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k) + (\mathbf{I} - \mathbf{D}_k \mathbf{Z}_i' \mathbf{W}_i \mathbf{Z}_i) \boldsymbol{\mu}_k.$$

Consequently, the posterior mean of $\mathbf{b}_i$ is

$$\hat{\mathbf{b}}_i = \mathbf{D}_k \mathbf{Z}_i' \mathbf{W}_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k) + (\mathbf{I} - \mathbf{D}_k \mathbf{Z}_i' \mathbf{W}_i \mathbf{Z}_i) \sum_{k=1}^{K} \pi_{ik}^{\boldsymbol{\theta}} \boldsymbol{\mu}_k.$$

The present study will use M*plus* for model estimation although other software programs have been developed in recent years to estimate GMMs (see, e.g., Open Mx, Latent Gold, or Flexmix in R). M*plus* is a statistical software package that estimates statistical models using observed and unobserved (latent) variables. It has a built-in estimation procedure for GMMs. As was previously demonstrated in Section 2.3.2, the growth mixture model based on linear mixed-effects model is statistically equivalent to the latent growth model and thus it is convenient to carry the estimation through a well-established and widely used commercial software. The specific method used in M*plus* for latent growth mixture model is called MLR (Muthén, 1998-2010), which uses a more robust method to calculate standard errors for the MLE estimates. In addition, M*plus* uses a quasi-Newton method under the full-information maximum likelihood (FIML) framework instead of the Newton Raphson procedure in

the M-step (Muthén, 2004). Maximum likelihood estimation for GMMs in M*plus* is a two stage analysis. In the first stage, the program generates specified number of sets of random starting values and run through a smaller number of iterations with each set using EM algorithm for more stable estimation. In the second stage, the program takes a number of sets with the highest likelihood and continues to iterate through a quasi-Newton algorithm until convergence criteria are met.

It is well known that the estimation of mixture models often encounters local maxima in likelihood function, which may result in biased parameter estimates (Hipp & Bauer, 2006; McLachlan & Peel, 2000). In the case of heteroscedastic normal components, Σ_i are unequal covariance matrices and the loglikelihood of the above function is unbounded. Thus, the global maximizer of the loglikelihood function does not exist. This has brought difficulties in maximum likelihood estimation of multivariate normal mixture distributions. The consistency of MLE solutions for normal components with unrestricted component covariance matrices is yet not verified mathematically (McLachlan & Peel, 2000). In real data analysis, the component covariance matrices $\mathbf{\Sigma}_i$ are often restricted to being the same.

$$\mathbf{\Sigma}_k = \mathbf{\Sigma} \qquad \text{for } k = 1, \ldots, K$$

where $\mathbf{\Sigma}$ is unspecified. Then the maximum likelihood estimation has a global maximization and is strongly consistent.

The focus of the present study does not allow any such restriction of covariance matrices, thus special attention should be paid on model estimation issues. Nitysuddhi and Bohning (2003) investigated the asymptotic properties of estimates computed using the EM algorithm for normal mixture models with component

specific variances empirically through a simulation study. They found that EM algorithm estimates were consistent and had small bias and mean square error except when the subgroup means were close to each other or the variance differences among components were large. As McLachlan and Peel (2000) pointed out, even though the likelihood for these models is unbounded, "there may still, under regularity conditions, a sequence of roots of the likelihood equation corresponding to local maxima with the properties of consistency, efficiency and asymptotic normality" (p. 41). The EM algorithm requires the specification of starting values which to a certain degree will affect the parameter estimates. A way to evaluate whether the estimates possess the above properties is to run the estimation from different starting values and compare the likelihood from different runs. The software M*plus* allows model estimation using a set of permutated random starting values.

2.5.3 Enumeration of Possible Subpopulations. An important issue in mixture distribution models is how to determine the number of mixture components. In the growth analysis case, the question "how many latent trajectory classes exist" needs to be addressed. Sometimes a researcher may have an a priori theory about the number of sub-populations, but in many cases firm knowledge about either the existence of the sub-populations let alone the number of sub-populations is tenuous. Similar to the field of exploratory factor analysis, researchers and scholars have developed a series of statistical tests and model fit indices to facilitate choosing the *correct* number of classes. Currently, many simulation studies have shown that the Bayesian information criterion (BIC) performs better than other information criteria across a variety of modeling settings (Jedidi, Jagpal, & Desarbo, 1997; Nylund,

Asparouhov, & Muthén, 2006; Tofighi & Enders, 2006; Yang, 2006). Some studies also found that Akaike's information criterion (AIC) tends to overestimate the number of components in finite mixture models (Celeux & Soromenho, 1996; Nylund, Asparouhov, & Muthén, 2006). In addition to model fit indices, two type of likelihood ratio tests, the Lo-Mendell-Rubin (LMR) test (Lo, Mendell, & Rubin, 2001) and bootstrap likelihood ratio test (BLRT) (McLachlan & Peel, 2000) have also been shown to be quite effective in determining the number of correct classes (Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008). A major disadvantage of BLRT is that it requires much longer running time than other tests or indices. For the practitioner who is comparing several, yet finite, number of models, the time to run BLRT is not as much of a concern. For methodological simulation studies, however, this is a major drawback, unless of course, the focus of the study is to evaluate the BLRT. Nonetheless, for testing competing models Nylund, Asparouhov and Muthén (2007) and Liu (2011) suggested only using BLRT when other tests or indices, like BIC, pared down the number of potential models to just a small number.

Another method to assess the number of classes in a mixture model is the normalized entropy criterion (NEC) proposed by Celeux and Soromenho (1996). This criterion measures how well separated the classes are from a specific mixture model. It aims to quantify the uncertainty of classification of subjects into latent classes. The entropy values range from 0 to 1, with 0 corresponding to random assignment of class membership and 1 to a perfect model-based classification (Celeux & Soromenho 1996).

As pointed out by many researchers (Bauer & Curran, 2003a; Jung & Wickrama, 2008; Muthén, 2003), besides statistical tests and model fit indices, the number of components of a mixture model should be determined by a series of factors including research question, theoretical support, interpretability of components, and the rule of parsimony. For the current simulation study, and based on evaluation from previous studies, BIC, LMR, and NEC will be the criteria for selecting the number of classes in growth mixture models.

2.6 Previous Simulation Studies in GMM

Previous studies about growth mixture models have focused mainly on model estimation and model selection. Muthén and Shedden (1999) described in detail how the EM algorithm worked in estimating latent growth mixture models. Hipp and Bauer (2006) investigated the local maxima problems involved in GMM estimation through maximum likelihood. Their simulation study found that the MLE estimates of GMM through the EM algorithm were very sensitive to starting values assigned in the beginning of the process. They further proposed a system to select starting values for better model convergence and fewer occurrences of local maxima of the likelihood.

Nylund, Asparouhov and Muthén (2007) and Tofighi and Enders (2008) investigated the performance of a variety of model fit indices and statistical tests on identifying the correct number of classes in growth mixture models. Both simulation studies adopted relatively simple GMM structures and manipulated such factors as class separation, sample sizes and mixture proportions. Nevertheless, in both studies the concept of class separation was not well specified and lacked systematic definition. The standardized difference between the means of two subpopulations is

not necessarily the best way to clarify how the two component distributions are separated from each other.

Another issue that has been overlooked by previous studies is the roles that both within- and between-subject variability play in GMM. The overlap among subgroups of growth mixture data depends on the fixed parameters (component specific to defining the functional form of growth) as well as the variance-covariance structure of the data. As was shown in Section 2.3 and Subsection 2.5.1, the variance-covariance structure of the data is a composition of the random effects variance-covariance structure and the within-subject error variance structure and the sub-population distributions may vary in either or both of these structures. A scientific way of measuring mixture distribution overlap taking into account of the variance-covariance structure is necessary if one wishes to systematically investigate the impact of the variance-covariance structures in the GMM framework. After reviewing a series of articles in the methodological literature and studies of class separation and mixture distribution generation algorithms in a variety of fields, the present study will use multiple indices and decompose mixture structures into different layers to show a more holistic picture of growth mixture data.

2.6.1 Measures of Distance between Component Distributions. An important factor that influences parameter estimates and class membership recovery for mixture distributions is how the component distributions in a mixture distribution are separated from (or in other words, overlapped with) each other. Several statistical indices have been proposed to measure the distance between mixture components. Ideally these measures of distance should satisfy the properties of statistical distance.

Let $g(x)$, $f(x)$, and $h(x)$ be three proper density functions and let $D(g, f)$ be the distance between $g(x)$ and $f(x)$. It should then follow that

a. $D(g, f) \geq 0$

b. $D(g, f) = 0$ if and only if $g = f$

c. $D(g, f) = D(f, g)$

d. $D(g, f) \leq D(g, h) + D(h, f)$

The first approach defines the distance between two densities as:

$$D_p(g, f) = (\int | g(x) - f(x) |^p \, dx)^{1/p},$$

where *p* is commonly set to be 1 or 2. When *p* is equal to 1, it is called Kolmogorov's distance (Ullah, 1996). This family of distance measures satisfies all four distance properties but its computation can become unwieldy as the number of dimensions increases.

The second approach is the family of relative entropy or divergence. Among approaches within this category, Kullback-Leibler (KL) distance is the one of great interest and is regularly used across many disciplines including engineering, economics and educational measurement. The KL distance from density $f(x)$ to density $g(x)$ can be defined by

$$D(f \parallel g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx.$$

KL distance does not satisfy the last two properties of symmetry and triangle inequality (c and d from the above list) and therefore is not referred to as a true metric of distance. That is, the distance from $f(x)$ to $g(x)$ may not be the same as distance

from $g(x)$ to $f(x)$. In practice, to make this measure symmetric, KL distance is often redefined as

$$D(f, g) = D(f \parallel g) + D(g \parallel f).$$

If *g* and *f* belong to certain parametric families, for instance the family of Gaussian distributions, an analytic expression for KL distance is available. Assume $f(x) = N(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$ and $g(x) = N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, then the symmetric version of KL distance between *f*(*x*) and *g*(*x*) is computed as

$$D(f, g) = \frac{1}{2}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g)'(\boldsymbol{\Sigma}_f^{-1} + \boldsymbol{\Sigma}_g^{-1})(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g) + \frac{1}{2} tr[\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_f - 2\mathbf{I}_d]$$

where $tr[\cdot]$ denotes the trace of a square matrix.

Another approach that has been regularly used to measure distances between Gaussian densities is Mahalanobis' distance (MD) proposed by Mahalanobis (1936). To calculate the distance between two probability densities $f(x)$ and $g(x)$, this measure can be written as

$$D_M = (\boldsymbol{\mu}_f - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_{fg}^{-1}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g),$$

where $\boldsymbol{\Sigma}_{fg}^{-1}$ is the pooled covariance between $f(x)$ and $g(x)$. A major advantage of Mahalanobis distance is that it satisfies all four properties of distance. However, this index is only valid to measure distance between two distributions with different means with the same or pooled covariance matrix.

The indices introduced above have been regularly applied in the field of psychology and social sciences. Nevertheless, KL distance and MD are not suitable in of themselves to the present study. The major purpose of this study is to investigate the influence of differences in covariance structures on mixture models and

Mahalanobis' distance fails this purpose by assuming consistent covariance structures across different mixture components. KL distance can be used to quantify the distance between two probability distributions assuming one of them is the true distribution of data. It is not straightforward to generate a mixture of distributions based on this measure. In order to investigate how parameters from mixture distributions with different amount of overlap among components affect the estimation results and class membership recovery, it is crucial to adopt an index that can define the separation/overlap of components in mixture distribution and an algorithm for generating artificial mixtures of univariate or multivariate normal distribution with controlled overlap quantified by the index.

With the fast development of studies on data clustering and finite mixture modeling, many different algorithms have been proposed to generate mixture distributions according to pre-specified amount of overlap in statistical literature. These methods attempt to manipulate group covariance matrices and intra-class correlation, changing standard deviations of mixtures, adding random variables with different expectations to data from the primary population, or altering the means of different distributions iteratively to reach desired overlap between generated mixture components (see, e.g., Atlas & Overall, 1994; Blashfield, 1976; Gold & Hoffman, 1976; McIntyre & Blashfield, 1980; Waller et al., 1999). However, these methods either fail to provide a precise and meaningful definition of population mixture overlap or cannot be extended to multivariate normal mixtures.

Recently there has been great improvement on cluster separation or mixture overlap indices. Various algorithms have been developed according to the definition

of the indices. Aitnouri, Dueau, Wang, and Ziou (2002) used the rate of overlap to describe how much two univariate Gaussian components of a mixture are separated from each other. The rate of overlap was defined as the ratio of the height of the intersection point of the two components to the height of the intersection point of the two components with maximum overlap. The maximum overlap happens when the height of the intersection point of the two components is equal to the minimum value of the standard deviations of the two component distributions. They proposed two algorithms to generate multivariate normal mixture distributions by controlling overlap using the widths of components or using the component means. Even though their definition of overlap is straightforward in the univariate cases, it is hard to visualize the intersection points in multivariate normal mixtures. Moreover, their method of actually simulating data is not done with a stand-alone program, but instead, must to be combined with Milligan's (1985) algorithm to generate multivariate mixture data.

Qiu and Joe (2006) defined the degree of separation of an univariate mixture as the difference between the biggest lower quintile of cluster 2 and the smallest upper quantile of cluster 1 divided by the difference of the biggest upper quantile of cluster 2 and smallest lower quantile of cluster 1. The ratio of the difference ranges from 1 when there is considerable gap between two clusters to -1 when the two clusters overlap substantially. However, like Aitnouri et al. (2002), the index and data generation algorithm put forth by Qiu and Joe became complicated if extended to multidimensional clusters greater than two. It can be incomplete and even problematic when multivariate clusters should be found through one dimensional

projection with the highest separation while the set of pairwise separation indices among neighboring clusters reach the requirement of minimum overlap. This algorithm is now implemented in R package *GenClus*.

Another data cluster generation procedure called "OCLUS" was developed by Steinley and Henson (2005). The OCLUS procedure was designed to generate multivariate data from a variety of distributions with certain amount of overlap which was quantified as the percentage of shared density between clusters. The corresponding data generation algorithm first assumes all dimensions are independent and all clusters are independent. Parameters of each clusters are then computed based on the provided overlap, distribution type and covariance or correlation information. The data will be generated from the computed distributions. To generate correlated variables or data with unequal variances among clusters, the clusters generated from uncorrelated space and equal variance distributions can be transformed to get correlated or unequal variance distributions. Although the overall overlap will be retained and the desired correlation and variances can be achieved, the means of transformed clusters can be shifted due to the oblique rotation of the data.

Maitra and Melnykov (2010) proposed a new method to generate sample multivariate Gaussian mixture distributions. In their approach, overlap between two mixture components is defined as the sum of their misclassification probabilities. If two *p* dimensional Gaussian components follow the distribution of $\phi(\mathbf{X};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ and $\phi(\mathbf{X};\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)$ with mixture proportion of π_i and π_j, the two misclassification probabilities are:

$$\omega_{j|i} = \Pr\left[\pi_i\phi(\mathbf{X};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) < \pi_j\phi(\mathbf{X};\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j) \mid \mathbf{X} \sim N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\right]$$

$$= \Pr_{N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)} \left[(\mathbf{X}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i) - (\mathbf{X}-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\mathbf{X}-\boldsymbol{\mu}_j) < \log \frac{\pi_j^2 \mid \boldsymbol{\Sigma}_i \mid}{\pi_i^2 \mid \boldsymbol{\Sigma}_j \mid} \right], \qquad (11)$$

and similarly,

$$\omega_{i|j} = \Pr_{N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \left[(\mathbf{X}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i) - (\mathbf{X}-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\mathbf{X}-\boldsymbol{\mu}_j) < \log \frac{\pi_i^2 \mid \boldsymbol{\Sigma}_j \mid}{\pi_j^2 \mid \boldsymbol{\Sigma}_i \mid} \right].$$

Thus the overlap ω_{ij} is just the sum of $\omega_{i|j}$ and $\omega_{j|i}$.

When covariance structures are not the same between two clusters, the misclassification probabilities are not easy to calculate analytically. The *p*-dimensional Gaussian components are decomposed into *p* independent non-central chi-square distributed random variables with one degree of freedom and *p* independent standard normal variables multiplied by mean differences, eigenvalues and eigenvectors. The probabilities are then computed using Davies' (1980) algorithm AS 155. The overlap index will guide the simulation of Gaussian components to generate mean and dispersion parameters for clusters to satisfy the overlap characteristics of mixture distributions. The dispersion matrices will be scaled iteratively to ensure the resultant distribution match the desired overlap properties. Both the average overlap and maximum overlap among clusters are accounted for in the data simulation process. This method has been implemented in R package *MixSim*.

For the present study, the method created by Maitra and Melnykov (2010) and outlined above will be adopted to generate multivariate normal mixtures of model parameters because of the simplicity in their definition of distribution overlap, the flexibility to specify a large variety of covariance structures in different clusters as

well as the convenience to simulate data using the existing program package in R. KL distance will also be used to indicate the degree of separation of the means of the intercepts and slopes. The overlap of subgroups of data will be dictated in terms of random effects, residuals and marginal data. The magnitude of overlap will be quantified by the index defined by Maitra and Melnylkov, which will be calculated based on certain degree of mean structure separation and specific variance-covariance structure listed previously.

Equation 11 shows that the overlap in the data is a function of mean structure separation among subgroups as well as how different the variance-covariance matrices of mixture components are. Thus, the key issue in a simulation study to investigate effect of variance-covariances on growth mixture model is to separate the effect of mean differences and variance-covariance differences and relate them to the overall data overlap. A small scale pilot study was conducted to evaluate possible separation indices for means and variances and their relation with overlap in the data. Chapter 3 will outline the specifics of the simulation study, the results of the small pilot study, as well as define the outcome measures.

Chapter 3: Methodology

The major research question of the current study is how within-subject level and between-subject level variability affect the model estimation of growth mixture models. As mentioned in Chapter 2, both mean structure differences and variance-covariance structure of the random effects (between-subject variability) and residuals (within-subject variability) affect the overall data overlap among mixture components. It is more difficulty for growth mixture models to detect underlying subgroups if the data are less separated across subgroups. A simulation study was conducted to evaluate how variability of growth parameters and residuals impact a growth mixture analysis. In Section 3.1, the method that was used for estimating growth mixture models in the simulation study will be outlined and discussed. Section 3.2 will introduce the design and data generation processes of the simulation study. The criteria measures to evaluate the simulation results will be defined in Section 3.3.

3.1 Estimation Method

The current study estimated a growth mixture model using maximum likelihood via the EM algorithm. No constraints were made on the variance-covariance matrix of random effects and residuals across mixture components (i.e., the most unrestricted models were estimated) except that each will be positive definite for the data generation. M*plus* software was used for the model estimation process. Multiple maxima often exist for mixture models as introduced in Chapter 2. Multiple sets of starting values of from a large range are regularly utilized to find the global maximum in mixture model estimation. M*plus* has two stages in ML

estimation of mixture models. The initial stage runs several iterations of the same model using a designated number of starting values sets. A certain number of starting value sets with the highest loglikelihood values are selected for the final stage estimation which will iterate until converge to, hopefully, the same highest loglikelihood value. If the best loglikelihood value is not reached, a warning is given by Mplus that the solution may be at a possible local maximum. This warning statement appears in the output and can be tracked and recorded. The current simulation study adopted Muthén and Muthén (1998-2010)'s recommendations using 100 sets of random initial stage starting values and 10 for final stage optimizations for growth mixture models.

3.2 Data generation

3.2.1 Population Model. The model of interest in the current study is a linear GMM. The hypothesis is that there are two subgroups of subjects with different growth trajectories (assuming both trajectories are linear). Thus, the true number of classes for the growth mixture model is two. Intercepts and slopes of the population model are assumed to follow multivariate normal mixture distributions. The mean as well as the variance-covariance structure of the intercepts and slopes may vary across mixture components. The residuals' variance-covariance structure is fixed to be $\sigma^2\mathbf{I}$ and σ^2 is either component-invariant or component-variant.

The model with component-invariant residual variance can be written as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{12}$$

$$\mathbf{b}_i \sim \sum_{k=1}^{2}\pi_k N(\mathbf{0}, \mathbf{D}_k) \text{ and } \mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$$

where *k* is the number of subgroups or latent classes underlying the general population. This model corresponds to Case 2 in Section 2.5.1. The model with component-variant residual variance depicts the scenario in Case 3 in Section 2.5.1 and can be written as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i,$$

$$\mathbf{b}_i \sim \sum_{k=1}^{2}\pi_k N(\mathbf{0}, \mathbf{D}_k) \text{ and } \mathbf{e}_i \sim \sum_{k=1}^{2}\pi_k N(\mathbf{0}, \mathbf{R}_k).$$

The list of parameters that will be estimated in current study are included in Table 1. All of the data generated based on a 2-class growth mixture model will be fitted with a 1-, 2- and 3-class growth mixture model to investigate the accuracy of class enumeration under a variety of simulation conditions. The number of time points for growth is fixed to be six and are equally-spaced assuming all individual growth trajectories in each subpopulation start and end at the same point.

Table 1

List of Parameter Notations in Current Study

	Intercept	Slope	Proportion	Intercept Variance	Slope Variance	Intercept-Slope Covariance	Residual Variance	
Class1	$\beta_0^{(1)}$	$\beta_1^{(1)}$	π_1	$\varphi_{00}^{(1)}$	$\varphi_{11}^{(1)}$	$\varphi_{01}^{(1)}$	ε [a]	$\varepsilon^{(1)}$
Class2	$\beta_0^{(2)}$	$\beta_1^{(2)}$		$\varphi_{00}^{(2)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(2)}$		$\varepsilon^{(2)}$

a. ε is the residual variance in the first simulation when residual variance is the same in two subpopulations

3.2.2 Manipulated Factors. The first issue to consider for the current simulation is the mixture proportion of subgroups (or mixture components). Several previous studies have concluded that the mixing proportion plays an important role in

growth mixture analyses and other types of mixture data analysis (Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008). The current study investigated the research problem under three different mixture proportion conditions 0.1/0.9, 0.3/0.7 and 0.5/0.5. All other factors were evaluated under each of these mixture proportion conditions.

Previous studies appearing in the literature (Everitt, 1981; Lubke & Muthén, 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008) have concluded that the estimation and classification accuracy of growth mixture modeling analyses and other latent variable mixture models are largely affected by how well the data of subgroups are separated from one another. As mentioned in Section 2.6.1, a variety of measures of mixture distribution separation (or overlap) have been proposed. Previous simulation studies in growth mixture models have regularly used Mahalanobis distance as a measure of class separation. Only a few studies (Nityasuddhi & Bohning, 2003) used their own measures of separation. Mahalanobis distance is based on a standardized mean difference among subgroups assuming variances among subgroups are the same. This distance index does not specifically take into account of the differences of variability in subgroups. Nityasuddhi and Bohning's (2003) D index considered both mean differences and variance differences for a univariate normal mixture scenario. However, the D index was not conceived as a standardized measure, which then makes it difficult to quantify the differences. In their paper, a range of means and variances for two groups were selected and from the computation of D, were categorized as resulting in three coarse levels: low, medium, and high. The current study hypothesizes that both the mean structure and covariance

structure of subgroups can affect how much overlap there is among the data, and in turn, will necessarily affect the estimation of the growth mixture model. Therefore, the simulation requires two separate indices to measure structural differences in the mean vectors and the variance/covariance matrices among subgroups, respectively.

The separation of growth mixture data among subgroups can be separated into two sources: distribution of growth parameters (between-subject variability in growth) and distribution of residuals (within-subject variation). The current simulation study was a composite of two smaller simulation studies. The first simulation study held the distribution of residuals the same across subgroups and examined the effects of growth parameters' (intercept and slope) distribution on data overlap, class membership detection and parameter estimates. The second simulation chose some cases in the first simulation with specific interest and added error distribution differences to subgroups to examine the interaction of growth parameter distribution effects and residual distribution effects. Adding error distribution differences among subgroups significantly reduced the global data overlap. Even though the separation of error distribution among subgroups helped reduce the overlap in the data, the effect of error distributions would be entangled with the growth parameter distribution effect. The investigation of this effect was decided upon after an examination of the results of the first simulation study.

Of great interest in my study is to investigate how the variability structure and mean structure of data interact with each other resulting in different degrees of overlap among subgroup data distributions. For the measure of data overlap, the current study adopted the mixture distribution overlap index proposed by Maitra and

Melnylkov's (2010), which was introduced in Section 2.6.1. Both the mean structure, $\mathbf{X}_i\boldsymbol{\beta}_k$, and variance-covariance structures of $\mathbf{b}_i$ and $\mathbf{e}_i$ influence the overlap in the growth data. To quantify mean separation, squared multivariate Mahalanobis distance (SMD) which has been used as a measure of data separation in many studies in relation to mixture distribution analysis (see Section 2.6.1 for details about this index) was used here too. The measure of variance-covariance matrix difference is a revision of the likelihood ratio statistics proposed by Manly and Rayner (1987). The statistic for the standard likelihood ratio test for a difference between covariance matrices can be calculated as

$$T = \sum_{k=1}^{K} n_k \log\left(\frac{\left|\hat{\Omega}_0\right|}{\left|s_k\right|}\right)$$

where $\hat{\Omega}_0 = \sum_{k=1}^{K} \frac{n_k s_k}{n}$ is the maximum likelihood estimator of the pooled common covariance matrix and s_k is the sample variance-covariance matrix to compare. The current study is not interested to test whether two sample covariance matrices are statistically different from each other per se, but rather to quantify this difference between two variance-covariance matrices. The revised index does not account for the sample size. This index of covariance matrices differences (C_d) is thus

$$C_d = \sum_{k=1}^{K} \pi_k \log\left(\frac{\left|\hat{\Omega}_0\right|}{\left|s_k\right|}\right)$$

The pooled common covariance matrix uses the mixture proportions as weights in calculating $\hat{\Omega}_0 = \sum_{k=1}^{K} \pi_k s_k$. The relation between C_d and Manly and Rayner (1987)'s statistics is linear and has a one-to-one correspondence as shown in Figure 6.

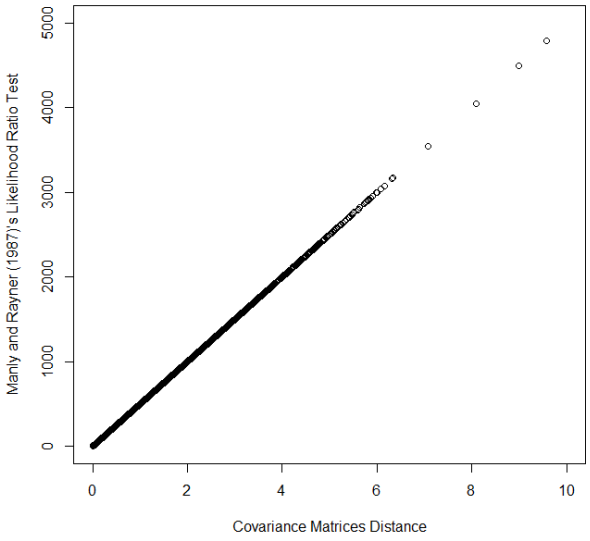

Figure 6. Relation *between* C_d *and Manly and Rayner (1987)'s statistics.*

3.2.3 Pilot study for relation between distance indices and data overlap.

To connect the mean structure difference and variance-covariance structure difference with the overlap in the data, a small-scale simulation was conducted to examine their relations. Because it was not known a priori how various differences in the mean structure and variance-covariance structure would be related to overlap, the design of the simulation was based on examining random values along a continuum instead of choosing particular values. Thus, the procedure started by generating a pool of mean

structures and covariance matrices with random differences among subgroups. Then, for each combination of generated means and variance-covariances, the overlap of the data was calculated. The size of the pool was 10,000 combinations of different mean structures and covariance matrices of each subgroup. For simplicity, the residual variance was not considered in this simulation. Class separation in the data is only a result of mean and variance-covariance differences of the growth parameters. The results showed that the mean structure and variance-covariance structure of subgroup growth parameters affected the overlap of the data quite differently. The interaction among the three indices also differed across different mixture proportion conditions.

Figure 7 illustrates the relation between Mahalanobis distance and distance between covariance matrices. Since the major purpose of the current study is to investigate how mean structure and covariance differences of subgroups affect the growth mixture model analysis, it is crucial to separate the two sources of differences. The graph suggests that there is no significant association between Mahalanobis distance and covariance matrices distance (C_d), which can support the design of the current simulation.

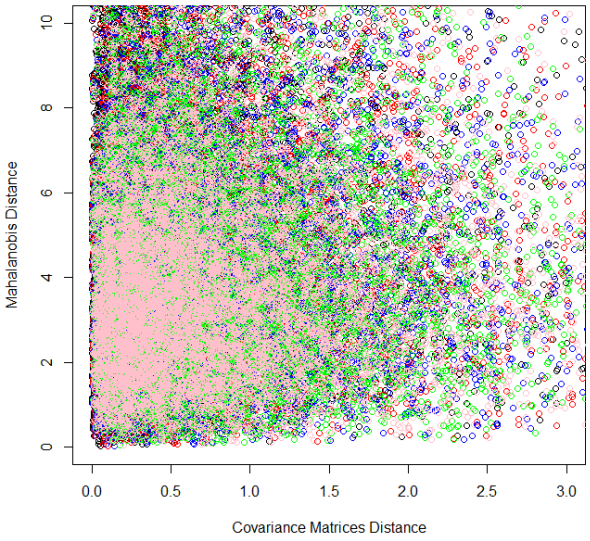

Figure 7. Relation between C_d and SMD.

The following graph (Figure 8) shows the relation between overlap of the data and the mean differences of intercepts and slopes as indicated through Mahalanobis distance. The graph suggests that at a certain level of Mahalanobis distance between subgroup growth parameters, the overlap of the data is limited and this limitation varies for different mixture proportion conditions. For example, given that the mean structure of subgroups are separated by Mahalanobis distance of 3, when the mixing proportion is in the ratio of 0.1/0.9, the maximum overlap of the data is approximately 0.3. However, when the mixing proportion ratio is 0.5/0.5, the maximum overlap of data is restricted to be less than 0.2.

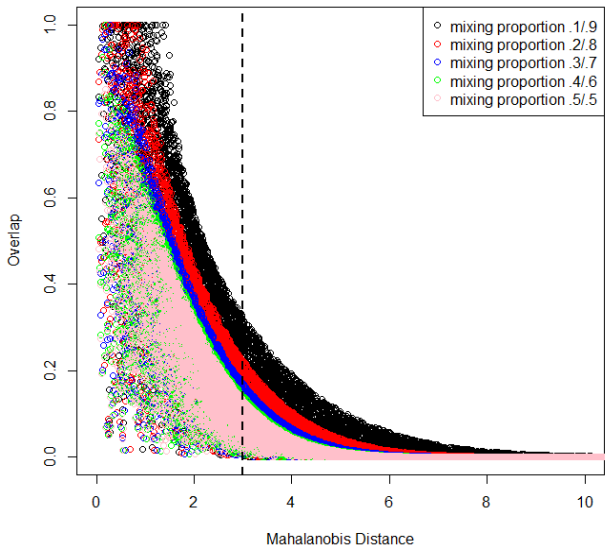

Figure 8. Relation between SMD and overlap in the data.

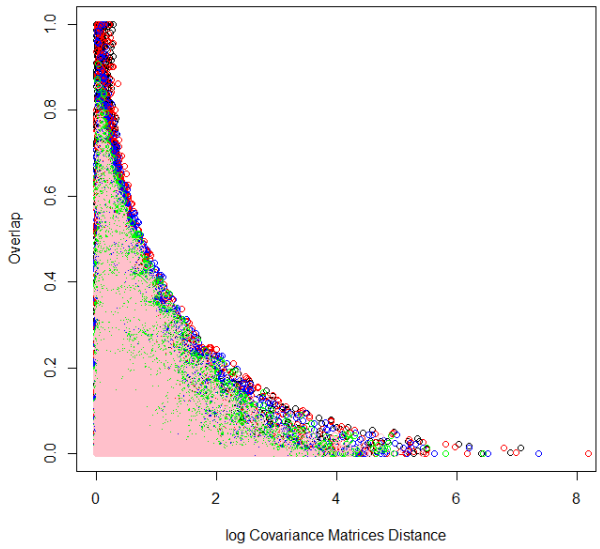

Figure 9. Relation between C_d and overlap in the data.

Figure 9 is a demonstration of the relation between covariance matrix distance among subgroups C_d and the overlap in the data. As distance between covariance

matrices becomes larger, the possibility of high overlap among data becomes smaller. Upon a closer examination of the simulated data, it was evident that as proportions of the two subgroups became more divergent, the overlap of data also depended on where the differences of subgroup variability occurred. When larger variance was associated with the subgroup with the larger proportion, even when mean structure difference and the covariance distance were the same, the overlap of data was smaller than when the larger variance was associated with the subgroup corresponding to the smaller proportion. This phenomenon was especially evident when class proportions were very different such as 0.9 and 0.1. Figure 10 is a contour plot depicting the relation between C_d, SMD and overlap of random effects when mixture proportion is 0.5/0.5. The figure shows that when C_d is larger than .6, even if the standardized mean differences of intercept and slope is zero, the overlap of random effects is less than 0.5.

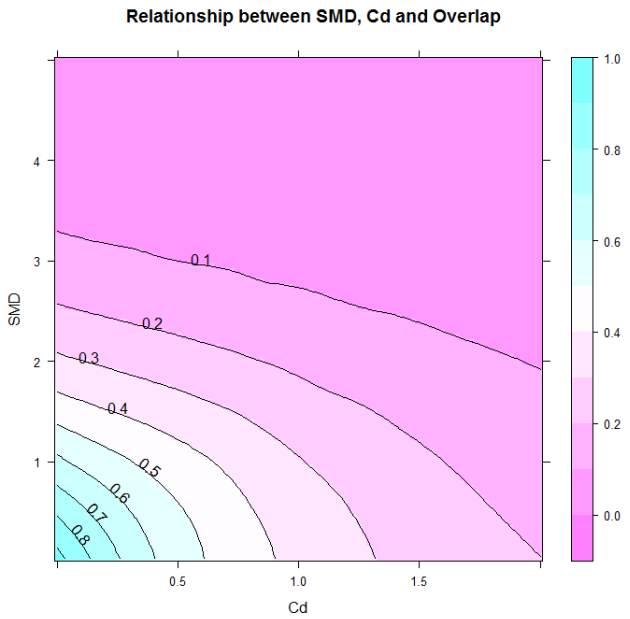

Figure 10. Relation between SMD, C_d and overlap in the data.

Table 2 summarizes the average overlap in the data generated using various mean structure and covariance matrices with different levels of separation. For each level of separation of SMD, the differences in mean structure can manifest in intercept differences or slope differences. Similarly, for each level of C_d, the separation in variance-covariance matrices can be a result of covariance differences or variance differences as well as where larger variances are located with mixture proportions that are unbalanced (as mentioned in the above paragraph). When mixing proportions are 0.5/0.5, within each level of SMD and C_d, how mean structure and variance-covariance matrices differ did not affect the overlap of the data very much. On the other hand, as the mixture proportions became more unbalanced, the variability of overlap becomes larger especially when mean structure differences were not large. Furthermore, unbalanced mixture proportions were associated with larger

overlap in the data across all levels of mean and covariance separation. As mean structure differences increased, especially when Mahalanobis distance was equal to 2.5, the overlap in the data was not significantly affected by other factors. These results formed the basis of the design structure of the current simulations study.

Table 2

Summary Statistics of Overlap by SMD and C_d under Different Mixture Proportions

Mix Proportion		0.5/0.5		0.7/0.3		0.9/0.1	
SMD	C_d	Mean	Std	Mean	Std	Mean	Std
0.5	.0.5	0.76	0.01	0.87	0.04	0.95	0.04
	0.3	0.60	0.03	0.66	0.03	0.67	0.04
	0.6	0.44	0.00	0.49	0.00	0.47	0.00
1	.0.5	0.60	0.01	0.70	0.03	0.87	0.05
	0.3	0.51	0.04	0.57	0.03	0.62	0.03
	0.6	0.37	0.00	0.43	0.00	0.44	0.00
1.5	.0.5	0.44	0.01	0.51	0.00	0.72	0.04
	0.3	0.39	0.03	0.46	0.03	0.55	0.03
	0.6	0.29	0.00	0.35	0.00	0.40	0.00
2	.0.5	0.31	0.01	0.36	0.01	0.54	0.01
	0.3	0.28	0.03	0.34	0.02	0.45	0.02
	0.6	0.20	0.00	0.27	0.00	0.35	0.00
2.5	.0.5	0.21	0.01	0.24	0.02	0.37	0.03
	0.3	0.19	0.02	0.24	0.01	0.37	0.02
	0.6	0.13	0.00	0.20	0.00	0.29	0.00

3.2.4 Population Parameters. Five levels of Mahalanobis distance (SMD) were examined in the current study to measure the mean structure distance of subpopulation growth trajectories, .5, 1, 1.5, 2 and 2.5. Several simulation studies related to growth mixture modeling analysis or latent class modeling (Everitt, 1981; Lubke & Muthén, 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders,

2008) regarded Mahalanobis distance of 2 as being indicative of well-separated classes. Figure 7 also suggests that when the Mahalanobis distance is at least 3, it is not possible for the overlap of data to be larger than 0.5. For each level of mean structure distance, there are two conditions: (1) intercepts are different across groups or (2) slopes are different across groups.

There were 3 levels of random effects (intercept and slope) covariance matrices distance (C_d), 0.05, 0.3 and 0.6 which indicate small, medium and large distances between two covariance matrices of subgroups. As shown in Figure 9, as C_d changes from 0.05 to 0.3, and then to 0.6, in most cases, there is a dramatic drop in the overlap of data. Under each level of C_d , two conditions will be considered: (1) keeping the variances of the intercepts and slopes the same and varying covariances between random intercepts and slopes across the subgroups or (2) vice versa. When varying variances across subgroups, the correlation between intercepts and slopes is set to be 0.2 for both subgroups. Further, the variances of the second subgroup is *d* times the variances of the first subgroup where *d* is a constant selected to make the distance between two variance-covariance matrices to have a certain level of C_d . The relation between C_d and *d* is

$$C_d = \pi_1 \log\left(\pi_1 + \pi_2 d\right)^2 + \pi_2 \log\left[\left(\pi_1 + \pi_2 d\right)^2 \Big/ d^2\right].$$

As explained previously, when proportions of the two subgroups are considerably different, whether larger variance is associated with the subgroup corresponding to the larger proportion or the subgroup with the smaller proportion can cause different overlap in the data given the same mean and covariance

differences. Empirical data from the small scale simulation suggested that this difference in overlap was negligible (about .02) when mixing proportions were 0.7/0.3 but comparatively large when mixing proportions were 0.9/0.1. However, when mixing proportions were 0.9/0.1 and larger variance was associated with the larger proportion, *d* had to be very large (>11) to reach the medium and large levels of C_d, which was not realistic in real-world applications. Therefore, for both conditions with mixture proportions of 0.7/0.3 and 0.9/0.1, a larger variance was assigned to the subgroup with the smaller proportion. When only covariances differed across subgroups, it was impossible to reach a C_d larger than 0.3. Therefore, there were only two levels of C_d under this condition.

Mean structure and variance-covariance matrices for each subgroup were set up to obtain the desired Mahalanobis distance and C_d. Combinations of mean structure difference, covariance matrices difference and the mixture proportion result in different overlap in the data--the overlap of data will be another factor to be evaluated for the simulation results. Figure 9 shows two examples of generated data under the simulation condition of mixture proportion ratio of 0.5/0.5, Mahalanobis distance of 1.5 (when slopes are different across subgroups) and C_d of 0.3. The graph on the left represents the situation in which the variances of intercept and slope differ in subgroups while the graph on the right represents the situation in which covariance of intercepts and slopes differ in subgroups.

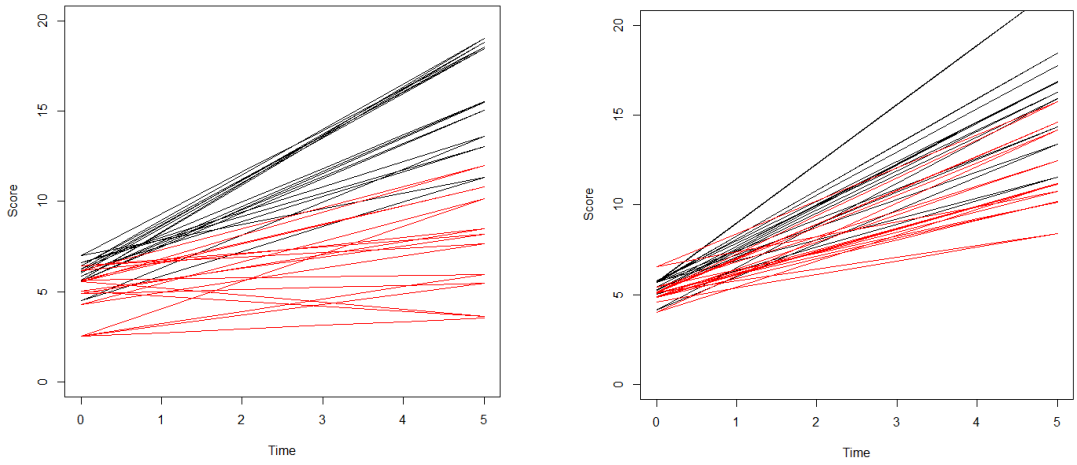

Figure 11. Examples of generated data.

For the first simulation study, the residual variance of the observed variables was held equal across classes in the data generation process. The magnitude of residual variance was selected specifically for each simulation condition to allow the intraclass correlation coefficient to be 0.45 for intercept and 0.15 for slope. For the second simulation study, the residual variance differed across classes.

Sample size is another factor that influences the estimation of mixture models. In the current study, the effect of the above mentioned factors on growth mixture analysis was evaluated under 3 choices of sample size: 200, 500 and 1000. Other simulation studies have incorporated sample sizes of these magnitudes (see, e.g., Nylund et al., 2007). Furthermore, the prevailing notion that mixture models do not operate well under smaller sample sizes has been amended to acknowledge that this conclusion could be mitigated by large class separation (see, e.g., Verbeke & Molenberghs, 2000).

Overall, the combination of all manipulated factors resulted in a Monte Carlo simulation with 540 cells. 20 replications were first run as a pilot study. One-hundred replications were generated within each design cell for full scale studies. Data used in the simulation were generated with R 2.14.1 (R Development Core Team, 2011) and estimated with M*plus* 6.2 (Muthén & Muthén, 2010). Details about parameters used for this simulation study are listed in Appendix A and sample M*plus* codes for estimating the growth mixture model are included in Appendix B.

3.3 Evaluation Criteria

The first step of evaluating a growth mixture model is to determine the number of latent classes in the data. As explained in Chapter 2, previous studies have shown that Bayesian information criterion (BIC) and sample size adjusted BIC (ABIC) performed better than other information criteria across a variety of modeling settings (Jedidi, Jagpal, & Desarbo, 1997; Nylund, Asparouhov & Muthén, 2007; Tofighi & Enders, 2006; Yang, 2006). In ABIC, the original sample size n was replace by $(n+2)/24$. Other studies also found that the Lo-Mendell-Rubin (LMR) test (Lo, Mendell, & Rubin, 2001) was effective in determining the number of correct classes (Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008). The current study used BIC, ABIC and LMR as criteria to select the model from fitting 1-, 2-, and 3-class growth mixture models. The true model is a 2-class model, i.e., the selection of 1- or 3-class model demonstrates under-extraction or over-extraction in model enumeration.

The next step is to evaluate parameter recovery under the proposed estimation scheme. The evaluation of parameter recovery only included those replications in

which the estimation converged without local maxima. The performance of model estimation was examined in terms of both estimation accuracy and estimation efficiency. Relative bias were used to assess the accuracy of parameter estimates over the 100 replications at various simulated conditions. They are computed by averaging each of the values over all parameter estimates across replications:

$$\hat{\theta}_{RB} = \frac{\sum_{r=1}^{R} (\hat{\theta}_r - \theta) \big/ \theta}{R}$$

where R is the total number of replications and $\hat{\theta}_r$ is the parameter estimate from a single replication sample and θ is the population parameter. In the above formula bias is divided by the true parameter, which implies that when the magnitude of true parameter is close to zero, the relative bias of parameter estimates could be artificially inflated. This issue did not affect the current study since no population parameters were set to be smaller than 0.2.

The efficiency of parameter estimates is measured as the standard deviation of the sample estimates from their average value, which is also known as the empirical standard error of estimates. The efficiency of parameter estimates is calculated as

$$SD(\hat{\theta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r - \frac{\sum_{r=1}^{R}\hat{\theta}_r}{R}\right)^2}\ .$$

The accuracy of the standard error estimates was evaluated by precision of estimates, which is defined as

$$\text{Precision } (\hat{\theta}) = \frac{SE(\hat{\theta})}{SD(\hat{\theta})},$$

where $SE(\hat{\theta}) = \sqrt{\frac{\sum_{r=1}^{R} se(\hat{\theta})_r^2}{R}}$ is the standard error of estimates $se(\hat{\theta})$ averaged across the 100 replications. If the estimated standard errors computed based on an approach are accurate, $SE(\hat{\theta})$ should be close to $SD(\hat{\theta})$ and the ratio close to 1 (Lee, Song, & Poon, 2004).

Entropy values were calculated for 2-class models, to quantify the uncertainty of classification of subjects into different subgroups. Entropy values range from 0 to 1, with 0 corresponding to assigning subjects completely randomly and 1 to a perfect certain classification (Celeux & Soromenho 1996). Another criterion of classification quality is the classification accuracy. The accuracy is evaluated by the proportion of subjects assigned to their true class according to the greatest posterior probability. In the current study the correct percentage of class membership assignment is calculated by averaging the correct classification rates of the two classes.

Finally, convergence rates have been recorded for each design cell. The impact of the manipulated factors was evaluated via factorial ANOVA to examine the effects of these factors under different simulation conditions. The model enumeration accuracy, classification quality as well as parameter estimation accuracy and efficiency were used as the dependent variables in separate ANOVAs and compared across different simulation conditions, sample size, mean structure separation, variance-covariance differences among subgroups, data overlap and mixture proportions. The interaction effect of these factors was also investigated.

3.4 Possible Problems in Simulation

Convergence and local maxima problems are regularly found in mixture model studies. Since the current study only examines parameter recovery in well-estimated cases, low convergence rates and high chance of local maxima will undermine the evaluation of parameter recovery and factorial ANOVA analyses of simulation results. The distribution of estimates from limited number of replications might not represent the true sampling distribution of population parameters. Unbalanced cell sizes within the factorial design may hinder the interpretation of ANOVA results. For the current simulations study, the pilot study provided preliminary information about difficulties in model estimation and certain simulation conditions were eliminated from full scale simulation due to high rates of non-convergence and local maxima. Cases with non-convergence and local maxima from conditions remaining in full scale simulation were excluded from final results and more replications were generated until the number of converged replications without local maxima reached 100. This process provided a balanced playing field to evaluate the simulation results systematically.

There are several possibilities for GMM to be identified as non-convergence in current studies. Naturally cases when maximum likelihood fails to find a solution to meet convergence criteria should be classified as not converged. It is also possible for results stemming from GMM analyses to have non-positive definite covariance structure for random effects as well as negative residual variances. These two situations are also considered as non-convergence in current study. As Wothke (1993) pointed out, many different situations can cause the violation of positive definiteness

and each situation requires different solutions to remove the possible cause. The reasons for nonpositive-definite covariance structures in GMM are most likely to be improper starting values and over-parameterization. Therefore, true parameters were used as starting values for the two-class GMM analysis and the weighted average of subgroup true parameters were used as starting values for one-class GMM analysis. For the three-class GMM analysis, there is no sensible way to assign appropriate starting values and the default starting values from M*plus* were used. The non-convergence rates have been documented and reported to provide some insights for practitioners.

Another possible problem that often interferes with simulation studies involving mixture models is label-switching. Label switching has been documented for mixture models when using MCMC estimation in a Bayesian framework. Since the current study uses maximum likelihood for estimating growth mixture models, the label-switching issues arising in a fully Bayesian analysis does not exist. However, as new research has pointed out (McLachlan & Peel, 2000; Tueller, Drotar & Lubke, 2011), the class labels are arbitrary in mixture models without previous knowledge of the subpopulations. In simulation studies, parameter estimates are aggregated over replications and from replication to replication the same classes may not be labeled the same. It is critical to avoid aggregating parameter estimates over mislabeled classes. The label-switching problem can be prevented by using true parameter values as the starting values, making model constraints or inspecting parameter estimates after estimation. Since two-class mixture models are the true model in current study,

inspecting parameters after estimation before aggregating estimates were used to ensure correct class labeling.

Chapter 4 Results

The current study explored many conditions involving differences of variance-covariance matrices among subpopulations of growth mixture data, which have not been evaluated by other studies. Due to the lack of guidance from the literature to inform the proposed simulation, an extensive preliminary pilot study was conducted to assist in selecting levels of the conditions for the current simulation. Some of results of the preliminary study, though based only on 20 replications of simulated data analyzed under 2-class GMM model, provided valuable insight for choosing levels of sample size and combination of mean structure differences (measured by SMD) and variance-covariance structure differences (measured by C_d). The preliminary results were also helpful in that they shed light on data analytic problems that researchers and practitioners alike may encounter when applying these methods in a substantive setting. In the remainder of the chapter, the preliminary study results will be discussed first followed by a discussion of the main simulation results.

4.1 Pilot Simulation Study Results

The main purpose of the pilot study was to investigate the convergence rates and frequency of local maxima in estimation. The assumption was that data generated from different combinations of simulation conditions would not have the same amount of difficulty in estimation. Some combinations of mean structure differences and variance-covariance structure differences in this simulation may result in data with a large degree of overlap between latent subpopulations. Analysis of data from these simulation conditions with medium to large sample size (N = 500 and N = 1000)

had large non-convergence rates. The pilot study results suggested that SMD of 0.5 and C_d of 0.2 and 0.4 had particularly large numbers of non-converged cases. The average non-convergence rate for cells from this combination of SMD and C_d was approximately 0.50. For those cells with C_d of 0.2, only 40% of the iterations converged. When SMD was equal to 1 and C_d was 0.2, the average convergence rate was also lower than 0.60. The 72 cells with SMD 0.5 or 1 and C_d of 0.2 or 0.4 (see Figure 12) encountered some level of estimation difficulty. Overall, 35 out of 72 cells in this combination have non-convergence rates larger than .40 and 3 of them had no converged cases at all. The pilot study results also showed that there were a large number of cases with possible local maxima for these cells. The average percentage of occurrence of local maxima was as high as 40%. Therefore, the combination of SMD and C_d as shown in Figure 12 was removed from full scale simulation.

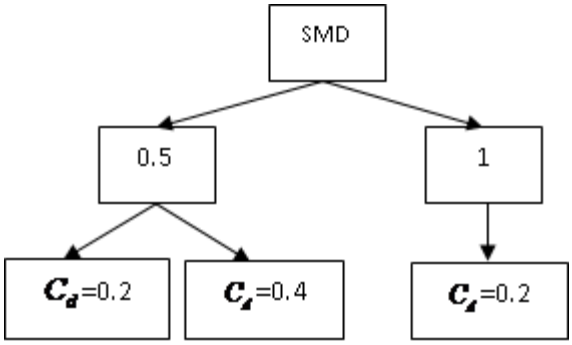

Figure 12. Combination of SMD and C_d which led to large data overlap.

In addition, the remaining cells with data of sample size 250 were also explored in the pilot study as it was thought based on the literature that computational

problems occurred with greater frequency when the sample size was small. The results suggested that a sample size of 250 might not be large enough to obtain stable parameter estimates for the majority of the conditions. When sample size was 250, across all other conditions, the average non-convergence rate was 0.42. Across mixing proportions, the average non-convergence rate of cells from different combinations of SMD and C_d are listed in Table 3. Forty-seven cells had a non-convergence rate higher than .40 and 6 of them had no converged replicates at all. The non-convergence rates when the mixing proportion was 0.9/0.1 (0.54) was much higher than when the mixing proportion was 0.5/0.5 (0.31) or 0.7/0.3 (0.32). The occurrence of solutions reaching local maxima was also more frequent under the smaller sample size condition than under the larger sample size condition. The average rate of local maxima was .15. Considering the high non-convergence rate as well as frequency of local maxima, it would appear difficult to obtain 100 converged replications for so many cells. Thus the full scale simulation will exclude the condition of small sample size, N =250.

Table 3

Non-Convergence Rates Across Levels of Latent Mean Differences (SMD) and Latent Variance-Covariance Differences (C_d) and Where the Sample Size N = 250

	C_d		
SMD	0.2	0.4	0.6
1	0.74	0.59	0.33
1.5	0.63	0.40	0.13
2	0.43	0.33	0.13
2.5	0.28	0.21	0.08

In summary, when either or both SMD and C_d were very small, the overall data overlap would be too large for the current model to be estimated without convergence or local maxima problems. Base on the pilot study results, the full scale simulation will no longer examine the combinations of SMD and C_d as shown in Figure 12. Smaller sample size like 250 also increased estimation difficulty and thus was be included in the final full scale simulation. After excluding the aforementioned simulation conditions, there were 228 simulation cells for the full scale simulation study and conditions of SMD will be nested within levels of C_d . The final decided conditions for the first simulation are listed in Table 3.

Table 4

Final Chosen Conditions for the First Simulation Study

Factor	Levels
Sample Size	500, 1000
Mixing Proportion	0..5/0.5, 0.7/0.3, 0.9/0.1
SMD (nested within C_d)	C_d =0.2, 1.5, 2, 2.5 C_d =0.4, 1, 1.5, 2, 2.5 C_d =0.6, 0.5, 1, 1.5, 2, 2.5
C_d (nested within Variance-Covariance Condition)	Variance Different, 0.2, 0.4, 0.6 Covariance Different, 0.2, 0.4
Mean Condition	Intercept Different, Slope Different
Variance-Covariance Condition	Variance Different, Covariance Different

4.2 Simulation Study-1 Results

Results of main simulation study are reported in two parts, Section 4.2 and Section 4.3. In Section 4.2.1, convergence rates and the chance of the occurrence of

local maxima are presented to provide a general picture of the efficacy of model estimation. Model enumeration results will then be introduced along with the performance of different model fit indices in Section 4.2.2. In Section 4.2.3, the results of parameter recovery in terms of relative bias, parameter estimation efficiency as well as the precision of standard error estimates will be discussed.

Effects of different factors that are of interests in the current simulation will be analyzed using a factorial ANOVA with a nested design. The criteria of judging the importance of an effect include a combination of statistical significance as measured by comparing the *p*-value to the significance level ($\alpha = 0.05$) and practical importance as measured by a variance accounted for effect size measure, η^2, with $\eta^2 > 0.06$. It has been recommended by scholars and researchers for over three decades that a measure of effect size should be used to interpret the results of hypothesis testing beyond a test of statistical significance (Cohen, 1988; Maxwell, 2000; Olejnik & Algina, 2000). Eta-squared, η^2, was chosen as a measure of effect size in the current study because of its additive property and comparability for the effects of different factors within the same study. Compared to another popularly used measure of effect size ω^2, η^2 is less sensitive to unequal sample size and heterogeneous variances which apply to the current study (Carroll & Nordholm, 1975). According to Cohen (1988), η^2 of 0.06 and 0.14 represent medium and large effect sizes for factorial ANOVA analysis, respectively. In the following sections, tables for the ANOVA results will only show those effects that meet these two criteria at the same time and omit the other effects that do not simultaneously meet these criteria.

4.2.1 Convergence and Local Maxima. Non-convergence has been a common problem when fitting growth mixture models or any general mixture model analysis. It is important to discuss the convergence rates of data estimation before making conclusions about parameter recovery, model enumeration or classification accuracy. As mentioned in previous sections, the criterion for a converged replication in current simulation study is that the estimation ended by meeting the desired convergence criterion as well as absence of non-positive definite variance-covariance estimates of random effects and residuals. The convergence rates of different simulation conditions are displayed in Table 5.

Table 5

2-Class Model Convergence Rates of Growth Mixture Model Estimation. Blank Cells Indicate Condition Combinations that were Omitted from the Main Simulation

		N=500			N=1000		
		C_d			C_d		
π_i	SMD	0.2	0.4	0.6	0.2	0.4	0.6
0.5	0.5			0.775			0.960
	1		0.865	0.885		0.988	0.970
	1.5	0.855	0.918	0.905	0.973	1.000	0.985
	2	0.935	0.970	0.940	0.998	1.000	0.995
	2.5	0.970	0.973	0.955	0.995	0.998	1.000
0.7	0.5			0.950			0.995
	1		0.845	0.985		0.985	1.000
	1.5	0.793	0.973	0.990	0.983	1.000	1.000
	2	0.930	0.988	1.000	0.990	0.998	1.000
	2.5	0.983	0.995	0.995	0.998	1.000	1.000
0.9	0.5			0.965			1.000
	1		0.545	0.985		0.735	1.000
	1.5	0.510	0.665	0.995	0.803	0.833	1.000
	2	0.633	0.723	1.000	0.888	0.908	1.000
	2.5	0.795	0.785	0.995	0.948	0.938	1.000

Recall that the population model used to generate data for the current simulation was a two-class growth mixture model as demonstrated by Equation 12 in section 3.2.1. To evaluate the accuracy of model enumeration, the generated data were estimated under 1-, 2- and 3-class growth mixture models. The convergence rates were high for estimating the 1-class growth mixture model. In this scenario, 100% of the replications across all simulation conditions had converged to a proper

solution. When fitting 2-class models, the convergence rates were also high for most of the cells. Across all conditions only 8% of the replications did not converge properly. Out of 228 full-scale simulation cells, eighty-nine of them had 100% convergence, fifty-five cells had convergence rates higher than 0.99 while only ten cells had convergence rates lower than 0.80. As could be foreseen, non-convergence increased when fitting 3-class GMMs. The average convergence rate for 3-class models was only 0.035 across all conditions. The convergence rate was slightly higher when the sample size was 1000. However, even in cells where conditions were deemed more ideal, the convergence rates were lower than 0.10. While somewhat disappointing, this result is understandable since the 3-class model was attempting to fit three variance-covariance matrices of the latent growth factors for data that were generated from a population model with only 2 classes. This "over-extracting" caused a large number of cases to converge to a solution where the variance-covariance matrix of random effects for at least one class was not positive-definite.

Sample size has been recognized as important factor in model convergence in previous studies (see e.g., Tolvanen, 2008). The current study also found similar results to that of Tolvanen. Of all replications using the 2-class GMM to fit the data, approximately 77% of non-converged replications had a sample size of 500 while only 24% of them had a sample size of 1000. The average convergence rate for cells with a sample size of 500 was 0.87 while the average convergence rate for cells with a sample size of 1000 was 0.96.

Convergence rates were also closely related to subpopulation overlap of the generated data. As described in section 2.6.1, the random effect overlap and overall

data overlap is the sum of misspecification probabilities of two subpopulations. The correlation between convergence rate and random effect distribution overlap was 0.52 and correlation between convergence rate and overall data overlap was 0.70. All of the conditions with convergence rates lower than 0.50 had overall data overlap larger than 0.50. Since the overlap of growth mixture data is determined by both the mean structure differences between subpopulations and variance-covariance structure differences, as expected, the convergence rates improved when SMD and/or C_d became larger. The convergence rates were similar when mixing proportions were 0.5/0.5 and 0.7/0.3 but lower when the mixing proportions were 0.9/0.1 as demonstrated in Table 6.

Table 6

Convergence Rate at Different Mixing Proportions

Mixing Proportion	Convergence Rate
0.5	0.93
0.7	0.94
0.9	0.73

As expected no local maxima problems were found for 1-class model estimation. For 2-class GMMs, the number of replications where the solutions reached local, not global, maxima was much lower than the number (rate) of non-converged replicates. The average percentage of model estimation with possible local maxima was only 2%. Detailed information about local maxima rates are shown in Table 7. The results indicated that data with unbalanced subpopulation sample sizes

were more likely to encounter local maxima problems. Increased sample size definitely decreased the number of local maxima. The average rate of local maxima for data with sample size of 500 was 0.035 while the rate of converging to a local maxima for data with a sample size of 1000 was only 0.008. As SMD and C_d increased, the number of replicates that converged to a local maxima decreased.

Table 7

Proportions of Replicates that Reached a Local Maxima in Fitting a 2-Class Growth Mixture Model. Blank Cells Indicate Condition Combinations that were Omitted from the Main Simulation.

		N=500			N=1000		
		C_d			C_d		
π_i	SMD	0.2	0.4	0.6	0.2	0.4	0.6
0.5	0.5			0.030			0.000
	1		0.010	0.000		0.000	0.000
	1.5	0.010	0.015	0.000	0.005	0.000	0.000
	2	0.010	0.000	0.000	0.000	0.000	0.000
	2.5	0.005	0.000	0.000	0.000	0.000	0.000
0.7	0.5			0.010			0.000
	1		0.040	0.000		0.005	0.000
	1.5	0.090	0.000	0.000	0.000	0.000	0.000
	2	0.015	0.000	0.000	0.000	0.000	0.000
	2.5	0.000	0.000	0.000	0.000	0.000	0.000
0.9	0.5			0.010			0.000
	1		0.285	0.010		0.105	0.000
	1.5	0.235	0.145	0.000	0.065	0.055	0.000
	2	0.115	0.130	0.000	0.020	0.035	0.000
	2.5	0.035	0.065	0.000	0.000	0.000	0.000

Because of the over-extraction problem, the number of solutions converging to a local maximum was much higher for 3-class model estimation. The average rate of local maxima across all conditions is 0.44. Increased sample size did not help reduce the number of local maxima as it did when fitting the 1- and 2-class GMMs. The number of solutions reaching local maxima was higher for replicates where the data were characterized by better class separation.

Replications that did not converge to a proper solution or those that reached local maxima or both were excluded from subsequent parameter recovery analysis. However, unlike the pilot study where convergence was much more problematic, additional replicates were generated and analyzed until the number of converged replications reached 100 for each simulation condition.

4.2.2 Identification of the Number of Latent Classes. One critical issue in GMM analysis is to decide the number of latent subpopulations in the data. This decision is typically made by fitting a GMM to the data with increasing number of latent classes; choosing the model with the best fit indicated by one of a number of model-fit indices. One research question of the current study was to examine the performance of several model fit indices in model enumeration of GMM. As defined in Section 2.5.3, the indices focused on here are BIC, ABIC and LMR, which all have been suggested to work well for mixture and latent class analyses in a series of previous methodological studies (Henson, Reise & Kim, 2007; Jedidi, Jagpal, & Desarbo, 1997; Nylund, Asparouhov & Muthén, 2007; Tofighi & Enders, 2008; Yang, 2006). The results of the current work suggested that both LMR and ABIC tend to over-extract the number of latent classes while BIC sometimes under-extracts

the number of latent classes. Overall, however, BIC had the highest rate of correct model enumeration (0.876) compared to ABIC (0.536) and LMR (0.532). The rate of over-extraction was not affected by differences in mixing proportions, levels of SMD or C_d. When sample size increased, this rate decreased, but not dramatically. When SMD and C_d increased (i.e., data were better separated), the rate of under-enumerating using BIC dropped significantly. Detailed information of correct class identification can be found in Table 8. Overall, BIC worked the best in detecting the correct number of latent classes of GMMs.

Table 8

Identification of Latent Classes Using ABIC, BIC and LMR

	Correct Identification	Over Extract	Under Extract
ABIC	0.536	0.450	0.014
BIC	0.876	0.002	0.122
LMR	0.532	0.424	0.044

4.2.3 Parameter Recovery. This section will initially discuss the factorial ANOVA results from analyzing outcome measures of relative bias, efficiency of parameter estimates and precision of standard error estimates. The results from the analysis will be used to inform and focus the discussion on only those condition combinations that demonstrated both statistical significance and practical importance. Bias is the difference between parameter estimates and population parameter values. Relative bias is bias divided by the true population parameter value. Compared to bias, relative bias provides a more relevant index that can be used as a basis of

comparison of estimates from true parameters when the true parameters are on different scales. In general, if the absolute value of relative bias is less than 0.10, the recovery of true parameter from the nominated model was considered to be acceptable. The efficiency of parameter estimates was measured as the standard deviation of the sample estimates from their average value, while the precision of standard error estimates was computed as the ratio of standard error estimates and efficiency of parameter estimates.

In Sections 4.2.3.1, 4.2.3.2 and 4.2.3.3, results of relative bias, efficiency of parameter estimates, and precision of standard error estimates will be discussed in detail. The factorial ANOVA results will be reported first to show the effects of different simulation factors on the outcome variables. In subsequent sections, details about relative bias, efficiency and precision under different simulation conditions will be presented. Only factors that showed significant effects on the outcome will be discussed.

4.2.3.1 Relative Bias of Parameter Estimates. Table 9 and Table 10 summarize the results of the ANOVA analysis on relative bias of parameter estimates where effects of manipulated factors that demonstrated simultaneous statistical significance (at $\alpha = .05$ level) and surpassed the $\eta^2 > 0.06$ threshold will be discussed. Table 9 shows the effects of different factors on relative bias of intercept, slope and mixing proportion estimates. For intercepts and slopes of two classes, the factor of variance-covariance condition nested within the levels of variance-covariance structure separation explains the largest proportion of variation of relative bias. The mean structure condition, in other words, whether the difference of mean

structure was on the intercept or slope, was another important factor that affected the relative bias of $\beta_1^{(1)}$, $\beta_1^{(2)}$ and $\beta_0^{(2)}$ but not $\beta_0^{(1)}$. In addition, there was a significant interaction effect of the mean structure condition and variance-covariance structure condition on relative bias of $\beta_0^{(1)}$, $\beta_0^{(2)}$ and $\beta_1^{(2)}$. The only factor that had significant influence on relative bias of mixing proportion π_1 was the mixing proportion condition itself which explained approximately 28.2% of the variation of relative bias of π_1.

Table 9

Factorial ANOVA Results on Relative Bias of Intercept, Slope and Mixing Proportion

Factors	$\beta_0^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$	π_1
Data Overlap					
π					28.2%
Sample Size					
C_d					
SMD(C_d)					
VarCond(C_d)	24.9%	19.0%	23.7%	11.9%	
MeanCond		14.9%	6.1%	13.4%	
VarCond×SMD (C_d)					
MeanCond×VarCond (C_d)	6.4%	14.4%		10.1%	

For the relative bias of variance-covariance structure estimates, the mean structure distance (SMD) nested within variance-covariance structure distance (C_d) had a significant effect on the relative bias of all variance-covariance components of the random effects except for class-2 slope variance, $\varphi_{11}^{(2)}$. Differences in the mean structure as well as their interaction with differences in variance-covariance structure

impacted the relative bias of class-2 intercept and slope variances but not class-1 estimates. Relative bias of residual variance estimates was affected by differences in the mixing proportion, C_d as well as variance-covariance structure condition. Intercept variance for class 2, $\varphi_{00}^{(2)}$, was the only parameter that was affected by the overall data overlap even though the effect size was barely above the evaluation criteria.

Table 10

Factorial ANOVA Results on Relative Bias of Variance-Covariance Estimates of the Random Effects

Factors	$\varphi_{00}^{(1)}$	$\varphi_{00}^{(2)}$	$\varphi_{11}^{(1)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(1)}$	$\varphi_{01}^{(2)}$	ε
Data Overlap		6.7%					
π					12.2%		13.2%
Sample Size						7.4%	
C_d	8.2%					7.7%	11.1%
SMD(C_d)	17.2%	8.3%	21.7%		16.5%	7.6%	
VarCond(C_d)						13.5%	6.2%
MeanCond		19.7%		16.2%			
VarCond×SMD (C_d)							
MeanCond×VarCond (C_d)		23.7%		20.3%			

The following illustrates the effects of different factors on relative bias of parameter estimates using graphical summaries and 5% and 95% quantiles. Figure 13 shows the variation of relative bias on intercept and slope estimates under different combinations of variance-covariance conditions and variance-covariance distance. When variances were the same and covariances were different across classes (represented by dash lines on the graph in Figure 13), relative bias was much smaller

than when covariances were different in the two classes than when variances were different. No significant difference was found for class-1 intercept and slope relative bias. In general, when variance-covariance matrices of the random effects of the two classes were more well-separated(i.e., C_d was larger), the relative bias of class-2 intercept and slope estimates was smaller. This difference was only manifest when variances were different acorss classes since when covariances were different across classes the relative bias of the intercept from both classes were very small and close to zero.

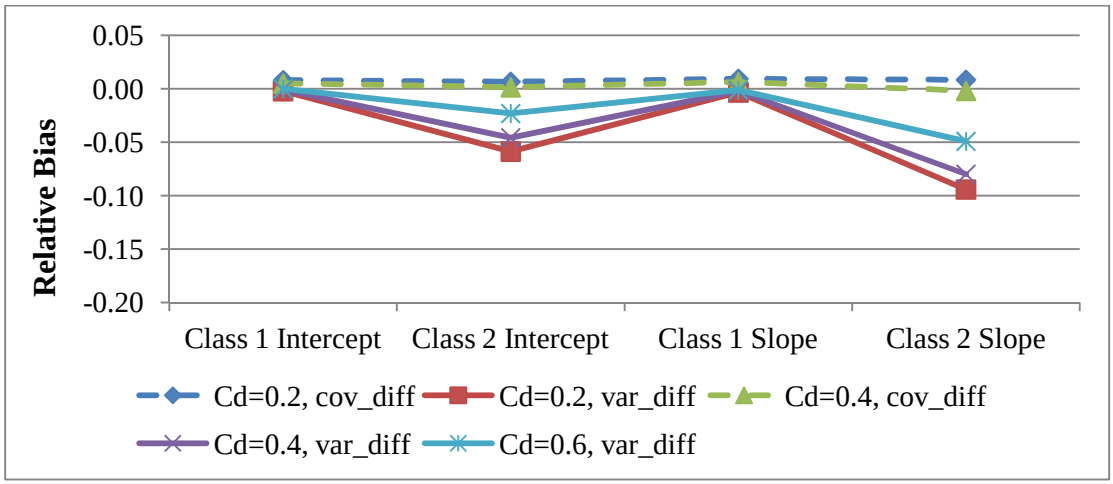

Figure 13. Relative bias of intercept and slope across different variance-covariance conditions.

Based on ANOVA results, the main effect of the mean structure condition significantly impacted the relative bias of the intercept and the slope. As demonstrated by Figure 14, differences in intercept or slope across classes led to larger relative bias of class-2 intercept or slope. To be specific, when intercepts were different acorss classes, which in current simulation design meant class-2 intercept

was smaller than class-1, the relative bias of class-2 intercept was larger than the class-1 intercept but no differences were found for slope estimates. In this situation, the model produced accurate (in terms of bias) estimates of the class-1 intercept and tended to underestimate the class-2 intercept. The same phenoma was found when examining the slope estimates. This pattern was also recognized across the mixing proportion conditions.

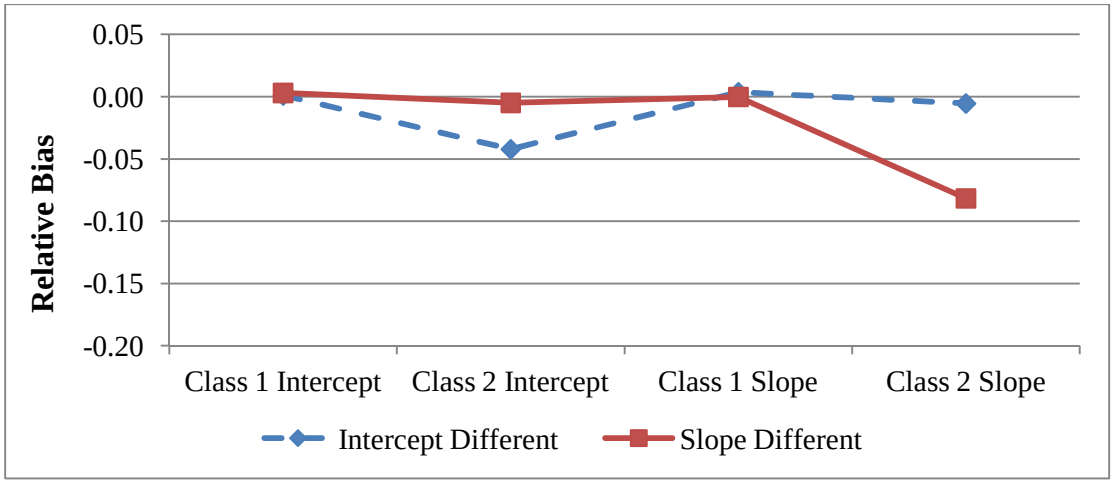

Figure 14. Relative bias of intercept and slope under different mean structure conditions.

The relative bias of mixing proportion estimates was only affected by the mixing proportion condition itself. Figure 15 shows that relative bias of the mixing proportion estimates was smaller when the percentage of subjects in the two classes were more similar (i.e., 0.50/0.50). Similarly to the bias of the intercept and slope estiamtes, the mixing proportions had larger bias especially when class sample sizes were not balanced. The second class proportion was constantly overestimated which in turn resulted in underestimation of the class-1 proportion.

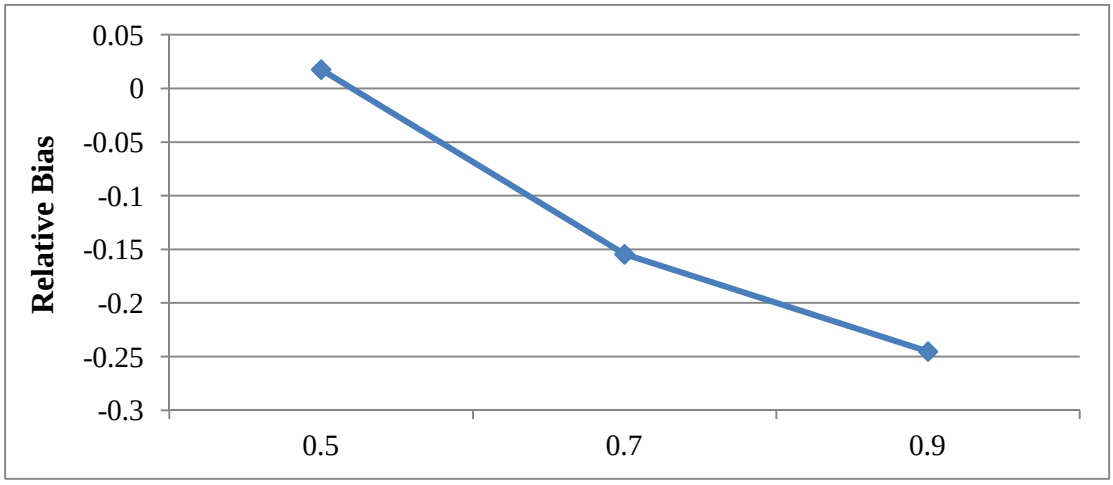

Figure 15. Relative bias of mixing proportion under different mixing proprotion conditions.

Figure 16 through Figure 18 demonstate relative bias of the random effects variances and covariances under different combinations of SMD and C_d. For the current simulation design, SMD was nested within C_d not crossed with C_d, which was due to the deletion of some combinations resulting in overly large data overlap and thus high non-convergence rates. Only SMD of 1.5, 2 and 2.5 were combined with all levels of C_d. Therefore, only these three levels of SMD were shown in the figures. As the results suggested, overall, the relative bias of the random effects variances and covariances decreased with increases across levels of SMD and C_d, especially for class-2 covariance estimates. In general, class-2 covariances had much larger and negative relative bias than the other random effects variance and covariances. For C_d = 0.60, the relative bias of class-2 covariance was much smaller especially when SMD was 1.5. However, we must keep in mind that there was only

variance difference in the random effects across two classes when C_d was 0.60. The model tended to underestimate the class-2 covariance, especially when data were more overlapped. To obtain estimates of the class-2 covariance with acceptable relative bias, SMD had to be larger than 1.5 or C_d had be to be larger than or equal to 0.40. The effects of SMD and C_d on other variance and covariance estimates were not as evident as the class-2 covariance.

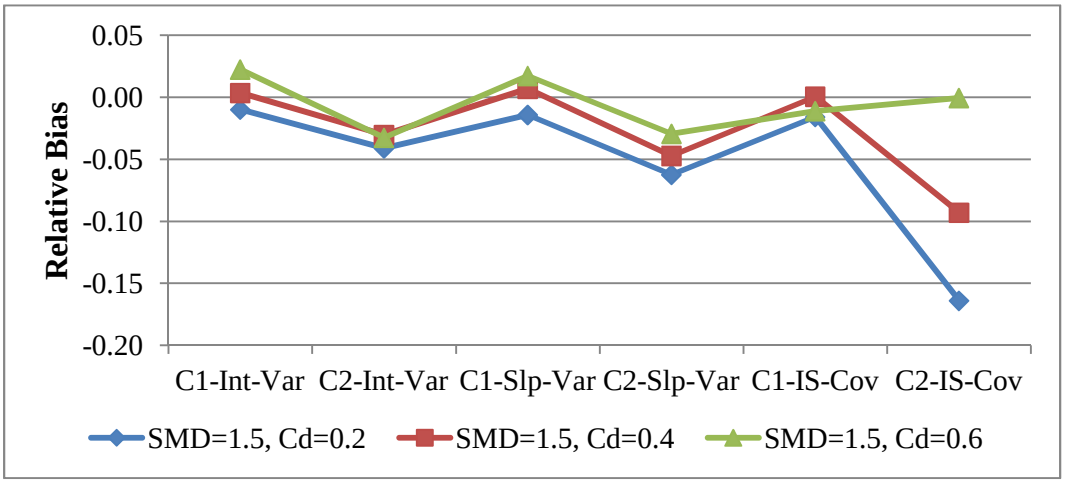

Figure 16. Relative bias of random effects variances and covariances when SMD = 1.5.

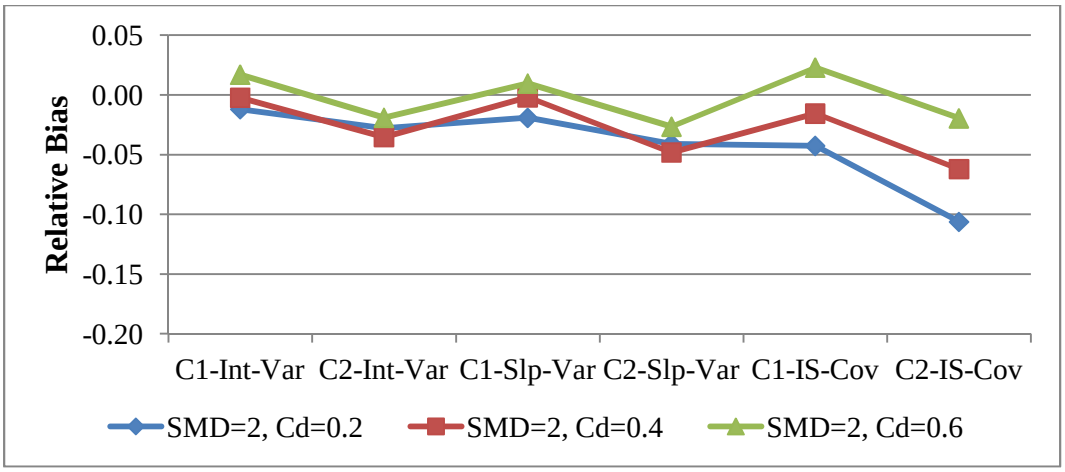

Figure 17. Relative bias of random effects variances and covariances when SMD=2.

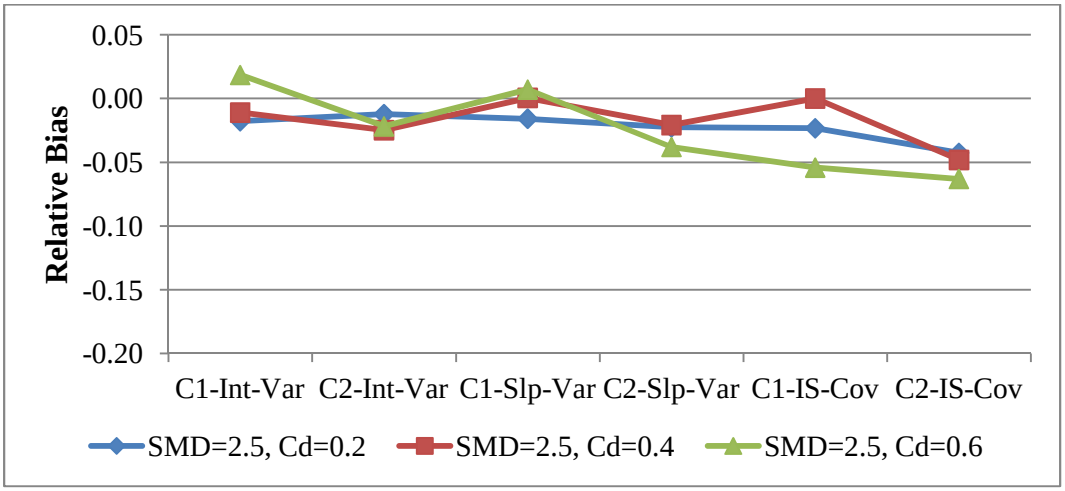

Figure 18. Relative bias of random effects variances and covariances when SMD=2.5.

The relative bias of the random effects variances and covariances under different combinations of the mean structure condition and the variance-covariance structure condition when $C_d = 0.20$ and $C_d = 0.40$ is shown in Figure 19 and Figure 20. The patterns of variation of relative bias when C_d is 0.20 or 0.40 were similar. When variances were different across classes, the effect of the mean structure

condition (intercept or slope different) on the random effect variance estimates were similar to the effect on mean structure estimates. When intercepts differed across classes, the class-2 intercept variance had larger relative bias; when slopes differed across classes, the class-2 slope variance had larger relative bias. However, when differences in the variance-covariance structure focused on covariance differences instead of variance differences, the effect of the mean structure was not apparent. In addition, under this situation, the relative bias of class-1 variances were larger than in the situation in which the variances were different.

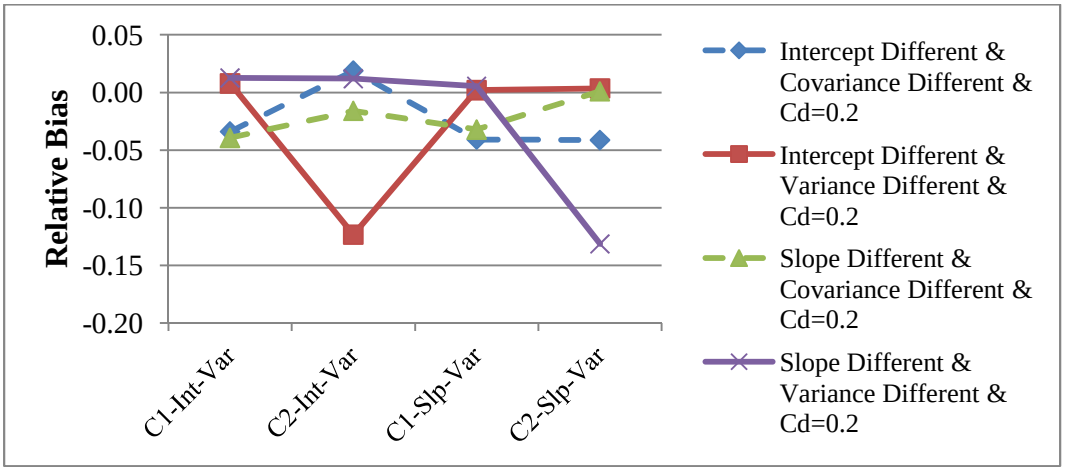

Figure 19. Relative bias of random effects variance under different combinations of mean structure and variance-covariance structure when C_d is 0.20.

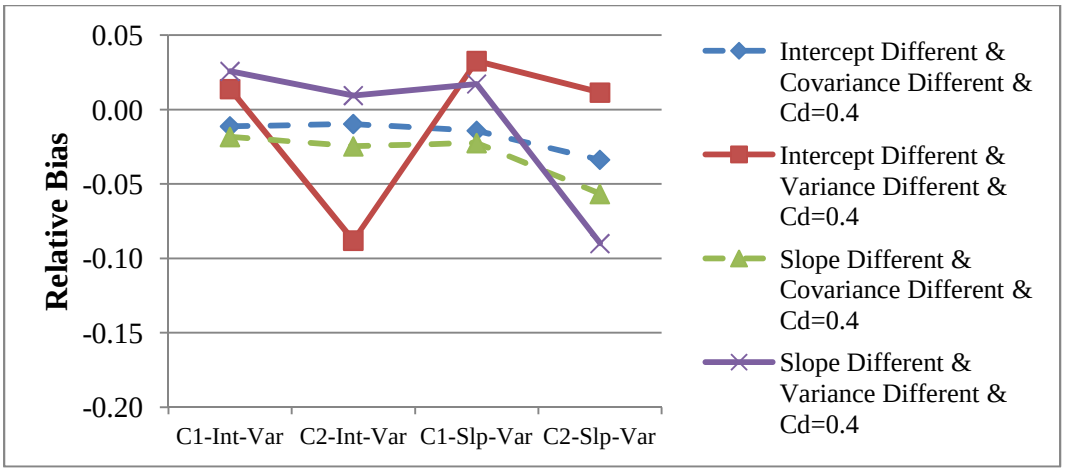

Figure 20. Relative bias of random effects variance under different combinations of mean structure and variance-covariance structure when C_d is 0.40.

For relative bias of the residual variance, even though the mixing proportion and C_d showed significant effects from the ANOVA analysis, the influence was not detectable in the graphical summaries shown in Figure 21 and Figure 22. Overall, the relative bias of the residual variance was quite small (close to zero).

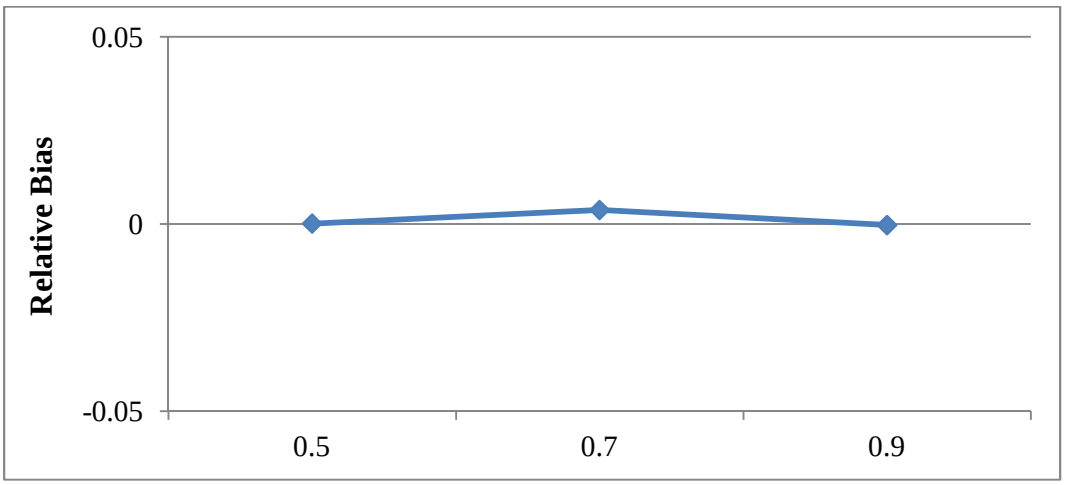

Figure 21. Relative bias of residual variance under different mixing proportion.

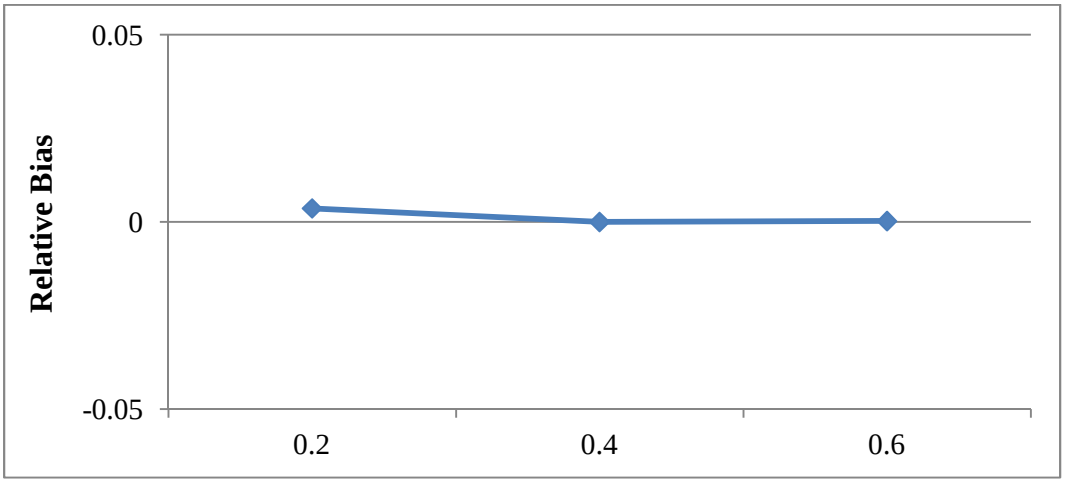

Figure 22. Relative bias of residual variance under different levels of C_d .

Table 11 shows the 5th and 95th percentile of relative bias of all parameters under different mixing proportions and sample size conditions. The range of relative bias was smaller when the sample size was larger (N = 1000). The range of relative bias was much larger when the mixing proportion was most disparate (i.e., 0.9/0.1) than when the mixing proportions were 0.5/0.5 and 0.7/0.3. Intervals capturing the range of relative bias of π_1 and the random effects variance and covariances were wider than that of relative bias of the mean structure estimates and residual variance.

Table 11

5th and 95th Percentile of Relative Bias Under Different Mixing Proportions and Sample Sizes

	0.5/0.5		0.7/0.3		0.9/0.1	
	N=500	N=1000	N=500	N=1000	N=500	N=1000
$\beta_0^{(1)}$	(-0.024, 0.014)	(-0.006, 0.008)	(-0.007, 0.013)	(-0.007, 0.008)	(-0.001, 0.018)	(-0.001, 0.008)
$\beta_0^{(2)}$	(-0.152, 0.005)	(-0.052, 0.005)	(-0.112, 0.021)	(-0.076, 0.013)	(-0.181, 0.049)	(-0.11, 0.017)
$\beta_1^{(1)}$	(-0.022, 0.017)	(-0.012, 0.009)	(-0.011, 0.014)	(-0.004, 0.008)	(-0.004, 0.026)	(-0.004, 0.014)
$\beta_1^{(2)}$	(-0.238, 0.024)	(-0.145, 0.146)	(-0.26, 0.019)	(-0.084, 0.015)	(-0.347, 0.037)	(-0.239, 0.026)
π_1	(-0.022, 0.094)	(-0.035, 0.04)	(-0.23, -0.069)	(-0.246, -0.086)	(-0.356, -0.147)	(-0.356, -0.121)
$\varphi_{00}^{(1)}$	(-0.072, 0.244)	(-0.054, 0.061)	(-0.081, 0.065)	(-0.06, 0.04)	(-0.056, 0.739)	(-0.023, 0.046)
$\varphi_{00}^{(2)}$	(-0.155, 0.053)	(-0.079, 0.028)	(-0.119, 0.066)	(-0.075, 0.007)	(-0.223, 0.129)	(-0.126, 0.077)
$\varphi_{11}^{(1)}$	(-0.066, 0.154)	(-0.037, 0.047)	(-0.067, 0.068)	(-0.042, 0.054)	(-0.069, 0.819)	(-0.043, 0.064)
$\varphi_{11}^{(2)}$	(-0.128, 0.047)	(-0.076, 0.035)	(-0.16, 0.061)	(-0.071, 0.038)	(-0.225, 0.056)	(-0.182, 0.052)
$\varphi_{01}^{(1)}$	(-0.219, 0.006)	(-0.155, 0.132)	(-0.118, 0.113)	(-0.117, 0.116)	(-0.07, 0.975)	(-0.04, 0.053)
$\varphi_{01}^{(2)}$	(-0.274, 0.179)	(-0.101, 0.042)	(-0.351, 0.087)	(-0.139, 0.075)	(-0.621, 0.122)	(-0.212, 0.053)
ε	(-0.008, 0.004)	(-0.003, 0.004)	(-0.006, 0.019)	(-0.003, 0.022)	(-0.006, 0.004)	(-0.003, 0.004)

Table 12 and Table 13 display the proprotions of cells with unacceptble relative bias of parameter estimates separated by different mixing proportion, SMD and C_d. Proportions larger than 0.30 are bolded in the table. In general, there were more cells with average relative bias of variances and covariances estiamtes greater than 0.10 than those with unacceptable relative bias of any mean structure estiamtes. For the mean structure estiamtes, no cells had unacceptable relative bias for class-1 parameter estimates. Among the 228 simulation cells, 100 cells had acceptable relative bias for all parameters estimates (except for the mixing proportion). Seventy-two of them were under sample size of 1000 and 55 of them had different covariances across classes. Sixty-five percent of cells with different covariances across classes

had acceptable relative bias of all parameters and only 31% of cells with different variances across classes had acceptable relative bias of all parameters.

Table 12

Percentage of Cells with Unacceptable Relative Bias of Parameter Estimates Under Different Simulation Conditions for Intercept, Slope and Mixing Proportion

Conditions	Level	$\beta_0^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$	π_1
Mixing Proportion	0.5	0.00	0.12	0.00	0.22	0.00
	0.7	0.00	0.05	0.00	0.13	**0.84**
	0.9	0.00	0.11	0.00	0.22	**0.99**
SMD	0.5	0.00	0.00	0.00	0.00	**0.67**
	1.0	0.00	0.06	0.00	0.11	**0.67**
	1.5	0.00	0.12	0.00	0.18	**0.67**
	2.0	0.00	0.12	0.00	0.25	**0.6**
	2.5	0.00	0.08	0.00	0.23	**0.52**
C_d	0.2	0.00	0.14	0.00	0.19	**0.61**
	0.4	0.00	0.08	0.00	0.18	**0.61**
	0.6	0.00	0.05	0.00	0.22	**0.60**

Table 13

Percentage of Cells with Unacceptable Relative Bias of Parameter Estimates Under Different Simulation Conditions for Variances and Covariances

Conditions	Level	$\varphi_{00}^{(1)}$	$\varphi_{00}^{(2)}$	$\varphi_{11}^{(1)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(1)}$	$\varphi_{01}^{(2)}$	ε
Mixing Proportion	0.5	0.12	0.08	0.08	0.08	0.25	0.21	0.00
	0.7	0	0.08	0.00	0.09	0.18	0.3	0.00
	0.9	0.05	0.22	0.07	0.3	0.05	**0.47**	0.00
SMD	0.5	**0.42**	0.08	**0.42**	0.17	**0.58**	0.25	0.00
	1.0	0.08	0.11	0.08	0.19	0.17	**0.31**	0.00
	1.5	0.02	0.15	0.03	0.18	0.12	**0.43**	0.00
	2.0	0.03	0.12	0.02	0.18	0.13	**0.33**	0.00
	2.5	0.03	0.13	0.00	0.08	0.15	0.25	0.00
C_d	0.2	0.03	0.19	0.00	0.17	0.11	**0.44**	0.00
	0.4	0.02	0.14	0.04	0.20	0.08	**0.35**	0.00
	0.6	0.15	0.03	0.12	0.08	**0.35**	0.15	0.00

4.2.3.2 Results of Effieciency of Parameter Estimates. Based on the results of factorial ANOVA analysis (see Table 14 and Table 15), efficiency of parameter estimates was significantly affected by sample size, especially for the class-1 intercept and slope estimates. Efficiency of the class-1 intercept and slope estimates was also affected by the mixing proportion condition and the distance of variance-covariance structure. The conditions of the variance-covariance structure and the mean structure had significant effects on the class-2 intercept and slope estimates but not the class-1 estimates. Efficiency of estimates of the mixing proportion was only affected by the levels of the mixing proportion itself.

Table 14

Factorial ANOVA Results on Efficiency of Intercept, Slope and Mixture Proportion Estimates

Factors	$\beta_0^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$	π_1
Data Overlap					
π	8.4%		7.7%		24.0%
Sample Size	21.0%	14.4%	16.9%	12.1%	
C_d	6.2%		9.7%		
SMD(C_d)					
VarCond(C_d)		11.5%	9.2%	12.4%	
MeanCond		12.3%		15.5%	
VarCond×SMD (C_d)					
MeanCond×VarCond (C_d)		6.8%			

The efficiency of the variance-covariance estimates was mostly impacted by the mixing proportion, sample size and SMD. The mixing proportion and C_d only affect the efficiency of the class-2 variance and covariance estimates. Overall data

overlap had some influence on class-2 variance estimates but not the covariance estiamtes. Efficiency of the residual variance was greatly affected by C_d and variance-covariance condition nested within C_d.

Table 15

Factorial ANOVA Results on Efficiency of Variance-Covariance Estimates

Factors	$\varphi_{00}^{(1)}$	$\varphi_{00}^{(2)}$	$\varphi_{11}^{(1)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(1)}$	$\varphi_{01}^{(2)}$	ε
Data Overlap		6.9%		9.2%			6.4%
π		25.9%		28.1%		24.0%	16.2%
Sample Size	7.4%	6.8%		6.2%	14.2%	10.1%	14.4%
C_d		11.9%		11.6%		9.2%	41.8%
SMD(C_d)	25.7%	9.0%	29.7%	10.9%	15.6%	6.4%	
VarCond(C_d)							26.3%
MeanCond							
VarCond×SMD (C_d)							
MeanCond×VarCond (C_d)							

Similarly to the relative bias of the mixing proportion estimates, the efficiency of π_1 was only affected by the mixing proportion itself. As shown in Figure 23, as the mixing proportions of the two classes become more unbalanced, the standard deviation of estimates becomes larger which means the efficiency of the parameter estimates decreases.

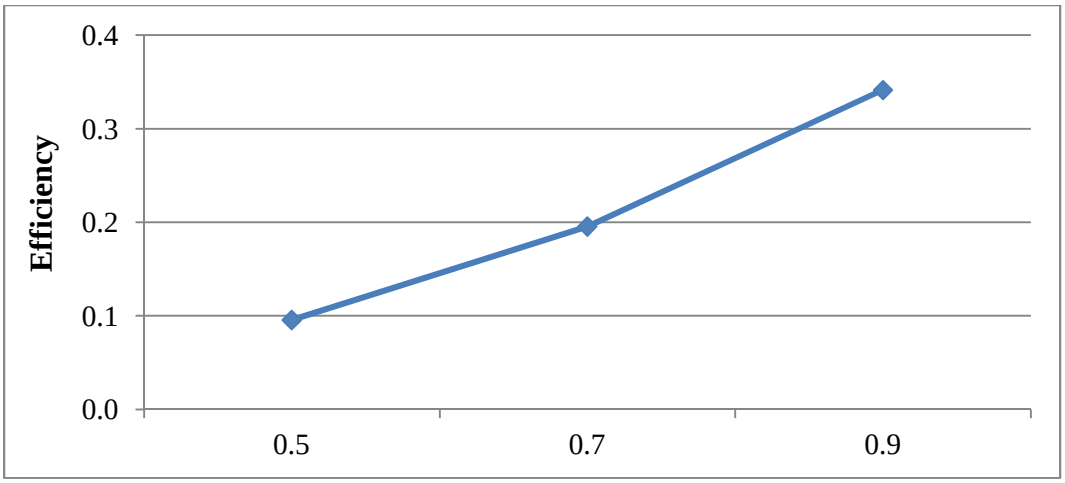

Figure 23. Efficiency of π_1 estimation under different mixing proportions.

For intercept and slope estimation, the standard deviation of class-2 estimates were larger than class-1 estimates as demostrated in Figure 24. In addition, an unsurprising result was that the larger sample size tended to lead to more efficient estimation of the mean structure parameters. Another important factor that impacted efficiency of the mean structure estimates was the variance-covariance condtions nested within different levels of C_d. In general, the standard deviation of the class-2 intercept and slope estimates were smaller when covariances were different across classes than when variances were different (see Figure 25 for detailed information). No significant impact was found for the class-1 estimates. Also, as C_d increased, the efficiency of the class-2 parameter estimates also increased. Mean structure differences also affected the efficiency of the class-2 intercept and slopes. When intercepts were different in the two classes, the efficiency of the intercept estimates was lower than when the slopes were different. When the slopes were different in the

two classes, the efficiency of the slope estimates was lower than when the intercepts were different. Please refer to Figure 26 for details.

Figure 24. Efficiency of intercept and slope estimation under different sample sizes.

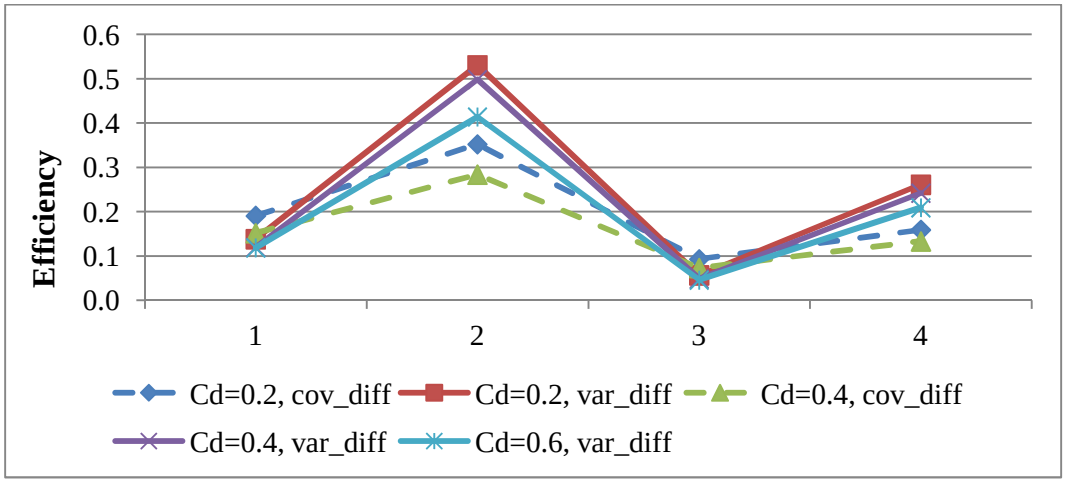

Figure 25. Efficiency of intercept and slope estimation under different variance-covariance conditions nested within C_d .

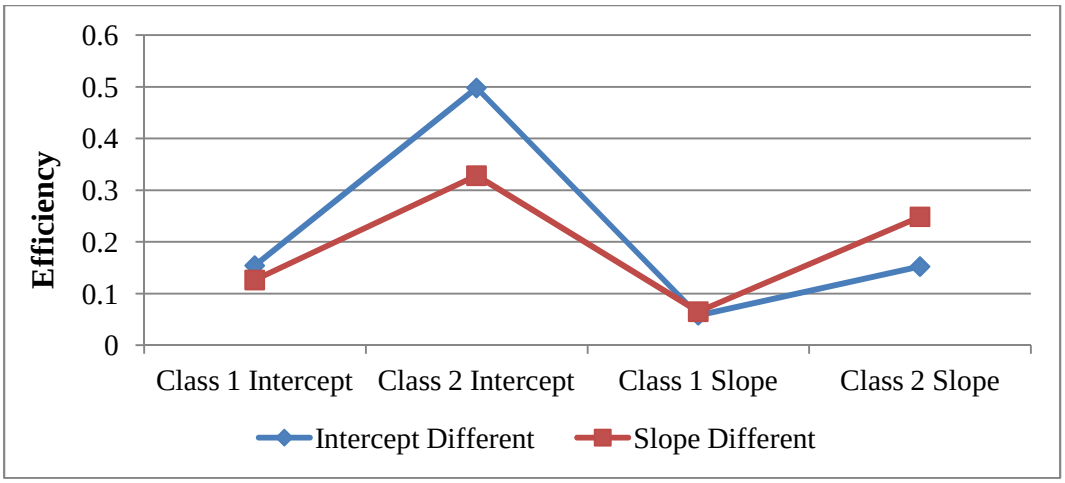

Figure 26. Efficiency of intercept and slope estimation under different mean structure condtions.

Due to the fact that the magnitude of the class-2 random effects variances changed significantly under different conditions of the mixture proportion and C_d, the standard deviations of the variance estimates were not comparable. However, the efficiency of residual variances, which were constrained to be the same across classes were on the same scale and comparable. As shown in Figure 27, as the level of C_d increased, the standard deviation of residual variance estimates increased as well.

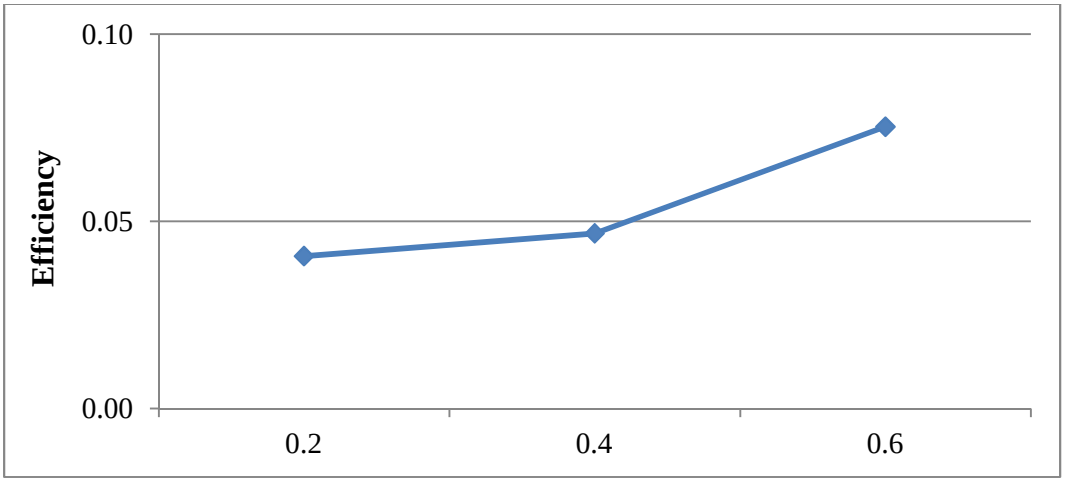

Figure 27. Efficiency of residual variance estimation under different C_d.

Table 16 shows the 5th and 95th percentile of efficiency of the mean structure parameters, the mixing proportions and residual variance under different mixing proportions and sample sizes. The range of the standard deviation of parameter estimates was smaller at the larger sample size level (N = 1000).

Table 16

5th and 95th Percentile of Efficiency Under Different Mixing Proportions and Sample Sizes

	0.5/0.5		0.7/0.3		0.9/0.1	
	N=500	N=1000	N=500	N=1000	N=500	N=1000
$\beta_0^{(1)}$	(0.141, 0.367)	(0.103, 0.237)	(0.107, 0.32)	(0.07, 0.174)	(0.073, 0.281)	(0.051, 0.107)
$\beta_0^{(2)}$	(0.169, 0.765)	(0.106, 0.461)	(0.246, 0.798)	(0.151, 0.567)	(0.428, 1.094)	(0.313, 0.827)
$\beta_1^{(1)}$	(0.061, 0.157)	(0.046, 0.109)	(0.043, 0.141)	(0.03, 0.079)	(0.03, 0.148)	(0.02, 0.113)
$\beta_1^{(2)}$	(0.08, 0.362)	(0.047, 0.291)	(0.11, 0.434)	(0.068, 0.307)	(0.186, 0.59)	(0.148, 0.414)
π_1	(0.057, 0.19)	(0.04, 0.133)	(0.14, 0.262)	(0.149, 0.23)	(0.266, 0.384)	(0.262, 0.403)
ε	(0.036, 0.133)	(0.024, 0.086)	(0.036, 0.096)	(0.024, 0.064)	(0.035, 0.067)	(0.024, 0.047)

4.2.3.3 Results of Precision of Standard Error Estimates. Compared to relative bias and efficiency of parameter estimates, precision of the standard error estimates was not as affected by the factors of interest in the current study. Table 17 and Table 18 summarize the factorial ANOVA results on precisions of standard error estimates. The effect sizes associated with the effects of factors on precision were only deemed of medium magnitude. Mean structure distance (SMD) nested within variance-covariance structure distance (C_d) had some effect on $\beta_0^{(2)}$, $\beta_1^{(1)}$, and the variance-covariance structure condition nested within C_d affected the efficiency of $\beta_1^{(2)}$. There was a significant interaction effect of the variance-covariance structure

condition and the mean structure distance (SMD) nested within C_d on $\beta_0^{(1)}$, $\beta_1^{(1)}$ and $\beta_1^{(2)}$.

Table 17

Factorial ANOVA Results on Precisions of Intercept, Slope Standard Error Estimates

Factors	$\beta_0^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$
Data Overlap				
π				
Sample Size				
C_d				
SMD(C_d)		7.0%	9.9%	
VarCond(C_d)				8.4%
MeanCond				
VarCond×SMD (C_d)	6.2%		9.4%	7.8%
MeanCond×VarCond (C_d)				

For the standard error estimates for the variance-covariance parameters, only precision of the class-1 variances was affected by the mixing proportion and mean structure distance nested within C_d, while precision of the class-1 covariance was affected by the interaction effect of the variance-covariance structure condition and the mean structure distance. No significant effects were found on precision of the class-2 variances and covariance standard error estimates or residual variance estimates.

Table 18

Factorial ANOVA Results on Precisions of Variance-Covariance Standard Error Estimates

Factors	$\varphi_{00}^{(1)}$	$\varphi_{00}^{(2)}$	$\varphi_{11}^{(1)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(1)}$	$\varphi_{01}^{(2)}$	ε
Data Overlap π	11.8%		6.3%				
Sample Size C_d							
SMD(C_d)	13.2%		13.9%				
VarCond(C_d)							
MeanCond VarCond×SMD (C_d)					6.8%		
MeanCond×VarCond (C_d)							

The precision of the standard error estimates, when close to one indcated, that the standard errors estimated by the proposed model reflected the variation in the population. Table 19 through Table 21 show 5th and 95th percentile of precision under different SMD, C_d levels, mixing proportions, and sample sizes. As shown in Table 21, the precision of standard error estiamtes was better (closer to 1) when sample sizes were large. Table 19 and Table 20 suggested that as SMD and C_d increased, precision tended toward 1. Even though several factors had significant effects on precision of the standard errors, the effect sizes were only moderate and no reasonable pattern were found when examining the relation between these factors and precision. Different parameters did not have the same level of precision on standard error estimates, however. Among all 228 simulation cells, only 60 of them had intercept and slope precision between 0.9 and 1.2. Among these cells with better

standard error precision, thirty cells belong to the medium high to high mean structure separation category, i.e., with SMD of 2 and 2.5. For SMD of 2.5. Half of these 60 cells had sample size of 1000 and the other half had smaple size of 500. Most of these cells with high precisionhad either high level of SMD or larger sample size or both. However, the distribution of C_d was not different among these cells.

Table 19

Precision of Standard Error Estimates Under Different SMD and C_d for Intercept and Slope

SMD	C_d	$\beta_0^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$
0.5	0.6	(0.795, 1.210)	(0.627, 1.081)	(0.899, 1.731)	(0.334, 1.218)
1	0.4	(0.525, 1.402)	(0.644, 1.132)	(0.446, 1.243)	(0.588, 1.373)
	0.6	(0.954, 1.215)	(0.691, 1.226)	(0.964, 1.168)	(0.631, 1.168)
1.5	0.2	(0.592, 1.417)	(0.648, 1.492)	(0.487, 1.38)	(0.612, 1.655)
	0.4	(0.855, 1.168)	(0.764, 1.931)	(0.683, 1.327)	(0.722, 1.352)
	0.6	(0.903, 1.334)	(0.766, 1.183)	(0.869, 1.223)	(0.722, 1.118)
2	0.2	(0.831, 1.282)	(0.876, 1.196)	(0.718, 1.411)	(0.744, 1.212)
	0.4	(0.91, 1.370)	(0.800, 1.207)	(0.906, 1.592)	(0.780, 1.245)
	0.6	(0.884, 1.189)	(0.869, 1.135)	(0.928, 1.174)	(0.761, 1.157)
2.5	0.2	(0.922, 1.183)	(0.793, 1.238)	(0.866, 1.216)	(0.778, 1.288)
	0.4	(0.949, 1.291)	(0.891, 1.284)	(0.888, 1.231)	(0.866, 1.295)
	0.6	(0.850, 1.214)	(0.846, 1.267)	(0.947, 1.163)	(0.817, 1.252)

Table 20

Precision of Standard Error Estimates Under Different SMD and C_d for Variances and Covariances

SMD	C_d	$\varphi_{00}^{(1)}$	$\varphi_{00}^{(2)}$	$\varphi_{11}^{(1)}$	$\varphi_{11}^{(2)}$	$\varphi_{01}^{(1)}$	$\varphi_{01}^{(2)}$	ε
0.5	0.6	(0.271, 1.823)	(0.652, 1.305)	(0.251, 1.246)	(0.668, 1.244)	(0.675, 1.392)	(0.66, 1.358)	(0.875, 1.232)
1	0.4	(0.576, 1.17)	(0.849, 1.417)	(0.476, 1.18)	(0.774, 1.248)	(0.794, 1.317)	(0.745, 1.309)	(0.871, 1.213)
	0.6	(0.377, 1.381)	(0.881, 1.191)	(0.415, 1.383)	(0.821, 1.214)	(0.931, 1.26)	(0.681, 1.247)	(0.938, 1.129)
1.5	0.2	(0.773, 1.666)	(0.827, 1.511)	(0.739, 1.245)	(0.789, 1.551)	(0.812, 1.179)	(0.767, 1.521)	(0.880, 1.098)
	0.4	(0.627, 1.192)	(0.857, 1.948)	(0.787, 1.168)	(0.849, 1.192)	(0.794, 1.181)	(0.816, 1.301)	(0.940, 1.096)
	0.6	(0.949, 1.166)	(0.859, 1.197)	(0.869, 1.246)	(0.745, 1.134)	(0.954, 1.134)	(0.845, 1.151)	(0.867, 1.075)
2	0.2	(0.831, 1.159)	(0.876, 1.31)	(0.823, 1.196)	(0.809, 1.304)	(0.819, 1.187)	(0.86, 1.262)	(0.956, 1.130)
	0.4	(0.934, 1.259)	(0.81, 1.263)	(0.920, 1.326)	(0.789, 1.273)	(0.973, 1.467)	(0.875, 1.231)	(0.925, 1.073)
	0.6	(0.83, 1.227)	(0.899, 1.232)	(0.904, 1.157)	(0.772, 1.182)	(0.951, 1.206)	(0.755, 1.137)	(0.883, 1.126)
2.5	0.2	(0.895, 1.154)	(0.836, 1.13)	(0.917, 1.151)	(0.821, 1.148)	(0.916, 1.147)	(0.848, 1.148)	(0.929, 1.113)
	0.4	(0.898, 1.268)	(0.898, 1.326)	(0.900, 1.332)	(0.858, 1.301)	(0.925, 1.13)	(0.933, 1.433)	(0.868, 1.139)
	0.6	(0.946, 1.315)	(0.854, 1.136)	(0.912, 1.291)	(0.836, 1.325)	(0.95, 1.133)	(0.917, 1.215)	(0.886, 1.117)

Table 21

Precision of Standard Error Estimates Under Different Mixing Proportions and Sample Sizes

	0.5/0.5		0.7/0.3		0.9/0.1	
	N=500	N=1000	N=500	N=1000	N=500	N=1000
$\beta_0^{(1)}$	(0.815, 1.273)	(0.91, 1.370)	(0.592, 1.447)	(0.832, 1.273)	(0.525, 1.245)	(0.762, 1.291)
$\beta_0^{(2)}$	(0.627, 1.267)	(0.78, 1.296)	(0.751, 1.275)	(0.78, 1.449)	(0.644, 1.78)	(0.672, 1.192)
$\beta_1^{(1)}$	(0.647, 1.411)	(0.863, 1.452)	(0.718, 1.195)	(0.899, 1.319)	(0.487, 1.731)	(0.368, 1.426)
$\beta_1^{(2)}$	(0.631, 1.321)	(0.713, 1.556)	(0.543, 1.285)	(0.628, 1.488)	(0.599, 1.548)	(0.612, 1.288)
$\varphi_{00}^{(1)}$	(0.851, 1.602)	(0.928, 1.823)	(0.773, 1.315)	(0.866, 1.202)	(0.396, 1.229)	(0.276, 1.183)
$\varphi_{00}^{(2)}$	(0.754, 1.31)	(0.881, 1.305)	(0.894, 1.195)	(0.846, 1.345)	(0.755, 4.963)	(0.849, 1.104)
$\varphi_{11}^{(1)}$	(0.778, 1.326)	(0.88, 1.332)	(0.645, 1.203)	(0.92, 1.308)	(0.419, 1.185)	(0.28, 1.19)
$\varphi_{11}^{(2)}$	(0.78, 1.275)	(0.886, 1.426)	(0.772, 1.325)	(0.783, 1.184)	(0.718, 1.551)	(0.821, 1.201)
$\varphi_{01}^{(1)}$	(0.839, 1.218)	(0.976, 1.392)	(0.819, 1.233)	(0.934, 1.187)	(0.794, 1.661)	(0.764, 1.212)
$\varphi_{01}^{(2)}$	(0.681, 1.215)	(0.865, 1.358)	(0.745, 1.284)	(0.849, 1.521)	(0.767, 1.73)	(0.831, 1.218)
ε	(0.867, 1.183)	(0.931, 1.129)	(0.875, 1.213)	(0.928, 1.117)	(0.88, 1.118)	(0.893, 1.158)

4.2.4 Classification Results. In this section, classification results of the GMM are provided. Two types of statistics will be applied to evaluate classification quality: entropy and classification accuracy. First, a factorial ANOVA model with nested design will be used to estimate the effect of different simulation factors on entropy and classification accuracy. Then results of these two statistics will be discussed in details, separately.

Table 22 shows how much of the variance of entropy and classification accuracy can be explained by each factor. Difference in the mixing proportions explains the majority of the variance (60.9%) of the entropy while mean structure differences nested within variance-covariance structure differences explaining the

largest proportion (59.9%) of the variance of classification accuracy. No interactions among the factors were found from the ANOVA analysis.

Table 22

Proportion of Variance Explained in Entropy and Classification Accuracy

Factors	Entropy	Classification Accuracy
Mixing Proportion	60.9%	23.9%
C_d		
SMD(C_d)	25.4%	59.9%
N		
Mean Condition		
Variance-Covariance Condition		

4.2.4.1Entropy. As an indicator of classification quality in mixture models, entropy is regularly used to evaluate the uncertainty in classifying subjects. Entropy, as defined in Chapter 3, values close to 1 suggest perfect classification and values around 0.8 are usually considered acceptable (Muthén et al., 2002). Across all simulation conditions, entropy values ranged from .227 to .791. As suggested by the factorial ANOVA results presented in Table 22, the mixing proportion condition had a significant effect on entropy. Data with larger differences in class proportion resulted in higher entropy values. Unsurprisingly, when subpopulations of data were more separated (larger SMD and C_d), the entropy values were higher as well (as shown in Figure 28 and 29). However, even for cells with the most optimistic condition combinations, the entropy values were barely acceptable. Data with larger sample size resulted in smaller entropy across all simulation conditions. Based on results from this simulation, the benchmark of 0.8 for entropy seems not realistic for

the proposed models. Entropy values larger than 0.6 were comparatively high for GMM models in current conditions. Detailed entropy information for different simulation conditions are listed in Table 23.

Table 23

Entropy under Different Simulation Conditions

π_i	SMD	C_d	Entropy	
			N=500	N=1000
0.5	0.5	0.6	0.296	0.227
	1	0.4	0.353	0.277
	1	0.6	0.340	0.291
	1.5	0.2	0.435	0.369
	1.5	0.4	0.418	0.368
	1.5	0.6	0.421	0.384
	2	0.2	0.503	0.463
	2	0.4	0.500	0.469
	2	0.6	0.521	0.503
	2.5	0.2	0.588	0.565
	2.5	0.4	0.596	0.576
	2.5	0.6	0.602	0.589
	0.5	0.6	0.345	0.330
	1	0.4	0.422	0.358
	1	0.6	0.421	0.384
	1.5	0.2	0.512	0.427
	1.5	0.4	0.477	0.434
	1.5	0.6	0.489	0.450
	2	0.2	0.531	0.504
	2	0.4	0.544	0.513
	2	0.6	0.567	0.528
	2.5	0.2	0.603	0.577
	2.5	0.4	0.618	0.598
	2.5	0.6	0.627	0.609
0.9	0.5	0.6	0.681	0.685
	1	0.4	0.645	0.633
	1	0.6	0.701	0.696
	1.5	0.2	0.662	0.655
	1.5	0.4	0.675	0.671
	1.5	0.6	0.727	0.729
	2	0.2	0.705	0.707
	2	0.4	0.712	0.716
	2	0.6	0.779	0.759
	2.5	0.2	0.747	0.744
	2.5	0.4	0.749	0.745
	2.5	0.6	0.791	0.785

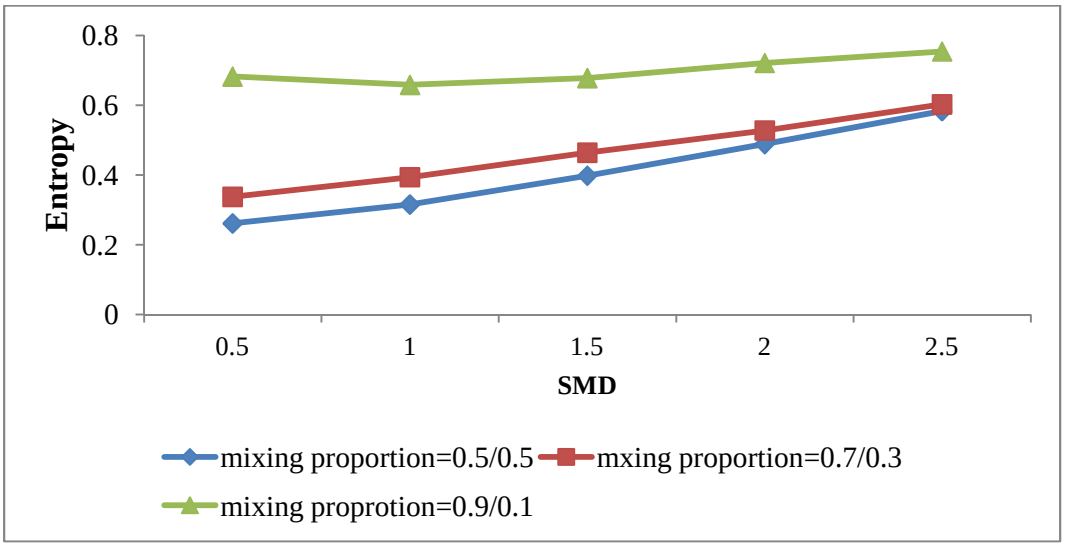

Figure 28. Entropy values under different SMD.

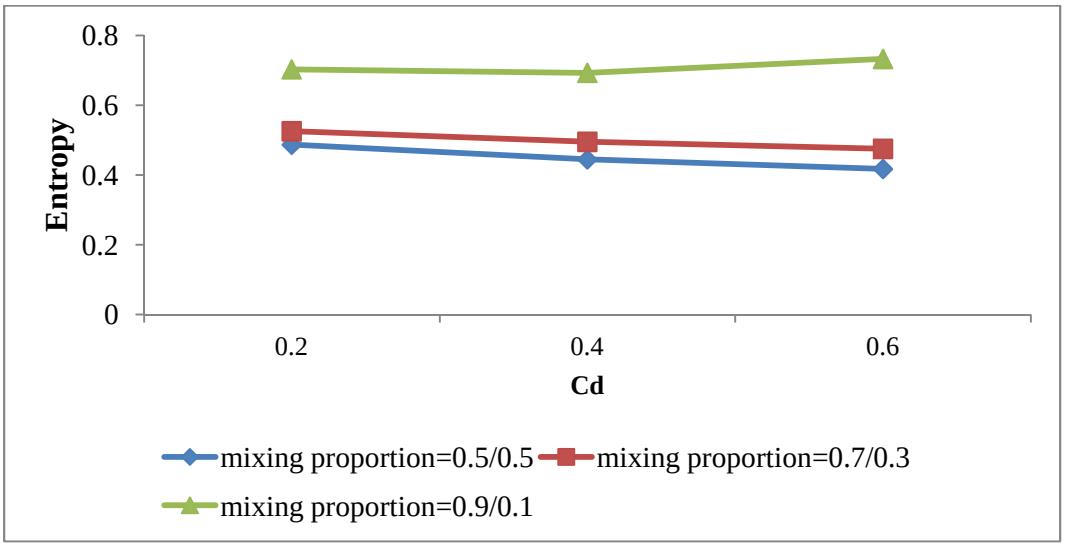

Figure 29. Entropy values under different C_d .

4.2.4.2 Classification Accuracy. Classification accuracy was defined in Chapter 3 as the percentage of correct assignment of subjects to the latent classes they

arose from. Across all simulation conditions, the percentage of correct assignment of class membership ranged from 0.625 to 0.875. The average classification accuracy was 0.742. Details about classification accuracy can be found in Table 24. As demonstrated in Figure 30, as SMD increased, classification accuracy tended to be higher across all other simulation conditions. Results also suggested that higher levels of C_d would also help improve correct assignment of class membership but the improvement was not as dramatic as the improvement caused by increased levels of SMD. Unlike entropy values, unbalanced sample size across subpopulations did not lead to better classification accuracy. In general, classification accuracy from data with mixing proportions of 0.9/0.1 and 0.7/0.3 was lower than those from 0.5/0.5, especially when SMD becomes larger. As shown in Figure 30 and Figure 31, increment of classification accuracy as increase of C_d was not as obvious as increment with SMD. Increasing the sample size from 500 to 1000 only improved classification accuracy slightly (0.739 for sample size 500 vs. 0.746 for sample size 1000). No significant sample size effect was found in the ANOVA analysis (see Section 4.2.4.1).

Table 24

Classification Accuracy across Different Simulation Conditions

			Classification Accuracy	
π_i	SMD	C_d	N=500	N=1000
0.5	0.5	0.6	0.701	0.720
	1	0.4	0.771	0.786
	1	0.6	0.836	0.847
	1.5	0.2	0.672	0.689
	1.5	0.4	0.742	0.757
	1.5	0.6	0.810	0.819
	2	0.2	0.648	0.637
	2	0.4	0.687	0.677
	2	0.6	0.736	0.729
	2.5	0.2	0.673	0.696
	2.5	0.4	0.739	0.751
	2.5	0.6	0.802	0.810
	0.5	0.6	0.855	0.861
	1	0.4	0.650	0.662
	1	0.6	0.706	0.714
	1.5	0.2	0.768	0.773
	1.5	0.4	0.823	0.828
	1.5	0.6	0.630	0.625
	2	0.2	0.662	0.651
	2	0.4	0.702	0.684
	2	0.6	0.730	0.727
	2.5	0.2	0.665	0.694
	2.5	0.4	0.720	0.734
	2.5	0.6	0.771	0.782
0.9	0.5	0.6	0.833	0.835
	1	0.4	0.871	0.875
	1	0.6	0.673	0.676
	1.5	0.2	0.702	0.708
	1.5	0.4	0.744	0.755
	1.5	0.6	0.791	0.800
	2	0.2	0.837	0.844
	2	0.4	0.652	0.689
	2	0.6	0.704	0.714
	2.5	0.2	0.731	0.735
	2.5	0.4	0.759	0.762
	2.5	0.6	0.792	0.796

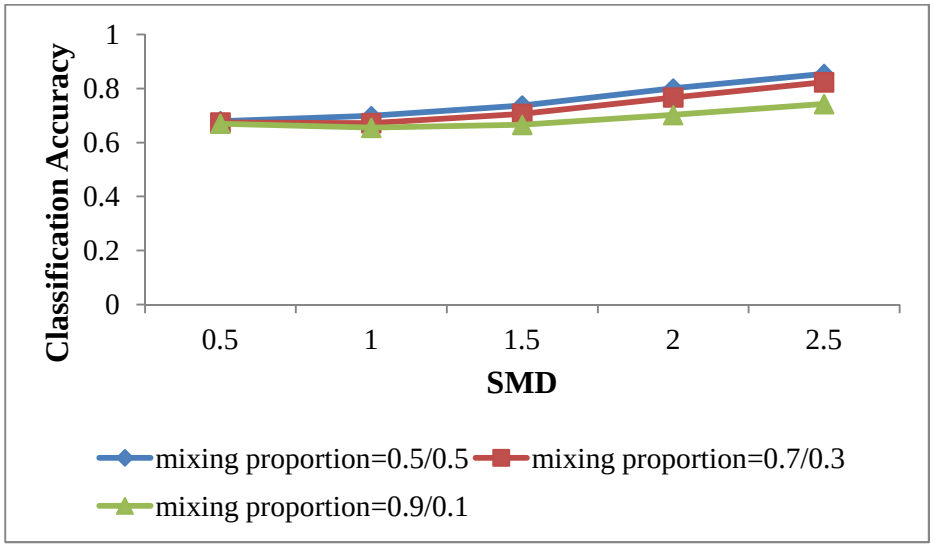

Figure 30. Classification accuracy under different levels of SMD and mixing proportions.

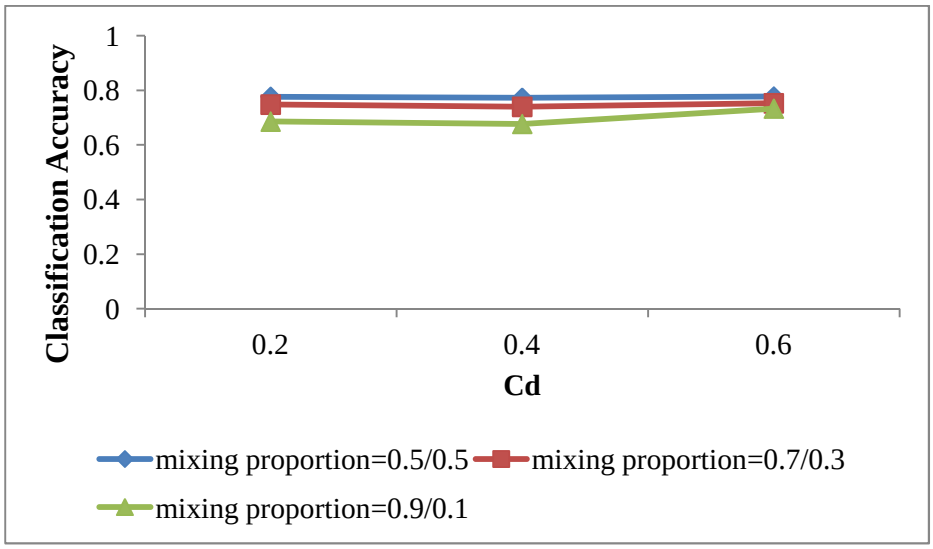

Figure 31. Classification accuracy under different levels of C_d and mixing proportions.

4.3 Simulation Study-2 Results

The purpose of the second simulation was to investigate the impact of residual variance on GMM estimation. The target simulation conditions were those that were not included in the first simulation study due to large overlap between data. The residual variances of these cells had been modified to examine whether class specific residual variances would enlarge the separation of data and improve convergence rates. Residual variance of the first class was designated to remain the same as in the first simulation while the residual variance of the second class was doubled. After adding differences in residual variances, the overall data overlap have been reduced significantly. The average overlap of data generated from these simulation conditions in the first simulation was 0.77. After residual variances were manipulated to be different across subpopulations, the average data overlap became 0.36. Please refer to Appendix B for parameters used in the second simulation study. The mean structure conditions for this simulation are 0.5 and 1 which indicate very small separation in mean structure across classes. The levels of variance-covariance distance include 0.2 and 0.4 as small to medium level of separation. Since the impact of mean structure condition had been evaluated in the first simulation study, this part of simulation will focus on the mean difference on intercept and keep the same slope across classes. In addition, the sample size for this simulation was fixed to be 500 since the first simulation had examined the effect of sample size.

4.3.1 Convergence and Local Maxima. The convergence rates for 1-class analysis were 100% for all 24 simulation cells. For two-class GMM analysis (true model in the current study), the convergence rates of the 24 simulation cells were all

within 98% to 100% except for the four cells with 0.9/0.1 mixing proportions as well as covariance differences in the two classes. Approximately 17% to 21% of the data in these four cells did not converge. No cases with possible local maxima were found for either 1-class or 2-class analysis.

Similar to the results in the first simulation, the convergence rates for 3-class analysis were extremely low with only a handful of converged cases across all 24 simulation cells. Furthermore, about 33% to 60% of the cases under these simulation conditions encountered local maxima problems.

4.3.2 Identification of Latent Classes. Similar to the results of first simulation study, BIC had the best performance in detecting correct number of classes. Decisions based on LMR were more likely to over-extract the number of latent classes. No significant relation was found between correct class enumeration rates and other simulation factors. Table 25 lists the class enumeration information of all three model fit indices. Since the overall data overlap of this simulation was on average higher than the first simulation, the correct model enumeration rates were higher for both ABIC and BIC. The over-extraction rate for ABIC dropped from 0.45 to 0.086 in the second simulation. The frequency of under-enumeration using BIC also dropped. No big difference was found for enumeration results from LMR.

Table 25

Identification of Latent Classes Using ABIC, BIC and LMR

	Correct Identification	Over Extract	Under Extract
ABIC	0.914	0.086	0.000
BIC	1.000	0.000	0.000
LMR	0.574	0.426	0.000

4.3.3. Parameter Recovery. Due to the simplicity of the design for this simulation, fewer factors will be evaluated on their effect on parameter recovery. ANOVA analysis results have shown only a couple of significant effects of any of the factors on relative bias, efficiency and precision of parameters based on the criteria used in the first simulation study. Therefore, this section will not discuss ANOVA results but present the results of parameter recovery in general.

In the second simulation, even though the distances between mean structures of the two classes were really small (0.5 and 1) and the separation of random effect variance-covariance structures was not large, the difference between residual variances between the two classes significantly reduced the overall data overlap. As a result, the relative bias of parameter estimates decreased accordingly. Out of 24 simulation cells, relative bias of all parameter estimates (except for mixing proportion estimates) from 16 of them were smaller than 0.1. Relative bias larger than 0.1 only occurred for variance and covariance estimates especially for those of the second class. In the first simulation, the relative bias for residual variances was small and not affected by any simulation conditions. In the second simulation, the relative bias for residual variances was larger when mixing proportions was more unbalanced especially for class-2 residual variance. Please see Table 26 for the 5^{th} and 95^{th} percentile of relative bias of parameter estimates.

Table 26

5th and 95th Percentile of Relative Bias under Different Levels of Mixing Proportions

	0.5/0.5	.07/.3	0.9/0.1
$\beta_0^{(1)}$	(-0.004, 0.003)	(-0.003, 0.002)	(0.000, 0.002)
$\beta_0^{(2)}$	(-0.019, 0.002)	(-0.006, 0.002)	(-0.013, 0.012)
$\beta_1^{(1)}$	(-0.003, 0.004)	(-0.002, 0.002)	(-0.002, 0.002)
$\beta_1^{(2)}$	(-0.004, 0.003)	(-0.004, 0.005)	(-0.006, 0.011)
π_1	(-0.017, 0.006)	(-0.342, -0.125)	(-0.401, -0.216)
$\varphi_{00}^{(1)}$	(-0.019, 0.03)	(-0.019, 0.035)	(-0.034, 0.009)
$\varphi_{00}^{(2)}$	(-0.086, 0.014)	(-0.047, 0.024)	(-0.065, 0.244)
$\varphi_{11}^{(1)}$	(-0.039, 0.021)	(-0.023, 0.017)	(-0.013, 0.011)
$\varphi_{11}^{(2)}$	(-0.055, 0.008)	(-0.053, 0.023)	(-0.103, 0.113)
$\varphi_{01}^{(1)}$	(-0.117, 0.111)	(-0.042, 0.102)	(-0.076, 0.049)
$\varphi_{01}^{(2)}$	(-0.094, 0.077)	(-0.099, 0.072)	(-0.078, 0.27)
$\varepsilon^{(1)}$	(-0.004, 0.012)	(-0.007, 0.006)	(-0.012, -0.002)
$\varepsilon^{(2)}$	(-0.003, 0.007)	(-0.012, 0.008)	(-0.041, 0.005)

The efficiency of parameter estimates at different mixing proportions are listed in Table 27. Similar to the first simulation, the standard deviation of parameter estimates were higher for class-2 parameter estimates than class-1 estimates. Since the variances parameters of the second class were much larger than the first class, the standard deviations of parameter estimates were not comparable and thus not listed in the table.

Table 27

5th and 95th Percentile of Standard Deviation of Parameter Estimates under Different Levels of Mixing Proportions

	0.5/0.5	.07/.3	0.9/0.1
$\beta_0^{(1)}$	(0.093, 0.129)	(0.070, 0.088)	(0.059, 0.074)
$\beta_0^{(2)}$	(0.123, 0.216)	(0.151, 0.268)	(0.257, 0.508)
$\beta_1^{(1)}$	(0.036, 0.052)	(0.029, 0.041)	(0.025, 0.03)
$\beta_1^{(2)}$	(0.05, 0.091)	(0.057, 0.101)	(0.107, 0.215)
π_1	(0.028, 0.042)	(0.164, 0.196)	(0.341, 0.396)

The precision of standard error estimates were better for all parameters in the second simulation with class specific residual variances. The ranges of precision were narrower and the values were closer to 1. The precision of standard error was more stable across all simulation conditions than that of the first simulation. Table 28 presents the 5th and 95th percentile of precision of standard error estimates under different mixing proportion conditions.

Table 28

5th and 95th Percentile of Precision of Standard Error Estimates under Different Levels of Mixing Proportions

	0.5/0.5	.07/.3	0.9/0.1
$\beta_0^{(1)}$	(0.856, 1.232)	(0.977, 1.221)	(0.909, 1.129)
$\beta_0^{(2)}$	(0.91, 1.146)	(0.959, 1.474)	(0.952, 1.217)
$\beta_1^{(1)}$	(0.928, 1.166)	(0.919, 1.15)	(0.919, 1.169)
$\beta_1^{(2)}$	(0.938, 1.052)	(0.932, 1.194)	(0.957, 1.12)
$\varphi_{00}^{(1)}$	(0.944, 1.093)	(0.926, 1.124)	(0.91, 1.236)
$\varphi_{00}^{(2)}$	(0.956, 1.167)	(0.981, 1.168)	(0.993, 1.266)
$\varphi_{11}^{(1)}$	(0.952, 1.215)	(0.89, 1.084)	(0.97, 1.173)
$\varphi_{11}^{(2)}$	(0.952, 1.133)	(0.946, 1.174)	(0.894, 1.161)
$\varphi_{01}^{(1)}$	(0.924, 1.112)	(0.946, 1.152)	(0.901, 1.161)
$\varphi_{01}^{(2)}$	(0.921, 1.22)	(0.886, 1.194)	(0.962, 1.29)
$\varepsilon^{(1)}$	(0.962, 1.183)	(0.995, 1.135)	(0.899, 1.197)
$\varepsilon^{(2)}$	(0.898, 1.147)	(0.936, 1.221)	(0.991, 1.156)

4.3.4 Classification Results. This section presents the classification results of the second simulation. The average entropy value for this simulation across all conditions was 0.660. The range of entropy was from 0.525 to 0.827. ANOVA indicated that the mixing proportion was the only factor that significantly affected the entropy values by explaining about 96.6% of the variation. As shown in Table 29, entropy values were higher when one class has much larger sample size than the other class.

Table 29

Entropy under Different Mixing Proportions

π_i	Entropy
0.5	0.560
0.7	0.621
0.9	0.799

Across all simulation conditions, the percentage of correct assignment of class membership ranged from 0.760 to 0.879 with an average classification accuracy of 0.834. As suggested by factorial ANOVA results (see Table 30), two factors that impacted the accuracy of class membership assignment were the mixing proportions and distances of mean structure between classes. Details about classification accuracy can be found in Table 30.

Table 30

Proportion of Variance Explained in Entropy and Classification Accuracy

Factors	Entropy	Classification Accuracy
Mixing Proportion	96.6%	71.31%
C_d		
SMD(C_d)		13.95%
Variance-Covariance Condition		

Table 31

Classification Accuracy across Different Simulation Conditions

π_i	SMD	Classification Accuracy
0.5	0.5	0.850
	1	0.866
0.7	0.5	0.845
	1	0.856
0.9	0.5	0.768
	1	0.817

Chapter 5 Discussion

Despite of the fast development of growth mixture model in the past twenty years, the influence of variance-covariance structures on growth mixture analysis have not been examined systematically. The focus of current study was the performance of growth mixture models with not only class-specific mean growth trajectories but also class-specific variance-covariance structures. The aim of this dissertation was to investigate how different conditions of variance-covariance of random effects and residuals affect the estimation of GMM with or without interaction with other factors like mean structure conditions, mixing proportion and sample size. Two simulation studies were conducted to evaluate the impact of random effects variance-covariance (between-subject variation) and residual variance (within-subject variation) separately. In both simulations, the performance of the linear growth mixture model under a variety of simulation conditions was assessed in terms of the model enumeration, membership classification as well as parameter recovery. In this chapter, major findings from the two simulations will be outlined and discussed, recommendations for researchers and practitioners will be addressed and limitations of current study as well as suggestions for future research will be presented.

5.1 Summary of Findings

5.1.1 Convergence Rates and Local Maxima. As shown in both of the simulation studies, convergence rates and the possibility of local maxima in GMM estimation were closely related to global overlap between subpopulation data distribution which is determined by mean structure separation (SMD) and variance-

covariance structure separation (C_d). Data with more unbalanced subpopulation sample sizes were more likely to encounter estimations problems like non-positive variance estimates and local maxima. The possible reason was that not enough information were given from the smaller size classes for the model to extract two classes from the population. Previous studies (Nylund, Asparouhov & Muthén, 2007 & Tofighi & Enders, 2008) suggested that over-extraction or in other word over-parameterization often causes model non-convergence. The 3-class GMM estimation encountered much more non-convergence and local maxima solutions than the 2-class or 1-class estimation. Increasing sample size and data separation definitely reduced the chance of these estimation problems. It is easier for the model to detect two classes when the subpopulations are further apart and there are enough data to provide information for each of the class.

5.1.2 Model Enumeration. The results of current simulation studies suggested that BIC again had the highest rates of correct model enumeration (0.876) compared to ABIC (0.536) and LMR (0.532). ABIC and LMR often over-extracted the number of latent classes while BIC sometimes led to model under-enumeration. The ABIC's adjustment on sample sizes seemed not suit the models in current study. Increasing the sample size and class separation (SMD and/or C_d) help lower the rates of under-enumerating using BIC and over-enumerating using ABIC.

5.1.3 Parameter Recovery. The relative bias values of most parameters in GMM of current study were acceptable for data generated from conditions of more than half of the cells in the two simulations studies. In general mean structure parameters, i.e. intercept and slope estimates, had smaller relative bias than variance

covariance estimates. Mixing proportion had the largest relative bias among all parameters. The first simulation results suggested that relative bias of mean structure parameters were affected mainly by variance-covariance structure condition, mean structure conditions as well as the their interaction. When the focus of variance-covariance difference between subpopulations was on covariances, the relative bias of both class-1 and class-2 intercept and slope were small. When variances of two subpopulations are different, however, the relative bias of class-2 intercept and slope were much larger than class-1 intercept and slope. The relative bias of variance-covariance parameters, on the other hand, were affected by the level of mean structure difference in two subpopulations, mean structure conditions as well as the interaction between mean structure condition and level of variance-covariance structure separation. The impact of variance-covariance condition on variances estimates of random effects is similar to that on mean structure estimates. The only factor that explained the variation of relative bias of mixing proportion was the mixing proportion itself. Larger differences in subpopulation sample sizes led to larger relative bias of mixing proportion estimates.

The first simulation results showed that sample size significantly affected the efficiency of most parameter estimates. Larger sample size would reduce the standard deviation of estimates significantly. Mean structure condition and variance-covariance condition also affect class-2 intercept and slope estimates but not class-1 estimates. In general, estimates of class-2 intercept and slope had lower efficiency but the efficiency improved when covariances instead of variances were different in subpopulations. The larger the difference between subpopulation sample sizes, the

larger the variation of mixing proportion estimates. Residual variance estimates had higher efficiency when variance-covariance structures of random effect in two subpopulations were less separated.

The precision of standard error for mean structure estimates was not affected by the simulation factors as much as the relative bias and efficiency. Only standard error of class-1 variances of intercept and slope were influenced by mixing proportion and level of mean structure separation between subpopulations. Overall, the precision of standard error estimates was not satisfactory for most simulation conditions. Only less than one third of the simulation cells had acceptable intercept and slope standard error precision. The level of mean structure separation played important role in precision. Conditions with medium high to high SMD or large sample size were easier to obtain more precise standard error estimates. No significant impact of C_d was found on precision of any parameter.

The second simulation study did not find any specific simulation factor with significant and systematic impact on any parameter recovery criteria. The relative bias of parameters and precision of standard errors were in general, acceptable. Class-specific residual variances reduced the overall data overlap significantly, which led to better model estimation results.

5.1.4 Classification Results. Results of entropy and classification rates were similar for the two simulations. Mixing proportion explained the majority of the variation of entropy and classification accuracy across different simulation conditions. The levels of mean structure separation in two subpopulations affected entropy and classification accuracy in the first simulation but only classification

accuracy in the second simulation. Entropy values were higher but classification accuracy was lower when mixing proportions were more unbalanced. Both statistics increased when SMD was higher.

5.2 Discussion

As expected prior to conducting the current study the performance of the proposed model was better when the overlap of the generated data is smaller, i.e., the distributions of the subpopulations were more separated. The possibility of non-convergence which mostly was caused by non-positive variance estimates and local maxima was lower when data were less overlapped. The overall data overlap is a result of class specific mean structures and variance-covariance structures but neither convergence rates nor local maxima were affected by where the differences of mean structure or variance-covariance structure were. It was also as expected that increasing sample size reduced model estimation difficulties especially the occurrences of local maxima. The results suggested that the most likely explanation for non-convergence problem may be over-extraction of parameters in GMMs. When the data overlapped too much, it was more difficult to extract two sets of mean and variance-covariance parameters for the subpopulations. In sum, the smoothness of the estimation of proposed model was more dependent on overall data separation and sample size which are two key factors for the model to mathematically build two classes based on the data, and less related to how the data were separated.

The impact of variance-covariance structures on parameter recovery was one of the major purposes of current study. The results suggested that not only the magnitude of how different the variance-covariance structures in two subpopulations

affect the accuracy of parameter estimates but also where the difference was located, i.e., variances differences or covariance differences. The average generated data overlap for cells with variances different across subpopulations was similar to those with covariances different across populations (0.37 vs. 0.35). However, the relative bias of class-2 parameter estimates from "variances different" conditions were larger than those from "covariances different" conditions. The reason of this scenario may that the large variation of parameter of the second class undermines the estimation accuracy.

The most difficult parameter to estimate for the proposed model was the proportion of sample size of each subpopulation. This parameter had the largest relative bias and big variation among iterations especially when sample sizes were quite different between two subpopulations. A possible explanation may be that the model may have more misclassification of subjects of class 1 to class 2 when class 2 has really small sample size, which in turn affects the estimation of mixing proportion. The random effects variance-covariance structures also had larger relative bias. The model slightly overestimated the first class intercept and slope variances but underestimated the second class variances. The current simulation design set the second class variances to be larger than the first class but the second class mean structure to be smaller than the first class. Apparently the model tended to magnify the differences on mean structures between two subpopulations but shrink the differences on variance-covariance structures.

Residual variances estimates were in general more accurate than the random effects variances and covariances. The current simulation also found many fewer

cases with negative residual variance estimates than non-positive definite variances of random effects. In both simulation studies, the residual variance has smaller relative bias and higher precision of standard errors. In the second simulation, two subpopulations were assigned different residual variances, which resulted in less overlap between subpopulation data. Even with very small mean structure and random effects variance-covariance structure differences, the convergence rates and parameter recovery have been largely improved. The relative bias of residual variances did not increase very much in the second simulation.

5.3 Recommendations

The current study extends the traditional focus of growth trajectories difference on GMM to variances and covariances among subpopulations. The results demonstrated that even when the mean growth trajectories are not much different between two subpopulations; it is still possible to discover latent classes among subjects based on differences on variance-covariance structures of subpopulations.

Data convergence and local maximum can be challenging for GMM estimation when data of two subpopulations are too overlapped especially when one subpopulation has much smaller sample size. It is important for practitioners and researchers to visually explore the data first and obtain some ideas about whether data overlap is small enough for GMM to detect two subpopulations. Not only the average growth trajectories of two subpopulations but also their variance and covariances should be evaluated. Not many methods for exploration of GMM data are available in literature. Researchers and scholars who are interested in GMM may extent the methods used for regression diagnostic analysis and build up a complete tool for

growth mixture data examination. The reason of non-convergence can be small sample size or large overlap. Current simulation results suggest that for sample size as small as 250, even data with medium level of separation often cannot converge.

The estimation of GMM with class specific random effects variance-covariance structures and residual variances provides a way to study subpopulation growth differences more thoroughly. The relative bias of intercept and slope parameters as well as residual variances was in general acceptable. However, researchers should be cautious about estimates of random effects variances and covariances especially when one class has really small sample size. If one class has really large variances of intercept or slope, the bias of its intercept or slope should also be larger. In this situation, the estimated class differences in growth trajectory might be larger population differences while the variances of growth trajectory may be underestimated. Sample size of 500 seems large enough to provide valid estimates. If data can be properly converged, increasing sample size does not seem to improve estimation accuracy.

Researchers may need to pay extra attention when they wish to assign subjects to different classes based on the model estimates. The entropy values were not very satisfying when the two classes were not well separated especially when mixing proportion was 0.5/0.5. The classification accuracy is also moderate. When mixing proportion is 0.7/0.3 or 0.9/0.1, the model had the tendency to assign more subjects to the smaller size class than the true sample size. Both entropy values and classification accuracy were not affected by variance-covariance structure differences very much but significantly influenced by the differences of mean growth trajectories. Therefore,

when two latent classes do not vary much in terms of mean growth trajectory, making inferences about subjects being in a particular latent class is not recommended. The current study used posterior probability to assign subjects to different latent classes. Researchers can explore other possible methods for membership assignment to see if they can improve classification accuracy.

5.4 Limitations of Current Study and Implications for Future Studies

The current study explored different variance-covariance structures on GMM which have not been studied systematically in previous studies. The research design intended to discover how these variance-covariance structures might affect the estimation of GMM models. Due to the lack of literature in qualifying the differences among variance matrices in two subpopulations, the current study modified Maitra and Melnykov (2010)'s index to generate a new index for measuring distance between variance matrices, C_d. The calculation of this index, taking into account of mixing proportions, may confound the mixing proportion factor in analyzing the simulation results.

Differences in variance-covariance structures can vary in a number of ways that may affect the estimation of model distinctively. There are a limited number of simulation conditions that can be accommodated in the current study within certain amount of time and only two possible patterns of differences have been examined in the simulation. The results suggested whether differences were on variances or covariance indeed led to different parameter estimation accuracy. Future studies can expand the scope on variance-covariance differences and evaluate their impact on GMM estimation.

Another limitation of this research is that the GMM model applied in the simulations was simplified to include only time as predictor and no covariates were incorporated. Previous study of Lubke and Muthén (2007) suggested that inclusion of covariates in a growth mixture model may help reduce the possibility of non-convergence and improve classification accuracy.

Appendix A

Data Generation Parameters for Simulation 1

mix proportion	Mean Condition	Variance Condition	Class 1		Class 2		Class 1		Class 2		Class 1	Class 2	SMD	C_d	Resid-var	Data Overlap
			i	s	i	s	i-var	s-var	i-var	s-var	is-cov	Is-cov				
0.5	i_diff	var_diff	5	2	0.72	2	1	0.25	5.1	1.275	0.2	0.2	2.5	0.6	3.7	0.27
0.5	s_diff	var_diff	5	2	5	-0.14	1	0.25	5.1	1.275	0.2	0.2	2.5	0.6	3.7	0.22
0.5	i_diff	var_diff	5	2	1.41	2	1	0.25	5.1	1.275	0.2	0.2	2	0.6	3.7	0.33
0.5	s_diff	var_diff	5	2	5	0.28	1	0.25	5.1	1.275	0.2	0.2	2	0.6	3.7	0.30
0.5	i_diff	var_diff	5	2	2.43	2	1	0.25	5.1	1.275	0.2	0.2	1.5	0.6	3.7	0.43
0.5	s_diff	var_diff	5	2	5	0.71	1	0.25	5.1	1.275	0.2	0.2	1.5	0.6	3.7	0.40
0.5	i_diff	var_diff	5	2	3.28	2	1	0.25	5.1	1.275	0.2	0.2	1	0.6	3.7	0.52
0.5	s_diff	var_diff	5	2	5	1.14	1	0.25	5.1	1.275	0.2	0.2	1	0.6	3.7	0.50
0.5	i_diff	var_diff	5	2	4.14	2	1	0.25	5.1	1.275	0.2	0.2	0.5	0.6	3.7	0.58
0.5	s_diff	var_diff	5	2	5	1.57	1	0.25	5.1	1.275	0.2	0.2	0.5	0.6	3.7	0.58
0.5	i_diff	var_diff	5	2	3.49	2	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.9	0.57
0.5	s_diff	var_diff	5	2	5	1.24	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.9	0.55
0.5	i_diff	var_diff	5	2	2.74	2	1	0.25	3.7	0.925	0.2	0.2	1.5	0.4	2.9	0.47
0.5	s_diff	var_diff	5	2	5	0.87	1	0.25	3.7	0.925	0.2	0.2	1.5	0.4	2.9	0.44
0.5	i_diff	var_diff	5	2	1.99	2	1	0.25	3.7	0.925	0.2	0.2	2	0.4	2.9	0.38
0.5	s_diff	var_diff	5	2	5	0.49	1	0.25	3.7	0.925	0.2	0.2	2	0.4	2.9	0.33
0.5	i_diff	var_diff	5	2	1.24	2	1	0.25	3.7	0.925	0.2	0.2	2.5	0.4	2.9	0.29
0.5	s_diff	var_diff	5	2	5	0.12	1	0.25	3.7	0.925	0.2	0.2	2.5	0.4	2.9	0.24
0.5	i_diff	cov_diff	5	2	4	2	1	0.25	1	0.25	0.55	-0.6	1	0.4	1.2	0.60
0.5	s_diff	cov_diff	5	2	5	1.5	1	0.25	1	0.25	0.55	-0.6	1	0.4	1.2	0.57
0.5	i_diff	cov_diff	5	2	3.5	2	1	0.25	1	0.25	0.55	-0.6	1.5	0.4	1.2	0.50
0.5	s_diff	cov_diff	5	2	5	1.25	1	0.25	1	0.25	0.55	-0.6	1.5	0.4	1.2	0.46
0.5	i_diff	cov_diff	5	2	3	2	1	0.25	1	0.25	0.55	-0.6	2	0.4	1.2	0.40
0.5	s_diff	cov_diff	5	2	5	1	1	0.25	1	0.25	0.55	-0.6	2	0.4	1.2	0.34
0.5	i_diff	cov_diff	5	2	2.5	2	1	0.25	1	0.25	0.55	-0.6	2.5	0.4	1.2	0.30
0.5	s_diff	cov_diff	5	2	5	0.75	1	0.25	1	0.25	0.55	-0.6	2.5	0.4	1.2	0.24
0.5	i_diff	var_diff	5	2	3.05	2	1	0.25	2.5	0.625	0.2	0.2	1.5	0.2	2.1	0.51
0.5	s_diff	var_diff	5	2	5	1.02	1	0.25	2.5	0.625	0.2	0.2	1.5	0.2	2.1	0.47
0.5	i_diff	var_diff	5	2	2.4	2	1	0.25	2.5	0.625	0.2	0.2	2	0.2	2.1	0.40
0.5	s_diff	var_diff	5	2	5	0.7	1	0.25	2.5	0.625	0.2	0.2	2	0.2	2.1	0.35
0.5	i_diff	var_diff	5	2	1.75	2	1	0.25	2.5	0.625	0.2	0.2	2.5	0.2	2.1	0.31
0.5	s_diff	var_diff	5	2	5	0.37	1	0.25	2.5	0.625	0.2	0.2	2.5	0.2	2.1	0.25
0.5	i_diff	cov_diff	5	2	3.5	2	1	0.25	1	0.25	0.46	-0.4	1.5	0.2	1.2	0.54
0.5	s_diff	cov_diff	5	2	5	1.25	1	0.25	1	0.25	0.46	-0.4	1.5	0.2	1.2	0.49
0.5	i_diff	cov_diff	5	2	3	2	1	0.25	1	0.25	0.46	-0.4	2	0.2	1.2	0.42
0.5	s_diff	cov_diff	5	2	5	1	1	0.25	1	0.25	0.46	-0.4	2	0.2	1.2	0.36
0.5	i_diff	cov_diff	5	2	2.5	2	1	0.25	1	0.25	0.46	-0.4	2.5	0.2	1.2	0.31
0.5	s_diff	cov_diff	5	2	5	0.75	1	0.25	1	0.25	0.46	-0.4	2.5	0.2	1.2	0.25
0.7	i_diff	var_diff	5	2	4.27	2	1	0.25	4.95	1.2375	0.2	0.2	0.5	0.6	2.7	0.63
0.7	s_diff	var_diff	5	2	5	1.63	1	0.25	4.95	1.2375	0.2	0.2	0.5	0.6	2.7	0.62
0.7	i_diff	var_diff	5	2	3.55	2	1	0.25	4.95	1.2375	0.2	0.2	1	0.6	2.7	0.57
0.7	s_diff	var_diff	5	2	5	1.27	1	0.25	4.95	1.2375	0.2	0.2	1	0.6	2.7	0.56
0.7	i_diff	var_diff	5	2	2.82	2	1	0.25	4.95	1.2375	0.2	0.2	1.5	0.6	2.7	0.49
0.7	s_diff	var_diff	5	2	5	0.91	1	0.25	4.95	1.2375	0.2	0.2	1.5	0.6	2.7	0.47
0.7	i_diff	var_diff	5	2	2.1	2	1	0.25	4.95	1.2375	0.2	0.2	2	0.6	2.7	0.41
0.7	s_diff	var_diff	5	2	5	0.55	1	0.25	4.95	1.2375	0.2	0.2	2	0.6	2.7	0.38
0.7	i_diff	var_diff	5	2	1.38	2	1	0.25	4.95	1.2375	0.2	0.2	2.5	0.6	2.7	0.33
0.7	s_diff	var_diff	5	2	5	0.19	1	0.25	4.95	1.2375	0.2	0.2	2.5	0.6	2.7	0.29
0.7	i_diff	var_diff	5	2	3.68	2	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.2	0.63
0.7	s_diff	var_diff	5	2	5	1.34	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.2	0.61
0.7	i_diff	var_diff	5	2	3.02	2	1	0.25	3.7	0.925	0.2	0.2	1.5	0.4	2.2	0.54
0.7	s_diff	var_diff	5	2	5	1.01	1	0.25	3.7	0.925	0.2	0.2	1.5	0.4	2.2	0.51
0.7	i_diff	var_diff	5	2	2.36	2	1	0.25	3.7	0.925	0.2	0.2	2	0.4	2.2	0.44
0.7	s_diff	var_diff	5	2	5	0.68	1	0.25	3.7	0.925	0.2	0.2	2	0.4	2.2	0.40

0.7	i_diff	var_diff	5	2	1.7	2	1	0.25	3.7	0.925	0.2	0.2	2.5	0.4	2.2	0.35
0.7	s_diff	var_diff	5	2	5	0.35	1	0.25	3.7	0.925	0.2	0.2	2.5	0.4	2.2	0.30
0.7	i_diff	cov_diff	5	2	4.03	2	1	0.25	1	0.25	0.62	-0.6	1	0.4	1.2	0.73
0.7	s_diff	cov_diff	5	2	5	1.51	1	0.25	1	0.25	0.62	-0.6	1	0.4	1.2	0.70
0.7	i_diff	cov_diff	5	2	3.54	2	1	0.25	1	0.25	0.62	-0.6	1.5	0.4	1.2	0.62
0.7	s_diff	cov_diff	5	2	5	1.27	1	0.25	1	0.25	0.62	-0.6	1.5	0.4	1.2	0.57
0.7	i_diff	cov_diff	5	2	3.06	2	1	0.25	1	0.25	0.62	-0.6	2	0.4	1.2	0.50
0.7	s_diff	cov_diff	5	2	5	1.03	1	0.25	1	0.25	0.62	-0.6	2	0.4	1.2	0.43
0.7	i_diff	cov_diff	5	2	2.58	2	1	0.25	1	0.25	0.62	-0.6	2.5	0.4	1.2	0.38
0.7	s_diff	cov_diff	5	2	5	0.79	1	0.25	1	0.25	0.62	-0.6	2.5	0.4	1.2	0.31
0.7	i_diff	var_diff	5	2	3.21	2	1	0.25	2.6	0.65	0.2	0.2	1.5	0.2	1.8	0.59
0.7	s_diff	var_diff	5	2	5	1.1	1	0.25	2.6	0.65	0.2	0.2	1.5	0.2	1.8	0.54
0.7	i_diff	var_diff	5	2	2.61	2	1	0.25	2.6	0.65	0.2	0.2	2	0.2	1.8	0.47
0.7	s_diff	var_diff	5	2	5	0.8	1	0.25	2.6	0.65	0.2	0.2	2	0.2	1.8	0.42
0.7	i_diff	var_diff	5	2	2.02	2	1	0.25	2.6	0.65	0.2	0.2	2.5	0.2	1.8	0.37
0.7	s_diff	var_diff	5	2	5	0.51	1	0.25	2.6	0.65	0.2	0.2	2.5	0.2	1.8	0.32
0.7	i_diff	cov_diff	5	2	3.54	2	1	0.25	1	0.25	0.51	-0.4	1.5	0.2	1.2	0.65
0.7	s_diff	cov_diff	5	2	5	1.27	1	0.25	1	0.25	0.51	-0.4	1.5	0.2	1.2	0.60
0.7	i_diff	cov_diff	5	2	3.05	2	1	0.25	1	0.25	0.51	-0.4	2	0.2	1.2	0.51
0.7	s_diff	cov_diff	5	2	5	1.02	1	0.25	1	0.25	0.51	-0.4	2	0.2	1.2	0.44
0.7	i_diff	cov_diff	5	2	2.57	2	1	0.25	1	0.25	0.51	-0.4	2.5	0.2	1.2	0.38
0.7	s_diff	cov_diff	5	2	5	0.78	1	0.25	1	0.25	0.51	-0.4	2.5	0.2	1.2	0.31
0.9	i_diff	var_diff	5	2	4.36	2	1	0.25	7.6	1.9	0.2	0.2	0.5	0.6	2.0	0.61
0.9	s_diff	var_diff	5	2	5	1.68	1	0.25	7.6	1.9	0.2	0.2	0.5	0.6	2.0	0.61
0.9	i_diff	var_diff	5	2	3.73	2	1	0.25	7.6	1.9	0.2	0.2	1	0.6	2.0	0.58
0.9	s_diff	var_diff	5	2	5	1.36	1	0.25	7.6	1.9	0.2	0.2	1	0.6	2.0	0.57
0.9	i_diff	var_diff	5	2	3.1	2	1	0.25	7.6	1.9	0.2	0.2	1.5	0.6	2.0	0.54
0.9	s_diff	var_diff	5	2	5	1.05	1	0.25	7.6	1.9	0.2	0.2	1.5	0.6	2.0	0.52
0.9	i_diff	var_diff	5	2	2.47	2	1	0.25	7.6	1.9	0.2	0.2	2	0.6	2.0	0.48
0.9	s_diff	var_diff	5	2	5	0.73	1	0.25	7.6	1.9	0.2	0.2	2	0.6	2.0	0.46
0.9	i_diff	var_diff	5	2	1.84	2	1	0.25	7.6	1.9	0.2	0.2	2.5	0.6	2.0	0.42
0.9	s_diff	var_diff	5	2	5	0.42	1	0.25	7.6	1.9	0.2	0.2	2.5	0.6	2.0	0.40
0.9	i_diff	var_diff	5	2	3.82	2	1	0.25	5.5	1.375	0.2	0.2	1	0.4	1.8	0.67
0.9	s_diff	var_diff	5	2	5	1.41	1	0.25	5.5	1.375	0.2	0.2	1	0.4	1.8	0.66
0.9	i_diff	var_diff	5	2	3.23	2	1	0.25	5.5	1.375	0.2	0.2	1.5	0.4	1.8	0.61
0.9	s_diff	var_diff	5	2	5	1.11	1	0.25	5.5	1.375	0.2	0.2	1.5	0.4	1.8	0.59
0.9	i_diff	var_diff	5	2	2.64	2	1	0.25	5.5	1.375	0.2	0.2	2	0.4	1.8	0.54
0.9	s_diff	var_diff	5	2	5	0.82	1	0.25	5.5	1.375	0.2	0.2	2	0.4	1.8	0.52
0.9	i_diff	var_diff	5	2	2.05	2	1	0.25	5.5	1.375	0.2	0.2	2.5	0.4	1.8	0.47
0.9	s_diff	var_diff	5	2	5	0.52	1	0.25	5.5	1.375	0.2	0.2	2.5	0.4	1.8	0.43
0.9	i_diff	cov_diff	5	2	4.24	2	1	0.25	1	0.25	0.8	-0.62	1	0.4	1.2	0.88
0.9	s_diff	cov_diff	5	2	5	1.62	1	0.25	1	0.25	0.8	-0.62	1	0.4	1.2	0.87
0.9	i_diff	cov_diff	5	2	3.87	2	1	0.25	1	0.25	0.8	-0.62	1.5	0.4	1.2	0.83
0.9	s_diff	cov_diff	5	2	5	1.43	1	0.25	1	0.25	0.8	-0.62	1.5	0.4	1.2	0.81
0.9	i_diff	cov_diff	5	2	3.49	2	1	0.25	1	0.25	0.8	-0.62	2	0.4	1.2	0.77
0.9	s_diff	cov_diff	5	2	5	1.24	1	0.25	1	0.25	0.8	-0.62	2	0.4	1.2	0.72
0.9	i_diff	cov_diff	5	2	3.11	2	1	0.25	1	0.25	0.8	-0.62	2.5	0.4	1.2	0.68
0.9	s_diff	cov_diff	5	2	5	1.05	1	0.25	1	0.25	0.8	-0.62	2.5	0.4	1.2	0.62
0.9	i_diff	var_diff	5	2	3.35	2	1	0.25	3.6	0.9	0.2	0.2	1.5	0.2	1.5	0.70
0.9	s_diff	var_diff	5	2	5	1.17	1	0.25	3.6	0.9	0.2	0.2	1.5	0.2	1.5	0.67
0.9	i_diff	var_diff	5	2	2.79	2	1	0.25	3.6	0.9	0.2	0.2	2	0.2	1.5	0.61
0.9	s_diff	var_diff	5	2	5	0.9	1	0.25	3.6	0.9	0.2	0.2	2	0.2	1.5	0.57
0.9	i_diff	var_diff	5	2	2.25	2	1	0.25	3.6	0.9	0.2	0.2	2.5	0.2	1.5	0.52
0.9	s_diff	var_diff	5	2	5	0.62	1	0.25	3.6	0.9	0.2	0.2	2.5	0.2	1.5	0.47
0.9	i_diff	cov_diff	5	2	3.7	2	1	0.25	1	0.25	0.63	-0.6	1.5	0.2	1.2	0.84
0.9	s_diff	cov_diff	5	2	5	1.35	1	0.25	1	0.25	0.63	-0.6	1.5	0.2	1.2	0.81
0.9	i_diff	cov_diff	5	2	3.27	2	1	0.25	1	0.25	0.63	-0.6	2	0.2	1.2	0.75
0.9	s_diff	cov_diff	5	2	5	1.13	1	0.25	1	0.25	0.63	-0.6	2	0.2	1.2	0.70
0.9	i_diff	cov_diff	5	2	2.84	2	1	0.25	1	0.25	0.63	-0.6	2.5	0.2	1.2	0.65
0.9	s_diff	cov_diff	5	2	5	0.92	1	0.25	1	0.25	0.63	-0.6	2.5	0.2	1.2	0.57

Appendix B

Data Generation Parameters for Simulation 2

mix proportion	Mean Condition	Variance Condition	Class 1		Class 2		Class 1		Class 2		Class 1	Class 2	SMD	C_d	Class 1	Class2	Data Overlap
			i	s	i	s	i-var	s-var	i-var	s-var	is-cov	is-cov			Resid-var	Resid-var	
0.5	i_diff	var_diff	5	2	4.24	2	1	0.25	3.7	0.925	0.2	0.2	0.5	0.4	2.87	11.49	0.25
0.7	i_diff	var_diff	5	2	4.34	2	1	0.25	3.7	0.925	0.2	0.2	0.5	0.4	2.21	8.85	0.27
0.9	i_diff	var_diff	5	2	4.41	2	1	0.25	5.5	1.375	0.2	0.2	0.5	0.4	1.77	7.09	0.32
0.5	i_diff	cov_diff	5	2	4.5	2	1	0.25	1	0.25	0.55	-0.6	0.5	0.4	1.22	4.89	0.28
0.7	i_diff	cov_diff	5	2	4.51	2	1	0.25	1	0.25	0.62	-0.6	0.5	0.4	1.22	4.89	0.30
0.9	i_diff	cov_diff	5	2	4.62	2	1	0.25	1	0.25	0.8	-0.62	0.5	0.4	1.22	4.89	0.39
0.5	i_diff	var_diff	5	2	4.35	2	1	0.25	2.5	0.625	0.2	0.2	0.5	0.2	2.14	8.56	0.27
0.7	i_diff	var_diff	5	2	4.4	2	1	0.25	2.6	0.65	0.2	0.2	0.5	0.2	1.81	7.24	0.30
0.9	i_diff	var_diff	5	2	4.45	2	1	0.25	3.6	0.9	0.2	0.2	0.5	0.2	1.54	6.16	0.37
0.5	i_diff	cov_diff	5	2	4.5	2	1	0.25	1	0.25	0.46	-0.4	0.5	0.2	1.22	4.89	0.29
0.7	i_diff	cov_diff	5	2	4.51	2	1	0.25	1	0.25	0.51	-0.4	0.5	0.2	1.22	4.89	0.32
0.9	i_diff	cov_diff	5	2	4.56	2	1	0.25	1	0.25	0.63	-0.6	0.5	0.2	1.22	4.89	0.41
0.5	i_diff	var_diff	5	2	3.7	2	1	0.25	2.5	0.625	0.2	0.2	1	0.2	2.14	8.56	0.26
0.5	i_diff	cov_diff	5	2	4	2	1	0.25	1	0.25	0.46	-0.4	1	0.2	1.22	4.89	0.28
0.7	i_diff	var_diff	5	2	3.8	2	1	0.25	2.6	0.65	0.2	0.2	1	0.2	1.81	7.24	0.28
0.7	i_diff	cov_diff	5	2	4.02	2	1	0.25	1	0.25	0.51	-0.4	1	0.2	1.22	4.89	0.30
0.9	i_diff	var_diff	5	2	3.9	2	1	0.25	3.6	0.9	0.2	0.2	1	0.2	1.54	6.16	0.36
0.9	i_diff	cov_diff	5	2	4.13	2	1	0.25	1	0.25	0.63	-0.6	1	0.2	1.22	4.89	0.39
0.5	i_diff	var_diff	5	2	3.49	2	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.87	11.49	0.23
0.5	i_diff	cov_diff	5	2	4	2	1	0.25	1	0.25	0.55	-0.6	1	0.4	1.22	4.89	0.27
0.7	i_diff	var_diff	5	2	3.68	2	1	0.25	3.7	0.925	0.2	0.2	1	0.4	2.21	8.85	0.26
0.7	i_diff	cov_diff	5	2	4.03	2	1	0.25	1	0.25	0.62	-0.6	1	0.4	1.22	4.89	0.29
0.9	i_diff	var_diff	5	2	3.82	2	1	0.25	5.5	1.375	0.2	0.2	1	0.4	1.77	7.09	0.31
0.9	i_diff	cov_diff	5	2	4.24	2	1	0.25	1	0.25	0.8	-0.62	1	0.4	1.22	4.89	0.38

Appendix C

Sample M*plus* Codes for Growth Mixture Models

```
title: two class GMM tryout
data: file=data_cond_5.3.txt;
variable: names are id class y1-y6;
        usevariables = y1-y6;
        classes=c(2);
analysis: type=mixture;
        starts= 100 10;
        stiterations=50;
        iterations=2000;
        miterations=5000;

Model: %overall%
      i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5;

      y1-y6*(resvar);

      [y1-y6@0];

      %c#1%

      i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5;
      [i*]; [s*];
      i*; s*; i with s*;

       y1-y6*(resv1);
       [y1-y6@0];

       %c#2%

       i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5;
      [i*]; [s*];
      i*; s*; i with s*;

       y1-y6*(resv2);
       [y1-y6@0];

output: samp tech1 tech11 tech13 tech14
```

Bibliography

Aitnouri, E., Dubeau, F., Wang, S., & Ziou, D. (2002). Controlling mixture component overlap for clustering algorithms evaluation. *Journal of Pattern Recognition and Image Analysis, 12*, 331-346.

Arellano-Valle R. B., Bolfarine, H., & Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science, 3*, 415-438.

Atlas, R., & Overall, J. (1994). Comparative evaluation of two superior stopping rules for hierarchical cluster analysis. *Psychometrika, 59*, 581-591.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for the overextraction of latent trajectory classes. *Psychological Methods, 8*, 338-363.

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? *Psychological Methods, 8*, 384-393.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis - Accuracy of 4 agglomerative hierarchical methods. *Psychological Bulletin, 83*, 377–388.

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

Bollen, K. A., & Curran, P. J. (2006). *Latent growth models: A structural equation perspective*. New Jersey: Wiley.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*, 195–212.

Chassin, L., Pitts, S. C., & Prost, H. (2002). Binge drinking trajectories from adolescence to emerging adulthood in a high risk sample: Predictors and substance abuse outcomes. *Journal of Consulting and Clinical Psychology, 70,* 67-78.

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70*, 976-985.

Collins, L. M., & Sayer, A. (2001). *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Connel, A. M., & Frye, A. A. (2006). Growth mixture modeling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behavior. *Infant and Child Development,15*, 609-621.

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

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*, 529-569.

Davies, R. (1980). The distribution of a linear combination of χ^2 random variables. *Applied Statistics, 29*, 323–333.

Demidenko, E. (2004). *Mixed models: Theory and applications.* New York:Wiley.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B, 39*, 1-38.

deRoon-Cassini, T. A., Mancini, A. D., Rusch, M. D., & Bonanno, G. A. (2010). Psychopathology and resilience following traumatic injury: A latent growth mixture model analysis. *Journal of Rehabilitation Psychology, 55*, 1-11.

Diebolt, J., & Robert C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. R. Statist. Soc. B, 56*, 363-375.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioral Research, 16*, 171-180.

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*. New York: John Wiley and Sons.

Fowlkes, E. B. (1979). Some methods for studying the mixture of two normal (lognormal) distributions. *Journal of the American statistical Association, 74*, 561-575.

Gold, E. M., & Hoffman, P. J. (1976). Flange detection cluster analysis. *Multivariate Behavioral Research, 11*, 217-235.

Gottman, J. M. (Ed.) (1995). *The analysis of change*. Mahwah, NJ: Lawrence Erlbaum.

Howell, D. C. (2007). The analysis of missing data. In Outhwaite, W. & Turner, S. (Ed.), *Handbook of Social Science Methodology* (pp. 208-224). London: Sage

Harring, J. R. (2005). *Nonlinear mixed effects mixture model: A model for clustering nonlinear longitudinal profiles*. Unpublished doctoral dissertation, University of Minnesota.

Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems, *Journal of the American Statistical Association*, *72*, 320-338.

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

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39-59.

Jung, T., & Wickrama, K. A. S. (2007). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *1*, 302-317.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*, New York: McGraw-Hill/Irwin.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, *7*, 305-315.

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

Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association, 83*, 1014–1022.

Lindstrom M. J., & Bates D. M. (1994) Corrections to Lindstrom and Bates (1988). *Journal of the American Statistical Association, 89*, 1572.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley and Sons.

Liu, M. (2011). Using latent profile models and unstructured growth mixture models to assess the number of latent classes in growth mixture modeling. (Unpublished doctoral dissertation). University of Maryland, College Park.

Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in an normal mixture. *Biometrika, 88*, 767-778.

Lubke, G., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 26-47.

Maitra, R., & Melnykov, V. (2010). Simulating Data to Study Performance of Finite Mixture Modeling and Clustering Algorithms. *Journal of Computational and Graphical Statistics, 19*, 354-376.

Manly, B. F. J., & Rayner, J. C. W. (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika, 74*, 841-847

McCullough, M. E., Enders, C. K., Brion, S. L., & Jain, A. R. (2005). The varieties of religious development in adulthood: A longitudinal investigation of religion and rational choice. *Journal of Personality and Social Psychology, 89*, 78-89.

McIntyre, R. M., & Blashfield, R. K. (1980). A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research, 15*, 225-238.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions* (2nd ed.). New York: Wiley.

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

Milligan, G. W. (1985). An algorithm for generating artificial test clusters. *Psychometrika, 50*, 123-127.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling. *Psychological Methods, 8,* 369-377.

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

Muthén, B., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143-165). Boca Raton: Chapman & Hall/CRC Press.

Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S., Yang C., Wang, C., Kellam, S. G., Carlin, J. B., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 9*, 456-678.

Muthén, B. O., Khoo, S. T., Francis, D., & Kim Boscardin, C. (2000). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In S. R. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 71-89). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B., & Muthén, L. (1998–2010). *Mplus User's Guide*. Los Angeles: Muthén and Muthén.

Muthén, B., & Muthén, L. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

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

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric group-based approach. *Psychological Methods, 4*, 139-157.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specificationand estimation of a nonparametric, mixed Poisson model. *Criminology, 31*, 327–362.

Nagin, D., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods, 6*, 18-34.

Nityasuddhi, D., & Böhning, D. (2003). Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances. *Computational Statistics and Data Analysis, 41*, 591-601.

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, 14*, 535–569.

Oberg, A., & Davidian, M. (2000). Estimating data transformations in nonlinear mixed effects models. *Biometrics, 56*, 65-72.

Odgers, C. L., Caspi, A., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Thomson, W. M., & Moffitt, T. E. (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of General Psychiatry, 64*, 476-484.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J. E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finish. *Journal of Educational Psychology, 97*, 299-319.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc. London A, 185*, 71-110.

Pinheiro, J. C., Liu, C., & Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution, *Journal of Computational and Graphical Statistics,10*, 249-276.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.

Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*, 565-576.

Qiu,W., & Joe, H. (2006). Generation of random clusters with specified degree of separation. *Journal of Classification, 23,* 315–334.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary acquisition in late talkers: Lexical development from 2.0 to 3.0. *Journal of Child Language, 27,* 293-311.

Segawa, E., Ngwe, J. E., Li, Y., & Flay, B. R. (2005). Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques. *Evaluation Review, 29,* 128-148.

Schafer, J. L., & Graham J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7,* 147-177.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Methods for Studying Change and Event Occurrence*. New York: Oxford University Press.

Steinley, D., & Henson, R. (2005). OCLUS: An Analytic Method for Generating Clusters with Known Overlap. *Journal of Classification, 22,* 221-250.

Stoolmiller, M., Kim, H. K., & Capaldi, D. M. (2005). The course of depressive symptoms in men from early adolescence to young adulthood: Identifying latent trajectories and early predictors. *Journal of Abnormal Psychology, 114,* 331-345.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances*

in latent variable mixture models (pp. 317-341). Charlotte, NC: Information Age Publishing, Inc.

Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study.* Unpublished doctoral dissertation, University of Jyvaskyla.

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., & Lesaffre E. (1997). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis, 23*, 541-556.

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

Waller, N. G., Underhill, J. M., & Kaiser, H. A. (1999). A method for generating simulated plasmodes and artificial test clusters with user-defined shape, size, and orientation. *Multivariate Behavioral Research, 34,* 123-142.

Wang, M., & Bodner, T. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods, 10,* 635-656.

West, B., Welch, K., & Galecki, A. (2007). *Linear mixed models: A practical guide using statistical software.* Boca Raton: Chapman & Hall.

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