2010-04-22

A Comparison of Modern Longitudinal Change Models with an Examination of Alternative Error Covariance Structures

Jaime Maerten-Rivera
University of Miami, jmaerten-rivera@miami.edu

UNIVERSITY OF MIAMI

A COMPARISON OF MODERN LONGITUDINAL CHANGE MODELS WITH AN EXAMINATION OF ALTERNATIVE ERROR COVARIANCE STRUCTURES

By

Jaime L. Maerten-Rivera

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2010

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

A COMPARISON OF MODERN LONGITUDINAL CHANGE MODELS WITH AN EXAMINATION OF ALTERNATIVE ERROR COVARIANCE STRUCTURES

Jaime L. Maerten-Rivera

Approved:

Randall D. Penfield, Ph.D.
Professor of Educational
and Psychological Studies

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Nicholas D. Myers, Ph.D.
Professor of Educational
and Psychological Studies

Soyeon Ahn, Ph.D.
Professor of Educational
and Psychological Studies

Okhee Lee, Ph.D.
Professor of Teaching
and Learning

MAERTEN-RIVERA, JAIME L. (Ph.D., Educational and Psychological Studies)
A Comparison of Modern Longitudinal Change Models with an Examination of Alternative Error Covariance Structures. (May 2010)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Randall D. Penfield
No. of pages in text. (71)

The purpose of this research was to compare results from two approaches to measuring change over time. The multilevel model (MLM) and latent growth model (LGM) were imposed and the parameter estimates were compared, along with model fit. The study came out of education and used data collected from 191 teachers as part of a professional development intervention in science, which took place over four years. There were missing data as a result of teacher attrition. Teachers reported use of reform-oriented practices (ROP) was used as the outcome, and teacher-level variables were examined for their impact on initial ROP and change in ROP from baseline to one year after the intervention.

Change in ROP was examined using a piecewise change model where two linear slopes were modeled. The first slope estimated the change from baseline to T1, or the initial change after the intervention while the second slope estimated the change from T1 to T3, or the secondary change. Parameter estimates obtained from MLM and LGM for a model using the error covariance structure commonly assumed in MLM (i.e., random slopes, homogeneous level-1 variance) were nearly identical. Models with various alternative covariance structures (commonly associated with the LGM framework) were examined, and results were nearly identical. Most of the model fit information was in agreement regarding the best fitting model being the model that assumed the typical

MLM error covariance structure with the exception of the standardized root mean square residual (SRMR) fit index.

The results from the models demonstrated that ROP increased after participating in the first year of the intervention and this level was sustained, though did not increase significantly in subsequent years. There was more variation in ROP at baseline. This information tells us that the intervention was successful in that after participating in the intervention the teachers' used ROP more frequently. The success of the intervention did not depend on any of the predictors that we assessed, and, as a group, the teachers became more similar in their use of reform-oriented practices over time.

DEDICATION

This work is dedicated to Wayne Maerten.

ACKNOWLEDGEMENTS

I would like to acknowledge the guidance I received from the members of my Dissertation Committee: Dr. Nicholas Myers, Dr. Randall Penfield, Dr. Okhee Lee, and Dr. Soyeon Ahn. Without their assistance, the study would not have been as thorough. I am appreciative of all of their efforts on my behalf.

I am very grateful to my family and friends for their support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Purpose of Research

Many studies are conducted where the variable of interest is measured from the same subject repeatedly over time in order to study growth or change.[1] The idea of measuring change has intrigued social scientists in various areas, including biostatistics (Laird & Ware, 1982; Rao, 1958), education (Goldstein, 1987; Raudenbush & Bryk, 2002; Rogosa & Willet, 1985), and psychometrics (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Tucker, 1958). Traditional approaches to studying change include ANOVA and regression, but these methods are limited because they analyze only mean change ignoring differences in initial status, and simply treating differences among individual change as error variance. Also, these methods cannot easily handle missing data. More recent methodological advancements have extended researchers' abilities to model individual differences and change over time using approaches such as the latent growth model or LGM (Duncan, Duncan, Strycker, Li & Alpert, 1999; Meredith & Tisak, 1990; Preacher, Wichman, MacCallum, & Briggs, 2008), which is a special case of the structural equation model (SEM), and the multilevel model or MLM (Raudenbush & Bryk, 2002; Singer & Willet, 2003).

The LGM is also referred to as a latent growth curve model (Preacher, et al., 2008). The MLM, is sometimes referred to as a hierarchical linear model (Raudenbush & Bryk, 2002), a random-coefficients model (Longford, 1993) or a random effects model (Laird & Ware, 1982). Although both approaches can be used for studying change, they

[1] Throughout this paper change will be used though others may use the term growth.

developed out of different fields and are based on different assumptions. Some missing data can be handled relatively easily in both approaches using full maximum likelihood estimation (Enders, 2001). In some research, the results from both approaches have provided similar fixed effects estimates after imposing certain constraints on the model (Chou, Bentler & Pentz, 1998; Rovine & Molenaar, 1998, 2000). In these studies, the covariance structure of the model was constrained to a structure commonly imposed in multilevel models (i.e., random slopes, homogeneous level-1 variance). These studies examined linear change or change that was quadratic in time.

MLM typically does not simply assume classical assumptions that residuals are identically distributed with homoscedastic variance across occasions and individuals, but rather the composite residuals have a quadratic dependence on time advancement, which allows for both heteroscedasticity and autocorrelation (Singer & Willet, 2003). Even so, studies (Chou, et al., 1998; Rovine & Molenaar, 1998, 2000) have noted that an advantage of the LGM model is the capability of allowing for a variety of covariance structures. However, recent advancements in MLM software allow for a number of covariance structures to be examined (Jenrich & Schluchter, 1986). The purpose of this research was to demonstrate that alternative covariance structures can be examined using the MLM approach and to compare results with those obtained using the LGM approach.

This research used data collected as part of a teacher professional development intervention in science, which took place over five years. At the end of each year, teachers were asked to complete a questionnaire, which included a group of items that formed a scale measuring reform-oriented practices (ROP), the outcome examined. As a result of the design of the study, each teacher could have four measurement occasions (baseline, T1, T2, T3). However, due to teacher mobility, there are missing data, which is an issue in many educational research settings. Both MLM and LGM will produce unbiased estimates assuming that the data are missing at random (MAR; Rubin, 1976) or that the missingness of a variable is related to other observed variables in the model, but is unrelated to that variable itself.

The intervention took place over three years. We hypothesized that much of the change in teaching practices would take place during the first year when the teacher was initially exposed to the intervention and that less (if any) changes would take place in the subsequent years; this is consistent with other research on science teaching professional development where large gains occurred during the first year but no change was seen in the second and third years (Supovitz, Mayer, & Kahle; 2000). Supovitz, Mayer, and Kahle (2000) describe this pattern as short-term growth and long-term stability. A piecewise change model was used to specify this change trajectory. Piecewise change models specify two or more linear slopes in one model. In this research, one linear slope estimated the change from baseline to T1, or the initial change after the intervention. A second slope estimated the change from T1 to T3, or the secondary change after the intervention. Predictors were then entered in the model. And finally, alternative error covariance structures were tested. The models throughout were examined using both MLM and LGM and parameter estimates were compared.

Significance of the Study

Several articles have compared the MLM and LGM approaches to modeling change specifically (Chou, Bentler, & Pentz, 1998; MacCallum et al., 1997; Mehta & West, 2000; Rovine & Molenaar, 1998, Rovine & Molenaar, 2000) and others have compared the SEM and MLM approaches more broadly (Curran, 2003; Muthen & Curran; 1997). Many of the articles offer a technical demonstration comparing computational similarities (Rovine & Molenaar, 1998; Rovine & Molenaar, 2000). While others focus on specific modeling issues such as modeling multivariate change (MacCallum et al., 1997) or the scaling of time (Mehta & West, 2000). Chou et al. (1998) took an applied approach to comparing the fixed and random effects from both approaches using data collected from high school students to model change in substance abuse and demonstrated that results for the estimates of fixed and random effects from both approaches were the same; however, this study used only the standard error covariance structure used in most MLM modeling.

Most of these past articles have come out of the field of psychology and were written by researchers who predominantly use SEM. When data were used to demonstrate the concepts the data were collected in the field of psychology or were simulated (Mehta & West, 2000; Rovine & Molenaar, 2000). In addition, these past studies examined linear change or change that was quadratic in time.

The current study is different from much of the existing literature for a number of reasons. First, this study came out of the field of education and used data collected from an education intervention where attrition was an issue. This is a content area to which many researchers can relate and provides a context for our applied study of these two approaches.

Second, much of the existing literature examined linear change models with the discussion of other types of change limited to models quadratic in time. We did not find that these change models fit our data and focused on a piecewise change model.

Therefore, we offer a practical application of a piecewise change model. In addition, it extends the previous literature on the similarities between MLM and LGM to models using non-linear change trajectories.

Finally, this study extends previous research comparing the two approaches by examining alternative error covariance structures in both approaches. Flexibility in defining the error covariance structure has long been attributed to LGM. However, there are other options available in MLM software, but research has not yet compared the alternative error covariance estimates obtained from these two approaches. In addition, it is meant to inform other researchers that the error covariance structure should be examined in change models as a source of model misspecification, regardless of the modeling approach used.

CHAPTER 2

Literature Review

Both the MLM and LGM offer advantages over other more traditional analyses used to model change. First, both MLM and LGM offer model-based approaches to handling missing data using maximum likelihood estimation (Enders, 2001). More traditional methods do not offer a model-based approach to handling missing data and in many analyses cases are simply deleted because they are missing data at one or more measurement occasion. Second, more traditional methods analyze only mean change ignoring differences in initial status and/or change trajectory, and simply treating differences among individual change as error variance. In LGM and MLM systematic changes due to individual differences (i.e., gender, race/ethnicity) on both initial status and change can help answer important research questions and the relationship between initial status and change can be examined.

Despite these similarities, there are differences between these two types of analyses. The literature review first discusses the history of each approach, then the model parameters and estimation, followed by the covariance structure for each approach. It then examines research comparing the two approaches.

History

Both MLM and LGM have long histories. Much of the work on the MLM framework developed out of the fields of education and biostatistics. While much of the work on the LGM framework developed out of the field of psychometrics.

MLM. The term *hierarchical linear model* was first introduced by Lindley and Smith (1972) and Smith (1973) who detailed a general framework for nested data with complex error structures. Dempster, Laird, and Rubin (1977), who worked in the field of biostatistics, made broader applications of Lindley and Smith's work possible by developing the expectation-maximation (EM) algorithm. The use of these approaches to study change was first applied by Laird and Ware (1982), and Strenio, Weisberg, and Bryk (1983). Raudenbush and Bryk (2002), working in the educational field, continued to develop on the hierarchical linear modeling approach and made the approach popular by creating software designed to analyze this type of data (Bryk, Raudenbush, Seltzer & Congdon; 1989). According to Kreft and de Leeuw (1998), soon after its introduction the HLM software was adapted as the 'official' software for educational multilevel analysis.

LGM. The first person to propose using latent variables within the factor analytic framework and trajectory modeling was Baker (1954). Not long after, Tucker (1958), who worked in the field of psychometrics, proposed a method for using latent factors to estimate known functional forms relating time to the repeated measures. Drawing on this work, Meredith and Tisak (1990), working in the field of psychometrics, proposed embedding change trajectory modeling within the confirmatory latent variable framework used in structural equation models demonstrating that the analytic interest is not specifically the repeated measures observed, but the unobserved latent trajectory factors. Around the same time, a multilevel structural equation model designed to handle hierarchically clustered observations was presented (Muthén & Satorra, 1989), which was later adapted for use with longitudinal data (Muthén, 1994).

Model Parameters and Estimation

The MLM and LGM have different underlying principles and therefore the estimation of the model parameters is performed differently. The following section first outlines the parameters used in the MLM, followed by an outline of the parameters used in LGM, and then compares the models in terms of parameters and estimation.

MLM. The MLM generalizes the simple regression model to circumstances in which the assumption of independent observations does not hold. Data with a hierarchical, or “nested”, structure are often found in educational research with the classic example of observations of students nested within schools. In this example, students within a school are more likely to be similar than a random sample of students. Ignoring the relationship among students within a school leads to violations of statistical assumptions and possibly erroneous results. A conventional student-level regression model may provide biased estimates, negatively biased standard error estimates, and inflated Type I error, while aggregating the data to the school-level results in loss of information and power (Raudenbush & Bryk, 2002). MLMs allow both levels of the data to be analyzed simultaneously, accounting for nonindependence of the data by allowing the regression coefficients to randomly vary. The MLM models the individual-level outcome variables as a linear function of the individual-level predictors within each group; then the coefficients from the first stage model become the dependent variables of the next level and are modeled as a linear function of the group-level predictors.

In a basic 2-level example, the level-1 model is shown as

$$y_{ij} = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} x_{pij} + r_{ij} \tag{1}$$

where y_{ij} represents the outcome variable for individual i in group j , β_{0j} is the level-1 intercept within group j , β_{pj} is the regression of y_{ij} on the p^{th} variable x within group j (or the slope), and r_{ij} is the residual for individual i within group j . Since the intercept and slope coefficients vary randomly over groups, these can be regressed upon one or more level-2 predictors denoted *w* such that the level-2 model can be displayed as

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{oq} w_{qj} + u_{0j}, \qquad (2)$$
$$\beta_{pj} = \gamma_{po} + \sum_{q=1}^{Q} \gamma_{pq} w_{qj} + u_{pj}$$

where γs represent the fixed coefficients for the regression of the random intercepts and slopes from the level-1 equation (e.g., β_{0j} and β_{pj}) on the q^{th} level-2 predictor w for each group j, and u_{oj} and u_{pj} represent the level-2 residuals.

MLM provides a method for estimating random effects change models by treating the data as nested, such that the repeated measures are nested within individuals (Raudenbush & Bryk; 2002, Singer & Willett; 2003). The level-1 model considers the within-person change over time (intraindividual), while the level-2 model considers the between-unit variation in change (inter-individual). In previous equations the subscripts to denote individual i in group j were used, whereas in the change model equations, time point t is nested within individual i. Similarly, β is replaced by π, and r by e since the model has been reduced by a level and the level-1 notation is retained for the individual level in the time-nested data. This research used teachers ROP as the dependent variable and the simplest linear change model with time as the only level-1 predictor can be expressed as

$$y_{ti} = \pi_{0i} + \pi_{1i} x_{ti} + e_{ti} \qquad (3)$$

where y_{ti} is the measure of ROP at the time t for the individual i, x_{ti} represents the time of measurement (e.g., $x_{ti} = t$) ,while π_{0i} and π_{1i} represent the change parameters of interest for individual i, that is, initial status and change trajectory, respectively; and e_{ti} is the time-specific residual. It is assumed that e_{ti} is normally distributed with mean 0 and variance σ^2. In most cases, time is coded to reflect the actual time interval between two observations.

Under the simple linear change assumption, the level-2 model for between-individual level contains two equations: one for the intercept, $\pi_{0i,}$ and the other for the change trajectory, π_{1i}. Using one explanatory variable, z_{1i}, for both π_{0i} and π_{1i}, the equation for level-2 is given as

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01} z_{1i} + r_{0i,} \\ \pi_{1i} &= \beta_{10} + \beta_{11} z_{1i} + r_{1i} \end{aligned} \tag{4}$$

where β_{00} and β_{10} are the mean intercept and slope values, β_{01} and β_{11} are the regression weights for the impact of z on initial status (π_0) and the change (π_1) respectively. The r_{0i} and r_{1i} are the individual deviations of each observation from the model-based expectation. The r_{0i} and r_{1i} are assumed to have means of 0 and that each person draws both level-2 residuals simultaneously from a bivariate normal distribution with variances $\tau_{00}(r_0)$ and $\tau_{11}(r_1)$, and covariance $\tau_{10}(r_1)$.

In the MLM framework, change is not limited to a single linear trajectory. Non-linear change can be modeled by adding additional time measurement variables. Similarly, piecewise change (Raudenbush & Bryk, 2002) can be specified such that change is treated in separate, although related pieces by modeling two linear slopes. Our data suggested that two linear slopes may be the most appropriate model for change in ROP.

LGM. The SEM framework uses a covariance structure analysis to assess whether a sample covariance matrix, $\mathbf{S}$, is consistent with a hypothetical matrix implied by the specification of a theoretical model, $\mathbf{\Sigma\theta}$. The basic statistical theory underlying SEM is based on examining the variances and covariances among the observed variables. Based on the covariance matrix being analyzed, a chi-square test of model fit and standard errors for the parameter estimates are obtained. This chi-square test provides an evaluation of the null hypothesis that the covariance matrix in the population is equal to that implied by the model (i.e., $\mathbf{\Sigma} = \mathbf{\Sigma\theta}$). The LGM is a special case of the mean and covariance structure model and requires not just information on the covariances of the variables, but also requires information regarding the means. In the LGM, in addition to assessing the covariance matrix, the model is assessed on whether the vector of sample means $\mathbf{m}$ is consistent with the vector implied by the theoretical model, $\boldsymbol{\mu}$. LGM evaluates the null hypothesis that the mean matrix in the population is equal to that implied by the model (i.e., $\boldsymbol{\mu} = \boldsymbol{\mu\theta}$).

The general SEM consists of the measurement equations and the latent variable equations. The measurement equations relate the observed variables to the latent factors via structural coefficients and residuals for each observed variable. The latent variable equations estimate the specified relationships among the latent variables, and allow researchers to incorporate other predictor variables into the equations.

The LGM is specified within the SEM by using the measurement model to specify latent factors pertaining to change while fixing some parameters. The longitudinal measures of the variable of interest are expressed as a function of two factors: an underlying measure of the variable, usually the initial status, and the developmental trajectory or change on the variable. The structural portion of the model is then used to specify relationships among the latent variables.

A simple linear change model with one time-invariant predictor variable in the model, an exogenous latent factor, ξ_1, is examined here. The measurement equation is

$$y_{ti} = \lambda_{t0}\eta_{0i} + \lambda_{t1}\eta_{1i} + \varepsilon_{ti} \tag{5}$$

where y_{ti} again refers to the measure of ROP at the time t for the individual i, and is predicted by η_{0i}, and η_{1i} which are the underlying factors representing the initial status and linear change trajectory, respectively.

The latent variable equation for the model is given as

$$\begin{aligned} \eta_{0i} &= \nu_0 + \gamma_{01}\xi_{1i} + \zeta_{0i}, \\ \eta_{1i} &= \nu_1 + \gamma_{11}\xi_{1i} + \zeta_{1i} \end{aligned} \tag{6}$$

where the η_{0i} and η_{1i} are factors representing the initial status and linear change trajectory; ν_0 and ν_1 are their corresponding intercepts represent the mean initial status and linear change values; γ_{01} and γ_{11} are the regression weights for the impact of ξ_1, on initial status (η_0) and linear change trajectory (η_1) respectively, and residuals ζ_{0i} and ζ_{1i} residuals are the individual deviations of each observation from the model-based expectation. Typical SEM assumptions are made, such as, and ζ are bivariate normally distributed with means equal to 0 and variance ψ. Random effects are represented by variances $\sigma^2(\zeta_0)$, $\sigma^2(\zeta_1)$, and $\sigma^2(\varepsilon)$. All λ parameters are fixed to constants, specifically, $\lambda_{t0} = 1$ and $\lambda_{t1} = t$, the time point at which the outcome was obtained.

The standard SEM model assumes that the means of all variables are zero. However, the LGM requires that the means of the latent variables measuring change be analyzed. The representation of means in SEM is based on the general principles of regression. When there are no predictors of the factor, the factor is regressed on a constant and the unstandardized coefficient is the mean of the predictor. When there are predictors of the factor, the factor is regressed on the predictor(s) and a constant to compute the intercept. Thus, the intercept equation for η_0 can be given as

$$\hat{\nu}_0 = \hat{\mu}_{\eta_0} - \hat{\beta}\hat{\mu}_{\xi_1} \tag{7}$$

where $\hat{\nu}_0$ represents the predicted intercept for η_0, $\hat{\mu}_{\eta_0}$ represents the predicted mean for η_0, $\hat{\beta}$ represents the regression weight for $\hat{\mu}_{\xi_1}$, the predicted mean for the predictor ξ_1.The intercept, is the average initial level of the outcome variable, after controlling for predictor variables, adjusted for measurement error, and is characteristic of the whole sample. The variance of the factor reflects the range of individual differences around the average initial level. The intercept of the η_1 factor can be computed in a similar way and interpreted similarly, but reflects the average change of the whole sample.

In the LGM framework change is not limited to a single linear trajectory. Non-linear change can be modeled by adding an additional term for time and adding another latent variable to represent the additional change. Similarly, piecewise change (Duncan, Duncan, & Strycker, 2006; Preacher et al., 2008) or discontinuity (Hancock & Lawrence, 2006) can be specified such that change is treated in separate, although related pieces by modeling two linear slope factors in one model. Our data suggested that two linear slopes may be the most appropriate model for change in ROP.

Model Parameter Equivalency. Although the MLM and LGM have different underlying principles, they are comparable in certain situations of change modeling. In the MLM the repeated measures design provides a two-level data structure with time nested within individuals, and time is entered as a predictor variable (e.g., x_{ti} in Equation 3). However, in LGM the level-1 measure of time is set to fixed values within the factor loading matrix (e.g., λ_{1t} in Equation 1). The set factor loadings disaggregate the level-1 and level-2 covariance structures in a single partitioned covariance matrix $\mathbf{S}$ which is used as the unit of analysis in the estimation, along with the mean structure $\mathbf{m}$. The latent factors in the LGM then represent the fixed and random effects associated with the stability and change of the repeated measures over time and the parameter estimates obtained from the LGM are comparable to those obtained from the MLM.

With all of the λ parameters fixed at constants in the LGM model, the parameter estimates associated with the LGM are the same as those for the MLM. Table 1 lists the corresponding parameters from the two approaches. The $y_{ti}, \pi_{0i}, \pi_{1i}, x_{ti}$, and e_{ti} in Equation 3 correspond to the $y_{ti}, \eta_{0i}, \eta_{1i}$, and ε_{ti} in Equation 7, respectively, with λ_{0t} fixed at 1. Additionally, the $\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}, z_{1i,} r_{0i,}$ and r_{1i} in Equation 4 correspond to $v_0, \gamma_{01}, v_1, \gamma_{11}, \xi_{1i}, \zeta_{0i}$, and ζ_{1i} in Equation 6, respectively.

Table 1

Corresponding Parameters from Two Approaches

MLM	LGM
β_{00}	v_0
β_{01}	γ_{01}
β_{10}	v_1
β_{11}	γ_{11}
σ^2	$\sigma^2(\varepsilon_{ti})$
τ_{00}	$\sigma^2(\zeta_0)$
τ_{11}	$\sigma^2(\zeta_1)$
τ_{10}	$\sigma^2(\zeta_0, \zeta_1)$

Covariance Structure

Most hypothesis tests focus on the fixed effects being estimated in a model since these usually provide answers to the research questions of interest. However, it is important to consider the error covariance structure of a model, which estimates the model's random effects. Refining hypotheses about the error covariance structure rarely changes the fixed effects parameter estimates obtained; however, it can affect the precision of estimates of the fixed effects, and therefore impacts hypotheses testing and confidence intervals (Singer & Willett, 2003). As the error covariance structure of a model is better represented, the standard errors may decline, increasing the likelihood of rejecting the null hypothesis. Thus researchers have noted that the analysis of longitudinal data is not complete without a close examination of the residuals (Fitsmaurice, Laird, & Ware, 2004; Weiss, 2005).

MLM. Singer and Willett (2003) review what they refer to as the "standard" multilevel model for change, which is the model using the typical error covariance structure specified in MLM analyses, and is the default provided by many MLM software programs.

By combining the level-1 and level-2 models given in Equations 3 and 4, and multiplying out the terms, the following equation can be given

$$y_{ti} = (\beta_{00} + \beta_{01}x_{ti} + \beta_{01}z_{1i} + \beta_{11}z_{1i} \times x_{ti}) + (e_{ti} + r_{0i} + r_{1i}x_{ti}). \quad (8)$$

The parenthesis separate out the model's structural and stochastic portions. The stochastic portion contains the composite residual, *s*. The value of *s* for an individual *i* on occasion *t* is

$$s_{ti} = [e_{ti} + r_{0i} + r_{1i}x_{ti}], \quad (9)$$

which is a weighted linear combination of the original three random effects from the level-1/level-2 specification (e_{ti}, r_{0i}, and r_{1i} with constants 1, 1, and x_{ti} acting as the weights).

In this model, there is a within-person error structure and a composite error structure. As a result of the time variable being included in the stochastic portion of the model, the within-person structure can be both heteroscedastic (with a minimum that will increase parabolically and symmetrically over time on either side of its minimum) and correlated over time. The composite residuals are expected to be normally distributed with zero means. In other words, the blocks of the error covariance matrix are identical across people as a result of a homogeneity assumption that the composite residuals may be heterscedastic and dependent within people, but the error structure is repeated identically across people.

The homogeneity assumption can be expressed by writing

$$\boldsymbol{s} \sim \boldsymbol{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_s & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_s & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_s & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_s \end{bmatrix}\right) . \qquad (10)$$

It is assumed that the complete vector of residuals $\mathbf{s}$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and a block-diagonal error covariance matrix made up of submatrices $\boldsymbol{\Sigma}_s$ and $\mathbf{0}$, where:

$$\boldsymbol{\Sigma}_s = \begin{bmatrix} \sigma^2_{s1} & \sigma_{s1s2} & \sigma_{s1s3} & \sigma_{s1s4} \\ \sigma_{s2s1} & \sigma^2_{s2} & \sigma_{s2s3} & \sigma_{s2s4} \\ \sigma_{s3s1} & \sigma_{s3s2} & \sigma^2_{s3} & \sigma_{s3s4} \\ \sigma_{s4s1} & \sigma_{s4s2} & \sigma_{s4s3} & \sigma^2_{s4} \end{bmatrix} . \qquad (11)$$

The HLM software (Scientific Software Inc., 2005) allows hypotheses testing of some alternative covariance structures through the use of the HLM2 component of the program. More recently, additional models (i.e., unrestricted, first-order autoregressive) can be compared using the hierarchical multivariate linear models (HMLM) component of the program, which enables the estimation of covariance pattern models (CPMs, see Bock 1975). Covariance pattern models are similar to SEM in that they are flexible in allowing a variety of assumptions about the variation and covariation of the repeated measures, however, in the standard application of these models every person must have a fixed time-series design, meaning that the number and spacing of time points must be invariant across people (Raudenbush & Bryk, 2002). However, the HMLM component of HLM uses the EM algorithm (Dempster et al.) to estimate models from incomplete data.

LGM. The LGM is a special case of the SEM and the flexibility of modeling the error covariance structure in SEM applies to the LGM. The SEM framework is a relatively broad, flexible approach which allows a variety of error covariance structures to easily be modeled using standard SEM software. The variance and covariance parameters can be estimated freely, set to be equal to another variance or covariance being estimated, or set to a specified value (including zero). This allows various models to be tested including autocorrelated level-1 residuals and level-1 residuals having different variances at different times. One widely recognized advantage of SEM is that any pattern of covariance among the residuals can be modeled by applying the correct constraints (Chou et al., 1998; Curran, 2003; Rovine & Molenaar, 1998, 2000). However, sample size may limit the complexity of the model being tested in the SEM framework since larger samples are needed in order to obtain adequate power, and more complex models may have difficulty converging (Heck & Thomas, 2000).

Common Alternative Covariance Structures

The properties imposed on a model's composite residual must match those required by the data. Therefore, it is best to evaluate some plausible alternative error covariance structures (Goldstein, Healy, & Rabash, 1994; Singer & Willett, 2003; Wolfinger, 1996). A nested model with more parameters estimated will always have a lower deviance; so a researcher must consider the relative fit of the model along with the degrees of freedom. Wolfinger (1996, p. 208) notes that, "a basic rationale is 'parsimony means power'; that is, to obtain the most efficient inferences about the mean model, one selects the most parsimonious covariance structure possible that still reasonably fits the data."

Some common alternative covariance structures are discussed below using the terminology from the MLM framework. The unrestricted error covariance structure is the least restrictive as the elements take on the value that the data demand. The other models place restrictions on either the variance or covariance components.

Unrestricted. In this model, there is a general variance-covaraiance matrix, Σ where each element of the error covariance structure takes on the value that the data demand. In our data with four time points, there are 10 unique variance-covariance parameters to be estimated, but as the number of time points increases, this model becomes much more complex. Nevertheless, with a small number of time points, this model can serve as a standard to test the fit of more parsimonious models.

Random slopes, homogeneous level-1 variance. This is the typical error covariance structure specified in MLM analyses described above. In a simple linear growth model, this model represents the variance-covariance parameters as a function of four underlying parameters ($\tau_{00}, \tau_{11}, \tau_{10}, \sigma^2$).

First-order autoregressive. In this model, the variances are the same at every time point. However, pairs of errors have identical covariances in bands parallel to the leading diagonal, which are the product of the residual variance, σ^2, and an error autocorrelation parameter, ρ. Since ρ is always fractional, error covariances decline the further out from the leading diagonal. In a simple linear growth model, this model represents the variance-covariance parameters as a function of three underlying parameters ($\tau_{00}, \rho, \sigma^2$).

Compound symmetry. This is the covariance structure commonly assumed in univariate repeated measures analysis of variance which requires that the variances be the same at every time point and also that all covariances are the same. This is equivalent to assuming that (a) the level-1 random effects, e_{ti}, are independent with homogeneous level-1 variance, σ^2, and (b) all participants have the same slope. In a simple linear growth model, this model represents the variance-covariance parameters as a function of two underlying parameters (τ_{00} and σ^2).

Research Purpose and Questions

Past research has demonstrated that models can be fit in both MLM and LGM using a standard error covariance structure and have criticized MLM for not being able to handle alternative error covariance structures. However, MLM is capable of estimating some alternative error covariance structures. This study extends existing research by comparing results from both approaches for models with alternative error covariance structures. The research highlights model building techniques, and demonstrates the use of a piecewise change model. First, the parameter estimates obtained from both approaches using the standard MLM covariance structure were compared. Then, models with various alternative error covariance structures were examined, and compared.

The research questions addressed were:

1) Does the model converge in both MLM and LGM given the moderate sample size and missing data?
2) How similar are the parameter estimates and the standard errors of these estimates from MLM and LGM when the error covariance structure is set to the “standard” error covariance structure used in MLM?

3) How similar are the parameter estimates and the standard errors of these estimates from MLM and LGM when examining alternative error covariance structures in the two approaches?
4) Do the fit statistics inform the researcher to accept the same final model using both MLM and LGM?

CHAPTER 3

Methodology

Research Setting

Longitudinal data obtained from a science teacher professional development project were used to compare the MLM and LGM. The research project aimed at improving science and English literacy achievement of English language learning (ELL) students at a time when science was becoming part of accountability policies across the nation. Over the course of its five-year period, the research implemented an educational intervention consisting of curriculum units and teacher workshops. The research involved teachers from grades 3 through 5 and their students at six elementary schools in a large urban school district. All of the schools enrolled relatively high proportions of ELL students and students from low socioeconomic status (SES) backgrounds, and had traditionally performed poorly according to the state's accountability policies.

A main goal of the intervention was to increase the use of reform-oriented teaching practices (ROP) in science and this was the outcome variable examined in this study. The change of the teachers on this measure was examined using both MLM and LGM. Teacher-level predictors were examined for their impact on both initial status in ROP and change over time in ROP. There were 191 teachers included in this study, with varying data points collected. The missing data were primarily the result of teacher attrition over the five years in which the professional development took place.

Data Collection

Participation. In the state in which this research took place, standardized testing in science was mandated in fifth grade starting in the 2006-2007 school year. The goal of the intervention was to prepare the first cohort of students to be tested in science by exposing the students to the third, fourth, and fifth grade science curriculum developed by the intervention. Therefore, the teachers were phased into the intervention, and later phased out of the intervention. During the first year (2004-2005), third grade teachers and their students began the intervention. In the second year (2005-2006), the intervention was started with the fourth grade teachers, while third grade was continued. In the third year (2006-2007), the intervention was started with the fifth grade teachers, while both third and fourth grade were continued. In the fourth year (2007-2008), the intervention was continued with the fourth and fifth grade, while the third grade teachers were phased out from the professional development intervention. The third grade teachers were encouraged to use the curriculum, but the professional development was not available to these teachers. In the fifth year (2008-2009), the intervention continued for fifth grade teachers, while the fourth grade teachers were phased out from the professional development. Table 2 below demonstrates which grades participated in each year of the intervention.

Table 2

Grades Participating in Each Year of the Intervention

	Year 1 2004-2005	Year 2 2005-2006	Year 3* 2006-2007	Year 4 2007-2008	Year 5 2008-2009
Grade 3	X	X	X		
Grade 4		X	X	X	
Grade 5			X	X	X

* This is the year in which standardized testing in science at fifth grade began.

Questionnaire collection. At the end of each school year, all third, fourth, and fifth grade teachers were asked to complete a questionnaire. The questionnaire took 30 – 45 minutes to complete. A small compensation was offered to participating teachers. Teachers whose grade levels were participating in the intervention completed the questionnaire at the final workshop of the year. Teachers who were not participating in the intervention were asked to complete the questionnaire at their school site. All teachers were asked to complete the questionnaire, but only teachers who taught science were included in this research. The number of teachers per school ranged from 17 to 43 with a mean of 31 teachers per school.

For the purpose of this research, when a teacher completed the questionnaire prior to beginning the intervention, the time was coded as Baseline. When a teacher completed the questionnaire at the end of his/her respective first year of participating in the intervention, time was coded as Time 1. Similarly, at the end of his/her respective second year of participating in the intervention, time was coded as Time 2; and at the end of his/her respective third year of participating in the intervention, time was coded as Time 3. A teacher could start participation during any time of the five-year intervention. For example, if a third grade teacher started teaching at a school during Year 3 of the intervention, when the teacher completed the questionnaire at the end of the year, the time would be coded as Time 1 since it was his/her first year of participating in the intervention.

Table 3 shows the number of teachers who completed the questionnaire at each Time. The maximum number of time points that a teacher could have is four (Group 1), in which case the teacher completed a baseline questionnaire, then participated in three years of the intervention and completed the questionnaire at the end of each year. Groups 8 through 11 represent cases where a teacher was unavailable at a scheduled time to complete the questionnaire. The other groups (Groups 2 through 7) represent missing data due to teacher attrition, (i.e., either leaving a school prior to participating in all three years of the intervention or entering a school after the baseline data had been collected).

Table 3

Patterns of Data Collected (N = 191)

	N	Baseline	Time 1	Time 2	Time 3
Group 1	18	X	X	X	X
Group 2	19	X	X	X	
Group 3	17	X	X		
Group 4	31	X			
Group 5	15		X	X	X
Group 6	28		X	X	
Group 7	58		X		
Group 8	1			X	
Group 9	1		X		X
Group 10	2	X		X	
Group 11	1	X		X	X
Totals	191	88	156	84	35

Variables

Outcome variable. The ROP scale consisted of ten items asking the teacher the frequency in which he/she used reform-oriented practices for understanding and inquiry while teaching science. A complete list of the questions that comprised the scales is given in the Appendix. The score for the ROP scale was computed using the average of the responses to the items that comprised the scale. There were 3 teachers missing baseline data; 3 teachers missing Time 1 data; and 1 teacher missing Time 3 data due to non-response of the questionnaire items pertaining to the ROP scale.

The descriptive statistics for each of the time points are displayed in Table 4. The reliability of the obtained scale scores for each time was estimated using Cronbach's alpha. Internal reliability estimates ranged from 0.83 to 0.85.

Table 4

Descriptives for Reform Oriented Practices at Each Time Point

Time	*N*	Alpha	*M*	*SD*
Baseline	88	0.85	2.54	0.55
Time 1	156	0.85	3.13	0.50
Time 2	84	0.83	3.21	0.45
Time 3	35	0.84	3.16	0.49

Predictor variables. The teachers at the school were also asked to complete a background information form when they first started teaching at the school. For the purpose of this analysis, race or ethnicity, highest degree, number of years teaching, number of science classes taken in college, and grade taught were examined as predictor variables. Research has shown that teachers have difficulty promoting students' scientific understanding and inquiry because of their insufficient knowledge of science content and content-specific teaching strategies (Garet, Porter, Desimone, Birman & Yoon, 2001; Kennedy, 1998; Loucks-Horsley, Hewson, Love & Stiles, 1998). In addition, some of these predictors were examined in our previous research examining ROP prior to the intervention (Maerten-Rivera, Penfield, Myers, Lee, & Buxton, 2009). The teacher background predictors (e.g., race, highest degree, number of years teaching, and number of science classes) may affect a teachers' knowledge of science and his/her ability to teach it. Similarly, they may be related to the change in teaching practices.

There are a number of reasons to hypothesize that differences by grade may exist. First, the intervention curriculum was developed such that the science topics and level of inquiry gets progressively more involved, placing more of an emphasis on ROP at higher grades. Thus differences may exist by grade level, especially in change rates. Second, fifth grade teachers may feel more pressure to teach science and to use the recommended reform practices since their students are required to take high-stakes tests in science and the students' test scores will reflect upon the teachers. This may affect initial ROP, and may also be reflected in different change rates.

Table 5 displays the descriptive statistics for the categorical predictor variables. There were no missing data on the predictor variables. The categorical predictors were dummy coded as predictors in the models; the reference group for each set of predictors is noted in Table 5. The number of years teaching ranged from 1 to 39 with a mean of 8.34 years (*SD* = 8.79). The number of science courses taken in college ranged from 0 to 27 with a mean of 4.94 (*SD* = 4.36).

Table 5

Summary of Teacher Background Predictor Variables (N = 191)

Variable	Group	Abbreviation	*N*	%
Ethnicity	White and Other	*	42	21.9
	Black	BLK	79	41.3
	Hispanic	HSP	70	36.6
Grade	Third	*	70	36.6
	Fourth	GR4	61	31.9
	Fifth	GR5	60	31.4
Highest Degree	Bachelor's	*	116	60.7
	Master's or higher	DEG	60	31.4

* Denotes reference group in dummy coded predictors.

The correlation matrix for the predictors is given in Table 6. There are no high correlations between these predictors.

Table 6

Intercorrelations among Teacher Predictor Variables

	TYRS	CLASS	BLK	HSP	GR4	GR5	DEG
TYRS	--	0.25**	-0.09	-0.08	0.03	0.06	0.39**
CLASS	--	--	0.00	-0.04	-0.02	0.09	0.15*
BLK	--	--	--	-0.64**	0.06	0.00	0.15*
HSP	--	--	--	--	-0.01	-0.02	-0.12
GR4	--	--	--	--	--	-0.04	0.09
GR5	--	--	--	--	--	--	-0.04
DEG	--	--	--	--	--	--	--

* $p < .05$
** $p < .01$

Data Analysis

As the first step in this analysis, graphs of the ROP variable over time were examined to explore the type of change displayed. The graph trajectories were examined for displaying no change, linear change, quadratic change, or piecewise change. This information, in addition to a priori expectations, led us to focus on the piecewise change model.

In the piecewise growth model two linear slope factors were modeled. The first growth factor examined the initial change in ROP after first participating in the intervention, while the second examined the change that followed after completing the first year of the intervention. This was done in both MLM and LGM by adding an additional time parameter. In MLM, the first time variable, x_1 had the values of 0, 1, 1, 1 representing baseline, T1, T2, and T3 respectively. The second time variable, x_2 had the values of 0, 0, 1, 2 representing baseline, T1, T2, and T3 respectively. In LGM the λ parameters were fixed as shown below:

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \tag{12}$$

The uncombined (by level) equation below summarizes this a priori MLM model with one level-2 predictor, *z*. For parsimony, the LGM equation is not displayed, but the equivalent parameters can be found by referring back to Table 1.

Level-1 model:

$$ROP_{ti} = \pi_{0i} + \pi_{1i} time_{ti} + \pi_{2i} time_{ti} + e_{ti}$$

Level-2 model:

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01} z_{1i} + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} z_{1i} + r_{1i} \\ \pi_{2i} &= \beta_{20} + \beta_{21} z_{1i} + r_{2i} \end{aligned} \tag{13}$$

Throughout the analyses, models were examined using full maximum likelihood estimation. The LGM models were examined using the Mplus software (Muthén & Muthén, 1998-2007), while the MLM models were examined using the HLM software (Raudenbush, Bryk, & Congdon, 2004). The parameter estimates obtained from the MLM approach and LGM approach were presented and compared for relevant models. Differences were evaluated and discussed as being due to either a rounding issue (i.e., rounding within two decimal places) or a more substantial difference.

In this research, a model building approach was generally employed as suggested by Raudenbush and Bryk (2002) when the level of a priori information does not allow for a strictly confirmatory approach. Broadly, a growth model was built, and then person-level predictors were added to the model to evaluate their effect on initial status and the growth parameters. Typical information used by each approach for evaluating the adequacy of a model was presented.

In MLM it is common to compare competing models instead of indicating the global fit directly. In this research, the deviance[2] (*D*) and number of parameters estimated was presented. Relative fit of models were evaluated using a likelihood ratio chi-square statistic, $\chi^2_{LR} = D_{baseline} - D_{fitted}$. Degrees of freedom for the test statistic was equal to $r_{fitted} - r_{baseline}$ where *r* was the number of parameters estimated. If the chi-square statistic was not statistically significant, then the baseline model (i.e., the simpler model) was retained. In addition, proportion of variance accounted for (PVAF)[3] or pseudo-R^2 (Raudenbush & Bryk, 2002; Singer & Willett, 2003) is often used as an index of goodness of fit of the marginal mean structure since it measures the agreement between the observed responses and model estimated marginal means (Wu, West & Taylor, 2009). Thus, the PVAF in ROP by predictors was presented if the model showed improvement in fit over the previous model. The R^2 can be interpreted as values of less than .09 having a small effect, between .09 and .25 as having a medium effect, and greater than .25 as having a large effect (Cohen, 1977), and therefore, the PVAF will be interpreted in the same way. Both

---

[2] The deviance is a measure of lack of fit of a model: -2ln(likelihood ratio)

[3] Congruent with Raudenbush and Bryk (2002), proportion of variance explained estimates for the level-1 model were determined from the following formula:

$$\frac{\sigma^2 \text{ (baseline model)} - \sigma^2 \text{ (fitted model)}}{\sigma^2 \text{ (baseline model)}}$$

and proportion of variance explained estimates for the level-2 model were determined from the following formula:

$$\frac{\tau_{qq} \text{ (baseline model)} - \tau_{qq} \text{ (fitted model)}}{\tau_{qq} \text{ (baseline model)}}$$

relative fit of models and the PVAF can be examined using the information from the LGM also.

A number of fit indices are available for assessing the LGM model. The model chi-square, χ^2 tests the null hypothesis that the model is correct by measuring the discrepancy between the observed and model implied mean vector and the observed and model implied covariance matrices. Failure to reject the null hypothesis (i.e., $p > .05$) supports the researcher's model. There are limitations with relying solely on the χ^2 statistic since the hypothesis tested by the χ^2 statistic is that the model has perfect population fit which may be implausible (Kline, 2005).

As a result, other fit indices have been used to examine the approximate fit of models. The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are incremental fit indices which measure improvement in model fit of the hypothesized model relative to a baseline model (Bollen & Curran, 2006). They differ because the TLI adjusts for model complexity. The CFI has a 0-1 range, while values of the TLI can exceed 1. The standardized root mean square residual (SRMR) and the root mean square error of approximation (RMSEA) are residual-based fit indices which assess how well a model reproduces the sample data by comparing it to a saturated model that exactly reproduces the sample covariance matrix. The SRMR is a measure of the mean absolute correlation residual, the overall difference between the observed and predicted correlations. The RMSEA estimates the amount of error of approximation per model degree of freedom and takes sample into account. Of these fit indices, the TLI, CFI, and RMSEA have been found to be more sensitive to misspecification of the factor loadings, while SRMR was more sensitive to misspecifications of factor covariance (Hu & Bentler, 1998; Yu, 2002).

Hu and Bentler recommended using SRMR in combination with one of the other fit indices to evaluate model fit, along with using the cutoff of CFI and TLI > .95, SRMR < .08; and RMSEA < .06, along with its 90% confidence interval. The SRMR has been shown to be sensitive to sample size (Sivo, Fan, Witta & White, 2006; Marsh & Bella, 1994), and research examining a correctly specified model demonstrated that with a sample size of 150, the optimal cut-off value for the SRMR was 0.12 and with a sample of 250 it was 0.10 (Sivo et al., 2006).

Examination of time. First, an unconditional multilevel model in which no predictors were included, except individual, was imposed using both approaches. This provided the intraclass correlation coefficient (ICC)[4], which was an estimate of the variance in ROP attributed to individual differences. Next, an unconditional change model was examined in which only the time variables were entered as predictors. If the unconditional variance of a random effect, τ_{qq}, was not statistically significant ($p > .05$), which suggests that there was not significant variation between individuals for a particular Level-1 coefficient, then a baseline model with the relevant random effect fixed to 0 across individuals was examined.

School as a level-2 predictor. The limited sample size did not allow for a three-level model to be analyzed though the data may have an additional level of nesting since the teachers were clustered within six different schools. In order to test the degree to

[4] Congruent with Raudenbush and Bryk (2002), the interclass correlation coefficient was determined from the following formula:

$$\frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

where τ_{00} = the estimated Level-2 variance for the model and σ^2 = the estimated Level-1 variance for the model.

which the assumption of independent observations was violated, in the first step of the analyses school was dummy coded into five variables and examined as a person-level predictor. Separate models were examined with the school predictors entered as a predictor of each random coefficient. A model comparison was conducted as described above. In addition, the PVAF by the school predictors was computed. If the PVAF was small (e.g., less than .09) then the school predictors were not kept in the model. Note that these variables were not retained because they are of direct interest, but because they are control variables.

Background predictors at level-2. In the next step, the background predictors (e.g., ethnicity, highest degree, years teaching, number of science classes, and grade taught) were examined. Ethnicity was dummy coded into two predictors representing Blacks (BLK) and Hispanics (HSP) while Whites served as the reference group. For highest degree, teachers with a bachelor's degree were coded with a 0 while those with a degree higher than the bachelor's were coded with a 1. Grade was dummy coded into two predictors representing Grade 4 (GR4) and Grade 5 (GR5) while Grade 3 served as a reference group. Separate models were examined with these predictors entered as predictors of each random coefficient. In each model, if a predictor was significant it was retained, and the PVAF of the group of predictors for that random effect was calculated along with the PVAF for each individual significant predictor or group of predictors if dummy codes had been used.

Alternative error covariance structures. Exploratory comparisons of the error covariance structure were conducted beginning with the unrestricted error covariance structure since it is the least restrictive and will always have the lowest deviance statistic

(Singer & Willett, 2003). Using both approaches, the models were examined using the following error covariance structures in this order: (1) unrestricted (2) random slopes, homogeneous level-1 variance, (3) first-order autoregressive, and (4) compound symmetric.

Since the models were nested, model comparisons were conducted as discussed above using a likelihood ratio chi-square statistic. In addition, the model fit indices discussed above from the LGM analyses were presented. Finally, for this portion of the analyses, the Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwartz, 1978) were used to compare models as they offer a simple and straightforward way to compare alternative models. Each of these fit statistics compare models based on the log likelihood statistic, but the AIC penalizes based upon the number of model parameters and the BIC penalizes based on the number of parameters and sample size (i.e., in larger samples, the researcher needs a larger improvement before the complex model is preferred over a simpler model). The model with the smaller information criterion fits "better."

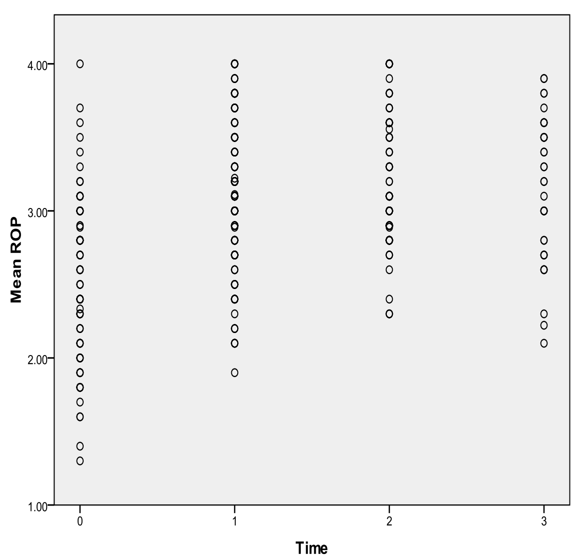
CHAPTER 4

Results

The graphs of the data suggested that there was change in the ROP variable across time. However, most of this change occurred during the first year of the intervention (e.g., from Baseline to T1), with little change in the following years (T1 to T3). As a result, we examined a piecewise change model[5] in which two linear segments of change were measured: from Baseline to T1 (initial change) and from T1 to T3 (secondary change). Figure 1 displays the scatterplot of observed ROP for by time. Figure 2 displays the change trajectories over the four possible time points for all teachers. Overall, there is a fairly steep increase from Baseline to T1, followed by a relatively flat trend from T1 to T3. The terminology used in the MLM framework will be used in the discussion of models from this point forward.

Figure 1

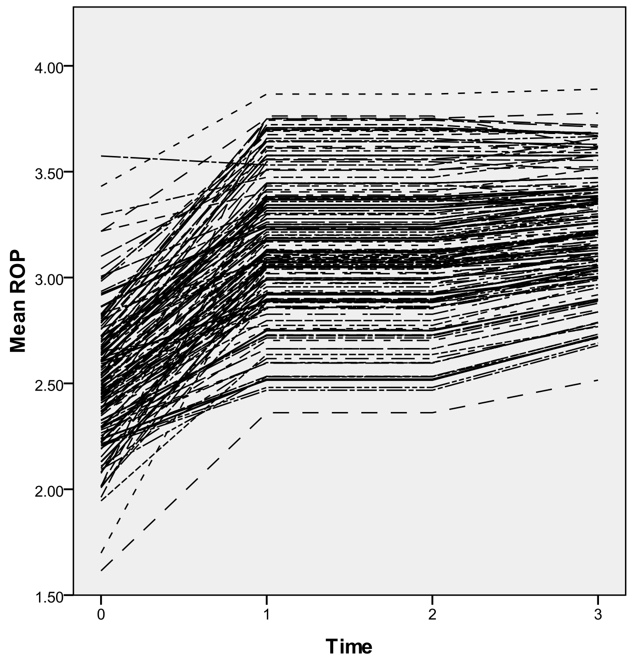
Scatterplot of Observed Values



[5] We examined models with linear and quadratic forms also, but the piecewise model fit the data best and therefore we focus on this model throughout.

Figure 2

Observed Change Trajectories (Emperical Bayes Estimates)



Examination of time. First, the unconditional multilevel was examined. Results from the model are provided in Table 7. The parameter estimates from both approaches were the same.

Table 7

Results of Fully Unconditional Model

		MLM	LGM
Parameter		*Coefficient (SE)*	*Coefficient (SE)*
β_{00}	$[v_0]$	3.01 (0.03)**	3.01 (0.03)**
σ^2	$[\sigma^2(\varepsilon_{ti})]$	0.25 (0.03)**	0.25 (0.03)**
τ_{00}	$[\sigma^2(\zeta_{0i})]$	0.07 (0.02)**	0.07 (0.02)**
		Deviance = 598.13 Estimated Parameters = 3	$\chi^2(11) = 155.03, p < .001$ CFI = 0.00 TLI = -0.19 SRMR = 0.78 RMSEA = 0.26 (0.23-0.30)

* $p < .05$ ** $p < .01$

Using the information provided from the model, the ICC was computed as 0.22, suggesting that 22% of the variance in ROP was due to individual differences while 78% of the variance in ROP was due to within-individual differences. The goodness-of-fit χ^2 test statistic indicated that the unconditional model did not fit the data. The CFI, TLI SRMR, and RMSEA suggested that the model did not exhibit approximate fit to the data.

In the next step an unconditional piecewise change model was examined. The results from this model are displayed in Table 1C of Appendix C. Both approaches yielded the same results for the model parameters. The variance of the random effect for the secondary change slope (τ_{22}) was not significant, so a model comparison was conducted to determine if a model with the secondary change slope constant across individuals fit the data significantly worse. The decrease in model deviance from the unconditional piecewise change model with the variance of secondary change fixed (i.e., baseline model) to the unconditional piecewise change model with the variance of secondary change random (i.e., fitted model) was not statistically significant, $\chi^2_{LR}(3) = 1.62$, $p > .500$. Therefore the unconditional piecewise change model with the variance of secondary change fixed (i.e., baseline model) was retained. The results from this model are presented in Table 8.

The PVAF by the time predictors was 0.64 suggesting that the time predictor variables accounted for 64% of the within-teacher variance in ROP scores. The goodness-of-fit χ^2 test statistic obtained from the LGM indicated that the change model fit the data. The CFI, TLI, and RMSEA also suggested that the model exhibited approximate fit to the data. However, the SRMR suggested that the model did not exhibit approximate fit to the data. This may be due to the small sample size.

Table 8

Results of Unconditional Piecewise Change Model with Variance of Secondary Change Fixed

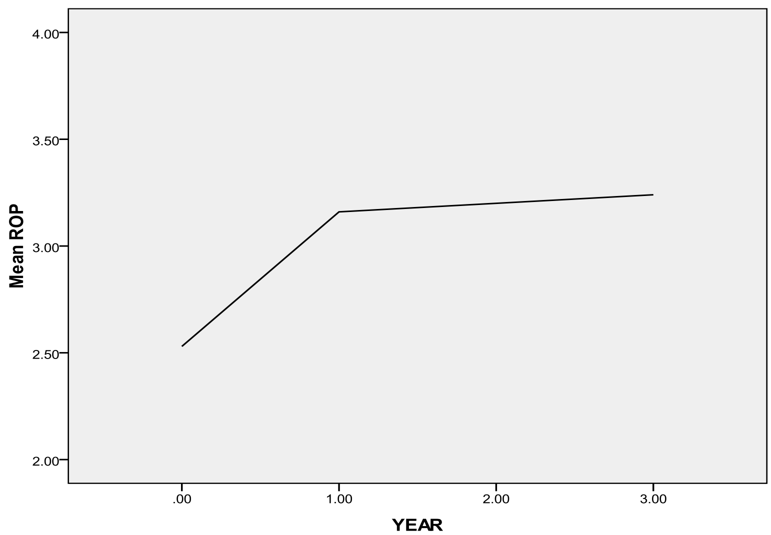
		MLM	LGM
Parameter		*Coefficient (SE)*	*Coefficient (SE)*
β_{00}	$[v_0]$	2.53 (0.06)**	2.53 (0.06)**
β_{10}	$[v_1]$	0.63 (0.06)**	0.63 (0.06)**
β_{20}	$[v_2]$	0.04 (0.03)	0.04 (0.03)**
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09 (0.01)**	0.09 (0.01)**
$\tau_{00}(r_0)$	$[\sigma^2(\zeta_0)]$	0.21 (0.05)**	0.21 (0.05)**
$\tau_{11}(r_1)$	$[\sigma^2(\zeta_1)]$	0.19 (0.06)**	0.19 (0.06)**
$\tau_{10}(r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.13 (0.04)**	-0.13 (0.05)**
		Deviance = 449.83 Estimated Parameters = 7	$\chi^2(7) = 6.73$, $p = .458$ CFI = 1.00 TLI = 1.00 SRMR = 0.20 RMSEA = 0.00 (0.00-0.09)

* $p < .05$ ** $p < .01$

Parameters of interest are the regression weights, βs, and the variances and covariances of r, τ_{00}, τ_{11}, and τ_{01} at the between-individual level, and variance of *e*, σ^2 at the within-individual level. Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, and covariances. The coefficient for the intercept (β_{00}) was 2.53 which was statistically significant indicating that the average true ROP across teachers at baseline was 2.53 points. The coefficient for initial change (β_{10}) was 0.63 which was statistically significant indicating that the average true change in ROP across teachers from baseline to one year after the intervention was 0.63 points. The coefficient for secondary change (β_{20}) was 0.04 which was not statistically significant indicating that the average true change in ROP across teachers from one year after the intervention to three years after the intervention was not different from zero. Figure 2 graphically depicts the change in ROP based on the model results.

Figure 2

Change in ROP Over Time



The estimate of the variance component σ^2 was 0.09, which was statistically significant suggesting that there was a significant amount of within-teacher variance in ROP from baseline to Time 3 even after controlling for both initial change and secondary change. The estimate of the variance component τ_{00} was 0.21, which was statistically significant suggesting that there was a significant amount of between-teacher variation around the average true ROP scores at baseline. The estimate of the variance component τ_{11} was 0.19, which was statistically significant suggesting that there was a significant amount of between-teacher variation around the average initial true change in ROP. The estimate of the covariance component τ_{10} was -0.13 (the correlation was -0.64), which was statistically significant suggesting that there was a significant amount of between-teacher covariation between the within-teacher estimates of true ROP at baseline, and the within-teacher estimates of initial true change in ROP across teachers. On average, teachers with lower ROP at baseline experienced a larger initial increase in ROP after one year of the intervention.

School as a level-2 predictor. In the next model, the school dummy predictors were included as predictors of the intercept, π_{00}. The results of this model are presented in Table 9. Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, and covariances. The deviance from the unconditional piecewise change model (i.e., baseline model) to the model with the school as a level-2 predictor of the intercept (i.e., fitted model) was statistically significant, $\chi^2_{LR}(5) = 19.78$, $p = .002$. Therefore the model with the school as a level-2 predictor of the intercept was retained.

The PVAF in the intercept by the group of school predictors was 0.05 suggesting that the school predictors accounted for 5% of the variation in the average true ROP across teachers at baseline. The goodness-of-fit χ^2 test statistic obtained from the LGM, indicated that the retained model fit the data. The CFI, TLI, and RMSEA also suggested that the model exhibited approximate fit to the data. The SRMR was reduced from 0.20 in the unconditional change model to 0.12 in the conditional model suggesting improvement in approximate fit to the data. Overall, there were small differences in initial ROP based on the school in which the teacher taught.

Table 9

Results of Unconditional Piecewise Change Model with School Variables in Intercept

		MLM	LGM
Parameter		*Coefficient (SE)*	*Coefficient (SE)*
β_{00}	$[\nu_0]$	2.54 (0.06)**	2.54 (0.06)**
β_{01}	$[\gamma_{01}]$	0.23 (0.13)	0.23 (0.13)
β_{02}	$[\gamma_{02}]$	0.40 (0.12)**	0.40 (0.12)**
β_{03}	$[\gamma_{03}]$	0.19 (0.12)	0.19 (0.12)
β_{04}	$[\gamma_{04}]$	0.04 (0.12)	0.04 (0.12)
β_{05}	$[\gamma_{05}]$	0.05 (0.12)	0.06 (0.12)
β_{10}	$[\nu_1]$	0.61 (0.06)**	0.62 (0.06)**
β_{20}	$[\nu_2]$	0.04 (0.03)	0.04 (0.03)
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09 (0.01)**	0.09 (0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.20 (0.04)**	0.20 (0.05)**
$\tau_{11}\ (r_1)$	$[\sigma^2\ (\zeta_1)]$	0.19 (0.06)**	0.19 (0.06)**
$\tau_{10}\ (r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.13 (0.04)**	-0.13 (0.04)**
		Deviance = 430.05 Estimated Parameters = 12	$\chi^2(22) = 24.23, p = .335$ CFI = 0.97 TLI = 0.97 SRMR = 0.12 RMSEA = 0.02 (0.00-0.07)

* $p < .05$ ** $p < .01$

β_{01}=SCHOOL1, β_{02}=SCHOOL2, β_{03}=SCHOOL3, β_{04}=SCHOOL4, β_{05}=SCHOOL5

Next, the school predictors were included as predictors of the initial change slope, π_{10}. The results of this model are presented in Table 2C of Appendix C. Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, and covariances. The decrease in model deviance from the model with the school as a level-2 predictor of the intercept only (i.e., baseline model) to the model with the school as a level-2 predictor of both the intercept and initial change slope (i.e., fitted model) was not statistically significant, χ^2_{LR} (5) = 10.90, *p* = .053. Therefore, the simpler model with school as a level-2 predictor of the intercept only was retained.

Background predictors at level-2. The teacher background predictors were entered as predictors of the intercept, π_{00}. The results of this model are presented in Table 3C of Appendix C. Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, and covariances. The decrease in model deviance from the model with only the school as a level-2 predictor of the intercept (i.e., baseline model) to the model with the school and teacher background variables as level-2 predictors of the intercept (i.e., fitted model) was not statistically significant, χ^2_{LR} (7) = 8.38, *p* = .300. Therefore, the simpler model with only the school as a level-2 predictor of the intercept was retained.

Next, the teacher background predictors were entered as predictors of the initial change slope, π_{10}. The results of this model are presented in Table 4C of Appendix C. Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, and covariances. The decrease in model deviance from the model with only the school as a level-2 predictor of the intercept (i.e., baseline model) to the model with school as a level-2 predictor of the intercept and teacher background variables as level-2 predictors of the slope (i.e., fitted model) was not statistically significant, χ^2_{LR} (7) = 13.22, *p* = .066. Therefore, the simpler model with only school as a level-2 predictor of the intercept was retained.

Alternative error covariance structures. The results from the models examining the alternative error covariance structures are presented in Tables 10 through 13. Table 10 displays the models that were examined along with their fit indices and model deviance comparisons. Table 11 displays the fixed effects estimates obtained from each of the models while Table 12 displays the estimates of the variance components.

Both approaches yielded the same results for the estimates of intercepts, regression weights, standard errors, variances, covariances, and autocorrelation parameters (when applicable) for each model. The unrestricted model (Model 1) had the lowest deviance; however, this model estimated the most parameters. Results from the model comparisons displayed in Table 10b show the random slopes, homogeneous level-1 model (Model 2) did not fit the data significantly worse than Model 1 (χ^2_{LR} (6) = 3.18, p > .500). The first-order auto regressive model (Model 3) was slightly more parsimonious than Model 2 and did fit the data significantly worse than both Model 2 (χ^2_{LR} (1) = 16.06, p < .001) and Model 1(χ^2_{LR} (7) = 19.24, p = .008). The compound symmetry model (Model 4) did fit the data significantly worse than Model 1 (χ^2_{LR} (8) = 19.93, p = .011) and Model 2 (χ^2_{LR} (2) = 16.75, p < .001), but did not fit the data significantly worse than Model 3 (χ^2_{LR} (1) = 0.69, p > .500).

The AIC and BIC both favor Model 2, as do the CFI, TLI, and RMSEA from the LGM framework. However, the SRMR is lowest for the unrestricted model (0.06) and increases for Model 2 (0.12), Model 3 (0.17), and Model 4 (0.17); recall that based on Hu and Bentler's (1998) criteria values <.08 of the SRMR suggests approximate model fit; however this index is especially sensitive to sample size and later research suggests the cut-off value of 0.12 with a sample of 150 (Sivo et al., 2006). Based on all of the fit indices, it would seem reasonable to accept Model 2, which is the structure commonly used in multilevel models.

Table 10

Comparison of Fit Indices across Alternative Models

(a) Models Summary	*Deviance*	*# of parameters estimated*	*AIC*	*BIC*	χ^2	*CFI*	*TLI*	*SRMR*	*RMSEA*
(1) Unrestricted	428.71	18	464.71	523.25	$\chi^2(16) = 21.05$, $p = .177$	0.94	0.90	0.06	0.04 (0.00, 0.08)
(2) Random slope, homogeneous level-1	431.89	12	455.89	494.92	$\chi^2(22) = 24.23$, $p = .335$	0.97	0.97	0.12	0.02 (0.00, 0.07)
(3) First-order autoregressive	447.95	11	469.96	505.73	$\chi^2(23) = 40.30$, $p = .014$	0.79	0.77	0.17	0.06 (0.03, 0.09)
(4) Compound symmetry	448.64	10	468.64	501.16	$\chi^2(24) = 40.98$, $p = .017$	0.80	0.78	0.17	0.06 (0.03, 0.09)

(b) Model Comparisons	*Difference in Deviance*	*df*	*p*
Model 1 versus Model 2	3.18	6	>.500
Model 1 versus Model 3	19.24	7	.008
Model 1 versus Model 4	19.93	8	.011
Model 2 versus Model 3	16.06	1	<.001
Model 2 versus Model 4	16.75	2	<.001
Model 3 versus Model 4	0.69	1	<.500

Table 11

Comparison of Fixed Effects and Standard Error Estimates for Five Models

	Model 1		*Model 2*		*Model 3*		*Model 4*	
	MLM	LGM	MLM	LGM	MLM	LGM	MLM	LGM
β_{00} [v_0]	2.54	2.54	2.54	2.54	2.52	2.52	2.52	2.52
	(0.06)**	(0.06)**	(0.06)**	(0.06)**	(0.05)**	(0.05)**	(0.05)**	(0.05)**
β_{01} [γ_{01}]	0.22	0.22	0.23	0.23	0.23	0.23	0.23	0.23
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.14)
β_{02} [γ_{02}]	0.40	0.40	0.40	0.40	0.41	0.41	0.40	0.40
	(0.12)**	(0.12)**	(0.12)**	(0.12)**	(0.12)**	(0.12)**	(0.12)**	(0.12)
β_{03} [γ_{03}]	0.18	0.18	0.19	0.19	0.20	0.20	0.20	0.20
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
β_{04} [γ_{04}]	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
β_{05} [γ_{05}]	0.07	0.07	0.05	0.06	0.07	0.07	0.07	0.07
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
β_{10} [v_1]	0.61	0.61	0.61	0.62	0.64	0.64	0.64	0.64
	(0.06)**	(0.07)**	(0.06)**	(0.06)**	(0.05)**	(0.05)**	(0.05)**	(0.05)**
β_{20} [v_2]	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)

* $p < .05$ ** $p < .01$ Standard errors are in ()

Table 12

Comparison of Variance-Covariance Estimates for Five Models

		MLM		LGM	
		Coefficient (SE)		*Coefficient (SE)*	
Model 1					
$\sigma^2(e_{t_0})$	$[\sigma^2(\varepsilon_{t_0})]$	0.28	(0.04)**	0.28	(0.04)**
$\sigma^2(e_{t_1})$	$[\sigma^2(\varepsilon_{t_1})]$	0.22	(0.03)**	0.22	(0.03)**
$\sigma^2(e_{t_2})$	$[\sigma^2(\varepsilon_{t_2})]$	0.18	(0.03)**	0.18	(0.03)**
$\sigma^2(e_{t_3})$	$[\sigma^2(\varepsilon_{t_3})]$	0.21	(0.05)**	0.21	(0.05)**
$\sigma^2(e_{t_0}, e_{t_1})$	$[\sigma^2(\varepsilon_{t_0}, \varepsilon_{t_1})]$	0.06	(0.03)	0.06	(0.03)
$\sigma^2(e_{t_0}, e_{t_2})$	$[\sigma^2(\varepsilon_{t_0}, \varepsilon_{t_2})]$	0.05	(0.03)	0.05	(0.03)
$\sigma^2(e_{t_0}, e_{t_3})$	$[\sigma^2(\varepsilon_{t_0}, \varepsilon_{t_3})]$	0.10	(0.04)*	0.10	(0.05)*
$\sigma^2(e_{t_1}, e_{t_2})$	$[\sigma^2(\varepsilon_{t_1}, \varepsilon_{t_2})]$	0.12	(0.02)**	0.12	(0.02)**
$\sigma^2(e_{t_1}, e_{t_3})$	$[\sigma^2(\varepsilon_{t_1}, \varepsilon_{t_3})]$	0.13	(0.03)**	0.13	(0.03)**
$\sigma^2(e_{t_2}, e_{t_3})$	$[\sigma^2(\varepsilon_{t_2}, \varepsilon_{t_3})]$	0.09	(0.03)**	0.09	(0.03)**
Model 2					
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09	(0.01)**	0.09	(0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.20	(0.04)**	0.20	(0.05)**
$\tau_{11}\ (r_1)$	$[\sigma^2(\zeta_1)]$	0.19	(0.06)**	0.19	(0.06)**
$\tau_{10}\ (r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.13	(0.04)**	-0.13	(0.04)**
Model 3					
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.14	(0.03)**	0.14	(0.03)**
ρ	$[\rho]$	0.14	(0.03)**	0.14	(0.03)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.09	(0.03)**	0.09	(0.03)**
Model 4					
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.13	(0.01)**	0.13	(0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.10	(0.02)**	0.10	(0.02)**

* $p < .05$ ** $p < .01$

Finally, Table 13 displays the fitted error covariance matrices, and the correlation matrices. These can be helpful when examining the structure being hypothesized by the alternative model, and how well they replicate the data. The covariance matrix for the unrestricted model (Model 1) represents the values determined by the data without restrictions; the variance of the baseline measurement (0.28) was higher than that of the other three measurements, which were very similar (0.22, 0.18, 0.21). This suggests that over time, the teachers' use of ROP became more similar.

Additionally, the correlation matrix shows that generally the correlations between the baseline measurement and the other measurements were lower than the correlations between the latter measurement occasions, with the interesting exception of the baseline measure with the T3 measure. This general pattern was captured fairly well by Model 2. Recall that the secondary change slope was fixed, while the initial change slope was random. Since the time predictor pertaining to the initial change slope had the values of 0, 1, 1, 1, the covariance and correlation matrices of Model 2 reflect this pattern, as the error covariance components are weighted by this time variable (see Equation 9). In Model 2 the variance for the baseline time point differs from the variance estimate pertaining to times 1, 2, and 3. Similarly, the covariance and correlation estimates are aligned such that those including the baseline measure differ than those from the other time points.

Inspection of the correlation matrix for Model 1 and Model 3 shows that the most disagreement between the models pertains to the correlation estimates for Baseline with T3 and T2 with T3. It is difficult to speculate why this pattern would appear in the unrestricted model. It may instead be a result of missing data, particularly at T3, causing these estimates to be less precise.

Table 13

Fitted Within-Person Covariance and Correlation Matrices

(a) Covariance Matrix			
Model 1	Model 2	Model 3	Model 4
$\begin{bmatrix} 0.28 & 0.06 & 0.05 & 0.10 \\ & 0.22 & 0.12 & 0.13 \\ & & 0.18 & 0.09 \\ & & & 0.21 \end{bmatrix}$	$\begin{bmatrix} 0.28 & 0.07 & 0.07 & 0.07 \\ & 0.21 & 0.12 & 0.12 \\ & & 0.21 & 0.12 \\ & & & 0.21 \end{bmatrix}$	$\begin{bmatrix} 0.23 & 0.11 & 0.09 & 0.09 \\ & 0.23 & 0.11 & 0.09 \\ & & 0.23 & 0.11 \\ & & & 0.23 \end{bmatrix}$	$\begin{bmatrix} 0.23 & 0.11 & 0.11 & 0.11 \\ & 0.23 & 0.11 & 0.11 \\ & & 0.23 & 0.11 \\ & & & 0.23 \end{bmatrix}$
(b) Correlation Matrix			
$\begin{bmatrix} 1.00 & 0.25 & 0.21 & 0.41 \\ & 1.00 & 0.60 & 0.60 \\ & & 1.00 & 0.45 \\ & & & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.27 & 0.27 & 0.27 \\ & 1.00 & 0.58 & 0.58 \\ & & 1.00 & 0.58 \\ & & & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.47 & 0.39 & 0.38 \\ & 1.00 & 0.47 & 0.39 \\ & & 1.00 & 0.47 \\ & & & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.44 & 0.44 & 0.44 \\ & 1.00 & 0.44 & 0.44 \\ & & 1.00 & 0.44 \\ & & & 1.00 \end{bmatrix}$

CHAPTER 5

Discussion

This study compared results obtained from the MLM and LGM approach to modeling change using a piecewise change model. After comparing a model using the standard multilevel modeling error covariance structure other models were examined using alternative error covariance structures. The discussion below focuses on how the results of the study add new insights to the existing literature while also confirming the results of previous research.

Discussion of Findings

The current study used data collected as part of a teacher professional development intervention in science, which took place over five years and consequently had missing data. The model was able to converge using both the MLM and LGM approach despite the moderate sample size and lack of complete data for all respondents at all time points (research question 1). In support of one widely noted advantage of both MLM and LGM to handle missing data, this research demonstrates that these modern approaches were able to conduct longitudinal analysis of missing data with relative ease, whereas it would be difficult using more traditional analysis approaches (e.g., ANOVA, regression).

This research provided an applied example of the use of a piecewise change model in an education setting. Three patterns of results about teacher change in ROP are noteworthy. First, analyses of the ROP variable showed that the change trajectory did not follow a single linear trend or a model quadratic in time. Rather two separate linear trajectories were modeled. The results from the models demonstrated that the teachers' use of reform-oriented practices increased after participating in the first year of the intervention and that this level was sustained, though did not increase significantly in subsequent years. This finding is similar to that of another study on teacher professional development in science where large gains occurred during the first year but no change was seen in the second and third years (Supovitz, Mayer, & Kahle; 2000) and the pattern was described as short-term growth and long-term stability.

Second, the standardized covariance estimate between the estimates of initial ROP and change in ROP suggests a relationship in which teachers with lower ROP at baseline experience a larger initial increase in ROP. Yet none of the predictor variables entered into the model were significant predictors of teacher change.

Finally, there appeared to be more variation in the use of reform-oriented practices at baseline and this variation decreased in subsequent years. This information tells us that the intervention was successful in that after participating in the intervention the teachers' tended to use reform-oriented practices more frequently. The success of the intervention did not depend on any of the predictors that we assessed, and, as a group, the teachers became more similar in their use of reform-oriented practices over time.

Research has demonstrated that the MLM and LGM approach calculate estimates in a similar way (Rovine & Molenaar, 1998, 2000). Previous research using an applied analysis to examine both a linear and quadratic change trajectory found that MLM and LGM provided similar fixed effects estimates, but slightly different estimates of the variances and covariances. The researchers reported that the HLM software that was used for the MLM analysis did not provide estimates of the standard errors of the variances and covariances, which they noted as a weakness (Chou et al., 1998). In our research using the HLM software for the MLM analysis the estimates of the fixed effects, and variances and covariances were nearly identical (research question 2). In addition, using full information maximum likelihood estimation (estimates of standard errors are not available using restricted maximum likelihood) we were able to obtain estimates of the standard errors for all parameters, including variances and covariances, and these again were nearly identical to those obtained from the LGM (research question 2). Thus this study extends previous research by demonstrating that similar estimates for all parameters and standard errors of these parameters can be obtained when modeling a piecewise change model.

Researchers have noted the inability to model various error covariance structures as a weakness of MLM (Chou et al., 1998; Rovine & Melenaar, 1998, 2000). Recent advancements in software allow certain pre-defined error covariance structures to be examined using the MLM framework. Our study demonstrated that estimates of the fixed effects, variances and covariances, and standard errors of these estimates were nearly identical under these models hypothesizing alternative error covariance structures (research question 3). This finding extends previous research that has been limited to comparing models using only the standard error covariance structure of MLM. Certain models can be examined using MLM, but we acknowledge that the LGM is more flexible, yet researchers must be cautious as sample size can limit model complexity, and the hypothesized model should have some theoretical basis.

Finally, a number of statistics were used to compare models. The deviance statistic can be obtained using either approach and models can be compared by examining the χ^2 difference statistic, AIC, and BIC computed from information from the deviance. Similarly the effect size or PVAF can be examined using information obtained from either approach, though this evaluation of models is more typically used in MLM. The model χ^2, CFI, TLI, SRMR, and RMSEA are fit statistics unique to LGM. In most models, generally these fit statistics were in agreement, and the models retained would have been retained regardless of the approach (i.e., MLM, LGM) used (research question 4).

The one fit statistic that did perform differently than the others was the SRMR. Using Hu and Bentler's (1998) criteria the SRMR suggested that the model did not fit the data for nearly all of the models involved in the model building process, while all of the other fit statistics suggested that the model did fit the data. Research has shown that the SRMR is more sensitive to sample size (Sivo et al., 2006; Marsh & Bella, 1994) and to misspecification of factor covariance (Hu & Bentler, 1998, Yu, 2002). Our research suggests that these both may play a factor. If Sivo et al.'s cut-off value for SRMR for a sample size of 150 (0.12) was used, the model retained based on the other fit indices would have fit the data. When comparing models with alternative error covariance structures, the SRMR suggested that the unrestricted model (i.e., the model that takes on the values of the data) had the best model fit (SRMR = 0.06), while all other fit indices suggested that the model with the traditional MLM covariance structure fit the data. Thus, it appears that the SRMR may be more sensitive to the misspecification in the error covariance structure pertaining to those covariance/correlation estimates related to T3. Additionally, the SRMR does not penalize for model complexity, while many of the other fit indices do.

Since LGM comes out of the SEM framework this analysis provides overall model fit information while MLM does not (Chou et al., 1998; Curran, 2003; Rovine & Molenaar, 1998, 2000). However, the PVAF can be used to examine model fit while the deviance statistic, χ^2 difference statistic, AIC, and BIC can be used to compare models. In addition, the many fit indices of SEM have been created with the claim that the fit index would more unambiguously point to model adequacy as compared to the χ^2 test, while no index has been able to support this (Hu & Bentler, 1999). Thus in most research, the fit statistic which best fits the model is reported. For a complete comparison of fit indices used in MLM and LGM see Wu et al. (2009) who notes that both frameworks provide information on model fit, but those used in MLM are less familiar. They also offer a full discussion on the strengths and weaknesses of each index in the context of change models where fit indices may not be evaluating all portions of the model.

Limitations of Study

This study suffered from a few limitations. First, we acknowledge that the sample size (specifically, the number of schools) was small, which limited the complexity of the model. The main limitation that this created was that we were not able to examine the effect of school as another level of dependency. Since we were not able to do this, we entered the school as a predictor at the teacher-level in order to control for its’ effect. Future research that collects data from enough schools to allow for a third level of modeling could advance understanding in this area.

Another limitation was in the number of time points in which data was collected. Data were collected at four different occasions. If more occasions had been collected we could have examined additional change that might have occured after being involved in the intervention for a longer period of time or even the effects of sustainability once the intervention had ended. Additionally, if more occasions had been included, additional error covariance structures could have been examined, such as a random slopes and heterogeneous level-1 variance model. Although the number of time points was relatively modest, the time over which the data were collected was spread out such that we could reasonably expect change on the outcome.

Even with these limitations, we feel that the results are useful as it does offer an applied example from an educational setting. A moderate sample size was used and the research suffered from missing data. This type of situation is often encountered when conducting research in educational settings.

Future studies could compare the two approaches using more complex models with larger sample sizes. For example, the MLM framework has been used to model outcomes with multiple measures, something that at one point only LGM was capable of. Similarly, LGM now offers ways to model different levels or nestings which is a concept that originated in the MLM framework (Raudenbush & Bryk, 2002; Singer & Willett, 2003). It seems that both approaches are borrowing from the other, but more research needs to be conducted to determine when similar results are obtained and when differences may exist. Applied studies of such similarities and differences are useful for researchers who may be deciding which technique to use.

Another area where more research is needed is in examining the similarities and differences in the fit indices when applied to change models. Wu et al. (2009) offer a starting point for this, and point out that different fit indices offer different information when applied to change models since the mean structure is included, and model fit can pertain to different portions of the model.

Implications

As the MLM and LGM are closely related the advantages, disadvantages, and compatibility of both approaches have been of interest to many researchers. It has been noted that differences in the two approaches are not due to either model's inabilities to produce the same estimates, but rather limitations in the software packages under each approach (Raudenbush & Bryk, 2002). Our study demonstrated that similar results can be obtained from the two frameworks. As a result, we encourage researchers to be aware that both methods can be used to examine change, to understand the advantages and disadvantages of each approach, and to consider which framework can better help them to answer their research questions.

REFERENCES

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle, In B.N. Petrov & F. Csaki (eds.), *Second international Symposium of information theory* (pp. 267-228). Akademiai Kiado, Budapest, Hungary.

Baker, G. A. (1954). Factor analysis of relative growth. *Growth*, *18*, 137-143.

Bock, R. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw Hill.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

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

Bryk, A., Raudenbush, S., Seltzer, M., & Congdon, R. (1989). *An introduction to HLM: Computer program and users' guide.* Chicago: Scientific Software.

Chou, C., Bentler, P. & Pentz, M.A. (1998). Comparisons of two statistcal approahces to study growth curves: the multilevel model and the latent curve analysis. *Structural Equation Modeling, 5*(3). 247-266.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Curran, P. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529-569.

Curran, P. & Peterman, M. (2005, October). *A curious discrepancy between multilevel and structural equation growth curve models with time-varying covariates.* Paper presented at annual meeting of the Society of Multivariate Experimental Psychology, Lake Tahoe, NV.

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

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, *76*(374), 341-353.

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

Duncan, T., Duncan, S., Strycker, L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Enders, C.K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8,* 128-141.

Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2004). *Applied longitudinal analysis.* New York: Wiley.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Griffin.

Goldstein, H., Healy, M.J.R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13,* 1643-1655.

Hancock, G., & Lawrence, F. (2006). Using latent growth models to evaluate longitudinal change. In G. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 171-196). Greenwich, CT: Information Age.

Heck, R., & Thomas, S. (2000). *An introduction to multilevel modeling techniques.* Mahwah, NJ: Lawrence Erlbaum.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jenrich, R., & Schluchter, M. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics, 42*, 805-820.

Kennedy, M. M. (1998). Education reform and subject matter knowledge. *Journal of Research and Science Teaching, 35*(3), 249-263.

Kline, R.B. (2005). *Principles and practices of structural equation modeling.* New York, NY: The Guilford Press.

Kreft, I., & de Leeuw (2004). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Laird, N., & Ware, H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Lindley, D. V. & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, *34*(1), 1-41.

Longford, N. (1993). *Random coefficients models.* Oxford: Clarendon.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics.* Thousand Oaks, CA: Corwin Press.

MacCallum, R., Kim, C., Malarkey, W., & Kiecolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariante Behavioral Research, 32*(3), 215-253.

Maerten-Rivera, J., Penfield, R., Myers, N., Lee, O., & Buxton, C. (2009). School and teacher predictors of science instruction practices with English language learners in urban elementary schools. *Journal of Women and Minorities in Science and Engineering, 15*(2), 93-118.

Marsh, H.W. & Balla, J. (1994). Goodness of fit in confirmatory facto analysis: the effects of sample size and model parsimony. *Quality & Quantity, 2,* 185-217.

McArdle, J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*(1), 110-133.

Mehta, P., & West, S. (2000). *Putting the individual back into individual growth curves. Psychological Methods, 5*(1) 23-43.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376-398.

Muthén, L., & Muthén, B. (2007). Mplus 5 [Computer software]. Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87-99). San Diego: Academic Press.

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

Rao, C. (1958). Some statistical methods for the comaparison of growth curve. *Biometrics, 14*(1), 1-17.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S., Bryk, A., & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Rogosa, D., & Willett, J. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*(2), 203-228.

Rovine, M., & Molenaar, P. (1998). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development, 22*(3), 453-473.

Rovine, M., & Molenaar, P. (2000). A structural modleing approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35*(1), 51-88.

Rubin, D.B. (1976) Inference and missing data. *Biometrika, 63,* 581-592.

Schwartz, G. (1978). Esimating the dimensions of a model. *Annals of Statistics, 6,* 461-464.

Singer, J., & Willet, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurence*. New York, NY: Oxford.

Sivo, S.A., Fan, X., Witta, A.L., & Willse, J.T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education, 74(3),* 267-288.

Smith. A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society B*, *35*(1), 67-75.

Strenio, J. L. F., Weisberg, H. I., & Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics*, *39*(1), 71-86.

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy, 14*(3), 331-356.

Tucker, L. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*(1), 19-23.

Weiss, R.E. (2005). *Modeling longitudinal data.* New York: Springer.

Wolfinger, R.D. (1996). Heterogeneous variance-covariance structure for repeated measures. *Journal of Agricultural, biological, and Environmental Statistics, 1*(2), 205-230.

Wu, W., West, S.G., & Taylor, A.B. (2009). Evaluating model fit for growth curve models: Intergration of fit indices from SEM and MLM frameworks. *American Psychological Association, 14*(3), 183-201.

Yu, C.Y. (2002) *Evaluating cutoff ctiteria of model fit indices for latent variable models with binary and continuous outcomes.* Unpublished doctoral dissertation, University of California, Los Angeles.

APPENDIX A

Scale Items

Reform-Oriented Practices Scale

(Item 1)	During the last month, how often did **YOU** do the following **in your science lessons**?
(Items 2, 3, 4, 6, 7, and 8)	During the last month, how often did you **ASK STUDENTS** to do following **in your science lessons**?
(Item 5, 9, and 10)	During the last month, how often did you **ASK STUDENTS** to do the following **in your science lessons for at least 10 minutes**?

All questions had the following response options:
Never or almost never, Some lessons, Most lessons, Every lesson

Item	Question
Scientific Understanding	
1	Use students' mistakes to generate class discussion
2	Explain the reasoning behind an idea
3	Apply science concepts to explain natural events or real world situations
4	Talk about things they do at home that are similar to what we do in science class (e.g., measuring, boiling water, freezing water)
5	Discuss their prior knowledge or experience related to the science topic or concept
Scientific Inquiry	
6	Use science process skills (e.g., hypothesize, organize, infer, analyze, evaluate, describe patterns, make models or simulations)
7	Use basic measurement tools (e.g., ruler, thermometer, scale/balance, timer, graduated cylinder)
8	Use everyday, household items (e.g., plastic cups or containers, food coloring, light bulbs, batteries)
9	Analyze relationships using tables, charts, or graphs
10	Write about what was observed and why it happened

APPENDIX B

Additional Literature

History

Statistical analysis of data over time can be traced back to early in the nineteenth century. Studies from as early as the beginning of the twentieth century attempted to address group-level growth. During the 1930's there was a major shift from focusing on estimating growth trajectories for entire groups to being able to estimate specific trajectories for each individual in the group (Bollen & Curran, 2006), which allowed predictors to be used to explain variability in individual growth trajectories. Until the early 1950's all growth models were estimated using analysis of variance techniques. However, several key developments in the middle of the twentieth century moved the estimation of trajectory models into the latent variable framework, where the growth process is not observed directly; instead, growth is observed indirectly through repeated measures.

LGM. The first person to propose using latent variables within the factor analytic framework and trajectory modeling was Baker (1954) concluding that factor analysis could be a useful tool for reducing complex repeated measures to relatively fewer latent factors that could result in a better understanding of the patterns of change. Not long after, Tucker (1958), who worked in the field of psychometrics, proposed a method for using latent factors to estimate known functional forms relating time to the repeated measures. However, he noted several issues such as communality estimation and rotation which remained to be resolved at the time.

Drawing on this work, Meredith and Tisak (1990), who also worked in the field of psychometrics, proposed embedding growth trajectory modeling within the confirmatory latent variable framework used in structural equation models. They demonstrated that the analytic interest is not specifically the repeated measures observed, but instead, the unobserved latent trajectory factors that lead to the repeated measures observed. Around the same time, a multilevel structural equation model designed to handle hierarchically clustered observations was presented (Muthén & Satorra, 1989), which was later adapted for use with longitudinal data (Muthén, 1994). The placement of growth curve models in the context of structural equations capitalized on the advantages of structural equations. Recent developments allow for various modeling strategies, including the use of multiple-indicator latent factors within each time point, the estimation of multiple-group models to evaluate interactions in development over time, the inclusion of mediating influences on growth processes, and the ability to model trajectories simultaneously in two or more variables (Bollen & Curran, 2006).

MLM. The term *hierarchical linear model* was first introduced by Lindley and Smith (1972) and Smith (1973) as part of their work on Bayesian estimation of linear models. Within the context of Bayesian estimation, Lindley and Smith detailed a general framework for nested data with complex error structures. The expectation-maximization (EM) algorithm developed by Dempster, Laird, and Rubin (1977), who worked in the field of biostatistics, made broader applications of Lindley and Smith's work possible and Dempster, Rubin and Tsutakawa (1981) demonstrated applicability of this approach to hierarchical data structures. The use of these approaches to study growth was first applied by Laird and Ware (1982), and Strenio, Weisberg, and Bryk (1983).

Raudenbush and Bryk (2002), working in the educational field, continued to develop on the hierarchical linear modeling approach and made the approach popular by creating software designed to analyze this type of data (Bryk, Raudenbush, Seltzer & Congdon; 1989). According to Kreft and de Leeuw (1998), soon after its introduction the HLM software was adapted as the 'official' software for educational multilevel analysis.

Error Covariance Structure

By combining the level-1 and level-2 models given in Equations 8 and 9, and multiplying out the terms, the following equation can be given

$$y_{ti} = (\beta_{00} + \beta_{01}x_{ti} + \beta_{01}z_{1i} + \beta_{11}z_{1i} \times x_{ti}) + (e_{ti} + r_{0i} + r_{1i}x_{ti}). \qquad (12)$$

The parenthesis separate out the model's structural and stochastic portions. The structural portion contains the hypotheses about the way that the outcome variable changes with time and depends on an explanatory variable z_{1i}. The stochastic portion contains the composite residual, *s*. The value of *s* for an individual *i* on occasion *t* is

$$s_{ti} = [e_{ti} + r_{0i} + r_{1i}x_{ti}]. \qquad (13)$$

The composite residual is a weighted linear combination of the original three random effects from the level-1/level-2 specification (e_{ti}, r_{0i}, and r_{1i} with constants 1, 1, and x_{ti} acting as the weights).

The composite residuals are expected to be independent across people and normally distributed with zero means, but within people are expected to be heteroscedastic and correlated over time, which makes more sense for longitudinal data and offers an advantage over cross-sectional designs. As a result, the residuals of the composite model have a multivariate normal distribution with zero means and block diagonal error covariance structure. The block diagonal means that all of the matrix's elements are zero except those within the blocks along the diagonal. There is one block per individual with the zero elements indicating that the residuals for each individual are independent of all others (e.g., the residuals for individual *i* has zero covariance with everyone else's residual), but the non-zero covariance parameters within each block allow the residuals to covary within person with the variances of the within-person residuals (e.g., the block's diagonal elements) allowed to differ across occasions. The blocks of the error covariance matrix are identical across people as a result of a homogeneity assumption that the composite residuals may be heteroscedastic and dependent within people, but the error structure is repeated identically across people.

The homogeneity assumption can be expressed by writing

$$\boldsymbol{s} \sim \boldsymbol{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_s & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_s & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_s \end{bmatrix}\right) \tag{14}$$

The complete vector of residuals $\mathbf{s}$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and a block-diagonal error covariance matrix made up of submatrices $\boldsymbol{\Sigma}_s$ and $\mathbf{0}$, where

$$\boldsymbol{\Sigma}_s = \begin{bmatrix} \sigma_{s1}^2 & \sigma_{s1s2} & \sigma_{s1s3} & \sigma_{s1s4} \\ \sigma_{s2s1} & \sigma_{s2}^2 & \sigma_{s2s3} & \sigma_{s2s4} \\ \sigma_{s3s1} & \sigma_{s3s2} & \sigma_{s3}^2 & \sigma_{s3s4} \\ \sigma_{s4s1} & \sigma_{s4s2} & \sigma_{s4s3} & \sigma_{s4}^2 \end{bmatrix} \tag{15}$$

In the standard multilevel model for change, the elements of the error covariance blocks possess a powerful dependence on time as can be seen in Equation 13. The composite residual variance in the standard model has a quadratic dependence on time such that it will have a minimum and will increase parabolically and symmetrically over time on either side of its minimum. The covariances of the composite residuals are also dependent on time, in that the covariance contains the product of pairs of times and this product dramatically affects the error covariance when times are large. Similarly, if the level-2 components are close to zero, the composite residual covariances will be close to zero and the error covariance matrix becomes diagonal. From the covariance composite residuals the correlations can be computed which allows the autocorrelations imposed to be examined. From the composite residual autocorrelation matrix, it can be determined how similar observations are based on temporal placement. The estimation of the autocorrelations extends beyond some analyses that assume zero autocorrelation.

APPENDIX C

Additional Tables

Table 1C

Results of Unconditional Piecewise Growth Model

		MLM	LGM
Parameter		*Coefficient (SE)*	*Coefficient (SE)*
β_{00}	$[v_0]$	2.53 (0.06)**	2.53 (0.06)**
β_{10}	$[v_1]$	0.63 (0.06)**	0.63 (0.07)**
β_{20}	$[v_2]$	0.04 (0.03)	0.04 (0.03)
$\tau_{00}(r_0)$	$[\sigma^2(\zeta_0)]$	0.21 (0.05)**	0.21 (0.05)**
$\tau_{11}(r_1)$	$[\sigma^2(\zeta_1)]$	0.22 (0.07)**	0.22 (0.08)**
$\tau_{22}(r_2)$	$[\sigma^2(\zeta_2)]$	0.00 (0.01)	0.00 (0.02)
$\tau_{10}(r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.13 (0.05)**	-0.13 (0.05)**
$\tau_{02}(r_0, r_2)$	$[\sigma^2(\zeta_0, \zeta_2)]$	0.00 (0.02)	0.00 (0.02)
$\tau_{12}(r_1, r_2)$	$[\sigma^2(\zeta_1, \zeta_2)]$	-0.02 (0.03)	-0.02 0.03
		Deviance = 448.21 Estimated Parameters = 10	$\chi^2(4) = 5.11$, $p = .277$ CFI = 0.98 TLI = 0.98 SRMR = 0.19 RMSEA = 0.04 (0.00 – 0.12)

* $p < .05$ ** $p < .01$

Table 2C

Results of Unconditional Piecewise Growth Model with School Variables in Intercept and Initial Slope

		MLM	LGM
Parameter		*Coefficient*	*Coefficient*
β_{00}	$[v_0]$	2.56 (0.05)**	2.56 (0.06)**
β_{01}	$[\gamma_{01}]$	0.07 (0.24)	0.07 (0.24)
β_{02}	$[\gamma_{02}]$	0.48 (0.25)	0.48 (0.25)
β_{03}	$[\gamma_{03}]$	0.14 (0.23)	0.14 (0.23)
β_{04}	$[\gamma_{04}]$	-0.07 (0.23)	-0.07 (0.23)
β_{05}	$[\gamma_{05}]$	0.36 (0.24)	0.36 (0.24)
β_{10}	$[v_1]$	0.60 (0.06)**	0.61 (0.06)**
β_{11}	$[\gamma_{11}]$	0.22 (0.24)	0.22 (0.25)
β_{12}	$[\gamma_{12}]$	-0.09 (0.25)	-0.09 (0.25)
β_{13}	$[\gamma_{13}]$	0.09 (0.25)	0.09 (0.24)
β_{14}	$[\gamma_{14}]$	0.15 (0.24)	0.15 (0.24)
β_{15}	$[\gamma_{15}]$	-0.36 (0.24)	-0.36 (0.25)
β_{20}	$[v_2]$	0.04 (0.03)	0.04 (0.03)
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09 (0.01)**	0.09 (0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.18 (0.04)**	0.18 (0.04)**
$\tau_{11}\ (r_1)$	$[\sigma^2\ (\zeta_1)]$	0.14 (0.05)*	0.14 (0.05)*
$\tau_{10}\ (r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.10 (0.04)*	-0.10 (0.04)*
		Deviance = 419.15 Estimated Parameters = 17	$\chi^2(17) = 13.33$, $p = .714$ CFI = 1.00 TLI = 1.07 SRMR = 0.11 RMSEA = 0.00 (0.00-0.05)

* $p < .05$ ** $p < .01$

β_{01}=SCHOOL1, β_{02}=SCHOOL2, β_{03}=SCHOOL3, β_{04}=SCHOOL4, β_{05}=SCHOOL5

β_{11}=SCHOOL1, β_{12}=SCHOOL2, β_{13}=SCHOOL3, β_{14}=SCHOOL4, β_{15}=SCHOOL5

Table 3C

Results of Piecewise Growth Model with Teacher Background Predictors in Intercept

		MLM	LGM
Parameter		*Coefficient*	*Coefficient*
β_{00}	$[v_0]$	2.54 (0.05)**	2.54 (0.06)**
β_{01}	$[\gamma_{01}]$	0.24 (0.13)	0.24 (0.13)
β_{02}	$[\gamma_{02}]$	0.36 (0.12)**	0.36 (0.12)**
β_{03}	$[\gamma_{03}]$	0.17 (0.13)	0.17 (0.13)
β_{04}	$[\gamma_{04}]$	0.01 (0.13)	0.01 (0.13)
β_{05}	$[\gamma_{05}]$	0.06 (0.12)	0.06 (0.12)
β_{06}	$[\gamma_{06}]$	0.00 (0.00)	0.00 (0.00)
β_{07}	$[\gamma_{07}]$	0.00 (0.01)	0.00 (0.01)
β_{08}	$[\gamma_{08}]$	-0.05 (0.07)	-0.05 (0.07)
β_{09}	$[\gamma_{09}]$	-0.01 (0.07)	-0.01 (0.08)
$\beta_{0,10}$	$[\gamma_{0,10}]$	0.09 (0.07)	0.09 (0.07)
$\beta_{0,11}$	$[\gamma_{0,11}]$	0.18 (0.08)*	0.18 (0.08)*
$\beta_{0,12}$	$[\gamma_{0,12}]$	0.16 (0.09)	0.16 (0.09)
β_{10}	$[v_1]$	0.61 (0.06)**	0.61 (0.07)**
β_{20}	$[v_2]$	0.04 (0.03)	0.04 (0.03)
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09 (0.01)**	0.09 (0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.20 (0.05)**	0.20 (0.05)**
$\tau_{11}\ (r_1)$	$[\sigma^2\ (\zeta_1)]$	0.19 (0.06)**	0.19 (0.06)**
$\tau_{10}\ (r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.14 (0.04)**	-0.14 (0.05)**
		Deviance = 421.67 Estimated Parameters = 19	$\chi^2(43) = 49.10$, $p = .242$ CFI = 0.93 TLI = 0.91 SRMR = 0.08 RMSEA = 0.03 (0.00-0.06)

* $p < .05$ ** $p < .01$

β_{01}=SCHOOL1, β_{02}=SCHOOL2, β_{03}=SCHOOL3, β_{04}=SCHOOL4, β_{05}=SCHOOL5, β_{06}=TEACHYRS, β_{07}=CLASSES, β_{08}= GR4, β_{09}=GR5, $\beta_{0,10}$=DEG, $\beta_{0,11}$=BLK, $\beta_{0,12}$=HSP

Table 4C

Results of Piecewise Growth Model with Teacher Background Predictors in Initial Change Slope

		MLM		LGM	
Parameter		*Coefficient*		*Coefficient*	
β_{00}	$[v_0]$	2.53	(0.06)**	2.53	(0.06)**
β_{01}	$[\gamma_{01}]$	0.23	(0.13)	0.23	(0.13)
β_{02}	$[\gamma_{02}]$	0.34	(0.12)**	0.34	(0.12)**
β_{03}	$[\gamma_{03}]$	0.14	(0.13)	0.14	(0.13)
β_{04}	$[\gamma_{04}]$	-0.02	(0.13)	-0.02	(0.13)
β_{05}	$[\gamma_{05}]$	0.02	(0.12)	0.02	(0.12)
β_{10}	$[v_1]$	0.62	(0.06)**	0.62	(0.07)**
β_{11}	$[\gamma_{11}]$	-0.01	(0.00)	-0.01	(0.00)
β_{12}	$[\gamma_{12}]$	0.00	(0.01)	0.00	(0.01)
β_{13}	$[\gamma_{13}]$	-0.10	(0.08)	-0.10	(0.08)
β_{14}	$[\gamma_{14}]$	0.02	(0.08)	0.02	(0.08)
β_{15}	$[\gamma_{15}]$	0.10	(0.07)	0.10	(0.08)
β_{16}	$[\gamma_{16}]$	0.19	(0.09)*	0.19	(0.09)*
β_{17}	$[\gamma_{17}]$	0.15	(0.09)	0.15	(0.09)
β_{20}	$[v_2]$	0.04	(0.03)	0.04	(0.03)
$\sigma^2(e_{ti})$	$[\sigma^2(\varepsilon_{ti})]$	0.09	(0.01)**	0.09	(0.01)**
$\tau_{00}\ (r_0)$	$[\sigma^2(\zeta_0)]$	0.20	(0.04)**	0.20	(0.05)**
$\tau_{11}\ (r_1)$	$[\sigma^2\ (\zeta_1)]$	0.19	(0.06)**	0.19	(0.06)**
$\tau_{10}\ (r_0, r_1)$	$[\sigma^2(\zeta_0, \zeta_1)]$	-0.14	(0.04)*	-0.14	(0.05)*
		Deviance = 417.76 Estimated Parameters = 19		$\chi^2(43) = 45.19$, $p = .380$ CFI = 0.98 TLI = 0.97 SRMR = 0.07 RMSEA = 0.02 (0.00-0.05)	

* $p < .05$ ** $p < .01$

β_{01}=SCHOOL1, β_{02}=SCHOOL2, β_{03}=SCHOOL3, β_{04}=SCHOOL4, β_{05}=SCHOOL5, β_{11}=TEACHYRS, β_{12}=CLASSES, β_{13}= GR4, β_{14}=GR5, β_{15}=DEG, β_{16}=BLK, β_{17}=HSP