Model Fit and Interpretation of Non-Linear Latent Growth Curve Models

by

Greg W. Welch

B.A., University of Wyoming, 2000

M.S., University of Wyoming, 2005

M.A., University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

the School of Education in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH
School of Education

This dissertation was presented

by

Greg W. Welch

It was defended on

July 2nd, 2007

and approved by

Suzanne Lane, Professor, Research Methodology in Education

Clement A. Stone, Associate Professor, Research Methodology in Education

Satish Iyengar, Professor, Statistics

Dissertation Advisor: Kevin H. Kim, Assistant Professor, Research Methodology in Education

Model Fit and Interpretation of Non-Linear Latent Growth Curve Models

Greg W. Welch, B.A., M.A., M.S.
University of Pittsburgh, 2007

This dissertation investigated the use of various techniques in modeling non-linear change in the context of latent growth modeling. A simulation study was conducted utilizing four between subjects factors: sample size (50, 75, 100, 150, 200, 300 and 500), slope variance (.15, .45 and .75), factor correlation (.15, .45 and .75) and growth curve (exponential, logarithmic and logistic). There was also a single within subjects factor: fit technique (quadratic, unspecified and spline). The outcomes of interest were the χ^2 model fit statistic and the following goodness-of-fit indices: CFI, GFI, AGFI, SRMR and RMSEA. Results indicated the unspecified technique provided the best statistical estimates of model fit while the quadratic technique provided the worst. This result was consistent across all of the between subject factor conditions. The spline technique performed very similarly to the quadratic technique. These results suggest applied researchers should pay very close attention when utilizing polynomial techniques and should also strongly consider the unspecified technique as either the model of choice or as a comparison to results obtained for another model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation would not have been completed without the support, patience and encouragement of many people. I would like to begin by thanking my committee members, Dr. Suazanne Lane, Dr. Clement Stone, Dr. Satish Iyengar and Dr. Kevin Kim, for their support and guidance during this process. I would also like to thank all of my colleagues at the University of Kansas, particularly Dr. William Skorupski and Dr. Bruce Frey, for their recommendations to "just get the damn thing done." I would also like to thank Dr. Skorupski for sharing his ideas on the simulation portion of this dissertation and Dr. Frey for reviewing the introduction. I must also include my good friend, Dr. Jason Castillo, for sharing the trials involved in the journey towards obtaining a Ph.D. and what it really means in the big picture of life. I would like to thank my sister, Becky, for looking up to me. I hope I have met all of your expectations. I thank my wife, Stephanie, for loving me during the times when I was unlovable and being there through this entire process. I would have never been able to do this without you by my side. I thank my children, Matt, Ashley and Breanna, for inspiring me to help make this world a better place through education. Lastly, I must thank my parents, John and Susan, for instilling in me the importance of education and the mindset to finish whatever I start. This is for you.

1.0 Introduction

Longitudinal research is increasingly common in the social and behavioral sciences. Regardless of discipline, incorporating time into the design of a study provides valuable information regarding the process of change. An educational researcher, for instance, assesses the development of mathematics ability by examining change in standardized assessment scores over a period of time. A psychologist studies the development of a particular mental illness or the long term effects of a treatment approach. A sociologist studies attitudes of a community towards a certain political decision. Regardless of the subject area, unit of analysis, or time interval, longitudinal research describes the process of change and the reasoning behind the process of change.

Traditionally, researchers have relied on two wave designs and difference score analyses to address questions regarding change. Two wave designs have garnered criticism from a methodological perspective and, more importantly, from a conceptual perspective (Willett, 1988). Measuring subjects at two points in time does not account for the important developments likely to occur between the measured time points. Rather the two wave design is only able to address the magnitude of change that has occurred between the measured time points. On the other hand, obtaining multiple measurements from a large number of individuals allows change to be appropriately treated as a continuous process (Rogosa, Brandt & Zimowski, 1982; Willett, 1988). A continuous process of change can be represented in various statistical frameworks via a growth model. By definition, growth models are statistical approaches to describing the shape of a developmental trajectory or the rate of change. Growth models are particularly effective for answering questions regarding baseline levels and the direction and magnitude of change in an outcome measure. While assessing baseline levels is straightforward in a growth model, choosing an appropriate method for describing the direction and magnitude of change is difficult.

Specialized approaches have been developed that extend growth models to the realm of measuring individual change. Individual growth modeling (Rogosa et al., 1982; Willett, 1988) provides researchers powerful tools for using continuous multiwave panel data to examine correlates and predictors of systematic interindividual differences in change. Individual growth models examine how the developmental process of each individual differs from the developmental process of the overall group. Furthermore, important predictors of individual differences can be identified. As mentioned previously for the general growth model, choosing among the methods for describing the direction and magnitude of change in individual growth modeling has also not been adequately defined in the literature.

To illustrate the process of individual growth modeling, consider the development of mathematics ability in elementary school children across five points in time. Three general questions could be addressed in such a design. First, how does mathematics ability change as a function of time for all elementary students involved in the study? Is the rate of change best explained by a linear function of time or is it best explained by a non-linear function? This question can be addressed via simple (repeated measures

Analysis of Variance) or advanced (Hierarchial Linear Modeling; Latent Growth Modeling) statistical approaches to modeling change. However, the latter approaches are necessary for addressing the remaining questions. That is, how much individual variability, if any, is there around the overall rate of change? Assuming each student will develop mathematics skills at the same rate is unrealistic. Thus, identifying interindividual differences in intraindividual change becomes very important and meaningful. Lastly, if significant differences in individual rates of change exist, what variables account for these differences? In other words, what factors contribute to individual differences in the development of mathematics ability? Differences in baseline ability can also be identified and explanatory variables can be used to explain these differences as well.

A general individual growth model consists of two levels. A Level-1, within-subjects, model is devised to describe each individual's initial status (intercept) and rate of change (slope) in relation to time. In general, intercept and slope parameters estimated at Level-1 are assumed to be of a similar form (e.g., linear, quadratic, cubic, etc.) across subjects but individual parameters may be heterogeneous. Each subject may display a different initial status and/or rate of change but the function used to model change is assumed to be the same across all subjects. For instance, mathematics ability in all elementary school students may develop linearly but each student's rate of change may not follow the exact same linear form. The Level-2, between-subjects model, describes heterogeneity in initial status and rate of change and relates this heterogeneity to important predictor variables, allowing for the identification of variables which explain the individual differences. In summary, while the Level-1 model describes the overall initial status and rate of change, the Level-2 model examines the deviations of each individual from the overall group means for initial status and rate of change.

When applying individual growth modeling techniques, the observed trajectories of change should be examined carefully. The linear model, where rate of change over time follows a constant pattern (trajectories follow a straight line), is common in applied settings due to ease of interpretation. Linear models have been useful in a variety of studies, including research on intelligence in children (Espy, Molfese & DiLalla, 2001), antisocial behavior in children (Curran & Hussong, 2002; Bank, Burraston & Snyder, 2004), parental involvement in student achievement (Fan, 2001; Hong & Ho, 2005), families at risk of maladaptive parenting, child abuse, or neglect (Willett, Ayoub & Robinson, 1991), development of height (Ghisletta, 2001), educational policy (Kaplan, 2002) and evaluation (Hess, 2000).

Various non-linear patterns of change (e.g., trajectories follow a curvilinear pattern) have also been reported in literature using individual growth modeling techniques. In a study of young children between the ages of 14-26 months conducted by Huttenlocher, Haight, Bryk and Seltzer (1991), vocabulary acquisition was found to follow a curvilinear pattern, where rate of new word acquisition increased over time. Similar non-linear rates of change were observed in studies on issues related to change in parent training therapy (Stoolmiller, Duncan, Bank & Patterson, 1993), recovery of cognitive function following pediatric closed head injury (Francis, Fletcher, Stuebing, Davidson & Thompson, 1991), and physical and psychological health (Aldwin, Spiro, Levenson & Cupertino, 2001). Other studies have revealed a mixture of linear and non-linear patterns of change. Muthén (1997) utilized data over four time points (grades 7-

10) from the Longitudinal Study of American Youth (LSAY), a national study of performance in and attitudes toward science and mathematics, to examine student attitudes towards mathematics. Findings from this study indicated 7th and 8th grade student attitudes towards mathematics changed linearly over time whereas mathematics achievement accelerated in higher grades.

The primary advantage of the linear growth model over the non-linear growth model is the straightforward interpretation of the slope parameter. Regardless of the mathematical function used to model the data, the intercept is interpreted as the "baseline" measure. On the other hand, the interpretation of the slope is complicated by the mathematical function chosen to model the data. If the process of change is modeled as a linear function, the slope represents a constant rate of change in the outcome measure. Interpretation of a non-linear slope is not as straightforward. Non-linear slopes can take various forms, all of which are designed to model non-constant change (e.g., a steep increase in the outcome measure up to a certain point followed by a leveling out period). Difficulties interpreting a non-linear slope may lead to the choice of a linear growth model even though it may not provide the best representation of the given process of change. However, the underlying process of change should be modeled as accurately as possible regardless of interpretation issues.

Modeling non-linear change can be a challenging process requiring the use of either a data transformation technique or selection of the most appropriate mathematical function for explaining the process of change. Transformation techniques are considered the "simplest" methods for addressing smooth, non-linear rates of change since transforming either the outcome measure, or time, makes fitting a linear growth model possible. In addition, identifying and tracking important predictors of change is not dependent upon the scale utilized so very little information is lost when transforming data from a non-linear to linear scale. Unfortunately, identifying the most appropriate transformation, and accurately interpreting the results, can be an arduous process.

The development of alternative techniques for modeling non-linear change has rendered the implementation of data transformation techniques unnecessary in many situations. Estimation of Level-1 and Level-2 model parameters can be conducted in statistical frameworks treating the intercept and slope as either random coefficients or latent variables. The random coefficient framework includes hierarchical linear modeling (HLM: Bryk & Raudenbush, 1987, 1992; Raudenbush & Bryk, 2002) or multi-level modeling (MLM; Bock, 1989; Goldstein, 1987, 1995). Methods approaching the analysis of change from the latent variable perspective take on various monikers but are all considered a special type of structural equation model (SEM; Bentler & Weeks, 1980; Jöreskog, 1973, 1977; McDonald, 1978; Sörbom, 1974) most commonly referred to as latent curve analysis (LCA; Meredith & Tisak, 1984, 1990) or latent growth modeling (LGM; Duncan, Duncan, Strycker, Li & Alpert, 1999). While each of these techniques are valuable, the latent variable approach is the focus of this dissertation due to it's generality and flexibility in modeling error terms. From this point forward, the LGM acronym will be used to identify the latent variable models considered in this dissertation.

Among the challenges inherent in modeling non-linear change are choosing the most appropriate technique to identify intercept and slope parameters and accurately interpreting parameter estimates. Selecting a technique is complicated by the possibility that various techniques may provide equally adequate statistical parameter estimates for empirical trajectories following a given process of change. In this instance, focus should switch from selecting a technique based on statistical guidelines to one providing the most accurate substantive interpretation of model parameters based on theoretical guidelines. However, as previously mentioned, interpretation of non-linear parameters is clouded by the curvature nature of empirical trajectories.

Given the lack of guidelines for choosing among the techniques for modeling non-linear change, this dissertation is designed to provide researchers a better understanding of the statistical performance and interpretability of parameters estimated by the techniques available for modeling non-linear change. In general, while there are defined situations for applying certain techniques, the boundaries for utilizing these techniques are not well understood. For instance, it is unclear whether certain non-linear techniques are more effective when modeling a curvilinear trajectory than others in terms of statistical estimates as well as substantive interpretations. Therefore, this dissertation is designed to address the following research questions:

(1) Which technique provides the best statistical estimates when modeling non-linear patterns of change?

(2) Which technique provides the most appropriate interpretation of parameter estimates when modeling non-linear patterns of change?

Addressing these research questions will provide researchers a better understanding of when to use a certain non-linear technique and which technique will provide the clearest interpretation of parameter estimates.

The remainder of this dissertation is organized as follows. Chapter 2 reviews the common approaches to modeling change and how each handles non-linear change. This chapter will include a brief discussion of data transformation approaches to modeling non-linear change as well as discussions of the repeated measures ANOVA, HLM and SEM approaches to modeling change. This chapter will also provide a detailed discussion of the techniques for modeling non-linear change specified in the hypotheses. Chapter 2 will conclude with an application of the repeated measures ANOVA and LGM approaches to modeling change in data collected by the National Center for Educational Statistics for the Early Childhood Longitudinal Study - Kindergarten Cohort of 1998-99 (ECLS-K; U.S. Department of Education, 2006). Chapter 3 provides a discussion of the simulation study designed to address the research questions. Chapter 4 summarizes the statistical results of the simulation study. Chapter 5 discusses the results of the simulation study and provides a discussion of the interpretation issues encountered for each technique. The dissertation concludes in chapter 5 with a discussion of the implications of the study and directions for future research.

2.0 Literature Review

In longitudinal studies, there are various methods available for dealing with data displaying nonlinear change. These methods include transformations of the data or direct modeling of the non-linear pattern of change. In the context of modeling individual growth, the latter procedures are preferable. This is primarily due to the ease in which non-linear change can be modeled via advanced techniques such as HLM or LGM. Traditional techniques, such as repeated measures analysis of variance (ANOVA), have also been utilized to model both constant and non-linear change. However, this technique is limited by an inability to model individual differences in change.

The following sections include a discussion of: (1) data transformation procedures; (2) repeated measures ANOVA; (3) Hierarchical Linear Modeling (HLM); and (4) Latent Growth Modeling (LGM). Each of the methods utilized in the simulation study are presented in detail within the section on LGM. To conclude the literature review, results from an application of the repeated measures ANOVA and LGM approaches to modeling change will be compared.

2.1 DATA TRANSFORMATION PROCEDURES

Data transformation techniques have long been utilized for research involving a non-linear relationship between variables. Among the most commonly utilized of the various data transformation techniques is the "ladder of powers" or "ladder of re-expressions" approach (Mosteller & Tukey, 1977). This approach can be used for transforming a variety of non-linear patterns of change (Daniel & Wood, 1971; Draper & Smith, 1981). Traditional statistical procedures making the assumptions of linearity and homoscedasticity of errors (e.g., linear regression) have benefited tremendously from the development of this and related techniques (e.g., Box-Cox transformation - Box & Cox, 1964).

The "ladder of powers" approach was presented by Singer and Willet (2003) as a viable option for dealing with non-linear change. This approach utilizes a numerical scale consisting of a center point representing the variable to be transformed, with positive powers greater than 1 (e.g., square, cube, etc.) above the center point and negative, logarithmic and fractional powers (e.g., square root, cube root, etc.) below the center point. Figure 1 is a reproduction of a graphical depiction of this approach provided by Singer and Willet (2003, p. 211).

To determine the most appropriate transformation, empirical trajectories are examined in relation to the "rule of the bulge" (Mosteller & Tukey, 1977). The four exemplars of change utilized by this method, shown in Figure 1, are matched to the observed empirical trajectories. After matching the observed trajectories to the exemplar, the variable is transformed to linearity by moving "up" or "down" the ladder of powers.

The transformations impact is determined by relative proximity to the center point, with transformations lying furthest in either direction have the most dramatic impact.

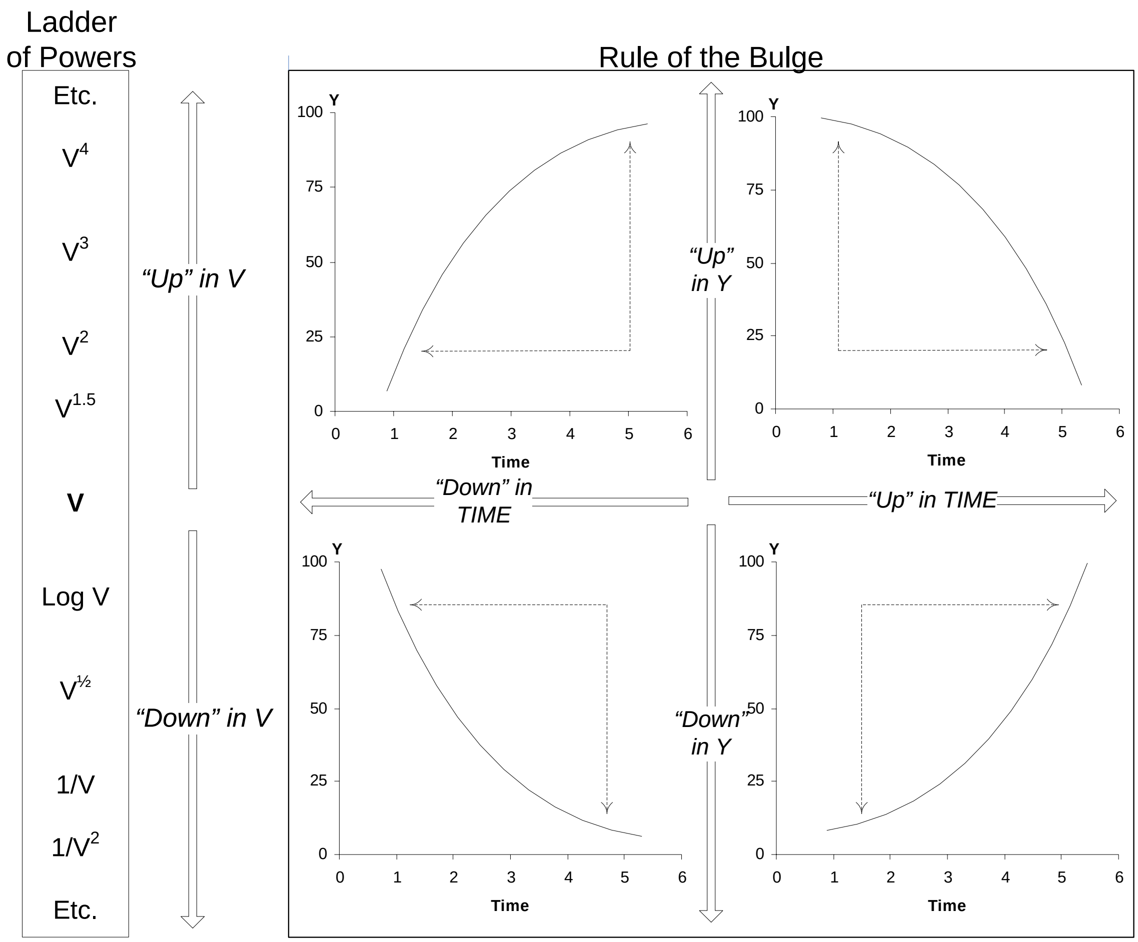

Figure 1. Ladder of Powers and Rule of the Bulge

Unfortunately, selecting an appropriate transformation for the plethora of individual trajectories typically encountered in applied settings requires compromise. In general, many individual trajectories may flow in a particular direction (e.g., growth or decline). However, the likelihood of observing identical trajectories is small, making it difficult to select a single transformation to account for all individual trajectories. For instance, squaring the outcome measure may linearize trajectories for some individuals whereas cubing the outcome measure may be necessary for linearizing other individual trajectories. Therefore, selecting the most appropriate transformation requires experimentation with viable alternatives.

There also exists the matter of interpretation. Substantively speaking, interpreting nonlinear data transformed to a linear scale is similar to interpretation in its original nonlinear metric. Unfortunately, similar interpretations do not hold numerically. To illustrate, suppose nonlinear trajectories observed for student development of mathematics ability during elementary years are transformed to a linear scale. A constant numerical value is calculated for the slope parameter, meaning rate of change is equal across all time points. Conversely, the slope trajectory becomes curved when transformed back to the original metric. Now, slope is no longer constant and, depending on the shape of observed trajectories, must be interpreted in terms of acceleration or deceleration at a given time point(s). In other words, while directional relationships can be gathered regardless of scale, the magnitude of change (e.g., numerical value of slope) is scale dependent. Also, interpreting non-linear trajectories through a linear slope parameter does not provide the most accurate representation of relationships in the data.

2.2 REPEATED MEASURES ANOVA

Analysis of variance (ANOVA) techniques are useful for quantifying the relationship between a dependent variable and one or more independent variables, also termed factors, with two or more levels. While the outcome variable in ANOVA is typically considered random (takes on values from a larger population), independent variables are either random or fixed (do not take on values from a larger population).

Independent variables are also considered either between-subjects factors (different subjects observed at each level of the IV) or within-subjects factors (same subjects observed at each level of the IV). A special type of ANOVA model which contains within-subjects factors is the repeated measures ANOVA. The repeated measures ANOVA model seeks to describe outcome measures from a random sample of subjects observed over several different fixed treatments or a single treatment at several fixed points in time.

By treating time as the independent variable, repeated measures ANOVA can be utilized for research questions involving change over time. Typically, research on time effects is designed to define the pattern of change rather than to just identify what is occurring at each time point. To illustrate, consider a repeated measures model consisting of a single factor, time, expressed as

$$y_{ij} = \mu + \alpha_j + \pi_i + (\alpha\pi)_{ij} + \varepsilon_{ij},$$

where i and j are used to distinguish individuals and time points, respectively. Thus, y_{ij} is the i^{th} ($i=1,\ldots,N$) response at the j^{th} time point ($j=1,\ldots,t$). The five parameters of interest are: (1) μ is the grand mean representing the average response over all individuals and time points; (2) π_i represents the individual effect or the individuals mean deviation from the grand mean; (3) α_j represents the time effect or the deviation of the time mean deviation from the grand mean; (4) $(\alpha\pi)_{ij}$ represents the idiosyncratic behavior of individual i at time j; (5) ε_{ij} represents errors or residuals when predicting y_{ij} from the aforementioned terms and are assumed to be independent and identically distributed $N(0, \sigma^2_{\varepsilon_{ij}})$ and also independent of π_i.

In the repeated measures model, an omnibus F-test is used to determine if mean response rates are different across time points. Typically, a significant F-test would be followed by a contrast procedure to identify which time points differ. However, with a quantitative factor such as time, researchers are generally more interested in describing the pattern depicted by observed changes. Polynomial trend analysis is a comparison method used for a quantitative variable in order to describe the pattern displayed by the means. For instance, a researcher interested in the development of mathematics ability could examine standardized assessment scores across multiple time points. Analyzing this data via repeated measures ANOVA would involve multiple steps. The first step would be determining if an overall time effect exists. If the omnibus F-test is significant, indicating a time effect, the next step becomes describing the pattern of change and/or identifying where the change is significant. This is done via the polynomial trend analysis, which uses orthogonal polynomials to identify linear, quadratic or higher order components in the pattern of change.

Polynomial trend analysis begins with the selection of contrast coefficients depicting the pattern of change to be tested. Given frequent occurrence and straightforward interpretation, a linear trend is initially examined. For instance, when the model consists of four time points, the coefficients $\{-3,-1,1,3\}$ could be utilized to determine if the pattern of change is linear. In many instances, the linear model does not account for all of the variability in the data. Under these circumstances, a set of coefficients which are orthogonal to those utilized for the linear trend are selected. If the quadratic trend is of interest, the coefficients $\{1,-1,-1,1\}$ are used in the comparison. This process could continue until the number of higher-order polynomials that can be tested (equal to the number of time points minus 1) is exhausted. However, linear or quadratic models typically account for an adequate amount of residual variability among the means. In addition, higher-order models such as the cubic or quartic (and sometimes even the quadratic) are not utilized due to difficulties with interpretation.

Application of repeated measures ANOVA to modeling change is limited by the assumption of compound symmetry (e.g., equal variances and covariances across fixed time factors), which when violated leads to an upwardly biased omnibus F-test and higher Type I error rates (Huynh & Mandeville, 1979). Identifying and correcting this bias does not imply the technique is correctly modeling the data (Chan, 2003). Nonlinear variation and covariation between time observations may reflect systematic interindividual differences in individual change. More importantly, repeated measures ANOVA and related techniques (e.g., regression, ANCOVA, MANOVA) are limited to group level analyses (assuming change in DV is same across all subjects), thus failing to provide sufficient information about individual differences in change (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002). The following approaches to be discussed are considered superior to the repeated measures ANOVA because they model change at the group level as well as the individual level.

2.3 HIERARCHICAL LINEAR MODELING

In many instances, data collected in the social and behavior sciences are hierarchical in nature. A research setting may consist of variables describing subjects and variables describing a larger unit in which subjects are grouped or "nested." Situations also arise where the larger unit in which subjects are nested is also considered nested within a larger unit. For instance, research questions in the field of education are typically geared toward identifying factors which effect student performance. Educational research is enhanced by examining the effect a teacher, school and/or school district has on student performance. In a design such as this, student is considered nested within teacher while teacher is nested within school and school is nested within school district.

Statistical modeling of nested designs is problematic due to assumptions made by traditional techniques. Techniques such as ANOVA and regression are limited by the assumption of independent observations. Relaxing this assumption is necessary for incorporating higher level variables, such as schools, when analyzing data at the individual student level. In educational research, student performance may differ by schools. Thus, students in different schools are generally considered to independent. On the other hand, students in the same school are generally not considered independent because they all attend the same school and some are even taught by the same teacher.

Hierarchical linear models can be used to model data containing dependent observations through the use of Level-1, Level-2 or even Level-3 models. In HLM, the Level-1 model is a within-subjects model represented as

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + r_{ij},$$

where y_{ij} represents the outcome measure for individual i nested within group j. The b_{0j} and b_{1j} terms are randomly varying intercept and slope parameters for each j group, x_{ij} is an indicator variable, and r_{ij} is the residual for individual i nested within group j. The residual terms are assumed to be independently and identically distributed $N\left(0, \sigma^2_{r_{ij}}\right)$.

The random intercept and slope parameters estimated at Level-1 become outcome variables in a Level-2, between-subjects, model. The Level 2 model, often referred to as a "slopes as outcomes" model (Burstein, 1980) without predictor variables, is represented as

$$b_{0j} = \gamma_{00} + \mu_{0j}$$

$$b_{1j} = \gamma_{10} + \mu_{1j},$$

where γ_{00} and γ_{10} represent the mean intercept and slope across all j groups in the sample. Each group's deviation from sample mean intercept and slope are represented by μ_{0j} and μ_{1j}. These unique effects are of primary interest since they represent the extent to which intercept and slope parameters for each group differ from overall intercept and slope parameters.

The HLM can be extended to individual growth modeling by considering time points nested within subjects rather than subjects nested within groups. Instead of estimating parameters at the group level, parameters are estimated at the subject level. To illustrate, consider the following model representing linear change over time for outcome variable y measured on i $(i=1,...,N)$ subjects at j $(j=1,...,t)$ occasions

$$y_{ij} = b_{i0} + b_{i1}x_{ij} + r_{ij}.$$

In the case of modeling change, the indicator x_{ij} contains the value of the observed time point *j* for individual *i*. Allowing b_{i0} and b_{i1} to vary across individuals provides the basis for the Level-2 model

$$b_{i0} = \gamma_{00} + \mu_{i0}$$

$$b_{i1} = \gamma_{01} + \mu_{i1},$$

where γ_{00} and γ_{10} represent the overall sample mean for the intercept (i.e., initial status) and slope (i.e., rate of change), respectively. The μ_{i0} and μ_{i1} terms are random effects representing individual deviations around the overall sample means. These terms are what distinguish each individuals intercept, b_{i0}, and slope, b_{i1}, from those of the overall sample mean and intercept. Heterogeneity in the individual intercept and slope parameters is determined by examining $\sigma^2_{\mu_{i0}}$ and $\sigma^2_{\mu_{i1}}$. Differences in individual intercepts and slopes exist if $\sigma^2_{\mu_{i0}}$ and $\sigma^2_{\mu_{i1}}$ are not equal to *0*.

Various nonlinear models are also available in the framework of HLM. The quadratic model involves an additional term describing the pattern of change. In this case, the Level-1 equation becomes

$$y_{ij} = b_{i0} + b_{i1}x_{ij} + b_{i2}x_{ij}^2 + r_{ij},$$

where the additional parameter, b_{i2}, captures the curvature or acceleration in the individual growth trajectories. To determine individual differences in the curvature parameter, b_{i2} is modeled at Level-2 as

$$b_{i2} = \gamma_{02} + \mu_{i2}.$$

Again, the amount of variability in the curvature parameter is measured by $\sigma^2_{\mu_{i2}}$.

The heterogeneity in individual parameter estimates can be explained by including indicator variables (e.g., gender, socioeconomic status, race, etc.) in the model. Indicators are introduced into the Level-2 equation of the model. For the quadratic model with a single predictor, the Level-2 equations become

$$b_{i0} = \gamma_{00} + \gamma_{10} z_{i1} + \mu_{i0}$$

$$b_{i1} = \gamma_{01} + \gamma_{11} z_{i1} + \mu_{i1}$$

$$b_{i2} = \gamma_{02} + \gamma_{12} z_{i1} + \mu_{i2}\,.$$

HLM provides a powerful framework for analyzing data with a nested structure. Flexibility with missing data is a primary advantage offered when applying HLM to repeated measures data. HLM allows both the number and timing of observations to vary randomly over participants, meaning subjects do not need to be measured at identical points in time nor do they need to be measured an equal number of times. The spacing between measurements on each individual also need not be equal. In addition, Level-1 predictors may be continuous and take on a different distribution for each member of the sample (Raudenbush & Bryk, 2002). A third level of the model could also be introduced to address effects of external environments (e.g., classroom, school, etc.) on individual change. However, HLM is limited by assumptions made regarding error terms and the inability to handle complex covariance structures, limitations overcome in the framework of latent growth modeling.

2.4 LATENT GROWTH MODELING

The latent growth model (LGM) is a special type of structural equation model (SEM). Karl Jöreskog is credited with bridging the gaps between path analysis, factor analysis and simultaneous equation modeling to create the modern day framework for SEM (Joreskog, 1973). SEM encompasses a family of techniques including regression, path analysis and confirmatory factor analysis. Arguably the most popular multivariate technique in the social and behavior sciences (Hershberger, 2003), SEM is perhaps best defined as a class of methodologies seeking to represent hypotheses about means, variances and covariances of observed data in terms of a smaller number of structural parameters defined by a hypothesized underlying model (Kaplan, 2000).

SEM provides a method for rigorously testing a hypothesized model of relations among variables through the use of a combination of manifest and latent variables. Latent variables are unobservable theoretical constructs, such as mathematics ability, intelligence, motivation, depression and anxiety, whose measurement relies on observable manifest variables. For instance, in educational research, mathematics ability can be measured via an achievement test designed to yield scores on various subsections of mathematics (e.g., algebra, geometry, calculus, etc.). Variables in SEM can also be considered either exogenous or endogenous. Exogenous variables are not explained in the model and either directly or indirectly influence the endogenous variables in the model.

A path diagram can be used to depict relationships under consideration in a SEM. A path diagram generally consists of ellipses or circles, rectangles or squares and single- and/or double-headed arrows. Rectangles or squares are used to represent manifest variables while ellipses or circles are used to represent latent variables. Single-headed arrows represent the impact of one variable on another, with the variable at the base of the arrow impacting the variable at the head. Double-headed arrows represent the correlation or covariance between two variables or, in some instances, may be used to represent the variance of an exogenous variable.

The path diagram in Figure 2 displays a full latent variable model containing an exogenous latent factor ξ_1 with two manifest indicators X_1 and X_2 predicting an endogenous latent factor η_1 with two manifest indicators Y_1 and Y_2. The γ_{11}, λ_{X11}, λ_{X21}, λ_{Y11}, and λ_{Y21} represent regression coefficients. A residual error in predicting η_1 from ξ_1 is represented by ζ_1 while δ_1, δ_2, ε_1 and ε_2 represent measurement errors for X_1, X_2, Y_1 and Y_2, respectively.

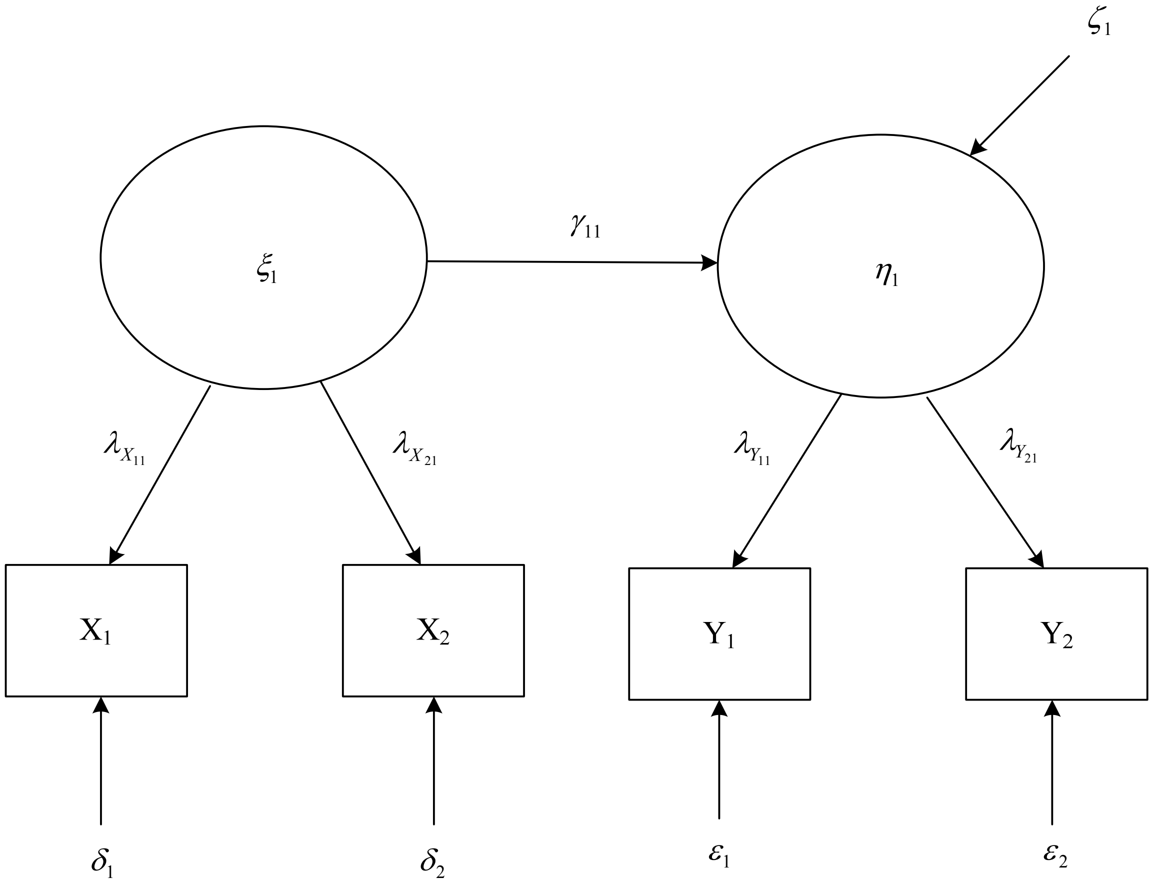

Figure 2. Full Structural Equation Model

SEM's can also be expressed in equation form using either the Bentler-Weeks model notation (Bentler & Weeks, 1979, 1980) or Lisrel notation (Jöreskg, 1973). Unless specified otherwise, the latter will be used throughout this manuscript. The full SEM presented in Figure 2 consists of a structural model representing relationships among latent factors and a pair of measurement models representing relationships between the exogenous and endogenous latent factors and their respective indicator variables. In matrix terms, the structural model is expressed as

$$\eta = v + B\eta + \Gamma\xi + \zeta,$$

where v is an intercept vector, η is a m x 1 vector of endogenous latent variables, ξ is a k x 1 vector of exogenous latent variables, B is a m x m matrix of regression coefficients relating the latent endogenous variables to each other, Γ is a m x k matrix of regression coefficients relating endogenous variables to exogenous variables, and ζ is a m x 1 vector of residual error terms.

The X and Y measurement models can be expressed in matrix terms as, respectively,

$$X = \tau + \Lambda_X \xi + \delta$$

and

$$Y = \alpha + \Lambda_Y \eta + \varepsilon.$$

Here, X and Y represent p x 1 and q x 1 vectors of p- and q-observed variables, respectively. The τ and α terms represent p x 1 and q x 1 intercept vectors, respectively, while Λ_X and Λ_Y are p x k and q x m matrices of factor loadings relating manifest variables to the k x 1 and m x 1 vectors of latent variables, ξ and η. The random error terms are contained in vectors δ and ε, respectively, which are the same dimension as X and Y.

The general SEM can be extended to longitudinal data, at the group and individual levels, through the latent growth modeling (LGM) framework. Meredith and Tisak (1984, 1990) are credited with extending the work of Tucker (1958) and Rao (1958) to modeling interindividual differences in change through SEM. LGM's assess change through multiple indicator latent variables, essentially representing Level-1 and Level-2 models from HLM as Y-measurement and structural models, respectively. Thus, for i ($i=1,2,...,N$) subjects measured at j ($j=1,2,...,t$) occasions, the Y-measurement model representing linear change can be expressed in equation form as

$$y_{ij} = \lambda_{0j}\eta_{0i} + \lambda_{1j}\eta_{1i} + \varepsilon_{ij},$$

where y_{ij} is the outcome measure for individual i at time j predicted by η_{0i} and η_{1i}, which represent individual parameters for initial status and rate of change. These individual parameters become outcome measures in the structural model, represented in equation form as

$$\eta_{0i} = \nu_0 + \zeta_{0i}$$

and

$$\eta_{1i} = \nu_1 + \zeta_{1i},$$

where η_{0i} and η_{1i} are functions of individual deviations, ζ_{0i} and ζ_{1i}, from sample mean initial status, ν_0, and rate of change, ν_1, respectively.

The path diagram displayed in Figure 3 represents a LGM with two latent factors, initial status (η_0) and linear slope (η_1), explaining outcomes at four time points (*Y1*-*Y4*). Each of the *Y* variables constitutes a score on the outcome measure at a specific point in time that contains random measurement error which is identified by ε_1, ε_2, ε_3 and ε_4. The Bentler-Weeks representation in EQS utilizes V999 as a constant (i.e., 1), thus providing the mechanism for transforming a pure covariance structure model into a mean and covariance structure model. Thus, the residual error terms ζ_0 and ζ_1 represent individual differences in η_0 and η_1 from the overall group means ν_0 and ν_1.

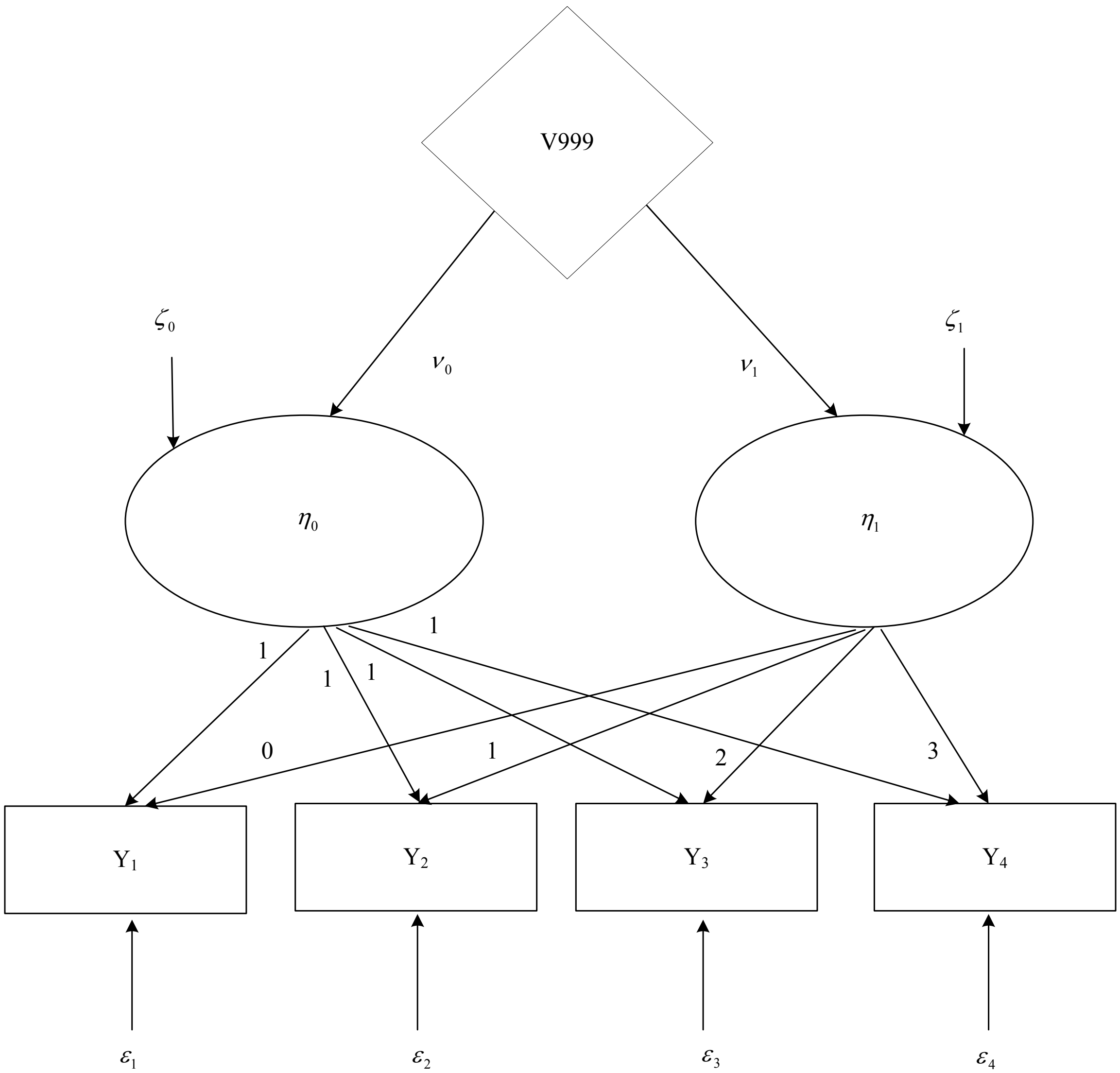

Figure 3. Unconditional Latent Growth Model

The model in Figure 3 is considered an unconditional LGM since it does not contain covariates. A conditional LGM containing a single covariate that influences η_0 and η_1 can be represented in equation form as

$$\eta_{0i} = \nu_0 + \gamma_{01}X_{1i} + \zeta_{0i}$$

and

$$\eta_{1i} = \nu_1 + \gamma_{11} X_{1i} + \zeta_{1i},$$

where γ_{01} and γ_{11} are regression coefficients relating X_{1i} to η_{0i} and η_{1i}, respectively. A path diagram for this model is on display in Figure 4.

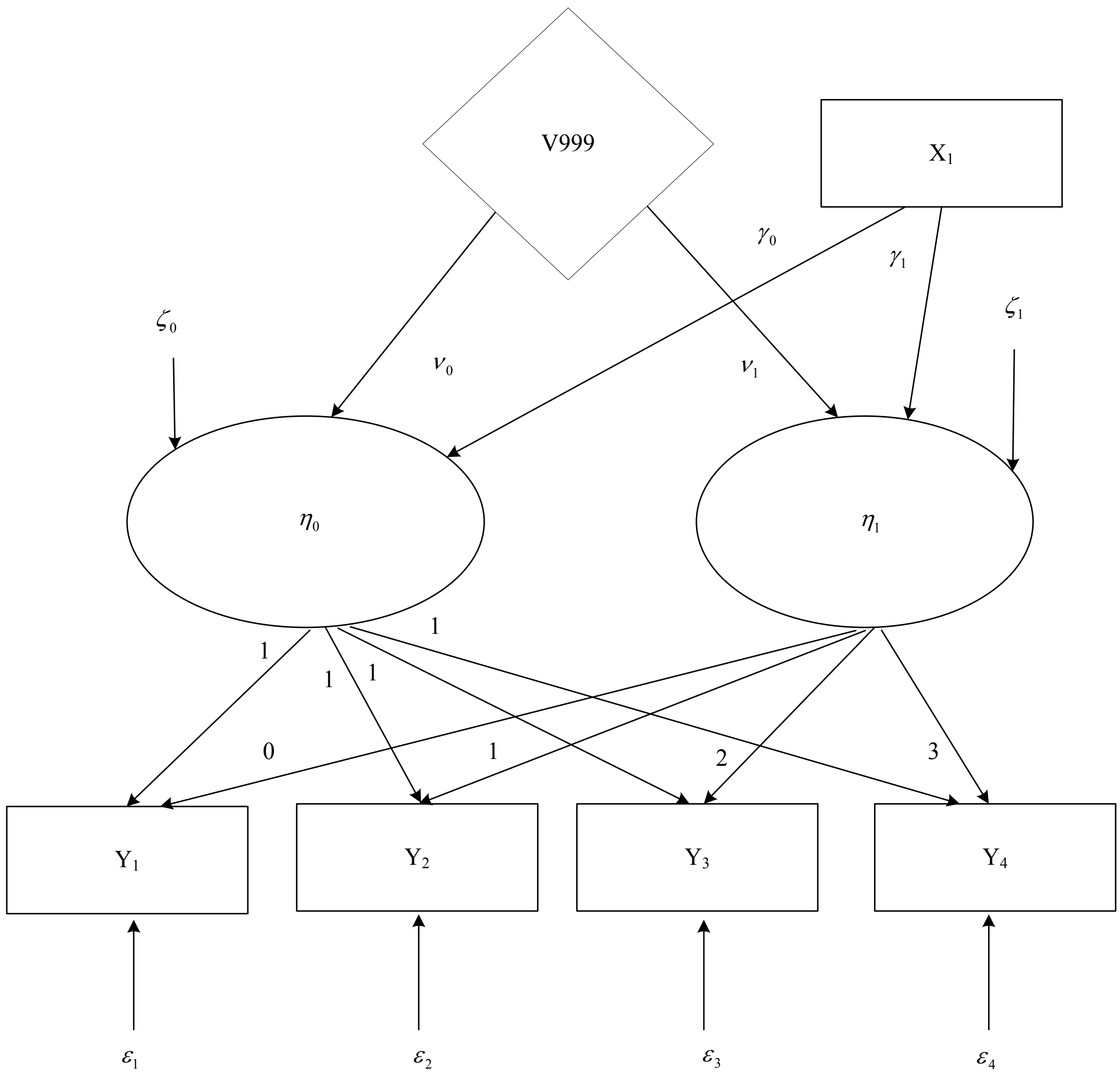

Figure 4. Conditional Latent Growth Model

In the LGM, λ_{0j} is fixed at 1 for all $j=1,...,t$. This being the case, there is one-to-one correspondence between parameters estimated in HLM and LGM. Of particular interest in these models is the observed variability in individual deviations from sample means. That is, the terms $\sigma^2_{\mu_0}$ and $\sigma^2_{\mu_1}$ from HLM and $\sigma^2_{\zeta_0}$ and $\sigma^2_{\zeta_1}$ from LGM indicate the degree of individual differences in the parameters describing initial status and rate of change. In addition, the covariance between the parameters describing initial status and rate of change provides valuable information regarding the manner in which individual trajectories are affected by the point at which they begin.

The LGM has the ability to address important hypotheses about individual differences in initial status and rate of change as well as allowing predictors of change to be incorporated into the model. For example, a researcher interested in the development of quantitative ability could assess student's quantitative skills across four equally spaced intervals and include gender as a variable in the model to determine if differences exist between males and females. In such a scenario, utilization of a LGM requires a continuous outcome measure and an adequate number of subjects to detect person level effects. In addition, conditions pertaining to missing data and distributional requirements of random effects in the *Y*-measurement model must also be met. Namely, an equal number and spacing of time points must be present for each subject and the random effects predictor variables must be identically distributed across all participants in each subpopulation (Raudenbush & Bryk, 2002). In comparison to the HLM, the LGM offers more flexibility in testing complex error structures (e.g., tests of homoscedasticity of measurement errors). In general, however, there is very little difference between parameter estimates obtained via HLM and LGM (Chou, Bentler & Petz, 1998). In fact, modeling growth in HLM has been shown to be a special case of the LGM (Curran, 2002).

2.5 NON-LINEAR LATENT GROWTH MODELING

Like the HLM approach to modeling change, non-linear trajectories are typically modeled in LGM through polynomial, unspecified or spline techniques. The implementation of these techniques in the LGM framework for specific types of non-linear trajectories will be discussed in the following sections.

2.5.1 Polynomial Models

The LGM framework utilizes polynomials to model non-linear trajectories. The linear model is considered a "first order" polynomial model due to time being raised to the 1^{st} power equaling itself. To illustrate, consider the *Y*-measurement model expressed in matrix form as

$$Y = \lambda_Y \eta + \varepsilon ,$$

where Y is a vector of values observed at given points in time which are defined in λ_Y, intercept (η_0) and slope (η_1) parameters are contained in η, and occasion specific measurement errors in ε, which are distributed $N\left(0, \sigma^2_{\varepsilon_{ij}}\right)$. In polynomial models, the coefficients used in the λ_Y matrix are chosen to be consistent with time of measurement. For instance, a linear model containing measures at four points in time could be represented in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} .$$

The column of *1*'s in λ_Y serve to fix η_1 to a value consistent with the initial measurement. The second column of λ_Y is used to define the time line of measurement. In general, X_1 is fixed to *0* while X_2, X_3, and X_4 are fixed to *1*, *2*, and *3*, respectively. The constant rate of change found in a linear slope is indicated by the equal difference between adjacent regression coefficients.

For non-linear trajectories, higher-order polynomials can be specified as the regression coefficients in the λ_Y matrix. The quadratic model can be represented in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ 1 & X_3 & X_3^2 \\ 1 & X_4 & X_4^2 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix},$$

where X_1, X_2, X_3 and X_4 are the same as before while X_1^2, X_2^2, X_3^2 and X_4^2 become *0*, *1*, *4* and *9*, respectively. The cubic model can be represented as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 & X_1^2 & X_1^3 \\ 1 & X_2 & X_2^2 & X_2^3 \\ 1 & X_3 & X_3^2 & X_3^3 \\ 1 & X_4 & X_4^2 & X_4^3 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}.$$

In the cubic model, X_1^2, X_2^2, X_3^2 and X_4^2 become *0*, *1*, *8* and *27*, respectively.

As the aforementioned models indicate, the number of parameters included in η is dependent upon the trend that is modeled. Higher-order polynomials can be used to model trajectories that display nonlinear change. However, while intercept is typically interpreted as initial status, interpretation of slope is dependent upon the number of parameters in η. In the cubic model, the linear slope, η_1, is interpreted as instantaneous rate of change while the quadratic and cubic slope, η_2 and η_3, describe the curvature in the observed trajectories. Thus, as the number of parameters in η increases, interpretation becomes more complex.

2.5.2 Unspecified Models

There are instances where complex non-linear trajectories of change make a priori specification of coefficients in Λ_Y difficult. In a two-factor model, LGM allows these coefficients to be estimated (Meredith & Tisak, 1990). To illustrate, consider a measurement model containing two-factors expressed in matrix notation as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

where X_1, X_2, X_3 and X_4 become *0*, *1*, f_a and f_b, respectively. The values f_a and f_b indicate freely estimated trajectories (from the data) of change at time points *3* and *4*. Unlike polynomial models, the unspecified model contains only intercept and slope factors, making interpretation fairly straightforward. Duncan, Duncan, Stryker, Li and Alpert (1999) indicate the slope, η_1, is better interpreted as a general "shape" factor in an unspecified model. Unless, of course, the values estimated for f_a and f_b follow linearly from the values used for the first two time points, *0* and *1*. In this case, η_1, would retain its original interpretation as constant rate of change and the unspecified model would essential be the linear model. Thus, the only difference between the unspecified model and the linear model are the values of coefficients to be estimated. There are no changes in the Y-measurement model.

2.5.3 Spline Models

Spline models break an observed curvature pattern of change into piecewise linear components. These models are especially useful for comparing rates of change at different periods in time. For instance, Frank and Seltzer (1990) found patterns of change in acquisition of reading ability for Chicago Public school students to differ between grades *1* thru *3* and *4* through *6*. Khoo (1997) found piecewise techniques to be useful for assessing the effectiveness of intervention programs, where rates of change are different before and after implementation of an intervention.

To illustrate a type of spline model, consider observed trajectories that are different between time points *0* thru *1* and *2* thru *3*. In matrix notation, this model is represented as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 & 0 \\ 1 & X_2 & 0 \\ 1 & X_2 & X_3 \\ 1 & X_2 & X_4 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix},$$

where X_1 is fixed to *0*, X_2 to *1*, X_3 to *2* and X_4 is fixed to *3*. Since rate of change is different prior to and following a particular time point, multiple slope parameters are needed to address the factor(s) causing curvature in observed trajectories. Thus, η_0 is the intercept, η_1 is the slope for period *1* (i.e., rate of change between time points *0* and *1*) and η_2 is the slope for period *2* (i.e., rate of change between time points *2* and *3*). The regression coefficients contained in Λ_Y define beginning and ending points for η_1 and η_2 as well as the magnitude of change between these points.

Another two-factor spline model could be implemented by specifying Λ_Y as

$$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \end{bmatrix} = \begin{bmatrix} 1 & \overline{Y}_1 \\ 1 & \overline{Y}_2 \\ 1 & \overline{Y}_3 \\ 1 & \overline{Y}_4 \end{bmatrix},$$

where $\overline{Y}_1$, $\overline{Y}_2, \overline{Y}_3$ and $\overline{Y}_4$ are the observed means at each time point. In other words, η_1 crosses through the mean of each time point. As in the two-factor unspecified model, rate of change parameters for intercept and slope(s) become dependent measures in spline models. However, since certain covariates may not be relevant to each rate of change parameter, separate structural models may be necessary.

2.6 EMPIRICAL EXAMPLE

Data collected for the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K; U.S. Department of Education, 2006) was used to demonstrate the application of latent growth models. For this example, a sample of *n=1650* students were selected. This sample consisted of students that were 1st time kindergartners at the beginning of the study who remained at the same school for the duration of the study. In addition, complete data was present for all of the students at each of the four time points of measurement (Fall and Spring of Kindergarten and Fall and Spring of 1st grade) and for the gender and ethnicity variables.

Investigation of the data revealed severe nonnormality and the presence of three outlying cases, which were ultimately removed. Table 1 displays descriptive statistics for the IRT-Scaled mathematics achievement scores for the *n=1647* students that were used in the analyses. The results reveal the presence of a nonlinear increase mathematics scores over time. The increase is scores during the Kindergarten year is 11.48 points whereas the increase during 1st grade is 18.76 points. The increase in scores between Kindergarten and 1st grade was 7.04 points. The standard deviation, on the other hand, increased at a relatively constant rate of approximately 2 points.

Table 1. Descriptive statistics for IRT-Scaled mathematics achievement scores

	Fall Kindergarten	Spring Kindergarten	Fall 1st Grade	Spring 1st Grade
Mean	25.18	36.64	43.68	62.44
Standard Deviation	9.16	11.69	13.36	15.88

Table 2 displays results for the overall model and the polynomial trend analysis obtained from a repeated measures ANOVA. The overall model results indicate a significant change in mathematics scores over the four measurement points and a large time effect. Results from the polynomial trend analysis indicate significant linear, quadratic, and cubic trends. However, only the linear trend produced a large effect size. Also, the cubic trend produced a larger effect size than the quadratic trend.

Table 2. Repeated Measures ANOVA results

	F-statistic	Df	P-value	Partial η^2
Linear	17361.72	(1,1646)	.000	.913
Quadratic	639.44	(1,1646)	.000	.280
Cubic	805.53	(1,1646)	.000	.329

The quadratic and cubic trends are slightly apparent in the line in Figure 5, which displays the empirical trajectories for four randomly selected students (labeled 1, 2, 3 and

4) and the overall group (labeled T). Between the fall of Kindergarten and the fall of 1st grade, mathematics scores appear to increase linearly. However, the change between spring of Kindergarten and fall of 1st grade is smaller than the change between fall of Kindergarten and spring of Kindergarten (this is the quadratic component). The cubic component is present due to the relative increase in mathematics scores in the spring of 1st grade.

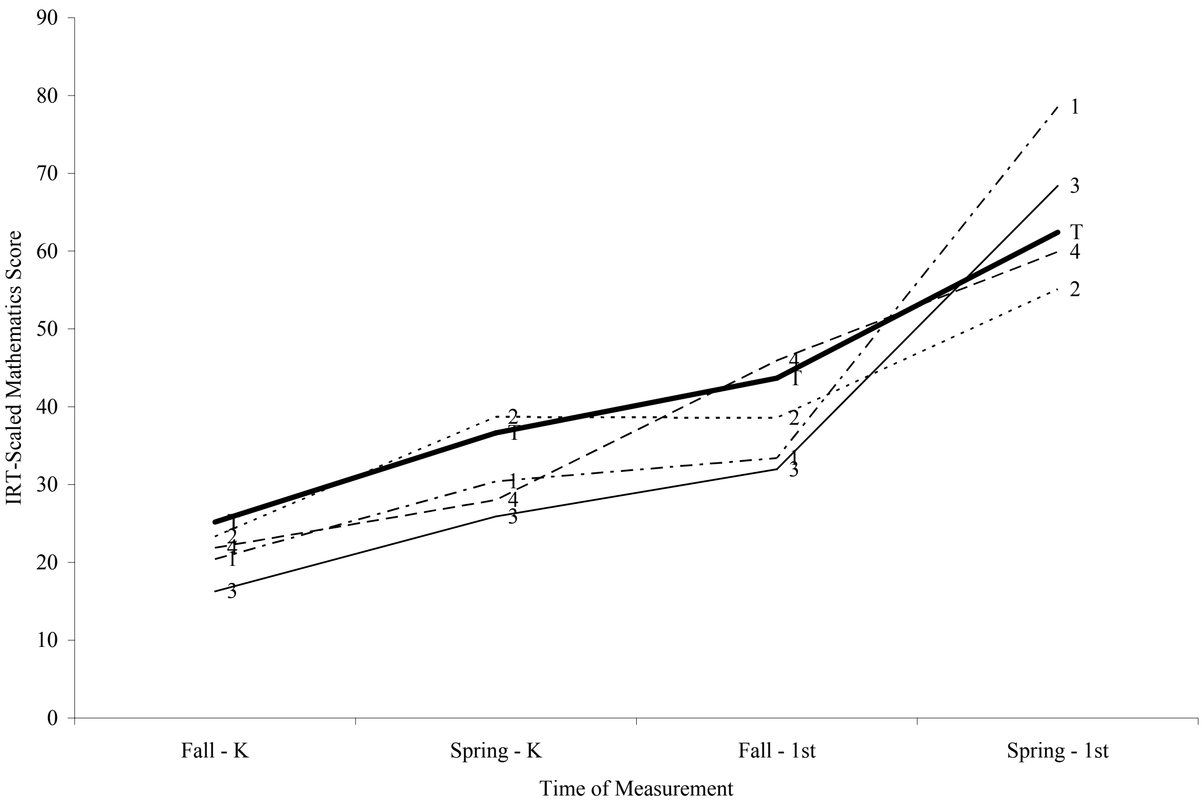

Figure 5. Empirical trajectories for 4 randomly selected students and the overall group

The repeated measures analysis of variance mentioned above is limited to describing group differences. In addition to identifying change in the overall group over time, LGM's are useful for identifying the presence of individual differences in change. Figure 5 displays growth trajectories for the entire sample of students (labeled T) and for four randomly selected students (labeled 1, 2, 3 and 4). As generally expected, none of the trajectories are the same nor do the individual student trajectories mirror the overall trajectory. While the trajectory for student 1 is similar to the overall trajectory, the increase in mathematics scores between fall of 1st grade and spring of 1st grade is greater for student's 1 and 3 than it is for the overall sample. The trajectory for student 2 is very similar to the overall trajectory while the trajectory for student 3 is different from the overall trajectory between the spring of kindergarten and the fall of 1st grade. Student 2 displays a greater increase between these time periods than is found overall and for students 1, 3, and 4.

Linear, mean-spline, and unspecified techniques were used to model change in the mathematics scores. Polynomial techniques were also utilized but did not provide adequate solutions and were not reported. Each of the models was specified to contain an intercept and a single slope factor. The difference between the specifications of the models was in the factor loading matrix, Λ_Y. The linear model contained the values 0, 1, 2 and 3 to represent the four equally spaced measurement occasions. The mean-spline model utilized the values 0, 1, 1.61 and 3.25 while the unspecified model utilized the same initial values of 0 and 1 but left the last two values to be estimated by the data. In each of the models, the variances of the error terms and factors were feely estimated as was the covariance between the intercept and slope factors.

Table 3 displays selected fit indices for each of the fit techniques. Since the data was nonnormally distributed (i.e., multivariate kurtosis = 32.31), the Satorra-Bentler rescaled χ^2_{SB} (Satorra & Bentler, 1988; 1994) statistic was utilized. The χ^2_{SB} downwardly adjusts the normal theory χ^2 according to the amount of nonnormality in the data. In the presence of nonnormal data, the Comparative Fit Index (CFI; Bentler, 1990) and Root

Mean Squared Error of Approximation (RMSEA; Steiger & Lind, 1980) are also available. All of the models were fit in EQS 6.1 (Bentler, 2004).

Table 3. Model fit indices for each technique

	χ^2_{SB}	Df	p	CFI	RMSEA	90% CI RMSEA
Linear	898.55	5	< .01	.935	.330	.311 - .348
Spline	59.93	5	< .01	.982	.082	.064 - .100
Unspecified	54.07	3	< .01	.983	.102	.079 - .126

As Table 3 indicates, the linear model was the least adequate in modeling the data. The χ^2_{SB} model test statistic was much higher in the linear model than the mean-spline and unspecified models. Likewise, the CFI and RMSEA fit indices indicated the linear model did not fit the data well while the mean-spline and unspecified models fit the data similarly. In comparison to the unspecified model, the χ^2_{SB} was slightly larger for the mean-spline while the RMSEA and the corresponding 90% confidence interval were slightly smaller.

Table 4 displays parameter estimates and standard errors obtained from each modeling technique.

Table 4. Parameter estimates and standard errors for each technique

	Linear	Mean-Spline	Unspecified
η_0	11.12 (0.09)	11.47 (0.09)	11.54 (0.17)
η_1	25.02 (0.23)	25.18 (0.23)	25.15 (0.23)
$\sigma^2_{\zeta_0}$	71.60 (4.40)	75.92 (4.47)	75.78 (4.46)
$\sigma^2_{\zeta_1}$	3.72 (0.58)	6.54 (0.61)	6.62 (0.63)
$\sigma_{\zeta_0\zeta_1}$	13.98 (1.03)	11.34 (1.00)	11.45 (1.04)
$\sigma^2_{\varepsilon_0}$	16.34 (1.60)	11.97 (1.56)	11.99 (1.59)
$\sigma^2_{\varepsilon_1}$	23.06 (1.68)	25.06 (1.60)	25.02 (1.59)
$\sigma^2_{\varepsilon_2}$	49.68 (1.94)	32.58 (2.03)	32.40 (1.99)
$\sigma^2_{\varepsilon_3}$	98.94 (3.74)	58.37 (4.25)	58.89 (4.22)

λ_{13}	-	-	1.62 (0.02)
λ_{14}	-	-	3.22 (0.05)

All parameter estimates and standard errors were significant at the $\alpha = .05$ level. As was the case with the fit indices, parameter estimates and standard errors from the spline and unspecified techniques were virtually identical. The linear model yielded particularly high estimates of error variance at all but one time point and low variance values for the disturbance terms. The similarities between the results for these techniques are due to the similarities in the factor loadings. The values estimated in the unspecified technique were very similar to the fixed values used in the mean-spline technique.

In summary, the linear model is not a good choice for modeling the trajectories found in this data. Moreover, it seems a trivial decision as to whether a spline or unspecified model should be used in this situation. Evaluated via the χ^2_{SB} difference test, these models were not found to be significantly different. In either case, this sample of subjects began with an IRT scaled mathematics score of approximately 11 which increased by approximately 25 points each assessment. In addition, significant individual differences in trajectories were also indicated by the variance in the disturbance terms (i.e., approximately 76 and 7, respectively). The Lagrange multiplier (LM) test (Chou & Bentler, 1990) was used to determine if either of these models could be improved. However, none of the suggested model additions (e.g., correlated error terms) were added due to the moderate amount of improvement in model estimates and the complexity they added to interpretation of the models.

3.0 Methodology

A simulation study was designed to examine the appropriateness of the quadratic, spline and unspecified techniques to modeling nonlinear trajectories in the framework of LGM. What follows is a discussion of each of the independent and dependent variables along with the procedure of the simulation study.

3.1 INDEPENDENT VARIABLES

There were 5 independent variables manipulated in the simulation study: sample size, trajectory pattern, slope variance, covariance between factor disturbances and non-linear technique.

3.1.1 Sample Size

The 7 levels of sample size used were: 50, 75, 100, 150, 200, 300, and 500. This covers the spectrum of sample sizes commonly seen in individual growth modeling via LGM. Statistical estimates for samples larger than 500 were not expected to change significantly, leaving little reason for the investigation of larger sample sizes. In general, minimum sample size requirements in SEM are determined by the ratio of subjects to parameters estimated. Kline (2005) has indicated this ratio to be at least 5:1, preferably 10:1, in order for statistical estimates to be accurate and meaningful. All of the models utilized in this study were simple LGMs without predictor variables, which, as seen in Table 5, resulted in between 14 and 19 parameter estimates. At the smallest sample sizes, *n=50* and *n=75*, the ratio of subjects to parameters was generally lower than the suggested 5:1, providing an evaluation of minimal sample size requirements in non-linear LGMs. Regardless of the non-linear technique implemented to fit the data, minimal sample size requirements were met in the remainder of the sample conditions.

Table 5. Parameters estimated by each fit technique

	Quadratic	Spline	Unspecified
Factor Means	ν_1, ν_2, ν_3	ν_1, ν_2	ν_1, ν_2
Factor Variances	$\sigma^2_{\varsigma_1}, \sigma^2_{\varsigma_2}, \sigma^2_{\varsigma_3}$	$\sigma^2_{\varsigma_1}, \sigma^2_{\varsigma_2}$	$\sigma^2_{\varsigma_1}, \sigma^2_{\varsigma_2}$
Covariance of Factor Variances	$\sigma_{\varsigma_1\varsigma_2}, \sigma_{\varsigma_1\varsigma_3}, \sigma_{\varsigma_2\varsigma_3}$	$\sigma_{\varsigma_1\varsigma_2}$	$\sigma_{\varsigma_1\varsigma_2}$
Error Variances	$\sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}$ $\sigma^2_{\varepsilon_4}, \sigma^2_{\varepsilon_5}$	$\sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}$ $\sigma^2_{\varepsilon_4}, \sigma^2_{\varepsilon_5}$	$\sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}$ $\sigma^2_{\varepsilon_4}, \sigma^2_{\varepsilon_5}$
Factor Loadings			$\lambda_{13}, \lambda_{14}, \lambda_{15}$

3.1.2 Trajectory Pattern

Various trajectories are found in longitudinal research designed to identify differences in individual growth. This dissertation focused on trajectories consistent with the exponential, logarithmic and logistic functions. Trajectories displaying a pattern of change consistent with either of these functions are generally observed in applied settings (Burchinal & Appelbaum, 1991). Logarithmic growth, where an attribute initially develops at a rapid pace but then levels off, essentially represents the "learning curve" found in many educational settings where knowledge of a concept is initially obtained rather quickly but then levels off as time goes on. Logistic growth, where an attribute takes some time to develop but then develops at a rapid pace before leveling off, is most often observed in developmental research on young children. Specifically, logistic growth is often seen in the vocabulary development of young children. Young children typically take a year or so to speak an initial word but once the initial word is spoken, additional words are acquired at an increasingly rapid pace until a certain age when acquisition of new words begins to slow considerably. Although not as common in the social and behavioral sciences, the exponential curve, in which growth of an attribute develops slowly but increases rapidly at later time points, was investigated due to the possibility of it also being observed in the applied realm.

Figure 6 displays the shape of each trajectory used for data generation.

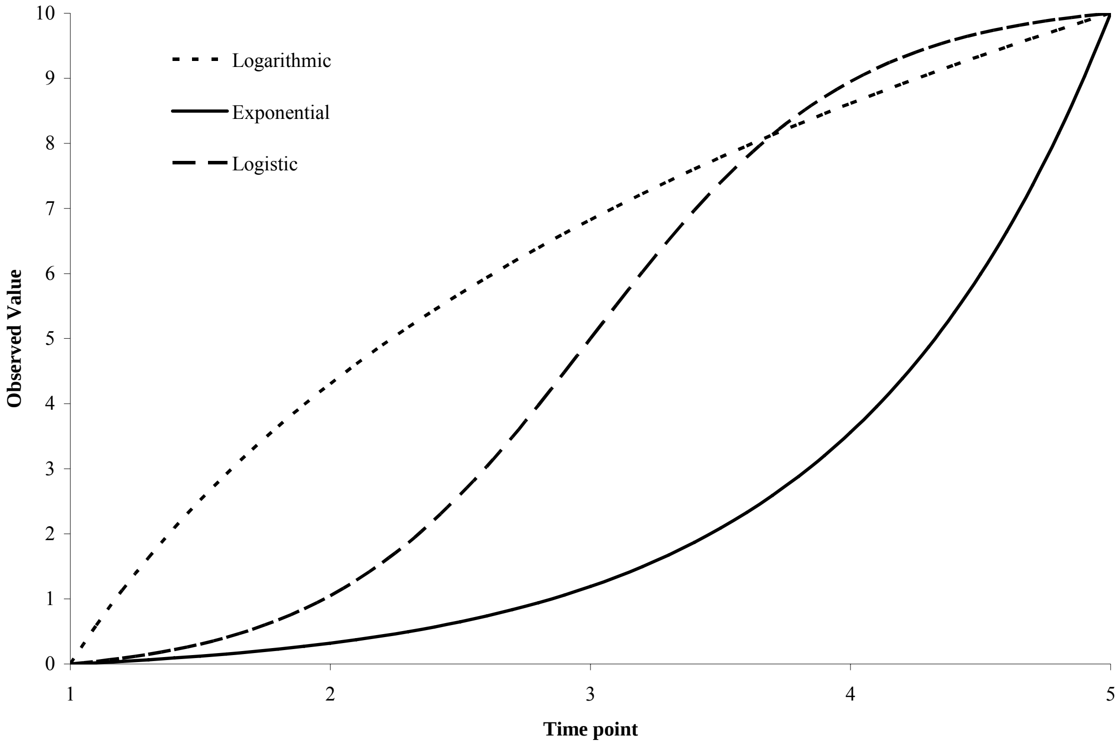

Figure 6. Trajectories of change used for generating data

These curves were created by applying exponential, logarithmic and logistic mathematical functions to the values representing each time point (i.e., *1*, *2*, *3*, *4*, *5*). After obtaining the transformed values, the curves were put on the same scale by applying a transformation that fixed the values of time points *1* and *5* at *0* and *10*, respectively. Table 6 displays the resulting means at each time point.

Table 6. Mean values at each time point for the growth trajectories

	Time 1	Time 2	Time 3	Time 4	Time 5
Exponential	*0*	*0.32*	*1.19*	*3.56*	*10*
Logarithmic	*0*	*4.31*	*6.83*	*8.61*	*10*
Logistic	*0*	*1.05*	*5.00*	*8.95*	*10*

3.1.3 Covariance Matrix Structure

Generating data in LGM is complicated, particularly for nonlinear trajectories, by the special covariance structure of these models. Thus, rather than utilizing a specific covariance matrix, data was generated according to the relationship between intercept and slope factors in a two factor model. Factor correlations and slope variances used for data generation were .15, .45 and .75, respectively, while the intercept and error variances were fixed at 1. Displayed within the cells of Table 7 are the values for the covariance between intercept and slope disturbances when the factor correlations and slope variances were fixed at the aforementioned values.

Table 7. Parameter values used for data generation

		Standardized $\sigma_{\varsigma_1\varsigma_2}$		
		0.15	0.45	0.75
$\sigma^2_{\varsigma_1}$	0.15	*0.06*	*0.17*	*0.29*
	0.45	*0.10*	*0.30*	*0.50*
	0.75	*0.13*	*0.39*	*0.65*

3.1.4 Model Fit Techniques

The polynomial, spline, and unspecified techniques were used to model the aforementioned nonlinear growth trajectories. For the polynomial technique, only quadratic and cubic models were implemented due to difficulties fitting and interpreting higher order polynomial models. In general, LGM requires at least three waves of data for a quadratic model and four waves of data for a cubic model (Singer & Willett, 2003). While five waves of data are commonly collected, cubic models are rarely utilized given the complexity involved in fitting and interpreting parameters. Including them in this study will serve to highlight the difficulty of fitting and interpreting cubic models.

The difference in the four analytic techniques lies in the number of factors and corresponding specification of λ_Y. Recall, a two-factor linear model is simply a special type of polynomial model (i.e., coefficient raised to the fist power which equals itself). Therefore, each step up the polynomial ladder leads to an additional factor and loadings consistent with the polynomial term. Thus, the quadratic model consists of three factors with λ_Y fixed to

$$\begin{bmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ 1 & X_3 & X_3^2 \\ 1 & X_4 & X_4^2 \\ 1 & X_5 & X_5^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}.$$

On the other hand, the spline and unspecified techniques consist of only two factors, leaving λ_Y as

$$\begin{bmatrix} 1 & X_1 & X_1 \\ 1 & X_2 & X_1 \\ 1 & X_3 & X_3 \\ 1 & X_3 & X_4 \\ 1 & X_3 & X_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix}$$

for the spline technique and

$$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \\ 1 & X_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & f_a \\ 1 & f_b \\ 1 & f_c \end{bmatrix}$$

for the unspecified. The EQS programs used to fit the data are displayed in Appendices A.1.2 thru A.1.4. Note that the notation in these programs does not match LISREL notation used throughout this manuscript developed by Karl Jöreskg (1973). EQS utilizes Bentler and Weeks (1980) notation where *V*, *F*, *E* and *D* are used in place of *Y*, η, ε and ς to denote variables, factors, errors and disturbances, respectively.

3.2 DEPENDENT VARIABLES

The goal of this study was to determine which analysis technique most accurately models the sample data. Like many traditional statistical techniques, SEM utilizes tests of statistical significance to determine if an implied theoretical model fits the sample data. Unlike many traditional statistical techniques, SEM supplements the overall test of significance with numerous goodness-of-fit indices which evaluate the fit of a model along a continuum. The statistical measures of fit chosen for this study will be discussed next.

3.2.1 Goodness-of-Fit Measures

Structural equation models rely on statistical measures for determining if the covariance matrix implied by the theoretical model is consistent with the sample covariance matrix. There are many statistical measures for determining fit in structural equation modeling. This dissertation will focus on the overall model goodness-of-fit statistic, denoted as χ^2 or *T*, Comparative Fit Index (CFI; Bentler, 1990), Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI; Jöreskog & Sörbom, 1984), Standardized Root Mean Squared Residal (SRMR; Bentler, 1995) and the Root Mean Squared Error of Approximation (RMSEA; Steiger & Lind, 1980).

The overall goodness-of-fit statistic is considered a global measure of model fit (i.e., exact fit statistic) because it assesses the magnitude of the discrepancy between the model implied covariance matrix $\Sigma(\theta)$ and the sample covariance matrix *S*. The χ^2 can be calculated as

$$\chi^2 = (N-1)F_{MIN},$$

where *N* is the number of subjects and F_{MIN} is the minimum fitting function obtained from an estimation method such as maximum likelihood (ML), the most commonly utilized fitting function in SEM (Gierl & Mulvenon, 1995). In practice, the χ^2 can range anywhere from *0* for a saturated model (i.e., the number of parameters estimated is equal to the number of elements in the variance/covariance matrix for the observed variables) to a maximum for the independence model (i.e., only the variance of the variables are estimated; all covariances are set equal to 0). A non-significant χ^2 value with associated degrees of freedom (df = $(p(p+1)/2)-q$, where *p* is the number of observed variables and *q* is the number of model parameters, indicates very little discrepancy between the model implied and sample covariance matrices. The difference between $\Sigma(\theta)$ and *S* is represented in a residual matrix which contains residual values close to zero when the χ^2 is not significant. Unfortunately, the χ^2 statistic is sensitive to sample size and departures from the multivariate normality assumption (e.g., Chou, Bentler & Satorra, 1991; Curran, West & Finch, 1996; Hu, Bentler & Kano, 1992).

Goodness-of-fit indices are often used to supplement the χ^2 model fit statistic. While the χ^2 model fit statistic utilizes a statistical distribution to make a distinction between a significant and non-significant difference, goodness-of-fit indices quantify the fit of the model along a continuum. Similar to the R^2 in regression analyses, fit indices quantify the extent to which the variation and covariation in the sample data are accounted for by the implied model. Fit indices are categorized as either absolute or incremental (Bollen, 1989; Gerbing & Anderson, 1993; Hu & Bentler, 1995, 1998; Marsh, Balla & McDonald, 1988; Tanaka, 1993). The GFI, AGFI, SRMR and RMSEA are considered absolute fit indices. Absolute fit indices directly assess how well the implied model reproduces the sample data by comparing (either implicitly or explicitly) the model implied covariance matrix to a saturated model which exactly reproduces the sample covariance matrix.

In contrast to absolute fit indices, incremental fit indices measure the proportionate improvement in fit by comparing the implied theoretical model to a more restrictive, nested baseline model which typically consists of observed variables that are allowed to vary but not covary. Incremental fit indices can be further defined as Type 1, Type 2 or Type 3 fit indices (Marsh et al., 1988; Hu et al., 1998). Unlike Type 1 fit indices, Type 2 and Type 3 fit indices assume the variables to follow a specific distributional form. Type 2 fit indices assume the expected value of the test statistic from the target model follows a central χ^2 distribution whereas Type 3 fit indices assume the expected value of the same statistic to follow a non-central χ^2 distribution. In addition, Type 3 fit indices also may use information from the baseline model that the target model is being compared to. The expected value of the baseline model is also assumed to follow a non-central χ^2 distribution. The CFI is the only incremental fit index to be examined in this dissertation. Table 8 summarizes the mathematical form and range of values for each of the fit indices to be used in this dissertation.

Table 8. Mathematical definitions and range of values for fit indices

Mathematical Definition	Range of Values
CFI = $\frac{1-\max[(T_T - df_T),0]}{\max[(T_T - df_T),(T_B - df_B),0]}$	0 to 1
GFI $=1-\left[\frac{tr(\Sigma^{-1}S-I)^2}{tr(\Sigma^{-1}S)^2}\right]$	Typically ranges from 0 to 1 but can be < 0 and > 1
AGFI $=1-\left[\frac{p(p+1)}{2df_T}\right](1-GFI)$	Same as GFI
SRMR $=\sqrt{\frac{2\Sigma_{t=1}^{p},\Sigma_{t}^{j=1}\left[(s_{ij}-\hat{\sigma}_{ij})/(s_{ii}s_{jj})\right]^2}{p(p+1)}}$	0 to 1
RMSEA $=\sqrt{\frac{\hat{F}_0}{df_T}}$, where $\hat{F}_0=\max\left[\frac{(T_T-df_T)}{(N-1)},0\right]$	Typically ranges from 0 to 1 but can be negative

Note. T_T is the test statistic for the target model; df_T is the df for the target model; df_B is the df for the baseline model; p is the number of observed variables; s_{ij} are the observed covariances; $\hat{\sigma}_{ij}$ are the reproduced covariances; s_{ii} and s_{jj} are the observed standard deviations.

3.3 DATA GENERATION AND PROCEDURE

Data was generated according to the mean values of each trajectory in Table 6 and the parameter specifications presented in Table 7. These values were input into EQS 6.1 (Bentler, 2004) under a two-factor LGM. This process was replicated *1000* times at each sample size (*7*) for each trajectory (*3*) with each type of factor correlation (3) and slope variance (3), resulting in a total of *189,000* (*7* x *3* x *3* x *3* x *1000*) raw datasets. The general program used to generate data is displayed in Appendix A.1.1. Each of the *189,000* raw datasets was fit to the quadratic, spline and unspecified models, resulting in *567,000* analyses. The programs used to fit the data are on display in Appendices A.1.2 thru A.1.4. Goodness-of-fit estimates from analyses which were free of convergence problems and/or condition codes (e.g., linear dependencies, negative variance) were imported into SPSS for analysis.

3.4 DATA ANALYTIC STRATEGY

The primary purpose of this dissertation was to identify the technique which provided the best model fit. To address this question, a series mixed ANOVA's were conducted to test for mean differences across conditions. An ANOVA on each of the outcome measures (i.e., p-value, GFI, AGFI, CFI, SRMR and RMSEA) was conducted with fit technique (quadratic, spline & unspecified) as the within-subjects factor and growth curve (exponential, logarithmic & logistic), sample size (50, 75, 100, 150, 200, 300 & 500), slope variance (.15, .45 & .75) and factor correlation (.15, .45 & .75) as the between-subjects factors. Given the large number of replications, particular attention was paid to effect sizes. The effect size of most interest was for the technique used to fit the data.

4.0 Results

The results of this dissertation are presented in four sections. Section one provides a verification of the data generation process. Section two provides a summary of convergence rates for each growth curve by the conditions of the study. Section three provides descriptive statistics for each growth curve by the conditions of the study and section four provides the results from the Mixed ANOVA.

4.1 VERIFICATION OF DATA GENERATION

Table 9 displays the means and standard deviations of the data values generated for each growth curve. The mean values are consistent with those on display in Table 6 with the only difference being that the generated values began with an initial time point of 1 rather than zero, thus causing the value at each of the corresponding time points to also be 1 unit higher. However, the trajectory of the curves was consistent with what was expected so the convergence failures and/or conditions codes to be mentioned in the next section can not be attributed to the data generation process.

Table 9. Mean and standard deviations of all generated data values (n=12,375,000) by growth curve.

	Time 1	Time 2	Time 3	Time 4	Time 5
Exponential	1.0001	1.3202	2.1899	4.5595	10.9994

	(1.4144)	(1.4943)	(1.8314)	(3.1618)	(7.3908)
Logarithmic	0.9994	5.3111	7.8314	9.6113	11.0017
	(1.4141)	(3.6308)	(5.2706)	(6.4554)	(7.3875)
Logistic	0.9995	2.0492	5.9984	9.9465	10.9969
	(1.4140)	(1.7673)	(4.0741)	(6.6855)	(7.3902)

As another means of verifying the data generation process, the syntax in Appendix A.1.1 was used to generate a single dataset containing 5000 cases. The factor loadings were set to be consistent with the logarithmic curve (i.e., 0, 4.31, 6.83, 8.62 and 10) while the slope variance and covariance between the disturbances was set to .75. This dataset was fit to a two factor model with the loadings fixed to the values defining the logarithmic curve (i.e., 0, 4.31, 6.83, 8.62 and 10) while the variance of the errors and disturbances and the covariance between the disturbances were free to be estimated. As seen in table 10, the estimates of these parameters are very close to the values used to generate the data. This supports the notion that any convergence issues are not related to the data generation process.

Table 10. Comparison of generated and fitted data values

	Generated	Fit
Factor Means	$\nu_1 = 1.013$	$\nu_1 = 1.013$
	$\nu_2 = 1.001$	$\nu_2 = 1.001$
Factor Variances	$\sigma^2_{\varsigma_1} = 1.000$	$\sigma^2_{\varsigma_1} = 0.958$
	$\sigma^2_{\varsigma_2} = 0.750$	$\sigma^2_{\varsigma_2} = 0.749$
Covariance of Factor Variances	$\sigma_{\varsigma_1\varsigma_2} = 0.750$	$\sigma_{\varsigma_1\varsigma_2} = 0.732$
Error Variances	$\sigma^2_{\varepsilon_1} = 1.000$	$\sigma^2_{\varepsilon_1} = 0.952$
	$\sigma^2_{\varepsilon_2} = 1.000$	$\sigma^2_{\varepsilon_2} = 1.028$
	$\sigma^2_{\varepsilon_3} = 1.000$	$\sigma^2_{\varepsilon_3} = 1.022$
	$\sigma^2_{\varepsilon_4} = 1.000$	$\sigma^2_{\varepsilon_4} = 1.007$
	$\sigma^2_{\varepsilon_5} = 1.000$	$\sigma^2_{\varepsilon_5} = 0.961$
Factor Loadings	Fixed to values for logarithmic curve – 0, 4.31, 6.83, 8.62 and 10.	

4.2 CONVERGENCE RATES

The 567,000 replications yielded 298,554 solutions which converged without a condition code. Tables 11 thru 13 display the number of admissible solutions for each growth curve by fit technique, slope variance, factor correlation and sample size. The cubic technique is not displayed in these tables due to convergence and/or condition codes problems resulting in not a single admissible solution. The majority of these condition codes were due to linear dependencies and further research needs to be conducted as to why this was the case.

Convergence rates were very high for the logarithmic curve across the other conditions of the study. This was not the case for the exponential and logistic curves. In both of these curve conditions, convergence rates were very low except when the unspecified model was fit to the data. In general, fitting the unspecified model resulted in the highest rates of convergence regardless of the curve it was being fit to. For the logarithmic curve, rates of convergence were of course highest for the unspecified fit technique followed by the quadratic technique and the spline technique. In addition, for the logarithmic curve, convergence rates slightly increased as the slope variance and factor correlation increased and also as the sample size increased. Again, this was the case for the other curves when the unspecified technique was fit to the data. However, when the other techniques were fit to the exponential and logistic curves, this pattern was not displayed. In some instances, the number of converging solutions decreased with sample size while in others a pattern across the conditions was not discernable.

Table 11. Number of admissible solutions for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	618	733	809	871	899	951	986
		.45	668	783	828	903	932	972	990
		.75	704	835	867	927	953	986	998
	.45	.15	752	882	944	976	993	993	1000
		.45	794	910	962	987	995	999	1000
		.75	850	946	975	998	1000	1000	1000
	.75	.15	784	739	923	964	983	995	1000
		.45	836	945	959	986	996	999	1000
		.75	905	958	983	991	1000	999	1000
Spline	.15	.15	502	591	674	785	829	893	963
		.45	583	644	746	813	865	930	973
		.75	599	721	787	846	889	939	992
	.45	.15	632	753	813	886	923	957	989
		.45	721	850	900	946	974	996	1000
		.75	773	904	934	987	991	997	1000
	.75	.15	614	704	760	814	852	876	937
		.45	751	843	886	942	965	985	998
		.75	826	907	948	982	993	998	1000
Unspecified	.15	.15	990	994	1000	1000	1000	1000	1000
		.45	987	998	1000	1000	1000	1000	1000
		.75	987	997	1000	1000	1000	1000	1000
	.45	.15	981	994	1000	1000	1000	1000	1000
		.45	983	999	1000	1000	1000	1000	1000
		.75	991	999	999	1000	1000	1000	1000
	.75	.15	984	997	1000	1000	1000	1000	1000
		.45	989	998	999	1000	1000	1000	1000
		.75	989	999	1000	1000	1000	1000	1000

Table 12. Number of admissible solutions for the exponential growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	49	98	14	2	0	0	0
		.45	321	300	24	5	5	1	0
		.75	358	324	298	390	232	230	330
	.45	.15	23	9	6	1	1	0	0
		.45	351	210	15	3	1	1	0
		.75	623	710	751	625	866	877	603

	.75	.15	2	0	0	0	0	0	0
		.45	14	3	0	0	0	0	0
		.75	208	381	271	317	183	1	0
Spline	.15	.15	72	118	98	102	56	45	14
		.45	108	84	118	94	62	47	19
		.75	144	157	92	95	51	21	18
	.45	.15	28	2	1	0	1	0	0
		.45	5	10	6	0	3	5	1
		.75	50	13	25	9	2	6	3
	.75	.15	2	0	0	0	0	0	0
		.45	15	2	0	0	0	0	0
		.75	10	3	2	2	8	0	0
Unspecified	.15	.15	662	681	751	807	846	891	948
		.45	698	721	752	803	853	881	946
		.75	654	757	796	855	873	929	968
	.45	.15	597	663	677	742	756	811	883
		.45	605	711	718	752	812	856	899
		.75	673	706	772	815	830	887	951
	.75	.15	611	627	656	685	700	771	837
		.45	578	663	671	696	760	812	846
		.75	615	692	716	774	808	845	885

Table 13. Number of admissible solutions for the logistic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	12	15	12	8	6	3	1
		.45	13	17	17	13	9	2	1
		.75	8	13	8	5	4	0	0
	.45	.15	6	8	10	2	2	0	0
		.45	11	6	5	6	0	0	0
		.75	1	3	2	0	0	0	0
	.75	.15	2	2	7	1	1	0	0
		.45	7	3	2	1	0	0	0
		.75	1	1	3	0	0	0	0
Spline	.15	.15	15	3	3	0	0	0	0
		.45	13	7	5	1	0	0	0
		.75	36	16	8	0	0	0	0
	.45	.15	1	3	0	0	0	0	0
		.45	11	5	2	0	0	0	0
		.75	9	9	1	0	0	0	0
	.75	.15	1	0	0	0	0	0	0
		.45	5	0	0	0	0	0	0
		.75	0	2	0	1	0	0	0

Unspecified	.15	.15	985	998	998	1000	1000	1000	1000
		.45	986	999	999	1000	1000	1000	1000
		.75	984	999	1000	1000	1000	1000	1000
	.45	.15	983	989	998	1000	1000	1000	1000
		.45	981	997	1000	1000	1000	1000	1000
		.75	990	996	1000	1000	1000	1000	1000
	.75	.15	975	989	999	1000	1000	1000	1000
		.45	975	995	1000	1000	1000	1000	1000
		.75	987	994	1000	1000	1000	1000	1000

4.3 DESCRIPTIVE STATISTICS

Tables 14 thru 19 provide means and standard deviations for each of the outcome measures (p-value for χ^2 model fit statistic, GFI, AGFI, CFI, SRMR and RMSEA) of the study broken down by fit technique (quadratic, spline and unspecified), slope variance (.15, .45 & .75), factor correlation (.15, .45 & .75) and sample size (50, 75, 100, 150, 200, 300 and 500) for the logarithmic growth curve. Separate tables were not created for the exponential and logistic growth curves due to the number of inadmissible solutions encountered in each. The inconsistency (in terms of convergence rates) across the cells of the factors for each of these curves makes comparisons difficult at best. However, Tables 20 thru 25 display means and standard deviations for the unspecified technique fit to the exponential and logistic curves. For these curves, this was the only fit technique which yielded an adequate number of admissible solutions for analysis.

The pattern of results for each of the outcome measures was consistent across the cells of the design. In general, better estimates of model fit were obtained (higher p-values for the χ^2 model fit statistic, higher values for the GFI, AGFI and CFI, and lower values for the SRMR and RMSEA) for the unspecified fit technique followed by the quadratic fit technique and then the spline fit technique. These results are consistent with the results found for the convergence rates. That is, convergence rates were higher for the unspecified technique followed by the quadratic technique followed by the spline technique. In addition, the values of the fit indices indicated better model fit as sample size increased but worse model fit as the slope variance and factor correlation increased. In many cases, these differences were very small but this was expected given the large number of replications (1000) conducted within each cell of the design. The pattern of results for the standard deviation was consistent with the pattern of results for the means.

Table 14. Descriptive statistics for the p-value of the model fit statistic for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	.08(.14)	.02(.06)	.01(.02)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.07(.14)	.03(.09)	.01(.04)	.00(.01)	.00(.00)	.00(.00)	.00(.00)
		.75	.07(.13)	.03(.08)	.01(.03)	.00(.01)	.00(.00)	.00(.00)	.00(.00)
	.45	.15	.06(.11)	.02(.06)	.01(.03)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.04(.09)	.01(.04)	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.75	.05(.10)	.01(.05)	.00(.02)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
	.75	.15	.02(.08)	.01(.03)	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.02(.06)	.01(.03)	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.75	.03(.07)	.01(.03)	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
Spline	.15	.15	.00(.02)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.75	.00(.01)	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
	.45	.15	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.75	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
	.75	.15	.00(.01)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.45	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
		.75	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)	.00(.00)
Unspecified	.15	.15	.48(.29)	.47(.29)	.48(.29)	.48(.29)	.49(.29)	.49(.29)	.50(.29)
		.45	.48(.30)	.49(.29)	.50(.29)	.49(.29)	.49(.29)	.50(.29)	.49(.29)
		.75	.47(.29)	.48(.29)	.49(.30)	.48(.29)	.47(.29)	.51(.28)	.49(.29)
	.45	.15	.47(.29)	.48(.29)	.50(.30)	.48(.29)	.49(.29)	.49(.29)	.50(.29)
		.45	.48(.29)	.48(.30)	.49(.28)	.49(.29)	.50(.29)	.50(.29)	.50(.29)
		.75	.47(.29)	.47(.29)	.48(.29)	.49(.29)	.49(.29)	.50(.29)	.51(.29)

	.75	.15	.47(.29)	.48(.29)	.49(.29)	.50(.29)	.49(.29)	.49(.29)	.50(.29)
		.45	.46(.29)	.50(.28)	.48(.29)	.49(.29)	.50(.29)	.49(.30)	.49(.29)
		.75	.48(.28)	.51(.29)	.49(.30)	.48(.29)	.50(.29)	.50(.29)	.49(.29)

Table 15. Descriptive statistics for the GFI for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	.95(.02)	.96(.02)	.97(.02)	.97(.01)	.98(.01)	.98(.01)	.98(.01)
		.45	.95(.03)	.96(.02)	.97(.02)	.97(.01)	.98(.01)	.98(.01)	.98(.01)
		.75	.95(.03)	.96(.02)	.97(.02)	.97(.01)	.97(.01)	.98(.01)	.98(.01)
	.45	.15	.94(.03)	.95(.02)	.96(.02)	.96(.02)	.96(.01)	.96(.01)	.97(.01)
		.45	.94(.03)	.95(.02)	.96(.02)	.96(.02)	.96(.01)	.97(.01)	.97(.01)
		.75	.94(.03)	.95(.02)	.95(.02)	.96(.02)	.96(.01)	.96(.01)	.97(.01)
	.75	.15	.92(.04)	.93(.03)	.93(.03)	.94(.02)	.94(.02)	.94(.01)	.94(.01)
		.45	.92(.03)	.93(.03)	.93(.03)	.94(.02)	.94(.02)	.94(.02)	.94(.01)
		.75	.92(.03)	.93(.03)	.93(.03)	.94(.02)	.94(.02)	.94(.01)	.94(.01)
Spline	.15	.15	.93(.03)	.93(.02)	.94(.02)	.94(.02)	.94(.01)	.95(.01)	.95(.01)
		.45	.92(.03)	.93(.02)	.94(.02)	.94(.02)	.94(.01)	.94(.01)	.95(.01)
		.75	.92(.03)	.93(.02)	.93(.02)	.94(.02)	.94(.01)	.94(.01)	.95(.01)
	.45	.15	.91(.03)	.91(.03)	.92(.02)	.92(.02)	.92(.02)	.92(.01)	.93(.01)
		.45	.90(.03)	.91(.03)	.92(.02)	.92(.02)	.92(.02)	.92(.01)	.94(.03)
		.75	.90(.03)	.91(.03)	.92(.02)	.92(.02)	.92(.02)	.92(.01)	.92(.01)
	.75	.15	.88(.03)	.88(.03)	.89(.02)	.89(.02)	.89(.02)	.89(.01)	.89(.01)
		.45	.87(.03)	.88(.03)	.88(.02)	.89(.02)	.89(.02)	.89(.01)	.89(.01)
		.75	.87(.03)	.88(.03)	.88(.02)	.89(.02)	.89(.02)	.89(.01)	.89(.01)
Unspecified	.15	.15	.95(.03)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.02)
		.45	.95(.03)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.75	.95(.02)	.97(.02)	.97(.01)	.94(.02)	.99(.01)	.99(.00)	.99(.00)
	.45	.15	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
	.75	.15	.96(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.96(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.96(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)

Table 16. Descriptive statistics for the AGFI for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	.83(.09)	.86(.07)	.88(.06)	.90(.05)	.91(.04)	.92(.03)	.93(.02)
		.45	.82(.09)	.86(.07)	.88(.06)	.89(.05)	.91(.04)	.92(.03)	.92(.02)

		.75	.81(.10)	.85(.08)	.87(.06)	.90(.05)	.90(.04)	.92(.03)	.92(.03)
	.45	.15	.78(.11)	.82(.09)	.83(.07)	.85(.06)	.86(.05)	.87(.04)	.88(.03)
		.45	.77(.11)	.81(.09)	.83(.07)	.85(.06)	.86(.05)	.87(.04)	.87(.03)
		.75	.77(.11)	.81(.09)	.83(.08)	.85(.06)	.85(.05)	.86(.04)	.87(.03)
	.75	.15	.70(.13)	.73(.11)	.74(.09)	.77(.07)	.77(.06)	.78(.05)	.79(.04)
		.45	.69(.13)	.73(.11)	.74(.09)	.76(.08)	.77(.07)	.78(.06)	.78(.04)
		.75	.69(.13)	.73(.11)	.74(.09)	.76(.07)	.77(.07)	.78(.05)	.78(.04)
Spline	.15	.15	.72(.11)	.74(.08)	.76(.07)	.78(.06)	.79(.05)	.79(.04)	.80(.03)
		.45	.71(.10)	.75(.09)	.76(.07)	.77(.06)	.78(.05)	.79(.04)	.80(.03)
		.75	.70(.11)	.74(.09)	.75(.08)	.77(.06)	.78(.05)	.79(.04)	.80(.03)
	.45	.15	.65(.12)	.67(.10)	.68(.08)	.69(.07)	.71(.06)	.71(.04)	.72(.03)
		.45	.64(.11)	.67(.10)	.68(.08)	.70(.07)	.70(.06)	.71(.05)	.72(.04)
		.75	.63(.11)	.66(.09)	.68(.08)	.70(.07)	.70(.06)	.71(.05)	.72(.03)
	.75	.15	.53(.12)	.55(.10)	.57(.09)	.58(.07)	.59(.06)	.59(.05)	.60(.04)
		.45	.52(.11)	.55(.10)	.57(.09)	.58(.07)	.58(.06)	.59(.05)	.59(.13)
		.75	.52(.12)	.55(.10)	.56(.09)	.57(.07)	.58(.06)	.58(.05)	.59(.04)
Unspecified	.15	.15	.81(.09)	.87(.06)	.90(.05)	.93(.03)	.95(.03)	.97(.02)	.98(.01)
		.45	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
		.75	.81(.09)	.87(.07)	.90(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
	.45	.15	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
		.45	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.96(.03)	.97(.02)	.98(.01)
		.75	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.02)	.97(.02)	.98(.01)
	.75	.15	.83(.08)	.89(.06)	.92(.04)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
		.45	.83(.09)	.89(.06)	.92(.04)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
		.75	.83(.08)	.89(.06)	.92(.05)	.94(.03)	.96(.02)	.97(.02)	.98(.01)

Table 17. Descriptive statistics for the CFI for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

			Sample Size						
Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	50	75	100	150	200	300	500
Quadratic	.15	.15	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
	.45	.15	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
	.75	.15	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.00)	.98(.00)
		.45	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.99(.00)	.99(.00)
		.75	.98(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
Spline	.15	.15	.97(.02)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)
		.45	.97(.02)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.00)
		.75	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.01)	.98(.00)
	.45	.15	.97(.02)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.00)
		.45	.97(.02)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.00)

		.75	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.01)	.97(.00)
	.75	.15	.96(.02)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)
		.45	.96(.02)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)
		.75	.96(.02)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)	.96(.01)
Unspecified	.15	.15	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
	.45	.15	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.03)	.99(.00)	.99(.01)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
	.75	.15	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)

Table 18. Descriptive statistics for the SRMR for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

			Sample Size						
Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	50	75	100	150	200	300	500
Quadratic	.15	.15	.05(.02)	.05(.02)	.05(.02)	.05(.02)	.04(.02)	.04(.01)	.04(.01)
		.45	.05(.02)	.05(.02)	.05(.02)	.05(.02)	.04(.02)	.04(.01)	.04(.01)
		.75	.05(.03)	.05(.02)	.05(.02)	.05(.02)	.05(.02)	.04(.01)	.04(.01)
	.45	.15	.06(.03)	.06(.02)	.06(.02)	.06(.02)	.06(.12)	.06(.01)	.06(.01)
		.45	.06(.03)	.06(.03)	.06(.02)	.06(.02)	.06(.02)	.06(.01)	.06(.01)
		.75	.07(.03)	.06(.03)	.06(.02)	.06(.02)	.06(.02)	.06(.01)	.06(.01)
	.75	.15	.07(.03)	.08(.21)	.07(.02)	.07(.02)	.07(.02)	.07(.01)	.07(.01)
		.45	.08(.03)	.08(.03)	.07(.02)	.07(.02)	.07(.02)	.07(.02)	.07(.01)
		.75	.08(.03)	.08(.03)	.08(.03)	.08(.02)	.08(.02)	.07(.02)	.07(.01)
Spline	.15	.15	.08(.03)	.08(.03)	.08(.02)	.08(.02)	.08(.02)	.08(.01)	.08(.01)
		.45	.09(.04)	.08(.03)	.08(.03)	.08(.02)	.08(.02)	.08(.02)	.08(.01)
		.75	.09(.04)	.08(.03)	.08(.03)	.08(.02)	.08(.02)	.08(.02)	.08(.01)
	.45	.15	.10(.03)	.10(.03)	.10(.02)	.10(.02)	.10(.02)	.10(.02)	.10(.01)
		.45	.11(.04)	.11(.03)	.11(.03)	.11(.02)	.11(.02)	.10(.02)	.10(.01)
		.75	.11(.04)	.11(.04)	.11(.03)	.11(.02)	.11(.02)	.11(.02)	.11(.01)
	.75	.15	.13(.05)	.13(.03)	.13(.03)	.13(.02)	.13(.02)	.13(.02)	.13(.01)
		.45	.14(.04)	.13(.04)	.13(.03)	.13(.03)	.14(.02)	.14(.02)	.14(.02)
		.75	.14(.05)	.14(.04)	.14(.04)	.14(.03)	.14(.03)	.14(.02)	.14(.02)
Unspecified	.15	.15	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)	.01(.01)
		.45	.05(.01)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)	.01(.01)
		.75	.04(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)	.01(.01)
	.45	.15	.04(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)
		.45	.04(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)
		.75	.03(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)
	.75	.15	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)	.01(.00)
		.45	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)	.01(.00)

		.75	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)	.01(.01)	.01(.00)

Table 19. Descriptive statistics for the RMSEA for the logarithmic growth curve by fit technique, slope variance, factor correlation and sample size

Model	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Quadratic	.15	.15	.09(.08)	.05(.02)	.07(.05)	.06(.04)	.10(.04)	.10(.03)	.10(.02)
		.45	.10(.08)	.10(.06)	.10(.06)	.10(.05)	.10(.04)	.11(.03)	.11(.02)
		.75	.11(.08)	.10(.07)	.10(.06)	.10(.04)	.11(.04)	.10(.03)	.11(.02)
	.45	.15	.13(.09)	.13(.07)	.14(.06)	.14(.04)	.14(.04)	.14(.03)	.14(.02)
		.45	.14(.08)	.13(.07)	.14(.06)	.14(.05)	.14(.04)	.14(.03)	.14(.02)
		.75	.14(.08)	.14(.07)	.14(.06)	.14(.04)	.15(.04)	.15(.03)	.15(.02)
	.75	.15	.19(.08)	.19(.07)	.19(.05)	.19(.04)	.20(.04)	.20(.03)	.20(.02)
		.45	.19(.08)	.19(.06)	.20(.05)	.20(.04)	.20(.04)	.20(.03)	.20(.02)
		.75	.19(.08)	.19(.06)	.19(.05)	.20(.04)	.20(.04)	.20(.03)	.20(.02)
Spline	.15	.15	.19(.08)	.20(.06)	.20(.05)	.20(.04)	.20(.03)	.20(.02)	.20(.02)
		.45	.20(.07)	.19(.06)	.20(.05)	.20(.04)	.20(.03)	.20(.02)	.20(.02)
		.75	.20(.08)	.20(.06)	.20(.05)	.20(.03)	.20(.03)	.20(.02)	.20(.02)
	.45	.15	.24(.08)	.24(.06)	.25(.05)	.25(.04)	.25(.03)	.25(.02)	.25(.02)
		.45	.25(.07)	.25(.06)	.25(.04)	.25(.04)	.25(.03)	.25(.03)	.25(.02)
		.75	.25(.07)	.25(.06)	.25(.05)	.25(.04)	.25(.03)	.25(.03)	.25(.02)
	.75	.15	.31(.07)	.31(.05)	.31(.05)	.31(.04)	.31(.03)	.31(.02)	.31(.02)
		.45	.32(.06)	.31(.05)	.31(.05)	.31(.04)	.32(.03)	.32(.03)	.31(.02)
		.75	.31(.07)	.31(.05)	.31(.05)	.32(.03)	.31(.03)	.32(.03)	.32(.02)
Unspecified	.15	.15	.10(.08)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.02(.02)
		.45	.10(.08)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.03(.02)
		.75	.10(.07)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.03(.02)
	.45	.15	.10(.07)	.08(.06)	.06(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
		.45	.09(.07)	.08(.06)	.06(.05)	.05(.04)	.04(.04)	.04(.03)	.03(.02)
		.75	.10(.08)	.08(.06)	.07(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
	.75	.15	.09(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)
		.45	.09(.07)	.06(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.03(.02)
		.75	.09(.07)	.06(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)

Tables 20 thru 25 display means and standard deviations for each outcome measure for the unspecified fit technique across all of the factors of the study for the exponential and logistic growth curves. The results for these curves are reported in this manner since the unspecified fit technique was the only technique to yield convergence rates at or above 60% within each cell of the other conditions. Therefore, mean comparisons were deemed to be meaningful. The results found in Tables 20 thru 25 are consistent with those found when fitting the unspecified model to the logarithmic curve.

Table 20. Descriptive statistics for the p-value of the model fit statistic for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Exponential	.15	.15	.47(.30)	.48(.29)	.48(.29)	.48(.29)	.49(.29)	.50(.30)	.50(.29)
		.45	.48(.28)	.48(.29)	.49(.29)	.48(.29)	.49(.29)	.51(.30)	.48(.29)
		.75	.46(.28)	.51(.29)	.48(.29)	.50(.30)	.50(.28)	.51(.29)	.49(.29)
	.45	.15	.49(.30)	.49(.30)	.47(.29)	.50(.29)	.48(.29)	.50(.29)	.51(.29)
		.45	.48(.29)	.49(.29)	.49(.29)	.50(.29)	.48(.28)	.52(.29)	.48(.28)
		.75	.48(.29)	.49(.29)	.51(.29)	.50(.29)	.51(.29)	.48(.29)	.51(.29)
	.75	.15	.48(.30)	.49(.28)	.50(.29)	.50(.29)	.48(.29)	.49(.29)	.50(.29)
		.45	.47(.06)	.50(.28)	.48(.29)	.50(.28)	.52(.30)	.51(.29)	.51(.30)
		.75	.48(.28)	.49(.28)	.49(.29)	.48(.29)	.50(.29)	.50(.28)	.49(.28)
Logistic	.15	.15	.48(.30)	.47(.29)	.47(.28)	.50(.30)	.50(.29)	.50(.28)	.50(.29)
		.45	.47(.30)	.48(.29)	.49(.29)	.50(.29)	.49(.29)	.49(.28)	.49(.28)
		.75	.48(.29)	.49(.29)	.49(.29)	.50(.29)	.50(.29)	.51(.29)	.51(.29)
	.45	.15	.47(.30)	.49(.30)	.50(.29)	.51(.29)	.50(.29)	.48(.29)	.49(.29)
		.45	.48(.30)	.47(.30)	.50(.29)	.48(.28)	.49(.29)	.51(.29)	.49(.29)
		.75	.48(.29)	.49(.29)	.49(.29)	.50(.29)	.49(.30)	.49(.28)	.48(.30)
	.75	.15	.48(.29)	.47(.30)	.50(.29)	.47(.29)	.48(.29)	.49(.28)	.50(.30)
		.45	.46(.29)	.46(.29)	.49(.29)	.49(.29)	.51(.29)	.50(.29)	.50(.29)
		.75	.48(.28)	.48(.29)	.48(.29)	.49(.29)	.52(.29)	.50(.29)	.51(.29)

Table 21. Descriptive statistics for the GFI for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

			Sample Size						
Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	50	75	100	150	200	300	500
Exponential	.15	.15	.95(.03)	.93(.03)	.97(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.45	.95(.02)	.97(.02)	.97(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.75	.95(.02)	.97(.02)	.97(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
	.45	.15	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.45	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.95(.02)	.90(.02)	.91(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
	.75	.15	.95(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.95(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.96(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
Logistic	.15	.15	.95(.03)	.97(.02)	.97(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.95(.03)	.97(.02)	.97(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.75	.95(.02)	.97(.02)	.97(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
	.45	.15	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.01)	.99(.00)
		.45	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.95(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
	.75	.15	.96(.02)	.97(.02)	.98(.01)	.98(.01)	.99(.01)	.99(.00)	.99(.00)
		.45	.96(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.96(.02)	.97(.02)	.98(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)

Table 22. Descriptive statistics for the AGFI for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

Table 4.14. Descriptive statistics for the AGFI for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size.

			Sample Size						
Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	50	75	100	150	200	300	500
Exponential	.15	.15	.81(.10)	.87(.07)	.90(.05)	.93(.04)	.95(.03)	.97(.02)	.98(.01)
		.45	.81(.09)	.87(.06)	.90(.05)	.93(.03)	.95(.03)	.97(.01)	.98(.01)
		.75	.81(.09)	.88(.07)	.90(.05)	.94(.04)	.95(.02)	.97(.02)	.98(.01)
	.45	.15	.82(.09)	.88(.06)	.90(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
		.45	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.02)	.97(.02)	.98(.01)
		.75	.82(.08)	.88(.06)	.91(.05)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
	.75	.15	.83(.09)	.89(.06)	.92(.05)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
		.45	.83(.09)	.89(.06)	.91(.05)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
		.75	.83(.08)	.89(.06)	.92(.05)	.94(.03)	.96(.02)	.97(.01)	.98(.01)
Logistic	.15	.15	.81(.09)	.87(.06)	.90(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
		.45	.81(.10)	.87(.07)	.90(.05)	.94(.03)	.95(.02)	.97(.02)	.98(.01)
		.75	.81(.09)	.88(.06)	.90(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)

	.45	.15	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
		.45	.82(.09)	.88(.07)	.91(.05)	.94(.03)	.95(.02)	.97(.02)	.98(.01)
		.75	.82(.09)	.88(.06)	.91(.05)	.94(.03)	.95(.03)	.97(.02)	.98(.01)
	.75	.15	.84(.08)	.89(.06)	.92(.05)	.94(.03)	.96(.02)	.97(.01)	.98(.01)
		.45	.83(.08)	.89(.06)	.92(.04)	.94(.03)	.96(.02)	.97(.02)	.98(.01)
		.75	.84(.08)	.89(.06)	.92(.05)	.95(.03)	.96(.02)	.97(.02)	.98(.01)

Table 23. Descriptive statistics for the CFI for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Exponential	.15	.15	.97(.03)	.98(.02)	.99(.01)	.99(.01)	.99(.01)	.99(.01)	.99(.00)
		.45	.98(.03)	.99(.02)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)
		.75	.98(.00)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)
	.45	.15	.98(.02)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)
		.45	.98(.02)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)
		.75	.98(.02)	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)
	.75	.15	.99(.02)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.02)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
Logistic	.15	.15	.99(.02)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
	.45	.15	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
	.75	.15	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.45	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)
		.75	.99(.01)	.99(.01)	.99(.00)	.99(.00)	.99(.00)	.99(.00)	.99(.00)

Table 24. Descriptive statistics for the SRMR for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Exponential	.15	.15	.08(.03)	.06(.02)	.05(.02)	.04(.02)	.04(.01)	.03(.01)	.02(.01)
		.45	.08(.03)	.06(.02)	.05(.02)	.04(.02)	.04(.01)	.03(.01)	.02(.01)
		.75	.08(.03)	.06(.02)	.05(.02)	.04(.02)	.04(.01)	.03(.01)	.02(.01)
	.45	.15	.07(.03)	.06(.02)	.05(.02)	.04(.01)	.04(.01)	.03(.01)	.02(.01)
		.45	.07(.03)	.05(.02)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)
		.75	.07(.03)	.05(.02)	.04(.02)	.04(.01)	.03(.01)	.03(.01)	.02(.01)
	.75	.15	.06(.02)	.05(.02)	.04(.02)	.04(.01)	.03(.01)	.03(.01)	.02(.01)

		.45	.06(.03)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)
		.75	.06(.02)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)
Logistic	.15	.15	.06(.02)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)
		.45	.06(.03)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)
		.75	.06(.03)	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)
	.45	.15	.05(.02)	.04(.02)	.03(.01)	.03(.01)	.02(.01)	.02(.01)	.02(.01)
		.45	.05(.02)	.04(.02)	.03(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)
		.75	.05(.02)	.04(.02)	.03(.02)	.03(.02)	.02(.01)	.02(.01)	.01(.01)
	.75	.15	.04(.02)	.03(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)
		.45	.04(.02)	.03(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)
		.75	.04(.02)	.03(.02)	.03(.01)	.02(.01)	.02(.01)	.02(.01)	.01(.01)
		.45	.09(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)
		.75	.08(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)

Table 25. Descriptive statistics for the RMSEA for the unspecified fit technique by growth curve, slope variance, factor correlation and sample size

Curve	$\sigma^2_{\varsigma_1}$	$\sigma_{\varsigma_1\varsigma_2}$	Sample Size						
			50	75	100	150	200	300	500
Exponential	.15	.15	.11(.08)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.03(.02)
		.45	.10(.07)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.03(.02)
		.75	.11(.08)	.08(.06)	.07(.05)	.06(.05)	.05(.04)	.04(.03)	.03(.02)
	.45	.15	.10(.08)	.08(.06)	.07(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
		.45	.10(.07)	.08(.06)	.07(.05)	.05(.04)	.05(.04)	.03(.03)	.03(.02)
		.75	.10(.07)	.08(.06)	.06(.05)	.04(.04)	.04(.03)	.04(.03)	.03(.02)
	.75	.15	.09(.08)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.03(.02)
		.45	.09(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.03(.02)
		.75	.09(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.03(.02)
Logistic	.15	.15	.10(.08)	.08(.06)	.07(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
		.45	.10(.08)	.08(.06)	.07(.05)	.06(.04)	.05(.04)	.04(.03)	.03(.02)
		.75	.10(.08)	.08(.06)	.07(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
	.45	.15	.10(.08)	.07(.06)	.06(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
		.45	.10(.08)	.08(.06)	.06(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
		.75	.09(.07)	.08(.06)	.06(.05)	.05(.04)	.05(.04)	.04(.03)	.03(.02)
	.75	.15	.08(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)
		.45	.09(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)
		.75	.08(.07)	.07(.06)	.06(.05)	.05(.04)	.04(.04)	.03(.03)	.02(.02)

4.4 MIXED ANOVA'S

A mixed ANOVA was conducted on each outcome measure with fit technique as the within-subjects factor, factor correlation, sample size and slope value as the between subjects factors. Due to the number of convergence problems encountered with the logistic curve it was not included in the analyses. In addition, the logarithmic and exponential curves were analyzed separately. In other words, the curve factor was not included as a between-subjects factor as was originally planned. Rather two separate Mixed ANOVAs were conducted.

As expected, the large number of replications resulted in significant p-values in virtually every condition. Therefore, interpretation of effect sizes measured via the partial η^2 became the focus of the analyses. The standards put forth by Cohen (1988) were utilized to identify the magnitude of the effect sizes. Tables 26 and 27 display the effect sizes from the Mixed ANOVAs for the logarithmic and exponential growth curves. As the tables show, the largest effect sizes were observed for the fit technique, sample size and the slope variance. There was also a large effect size for the interaction between the fit technique and the slope variance. All other effect sizes were small. In general, the effect sizes were larger for the logarithmic curve than for the exponential curve.

Table 26. Partial values for each of the outcome measures when data was fit to the logarithmic curve

	p-value	GFI	AGFI	CFI	SRMR	RMSEA
Fit	.745	.876	.876	.886	.808	.894
Sample	.007	.304	.304	.019	.031	.012
Slope	.003	.450	.450	.229	.224	.367
Factor Correlation	.000	.002	.002	.008	.004	.001
Fit*Sample	.002	.044	.044	.015	.028	.091
Fit*Slope	.001	.430	.430	.282	.269	.358

Fit*Factor Correlation	.000	.001	.001	.011	.006	.001
Sample*Slope	.002	.002	.002	.001	.004	.000
Sample*Factor Correlation	.000	.001	.001	.000	.001	.001
Slope*Factor Correlation	.000	.000	.000	.000	.001	.000
Fit*Sample*Slope	.001	.001	.001	.001	.001	.001
Fit*Sample*Factor Correlation	.000	.000	.000	.001	.001	.000
Fit*Slope*Factor Correlation	.000	.000	.000	.000	.001	.000
Sample*Slope*Factor Correlation	.000	.000	.000	.000	.001	.000
Fit*Sample*Slope*Factor Correlation	.000	.000	.000	.001	.001	.000

Table 27. Partial values for each of the outcome measures when data was fit to the exponential curve

	p-value	GFI	AGFI	CFI	SRMR	RMSEA
Fit	.084	.602	.602	.657	.244	.601
Sample	.055	.055	.055	.004	.052	.006
Slope	.005	.123	.123	.017	.084	.081
Factor Correlation	.060	.006	.006	.079	.007	.005
Fit*Sample	.055	.065	.065	.028	.047	.044
Fit*Slope	.005	.080	.080	.024	.064	.046
Fit*Factor Correlation	.060	.081	.081	.030	.032	.103
Sample*Slope	.028	.051	.051	.040	.013	.037
Sample*Factor Correlation	.038	.033	.033	.026	.008	.021
Slope*Factor Correlation	.013	.000	.000	.009	.002	.000
Fit*Sample*Slope	.028	.038	.038	.025	.032	.029
Fit*Sample*Factor Correlation	.038	.057	.057	.062	.043	.049
Fit*Slope*Factor Correlation	.012	.023	.023	.012	.020	.024
Sample*Slope*Factor Correlation	.000	.000	.000	.000	.000	.000
Fit*Sample*Slope*Factor Correlation	.000	.000	.000	.000	.000	.000

5.0 Discussion

The major purpose of this dissertation was to examine the effectiveness of the quadratic, spline and unspecified techniques in modeling nonlinear change in the framework of latent growth modeling. A second purpose was to provide a better understanding of the interpretation of these models. The sections to follow discuss important findings and how these findings relate to the interpretation of these models. The dissertation concludes with a discussion of the limitations of this research and directions for future research.

5.1 SUMMARY OF FINDINGS

The design of this simulation study resulted in the unspecified fit technique being the most effective approach to modeling the nonlinear change represented by the exponential, logarithmic and logistic curves as defined in this study. This result is not overly surprising, at least from a pure statistical perspective, given that the unspecified technique allows the shape of the empirical growth curves to be determined by the data. Thus, the "true" statistical model is being fit to the data. Unfortunately, the "true" statistical model may not provide a meaningful substantive interpretation. This issue will be discussed in more depth in the section to follow.

Results also confirmed the complexities involved in estimating quadratic and spline models. While quadratic models were difficult to estimate in this simulation study, they have been successfully utilized in the applied realm. These results support the notion that a theoretical basis is needed to estimate these models successfully.

The number of usable replications and resulting fit indices from the spline models were very similar to those for the quadratic models. This was particularly the case when both techniques were fit to the logistic curve. As is the case with the quadratic model, the spline model (i.e., linear spline as opposed to a mean spline) utilized in this study also appears to require a theoretical basis. Recall, linear spline models break observed empirical trajectories into pieces which typically identify change prior to and following the occurrence of an event (e.g., intervention). The data of this study was not generated in this manner so the spline model utilized was not necessarily appropriate, particularly for the logistic curve. On the other hand, a mean spline model may have generated very different results (at least in terms of model fit) due to the mean of each time point being utilized rather than predetermined values.

Another important point to make from the results of this study was the difficulty fitting the selected models to the logistic curve. The logistic curve in this study involved a slight increase to a certain point followed by a dramatic increase which was then followed by a leveling off period. The quadratic model did not fit well due to it being more appropriate for trajectories displaying non-constant change in a single direction. Likewise, the spline model utilized was also meant to model change in a single direction.

Perhaps a cubic model, which was designed for this type of non-constant change, would provide adequate model fit.

The effects (or lack thereof) of the other factors included in the simulation study also deserve mention. As is the case in most statistical procedures, sample size did play a role in obtaining better estimates of model fit regardless of the other conditions (i.e., in general, as sample size increased better model fit was achieved). However, the independent variables that were essentially ignored in the results section were the correlation between the factors and the variance in the slope parameter. Both of these factors were used to generate the random data and after carefully inspecting the results were not found to yield meaningful differences across the levels of each factor.

5.2 INTERPRETATION OF MODELS

Interpreting non-linear latent growth models is dependent upon the model that is utilized. For instance, interpretation of polynomial models is difficult due to the additional slope parameters included in the model. A quadratic model is much simpler to interpret than a cubic or quartic model. Unspecified and spline models are even easier to interpret than the quadratic model. Of course the linear model offers the easiest interpretation hence many researchers attempt to apply a linear transformation to the data rather than utilize one of the alternative non-linear techniques. However, care should be taken when interpreting the parameters and standard errors of non-linear data transformed to a linear scale.

The lack of parameter (and standard error) estimates obtained for this dissertation make interpretation of the models of interest difficult. These estimates were not examined since the primary purpose of this dissertation was to evaluate the model fit of the selected fit techniques. In addition, mathematical proofs for each of the parameters of interest (particularly the slope parameter) were also beyond the scope of this dissertation. However, the results of this dissertation highlight an important point in the interpretation of non-linear LGM's. That is, the unspecified model will likely fit (at least statistically) any type of non-linear curve. In addition, by comparing alternative models (such as the linear and quadratic) to the unspecified model, the general pattern of change can be depicted and is fairly easy to interpret. For instance, if the linear, quadratic or cubic models do not result in adequate model fit, the unspecified model can be adopted provided it offers reasonable statistical estimates. Then the empirical trajectory for the entire group can be interpreted as a general shape trajectory rather than having a linear and/or quadratic component.

Unspecified models can be interpreted in two different ways depending on the unspecified model utilized. The most common unspecified model is the one utilized in this study where the first and second factor loadings were fixed to 0 and 1, respectively. In this case, the estimated factor loadings are interpreted as the amount of change between time points in relation to the amount of change between the first two time points. On the other hand, an unspecified model where the first and last loadings are fixed to 0 and 1, respectively, is interpreted differently. In this case, the model is interpreted as the proportion of overall change between respective time points.

Interpretation of quadratic and spline models is difficult as they both require a sound theoretical basis. Spline models simply fit linear trajectories in between one or more transition points. Thus, these models are interpreted similarly to a linear model with the addition of the difference in change prior to and following the transition point. Polynomial models are much more complicated especially as the number of slope parameters increases. Bollen and Curran (2006) describe the complexities of interpreting cubic models as "…and the cubic model implies change in the change in the rate of change over time" and suggest applied researchers take great care in interpreting polynomial models.

5.3 LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

As with any simulation study, this dissertation was limited by the number of conditions that were included. This was particularly the case with the number of empirical growth trajectories that were examined. The exponential, logarithmic and logistic curves are all very common in the applied literature. However, future studies in the realm of non-linear LGM should examine differing degrees of non-linearity. For instance, it would be useful to determine at what point a non-linear model is more appropriate than a linear model. When does the exponential curve become 'flat' enough to warrant the use of a simple linear model? Surely at some point a difference in the statistical fit of the models will be realized. Determining this point is crucial to the selection of the appropriate model. Then, the even more important question of how the model is interpreted can be focused upon.

Future research should also further investigate the variability in the slope parameter and the correlation between the intercept and slope factors. This study investigated an adequate amount of variability in the slope parameter yet was unable to find a significant effect. At some point the individual differences in the slope parameter must have a bearing on the statistical estimates in the model. Could it be possible that it is more difficult to identify individual differences in the slope parameter as the data becomes more non-linear? And what role does the relationship between the intercept and slope play in estimating the model? Does this relationship have a different effect on non-linear models than linear models?

Future research should also examine the different types of spline models. The linear spline model utilized in this study was clearly inappropriate for the growth trajectories under investigation. A mean spline model may have yielded better statistical estimates of model fit in this study and may be a better general model for applied researchers.

5.4 SUMMARY

This dissertation investigated the application of various techniques for modeling non-linear change in the framework of latent growth curve modeling. Non-linear LGM's are becoming more common in the applied realm and as such the methodology behind these techniques deserves more attention. This dissertation contributes to research in this area by investigating the polynomial, spline and unspecified techniques. The major finding from this dissertation is that applied researchers would be wise to begin any analysis with the unspecified LGM unless a solid theoretical justification can be made for another model. The unspecified LGM provides solid statistical estimates of model fit and model parameters which can be used as a guide for determining the shape of the empirical trajectories or as the final model itself. The unspecified LGM is a also a viable alternative for applied researchers unwilling to face the difficulties associated with fitting higher order polynomial models or lacking the theoretical basis required for a linear spline model. These results are clearly based on a limited number of conditions that may or may not be applicable to real world datasets. More research must be done in this area to validate and build upon the conclusions made from this study.

PROGRAMS USED FOR THE SIMULATION STUDY

The following programs were used to generate and fit the data to the models of interest.

APPENDIX A

PROGRAMS USED FOR THE SIMULATION STUDY

A.1 EQS SIMULATION PROGRAMS

A.1.1 General Program for Generation Raw Data

```
/TITLE
 General Program for Generating Raw Data
/SPECIFICATIONS
VARIABLES=5; CASES=*;              !note 1
MATRIX=RAW; ANALYSIS = MOMENT;
METHOD=ML;
/EQUATIONS
V1 = 1F1 + *F2 + E1;               !note 2
V2 = 1F1 + *F2 + E2;
V3 = 1F1 + *F2 + E3;
V4 = 1F1 + *F2 + E4;
V5 = 1F1 + *F2 + E5;
F1 = a*V999 + D1;                  !note 3
F2 = b*V999 + D2;
/VARIANCES
E1 to E5 = 1;
D1 = 1;                            !note 4
D2 = *;
/COVARIANCES
D1,D2 = *;                         !note 5
/SIMULATIONS
POPULATION = MODEL;
SEED = 123456789;
```

```
REPLICATION = 1000;
SAVE = CONCATENATE;
DATA = 'NameOfRawData';
/END
```

note 1: 50, 75, 100, 150, 200, 300, & 500.
note 2: The slope estimates (*) are dependent upon the values in Table 3.2.
note 3: Intercept and slope parameters estimated from the data.
note 4: Values of slope variance (D2) taken from Table 3.3
note 5: Values of covariance between intercept and slope (D1,D2) taken from Table 3.3.

A.1.2 Program for Fitting Raw Data to the Quadratic Model.

```
/TITLE
 Fit Raw Data to Quadratic Model
/SPECIFICATIONS
 DATA='NameOfRawData';
 VARIABLES=5; LOOP=1000; CASES=*; MATRIX=RAW; !note 1
 ANALYSIS = MOMENT; METHOD=ML;
/EQUATIONS
 V1 = 1F1 + 0F2 + 0F3 + E1;
 V2 = 1F1 + 1F2 + 1F3 + E2;
 V3 = 1F1 + 2F2 + 4F3 + E3;
 V4 = 1F1 + 3F2 + 9F3 + E4;
 V5 = 1F1 + 4F2 + 16F3 + E5;
 F1 = *V999 + D1;
 F2 = *V999 + D2;
 F3 = *V999 + D3;
/VARIANCES
 E1 to E5 = *;
 D1 to D3 = *;
/COVARIANCES
 D1 to D3 = *;
/PRINT
FIT=ALL;
TABLE=EQUATION;
/OUTPUT
Parameters;
Standard Errors;
RSquare;
Codebook;
Listing;
DATA='ResultsOutput';
/TECHNICAL
ITER=500;
```

```
/END
note 1: Cases equal to 50, 75, 100, 150, 200, 300, & 500.
```

A.1.3 Program for Fitting Raw Data to the Unspecified Model.

```
/TITLE
 Fit Raw Data to Unspecified Model
/SPECIFICATIONS
 DATA='NameOfRawData';
 VARIABLES=5; LOOP=1000; CASES=*; MATRIX=RAW; !note 1
 ANALYSIS = MOMENT; METHOD=ML;
/EQUATIONS
 V1 = 1F1 + 0F2 + E1;
 V2 = 1F1 + 1F2 + E2;
 V3 = 1F1 + *F2 + E3;                    !note 2
 V4 = 1F1 + *F2 + E4;
 V5 = 1F1 + *F2 + E5;
 F1 = *V999 + D1;
 F2 = *V999 + D2;
/VARIANCES
 E1 to E5 = *;
 D1 to D2 = *;
/COVARIANCES
 D1 to D2 = *;
/PRINT
FIT=ALL;
TABLE=EQUATION;
/OUTPUT
Parameters;
Standard Errors;
RSquare;
Codebook;
Listing;
DATA='ResultsOutput';
/TECHNICAL
ITER=500;
/END
note 1: Cases equal to 50, 75, 100, 150, 200, 300, & 500.
note 2: Coefficents at time points 3, 4, and 5 are estimated from the data.
```

A.1.4 Program for Fitting Raw Data to the Spline Model.

```
/TITLE
 Fit Raw Data to Spline Model
/SPECIFICATIONS
 DATA='NameOfRawData';
 VARIABLES=5; LOOP=1000; CASES=*; MATRIX=RAW; !note 1
 ANALYSIS = MOMENT; METHOD=ML;
/EQUATIONS
 V1 = 1F1 + 0F2 + 0F3 + E1;
 V2 = 1F1 + 1F2 + 0F3 + E2;
 V3 = 1F1 + 2F2 + 2F3 + E3;
 V4 = 1F1 + 2F2 + 3F3 + E4;
 V5 = 1F1 + 2F2 + 4F3 + E5
 F1 = *V999 + D1;
 F2 = *V999 + D2;
 F3 = *V999 + D3;
/VARIANCES
 E1 to E5 = *;
 D1 to D3 = *;
/COVARIANCES
 D1 to D3 = *;
/PRINT
FIT=ALL;
TABLE=EQUATION;
/OUTPUT
Parameters;
Standard Errors;
RSquare;
Codebook;
Listing;
DATA='ResultsOutput';
/TECHNICAL
ITER=500;
/END
```

note 1: Cases equal to 50, 75, 100, 150, 200, 300, & 500.

BIBLIOGRAPHY

Bank, L., Burraston, B., & Snyder, J. (2004). Sibling conflict and ineffective parenting as predictors of adolescent boys' antisocial behavior and peer difficulties: additive and interactional effects. *Journal of Research on Adolescence, 14,* 99-125.

Bentler, P.M. (1979). Interrelations among models for the analysis of moment structures. *Multivariate Behavioral Research, 34,* 181-197.

Bentler, P.M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493-571.

Bentler, P.M. (1990). Comparative fit indexes in structural equation models. *Psychological Bulletin, 107,* 238-246.

Bentler, P.M. (1995). *EQS structural equations program manual*. Encino: Multivariate Software, Inc.

Bentler, P.M. (2004). *EQS 6 for windows program manual*. Encino: Multivariate Software, Inc.

Bentler, P.M. & Weeks, D.G. (1980). Linear structural equations with latent variables. *Psychometrika, 45,* 289-308.

Bock, R.D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

Box, G.E.P., & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, 26,* 211-252.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147-158.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Burchinal, M., & Appelbaum, M.I. (1991). Estimating individual developmental functions: methods and their assumptions. *Child Development, 62,* 23-43.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8,* 158-233.

Chan, D. (2003). Data analysis and modeling longitudinal processes. *Group & Organization Management, 28,* 341-365.

Chou, C-P., & Bentler, P.M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange Multiplier, and Wald tests. *Multivariate Behavioral Research, 25,* 115-136.

Chou, C-P., Bentler, P.M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347-357.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, NJ: Erlbaum.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147-167.

Curran, P.J. (2002). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529-569.

Curran, P.J., & Hussong, A.M. (2002). Structural equation modeling with repeated measures data: latent curve analysis. In D.S. Maskowitz & S. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: methods and applications,* (pp. 59-86). New York: Erlbaum Associates.

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16-29.

Duncan, T.E., Duncan, S.C., & Hops, H. (1996). The role of parents and older siblings in predicting adolescent substance use: Modeling development via structural equation latent growth curve methodology. *Journal of Family Psychology, 10,* 158-172.

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

Espy, K.A., Molfese, V.J., & Dilalla, L.F. (2001). Effects of environmental measures on intelligence in young children: growth curve modeling of longitudinal data. *Merrill-Palmer Quarterly, 47,* 42-73.

Fan, X. (2001). Parental involvement and students' academic achievement: A growth modeling analysis. *The Journal of Experimental Education, 70,* 27-61.

Francis, D.J., Fletcher, J.M., Stuebing, K.K., Davidson, K.C., & Thompson, N.M. (1991). Analysis of change: modeling individual growth. *Journal of Consulting and Clinical Psychology, 59,* 27-37.

Frank, K., & Seltzer, M. (1990, April). *Using the hierarchical linear model to model growth in reading achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Gerbing, D.W., & Anderson, J.C. (1993). Monte carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.

Ghisletta, P., & McArdle, J.J. (2001). Latent growth curve analyses of the development of height. *Structural Equation Modeling, 8,* 531-555.

Gierl, M.J., & Mulvenon, S. (1995, April). *Evaluation of the applications of fit indices to structural equation models in educational research: A review of literature from 1990 through 1994*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Goldstein, H.I. (1987). *Multilevel models in educational and social research*. London: Oxford University Press.

Goldsten, H.I. (1995). *Multilevel statistical models*. London: Edward Arnold.

Hess, B. (2000). Assessing program impact using latent growth modeling: a primer for the evaluator. *Evaluation and Program Planning, 23,* 419-428.

Hong, S., & Ho, H.Z. (2005). Direct and indirect longitudinal effects of parental involvement on student achievement: second order latent growth modeling across ethnic groups. *Journal of Educational Psychology, 97,* 32-42.

Hu, L., & Bentler, P.M. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 76-99). Newbury Park, CA: Sage.

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

Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351-362.

Huttenlocher, J.E., Haight, W., Bryk, A.S., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-249.

Huynh, H., & Mandeville, G.K. (1979). Validity conditions in repeated measures designs. *Psychological Bulletin, 86,* 964-973.

Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Academic Press.

Jöreskog, K.G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), *Applications of statistics* (pp. 265-287). Amsterdam: North-Holland.

Jöreskog, K.G. & Sörbom, D. (1984). *Lisrel VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods* [Computer Program]. Mooresville, IN: Scientific Software.

Kaplan, D. (2000). *Structural equation modeling: foundations and extensions*. Thousand Oaks, CA: Sage.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education, 77,* 189-215.

Khoo, S.T. (1997). *Assessing interactions between program effects and baseline: A latent curve approach*. Unpublished doctoral dissertation, University of California, Los Angeles.

Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness of fit indices in confirmatory factor analysis: The effect of sample size and model parsimony. *Quality and Quantity, 28,* 185-217.

McDonald, R.P. (1978). A simple comprehensive model for th analysis of covariance structure. *British Journal of Mathematical and Statistical Psychology, 31,* 59-72.

Meredith, W., & Tisak, J. (1984, October). *"Tuckerizing" curves*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

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

Muthén, B.O. (1997). Latent variable modeling of longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology,* (pp. 453-480). Boston: Blackwell.

Muthén, B.O. & Curran, P.J. (1997). General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 377-402.

Mosteller, F., & Tukey, J.W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

National Center for Educational Statistics (2004). *ECLS-K longitudinal kindergarten-third grade public use data files and electronic codebook*. Washington, DC: Author.

Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14,* 1-17.

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

Rogosa, D.R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 726-748.

Satorra, A., & Bentler, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 proceedings of the Business and Economics Sections* (pp. 308-313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Analysis of latent variables in developmental research* (pp. 399-419). Newbury Park, CA: Sage.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229-239.

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press, Inc.

Steiger, J.H., & Lind, J.C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting for the Psychometric Society, Iowa City, IA.

Stoolmiller, M., Duncan, T., Bank, L., & Patterson, G.R. (1993). Some problems and solutions in the study of change: Significant patterns in client resistance. *Journal of Consulting and Clinical Psychology, 61,* 920-928.

Tanaka, J.S. (1993). Multifaceted conceptions of fit in structure equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Tanaka, J.S. & Huba, G.J. (1985). How big is big enough?: Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,* 134-146.

Tisak, J., & Meredith, W. (1990). Descriptive and associative developmental models. In A. Von Eye (Ed.), *Statistical methods in longitudinal research* (Vol. 2, pp. 387-406). San Diego, CA: Academic Press.

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

U.S. Department of Education, National Center for Education Statistics. (2006). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: Fifth-Grade Public Use Data Files and Electronic Codebook* (NCES 2006-035). Washington, DC: U.S. Government Printing Office.

Willett, J.B. (1988). Questions and answers in the measurement of change. In E.Z. Rothkopf(Ed.), *Review of Research in Education, 15,* 345-422.

Willett, J.B. (1994). Measuring change more effectively by modeling individual change over time. In T. Husen & T.N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed.). Elmsford, NY: Pergamon Press.

Willett, J.B. & Sayer, A.S. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363-381.

Willett, J.B., Ayoub, C.C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. *Journal of Consulting and Clinical Psychology, 59,* 38-47.