COMPARISON OF LATENT GROWTH MODELS WITH DIFFERENT TIME CODING STRATEGIES IN THE PRESENCE OF INTER-INDIVIDUALLY VARYING TIME POINTS OF MEASUREMENT

By

BURAK AYDIN

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS IN EDUCATION

UNIVERSITY OF FLORIDA
2010

To
Mehmetali, Kıymet and Burçak Aydın

ACKNOWLEDGMENTS

First of all, I would like to thank Dr. Walter Leite and Dr. James Algina for guiding me in my thesis. I thank to faculty and students of the research and evaluation methodology program in the Educational Psychology Department. I would also like to thank Turkish Government for the financial support.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 7

LIST OF FIGURES 9

CHAPTER

1 INTRODUCTION 12

2 EMPIRICAL ILLUSTRATION 22

Sample 26
Measures 27
Mathematics Scores 27
Reading Scores 28
Learning Related Behaviors 28
Socio Economic Status (SES) 29
Age, Gender, Race, Kindergarten Retention, Disability Status: 29
The Final Data 29
Statistical Models 31
Hierarchical Linear Modeling 31
Latent Growth Curve Model 32
Time points 33
Four different time coding strategies 34
Results 37
Discussion 42

3 SIMULATION STUDY 44

Design 44
Data Generation 48
Data Analysis 49
Results 51
Convergence Rates and Improper Solutions 51
Relative Bias of Parameter Estimates and Standard Errors 51
Model Fit 53
Discussion 57

4 CONCLUSION 67

APPENDIX

CALCULATING THE NUMBER OF DAYS WITH SPSS 71

LIST OF REFERENCES 72

BIOGRAPHICAL SKETCH.. 77

LIST OF TABLES

Table page

2-1 Demographic Characteristic for the ECLS-K Full and Analytical Samples 30

2-2 SES and Math IRT scores for the ECLS-K Full and Analytical Samples 31

2-3 Descriptive Statistics for Number of Days between the Fall Kindergarten Assessment and the Other Assessments 34

2-4 Number of days between assessments divided by 100 35

2-5 Number of days between assessments divided by 358 36

2-6 Descriptive Statistics for Mathematics IRT Scores First through Fifth Grade 36

2-7 Parameter estimates of RRC model and the unconditional LGM model with different time coding combinations. 37

2-8 Variances of Parameter Estimates 40

2-9 Estimated Mean of IRT Based Math Scores ($y_{it} = \alpha_i + \lambda_t \beta_{1i} + \lambda_t^2 \beta_{2i} + \varepsilon_{it}$)........ 40

2-10 Model Fit Information for Unconditional LGM 40

2-11 Comparison of parameter estimates for the intercept between Morgan's ISAO model and the conditional LGM model with different time coding combinations. 41

2-12 Comparison of parameter estimates for the linear slope between Morgan's ISAO model and the conditional LGM model with different time coding combinations. 42

2-13 Model Fit Indices for Conditional LGM 42

3-1 Characteristics of assessment time points distributions 45

3-2 Population Parameters 48

3-3 Example of Mplus LGM Program 50

3-4 Percentages of convergence and improper solutions for each condition 62

3-5 Comparison of relative parameter bias estimates across conditions.................... 63

3-6 Relative bias estimates for standard errors. 64

3-7 Chi-square statistic results across conditions 65

3-8 Comparison of model fit information across conditions. 66

LIST OF FIGURES

Figure		page
2-1	Path model for an unconditional quadratic latent growth model.	25
2-2	Mathematics Growth Trajectory	36
3-1	Distributions for the third assessment wave time points in narrow range.	47

Abstract of Thesis Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Master of Arts in Education

COMPARISON OF LATENT GROWTH MODELS WITH DIFFERENT TIME CODING STRATEGIES IN THE PRESENCE OF INTER-INDIVIDUALLY VARYING TIME POINTS OF MEASUREMENT

By

Burak Aydin

August 2010

Chair: Walter L. Leite
Major: Research and Evaluation Methodology

Especially in the last two decades, there is a much greater availability of large and complex longitudinal data sets. As a structural equation modeling (SEM) approach, latent growth models (LGM) is one of the widely used methods to analyze longitudinal data sets. In a LGM individual growth is defined as a function of time. Therefore, it is important to correctly define time indicators and adequately implement them into a LGM. There have been some studies guiding researchers to code time adequately in a growth model and these studies generally used participants' ages as the time indicators. However, in the education discipline it is also common to use measurement waves to define time in a growth model. Furthermore, not many studies investigated the heterogeneity in assessment dates, and researchers tended to assume that each child was measured at the same time in a measurement wave. Thus, it is not clear how procedures perform in the presence of inter-individually varying time points of measurement. Consequently, the aim of this thesis was to determine if the results of a LGM is valid when heterogeneity in a measurement wave was omitted.

An empirical study to compare different time coding strategies in a LGM is presented in this thesis. A conditional quadratic growth model was constructed based on a published growth model in which mathematical development was investigated for a subsample taken from Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K). Following the empirical study, a simulation study was built on an unconditional linear growth model with three measurement occasions. The simulation study examined the heterogeneity effect through the manipulation of three factors: the sample size factor had three levels, 200, 2000 and 8000: the range of the data collection period for each measurement had three levels, narrow, moderate and wide: the distribution of measurement occasions had three levels, uniform, moderately skewed and extremely skewed.

The results of the empirical study showed that omitting the heterogeneity in assessment dates for the ECLS-K did not cause considerable differences. However, the simulation study revealed that residual variances were affected by the range factor. Residual variances increased as the range of the measurement occasions increased. An interaction between the range and distribution factor was associated with substantial negative bias for the slope variance estimates and the largest absolute value of the bias was found for the condition with a wide range and an extremely skewed distribution of measurement occasions. A sample size of 200 was associated with an increased number of improper solutions when heterogeneity was omitted. The chi square test statistic indicated lack of exact fit for many conditions. The fit indices indicated models fit adequately to the generated data sets, even though some parameters were estimated with substantial bias

CHAPTER 1
INTRODUCTION

Most of the researchers in the social science disciplines are interested in understanding the source of stability and change in variables. A psychologist might investigate a student population by collecting data in several time periods in order to see if they develop a particular skill. A sociologist, studying archives, might want to understand crime trend in a community. An economist, for example, might want to develop a model to examine changes in amounts of import/export products for a country over years. The common point in these examples is that in each case change occurs as a function of time. The time unit can be days, weeks, months, years and even decades. However, in the education discipline, time units are generally specified as months or years, and the unit of analysis is generally students. For example, an educational researcher might try to map the advancement of math aptitude of individuals for the first six years of school. In this case researcher needs to collect data at several grades. In this fashion, the collected data can be called a longitudinal data set. Since, in the past decades, there is much greater availability of large and complex longitudinal data sets, researchers are attracted to develop methods to analyze how change comes about, and how much change occurs. Analysis of variance(ANOVA) , multivariate analysis of variance (MANOVA), analysis of covariance (ANCOVA), multivariate analysis of covariance(MANCOVA), auto-regressive and cross lagged multiple regression are some of the traditional methods to analyze longitudinal data sets. Each of these methods has pros and cons (see e.g., Collins & Sayer, 2001; Gottman, 1995). Because researchers also tried to define how the change process differ among individuals and because above mentioned methods have assumptions which are rarely met in practice in the social sciences, such as sphericity, new methods have been developed to assess individual change. Actually, recent methodological research in the measurement of individual change has been reoriented from 'change' to 'growth' and has begun to focus directly on the individual trajectories. Hierarchical linear modeling (HLM) (see Raudenbush &Bryk, 2002) and a structural equation modeling (SEM) approach called latent growth modeling (LGM) have become the most frequently used methods to assess growth over time. Excellent introductions to LGM can be found in Meredith & Tisak, 1990; B. O. Muthén & Khoo, 1998; T. E. Duncan, Duncan, Strycker, Li, & Alpert, 1999; Willett & Keiley, 2000, Bollen & Curran 2006. Both LGM and HLM use the individual data approach that is the expectations regarding means and covariances are modeled at the level of individuals. Because both approaches are similar in terms of data structure and model structure, they are equivalent models in most conditions. Nevertheless, Mehta and West (2000) pointed out four reasons to use SEM over HLM approach. They stated that formulating a structural model involving growth factors as predictor or criterion variable is easier with SEM. The second reason is that it is possible to specify a growth model with a measurement structure in which latent variables are measured by several indicators at each time point. It is also easy to create models with different forms of growth and to investigate cohort effects when SEM is used.

LGM methods simultaneously focus on changes in covariance, variances, and mean values over time (Dunn, Everitt.&Pickles, 1993) and these methods use initial status and developmental trajectories (e.g., linear, quadratic) to describe change in an individual’s behavior. LGM can estimate the variability across individuals in both initial status and trajectories, as well as specify coefficients for testing the effect of other variables or constructs to explain variations in the initial status and trajectories (Hancock & Lawrence, 2006). In LGM approach, the intercept and slope terms are treated as latent variables and they are allowed to vary between individuals. As an oversimplified example, assume a theory dictates that reading scores for a first grade student population should follow a generally linear growth trajectory over 3 years. The initial point of measurement (time 1) can serve as a reference point for the development. Therefore, each individual's score can be written as;

$$y_{it} = \alpha_i + \lambda_t \beta_{1i} + \varepsilon_{it} \tag{1-1}$$

where y_{it} is the reading score for *i* th student at time *t*, α_i defines the intercept for case *i*, and β_{1i} the linear slope. The λ_t, is a constant where $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 2$ which scales the intercept to be interpreted as the initial status, ε_{it} is the error term. Note that interpretation of α_i depends on the coding of time, λ_t . Equation 1-1 implies that individual deviations in growth trajectories affected by the coding of time. In a LGM it is important to carefully specify how time is measured and defined.

The databases of Academic Search Premier and PsycINFO were searched in order to find applied LGM studies. Published journal articles using LGM showed that researchers tend to assume that each individual was assessed at the same time and so they use fixed time points for measurement occasions (e.g., Morgan, Farkas and Wu,2009; Perez 2008) Generally speaking, assessing every individual at the same time may be considered ideal. However, in longitudinal studies with a large sample of students (ECLS-K, LSAY) it is frequently impractical to conduct assessments at the exact same occasions for all students. These studies yield data which are partially homogeneous in the times of measurement, in that the reports cover the same period of time (e.g., fall kindergarten, spring first grade). Considering these 'partially homogeneous' time points and assuming an individual’s responses are dependent on the particular times of measurement, taking into account inter-individually varying time of measurement could be essential to have valid interpretations of the data (Blozis and Cho, 2008). Omitting the heterogeneity of time of assessment occasions may cause biased estimations of change.

My literature review showed that relatively few studies have been conducted on time coding issues in a LGM. One of the time coding issues is to decide whether age of the participants or the measurement dates will be used to scale time variable. Mehta and West (2000) published a methodological guideline in which they provided different scaling of time strategies in linear growth curve models. They used participants' ages to create time indicators and investigated three different time-structured measurement designs. In the first design, they built a LGM for participants who were identically aged at the beginning of the study. Occasions of measurement were assumed to be equally spaced, with participants measured at ages 12, 13, 14 and 15. The authors investigated two different time scaling schemes under the first design. The difference between these schemes was the origin of time which is often chosen by the researcher to meet a particular theoretical interest. Time points are assigned as 0, 1, 2 and 3 when the origin is chosen to be age 12. Then the average intercept represents the mean score in the population at age 12. Time points are assigned as -3,-2,-1, 0 when the origin set to be age of 15. Then the intercept represents the mean score in the population at age 15. As expected, estimations differed in two different schemes due to change in time scale.

The second design was a cohort sequential design and is not related to focus of current research. The third design was more realistic case in which age of the subjects varied between 12 and 14 at the beginning of the study. The authors suggested different scaling strategies in the presence of age heterogeneity in order to obtain more accurate estimates because the scaling for the first design fails to capture time dependence. They investigated two time scaling schemes in the presence of age heterogeneity. For the first model, they scaled time with respect to a common origin across all individuals, meaning ages of participants were centered to a common age. For example, assigning 12 as the common age and with 4 repeated measures, time points were 0, 1, 2 and 3 if the participant was 12 years old at the beginning of the study, and 2, 3, 4 and 5 if the participant was 14 years old at the beginning of the study. With this scheme, the average intercept interpreted as the mean values at age 12. Secondly, they scaled time with respect to an individual origin, meaning centering was unique to the individual, each participants age centered to his or her age at the beginning of the study. Doing so, each participant had the time points of 0, 1, 2 and 3 even though they varied at the starting age. In order to capture the heterogeneity effect, the authors used participant's age at the beginning of the study as an intercept predictor. The authors showed that these two schemes produced essentially same results. However, for ease of interpretation, they recommended using centering at a common age. Even though their work provided different designs and different schemes for time coding, all analyses conducted with the assumption that occasions of measurement were equally spaced.

Biesanz, Deeb-Sossa, Papadakis, Bollen & Curran (2004) conducted research to understand the role of time coding in estimating and interpreting growth curve models. For expository purposes, they investigated an unconditional linear growth model with three repeated measures where participants measured at ages of 5, 7 and 9. They changed the origin of time and reported results for 3 different time coding schemes. Then, they examined an unconditional quadratic growth model with five repeated measures and they modeled changes in children's weight. Data was collected for each child at ages of 5, 7,9,11 and 13. They used four different time coding strategies with an assumption of individuals assessed at the same time points. The first time coding scheme was to set the origin of time at age of 5, and then they assigned 0,2,4,6 and 8 for the ages 5, 7,9,11 and 13 respectively. For the second time coding scheme, they coded time by setting the last period of assessment time to zero. Therefore, they assigned the codes -8,-6,-4,-2 and 0 to the assessment periods. For the third time coding scheme, they centered time at the midpoint of assessments and assigned -4,-2, 0, 2 and 4to the measurement periods. These three different placement of time's origin produce information about individual differences in weight and individual differences in the rate of change in weight growth at specific ages. Changing the origin of time caused differences in estimates of some parameters (e.g. intercept, slope). The authors provided an analytical explanation to show how these differences in parameter estimates occurred and how they can be calculated based on the first scheme they used. This analytical explanation was based on the relationship between three different schemes. The authors basically created a transformation matrix and the size of this matrix was determined by the shape of the estimated growth function. For example, by creating the appropriate transformation matrix, the authors were able to change the time points from 0, 2, 4 to -4,-2, 0 within the same unconditional linear growth model. They used this transformation matrix to calculate differences in parameter estimates. Furthermore, the authors used orthogonal polynomial codes as the fourth time coding scheme for an unconditional quadratic growth model. On one hand the coefficients obtained with this scheme caused interpretation difficulties. On the other hand, it was stated that with the orthogonal polynomial codes, estimation of the overall growth curve model might be more readily achieved. The authors also examined a conditional quadratic growth model with different time coding schemes. They included a predictor for intercept, slope and quadratic term. Their result showed that models with different schemes fit the data equivalently, but parameter estimates were different. It was shown that differences in these estimates can also be calculated with using adequate transformation matrices.

In Mehta & West (2000) and Biesanz, Deeb-Sossa, Papadakis, Bollen & Curran. (2004), the coding of time points was based on the age of the participants. This approach is reasonable when the response is most sensitive to changes in age. Another time coding strategy is to set time points based on measurement waves. In the education discipline, researchers generally tend to use measurement waves as the basis for coding time. For example they set time points based on semesters (e.g. fall and spring). This preference is a more logical approach if we assume growth in a behavior for school kids is related to time they spend in classes rather than their ages. In the work of Blozis and Cho (2008) both measurement waves and participants' age were used to create time points. Their work built on two representative longitudinal studies. For the first study, data were gathered from National Longitudinal Survey of Youth (NLSY). Children aged between 6 and 14 were the subjects and their antisocial behavior scores were the dependent variables which were collected in four different assessment waves. They created seven different time coding schemes to estimate a LGM for NLSY data in which participants' ages varied at the beginning of the study and participants were assessed on different calendar dates within a measurement occasion. Thus, complete inter-individual time heterogeneity was one of the characteristics of the data. Even though they tried different time coding strategies (e.g., person-mean centered, group-mean centered) they reached same parameter estimates and overall model fit. The second longitudinal study came from a slightly different population. The incapacity of participants was assessed nearly annually over a five-year period. Years after first diagnosis and assessment waves were used to create time points. Using the relative standard deviation approach, they stated that the mean of the measures of time with regard to time since diagnosis in the second data was 12 times more heterogonous than the variability of mean ages in the NLSY data. The relative standard deviation measure was calculated based on the absolute value of the ratio of the standard deviation to the mean, which was then multiplied by 100. For example, the mean age of participants in the NLSY data was 10.12 with a standard deviation of .563; these values yielded a relative standard deviation of 5.56. Large heterogeneity caused remarkable differences in some of the parameter estimates with person-mean centered and grand-mean centered time coding strategies when using years after the first diagnosis as time points. The authors also used measurement waves to create time points. First, they assigned 0,1,2,3 and 4 for the 5-year annually assessments. Second they assigned -2,-

1, 0,1and 2 to set the origin as the midpoint of assessments. With these two coding strategies each participant had the same time points in each scheme. These two strategies produced exact same model fit. In order to capture the effect of heterogeneity, authors created a third coding scheme and they implemented the mean years since diagnosis as a predictor of the intercept and time effect into the second scheme. Doing so, they achieved a more adequate model fit.

In this study I attempted to investigate if taking into account heterogeneity of time of assessment would improve the fit of latent growth models. I am also interested in whether the parameter estimates would change due to time coding differences. I first decided to perform an analysis of real data using the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K) because this data set provides exact day of assessment for each individual. To provide a realistic context for my study, I chose to build my research on a published model. After a careful search of the applied literature, I decided to use Morgan, Farkas and Wu's (2009) work. These authors examined whether and to what extent the timing and persistence of mathematics difficulties (MD) in kindergarten predicted children's first through fifth grade math growth trajectories. They employed HLM with fixed time points to analyze the math growth in a subsample taken from ECLS-K.

In order to demonstrate the effects of ignoring the variability in measurement occasions, I tried to replicate Morgan, Farkas and Wu's (2009) analysis. Mplus software (Muthen & Muthen, 2008) was used because of its flexibility. It has been shown that, in most cases, HLM produces estimates that are equivalent to the estimates produced from a LGM (Raudenbush, 2001; Rovine & Molenaar, 2000; Mehta & West 2000;

Hertzog and Nesselroade, 2003). In order to conduct the replication study; I created a latent curve model and investigated four different time coding strategies. In two of these strategies the coding of time was based on measurement waves and resulted in fixed time points. Participants' ages was used as a predictor for the intercept and the linear slope. In two other strategies, individual-specific time points were used. Unlike the in previous research, I investigated the differences of time coding strategies using number of days between assessments; in other words, heterogeneity in assessment dates within the waves while including the age as a predictor for the intercept and the slope. It was hypothesized that models with individual specific time points would provide a more adequate fit. After the empirical illustration, a simulation study was conducted to explore differences between fixed time points approach and individual specific time points approach for an unconditional linear latent growth model. It was hypothesized that, using fixed time points produces biased estimates if assessment dates vary within a measurement occasion.

CHAPTER 2
EMPIRICAL ILLUSTRATION

Methodology

The aim of this study is to examine the effect of different time coding strategies for both unconditional and conditional nonlinear latent growth model including a quadratic (time-squared) term .The level 1 model equation is:

$$y_{it} = \alpha_i + \lambda_t \beta_{1i} + \lambda_t^2 \beta_{2i} + \varepsilon_{it} \tag{2-1}$$

where y_{it} is the value of the trajectory variable *y* for the *i* th case at time *t,* α_i defines the intercept for case *i*, β_{1i} the linear slope, and β_{2i} is the curvature. The λ_t , is a constant where a common coding convention is to have $\lambda_1 = 0$ and $\lambda_2 = 1$,which scales the intercept to be interpreted as the initial status. The variable λ_t^2 is simply the squared value of time at assessment *t*, and ε_{it} is the error term. These components combine additively to reproduce the value of *y* for individual *i* at time *t*. The LGM shown above, treats components of α_i , β_{1i} , β_{2i} as random variables, and level 2 equations can be written as:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha i} \tag{2-2}$$
$$\beta_{1i} = \mu_{\beta_1} + \zeta_{\beta_1 i} \tag{2-3}$$
$$\beta_{2i} = \mu_{\beta_2} + \zeta_{\beta_2 i} \tag{2-4}$$

where μ_α is the mean intercept across all cases, μ_{β_1} is the mean slope across all cases and μ_{β_2} is the mean curvature across all cases. Equation 2-2 represents the individual intercept α_i as a function of the mean of the intercepts for all cases (μ_α) and an error term $\zeta_{\alpha i}$. Equation 2-3 represents the individual linear growth component β_{1i} as a function of the mean of slopes for all cases (μ_{β_1}) and an error term. Equation 2-4 represents the individual quadratic growth component β_{2i} as a function of the mean of curvatures for all cases ($\mu_{\beta 2}$) and an error term. The $\zeta_{\alpha i}, \zeta_{\beta_1 i}$ and $\zeta_{\beta_2 i}$ are error terms with means of zero and variances of $\psi_{\alpha\alpha}, \psi_{\beta_1\beta_1}$ *and* $\psi_{\beta_2\beta_2}$. These error terms are assumed to be uncorrelated with ε_{it}. In the unconditional model the variance of α is equal to the variance of ζ_{α} , the variance of β_1 is equal to $\psi_{\beta_1\beta_1}$ and the variance of β_2 is equal to $\psi_{\beta_2\beta_2}$.

Within the structural equation framework, if we consider a *T* x 1 vector *y* that includes the set of *T* repeated measures of *y* for individual *i*, the level 1 Equation can be written in matrix terms;

$$y = \Lambda\eta + \varepsilon \tag{2-5}$$

where *y* is a *T* x 1 vector of repeated measures, Λ is a *T* x *m* matrix of factor loadings, η is an *m* x 1 vector of *m* latent factors, and ε is a *T* x 1 vector of residuals.

Since Morgan's model investigated students' math aptitude in 1st, 3rd and 5th grades, the elements of Equation 2-5 are;

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & T_1 & T_1^2 \\ 1 & T_2 & T_2^2 \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} \tag{2-6}$$

Each observation of *y* for individual *i* at time *t* is a weighted combination of a random intercept, a random linear trajectory component, a random quadratic trajectory component, and an individual and time-specific error. The level 2 equations can be written as

$$\eta = \mu_\eta + \zeta \tag{2-7}$$

where μ_η is m x 1 vector of factor means and ζ is a m x 1 vector of errors. For the quadratic model, the matrix form of Equation (2-7) is

$$\begin{pmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \mu_\alpha \\ \mu_{\beta_1} \\ \mu_{\beta_2} \end{pmatrix} + \begin{pmatrix} \zeta_{\alpha i} \\ \zeta_{\beta_1 i} \\ \zeta_{\beta_2 i} \end{pmatrix} \tag{2-8}$$

If we substitute Equation (2-7) into Equation (2-5), we could get reduced-form expression of *y*,

$$y = \Lambda\mu_\eta + (\Lambda\zeta + \varepsilon) \tag{2-9}$$

The model implied variance of the reduced form is

$$VAR(y) = \Lambda\psi\Lambda' + \Theta_\varepsilon \tag{2-10}$$

where Θ_ε represents the covariance structure of the residuals for the *T* repeated measures of *y*, ψ is the covariance matrix of the equation errors, ζ , among the latent trajectory factors.

Elements of ψ for the quadratic model are;

$$\psi = \begin{pmatrix} \psi_{\alpha\alpha} & \psi_{\beta_1\alpha} & \psi_{\beta_2\alpha} \\ \psi_{\beta_1\alpha} & \psi_{\beta_1\beta_1} & \psi_{\beta_2\beta_1} \\ \psi_{\beta_2\alpha} & \psi_{\beta_1\beta_2} & \psi_{\beta_2\beta_2} \end{pmatrix} \tag{2-11}$$

For an unconditional model the variance of η is equal to the variance of ζ . In the case of three waves of data, the quadratic model has 12 parameters to estimate from 9 means, variances and covariances of the observed variables. Without constraints three waves of data are not sufficient to identify the model. In order to identify the model, I assumed that the error variances of the observed variables are equal. With this assumption there were still 10 unknown parameters to estimate. Then I fixed the variance of the quadratic term at zero. The zero variance of the quadratic term reduced unknown parameter numbers to 7, since covariance of the slope-quadratic and intercept-quadratic was automatically set to zero.

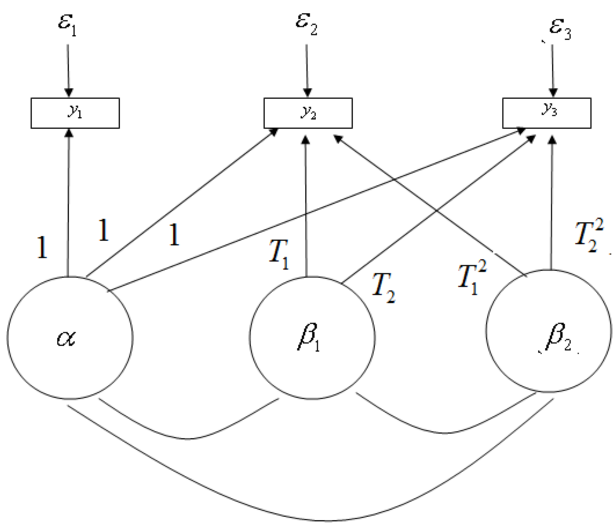

Figure 2-1: Path model for an unconditional quadratic latent growth model.

The unconditional model differs from the conditional model in the level 2 equations for the random intercept, linear slope and curvature. Since Morgan's model did not include predictors for the quadratic term, level 2 equations are:

$$\alpha_i = \mu_\alpha + \gamma_{\alpha 1} x_{1i} + \gamma_{\alpha 2} x_{2i} + \zeta_{\alpha i} \tag{2-12}$$

$$\beta_{1i} = \mu_{\beta_1} + \gamma_{\beta 1} x_{1i} + \gamma_{\beta 2} x_{2i} + \zeta_{\beta_1 i} \tag{2-13}$$

$$\beta_{2i} = \mu_{\beta_2} + \zeta_{\beta_2 i} \tag{2-14}$$

here μ_α and μ_{β_1} are the intercepts for the equations that predict the intercepts and linear slopes across all cases. Note that μ_α and μ_{β_1} are the mean intercepts and mean linear slopes when x_{1i} and x_{2i} are zero. The x_{1i} and x_{2i} are two predictors of the intercepts and linear slopes, $\gamma_{\alpha 1}$ and $\gamma_{\alpha 2}$ are the coefficients for x_{1i} and x_{2i} in the random intercept equation. The $\gamma_{\beta 1}$ and $\gamma_{\beta 2}$ are coefficients in the linear slope equation. These coefficients have the same interpretation as they have in a regression model. They provide the expected difference in the outcome for a 1-unit difference in the explanatory variable net of the other explanatory variable. Morgan's model included eleven time invariant predictors.

Within the structural equation framework, a quadratic conditional model with three time points can be written in matrix forms;

$$y_i = \Lambda \eta_i + \varepsilon_i \tag{2-15}$$

$$y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix}, \Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 1 & T_1 & T_1^2 \\ 1 & T_2 & T_2^2 \end{bmatrix}, \eta_i = \begin{bmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix}$$

$$\eta_i = \mu_\eta + \Gamma x_i + \zeta_i \tag{2-16}$$

$$\eta_i = \begin{bmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \mu_\eta = \begin{bmatrix} \mu_\alpha \\ \mu_{\beta_1} \\ \mu_{\beta_2} \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_{\alpha 1} & \gamma_{\alpha 2} & \cdots & \gamma_{\alpha K} \\ \gamma_{\beta_1 1} & \gamma_{\beta_1 2} & \cdots & \gamma_{\beta_1 K} \\ \gamma_{\beta_2 1} & \gamma_{\beta_2 2} & \cdots & \gamma_{\beta_2 K} \end{bmatrix}, x_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Ki} \end{bmatrix}, \zeta_i = \begin{bmatrix} \zeta_{\alpha i} \\ \zeta_{\beta_1 i} \\ \zeta_{\beta_2 i} \end{bmatrix}$$

Sample

The ECLS-K is a nationwide longitudinal study aimed to gather extensive data about children's cognitive and behavioral skill development and their learning experiences starting from kindergarten and ending with fifth grade. Students in this data attend both public and private schools. The ECLS- K study was conducted by the National Center for Educational Statistics (NCES). The ECLS-K study used a nationally representative sample. Consistent with the purpose of this study, mathematic assessment scores and exact assessment dates were used. These mathematics scores were obtained at five different time periods: fall of Kindergarten, spring of Kindergarten, spring of first grade, spring of third grade and spring of fifth grade. The original data included 17,565 cases coming from approximately 3500 classrooms in 1280 schools.

In order to replicate Morgan, Farkas and Wu's (2009) study, children of (a) White, non-Hispanic or (b) Black/African American, non Hispanic families were chosen. This initially defined a sample consisting of 12,385 children.

Measures

Mathematics Scores

The mathematics assessment for the ECLS-K study was designed to measure conceptual and procedural knowledge and also problem solving within particular content strands. For the five different math assessments, the content of the questions can be categorized as number sense, properties, operations, measurement, geometry, spatial sense, data analysis, statistics, probability, algebra and functions. The first three categories contained the largest number of items for all grades.

Mathematics scores were scaled using IRT procedures. The advantage of the IRT based scores is that it provides comparable scores in many cases. IRT scoring makes possible longitudinal measurement of gain in achievement over time IRT methods estimate the difficulty, discriminating ability and guessing values for the items. Using these values, individual item responses are used to estimate the respondents' latent abilities. The National Center of Education Statistics (NCES) chooses IRT based score as the most appropriate metric for growth modeling. The reliabilities of IRT scores ranged from .89 to .94(NCES). The correlations between IRT scaled scores and scores from *Woodcock-McGrew-Werder Mini Battery of Achievement* (Woodcock, McGrew, &

Werder, 1994) were compared to assess concurrent validity. High correlations were found (.i.e., fifth grade = .80 and third grade = .84) (NCES).

Reading Scores

Reading difficulties of children were one of the predictors of the mathematical growth. ECLS-K assessed children's reading skills at several time points. The content of the questions in this test can be categorized as print familiarity, letter recognition, decoding, sight word recognition, receptive vocabulary, and comprehension. The field test of the instrument showed no differential item functioning and adequate item level statistics. The reliability coefficient of fall kindergarten reading IRT scores was .91 (NCES, 2006). Reading test scores of the fall kindergarten were used to identify children with reading difficulties. Consistent with the literature 10% cut off was used. Scores in the lowest 10% were coded as 1 indicating reading difficulty, the rest of the scores (90%) coded as 0 indicating no reading difficulty.

Learning Related Behaviors

An instrument was created for ECLS-K based on the *Social Skills Rating System* (Gresham & Elliott, 1990) to measure a student's attentiveness, task persistence, eagerness to learn, learning independence, adaptability to changes in routine, and organization. Teachers rated children's learning related behaviors during the fall kindergarten. The scores for the fall of kindergarten yielded a split half reliability coefficient of .89(NCES). Scores in the lowest 10% were coded as 1 indicating behavior problems, the rest of the scores (90%) were coded as 0 indicating no behavior problems. This dummy coded variable was used as one of the predictors of children's initial knowledge of mathematics and mathematical growth.

Socio Economic Status (SES)

Individual SES values were used as one of the predictors. NCES assessed a household's SES by using parents' education level, occupation and household income. The continuous scale of SES (WKSESL) was chosen as the variable. The scores were gathered during the spring of kindergarten.

Age, Gender, Race, Kindergarten Retention, Disability Status

The children's age in months at the beginning of the fall kindergarten (September 1998) were used to create one of our independent variables. The dichotomous variables of race, gender and retention status were also independent variables where 1 indicated, White, female and retention, respectively, and 0 indicated Black/African American, male and no-kindergarten retention respectively. Individualized Education Plan (IEP) was used for the students with disabilities. Records for spring of kindergarten were used to create a dichotomous variable where 1 indicates that child had an IEP.

The Final Data

In the work of Morgan, Farkas and Wu's (2009) they excluded those children who had missing data on any child-level predictor (e.g. Race, gender, retention) or the Mathematics test at the kindergarten time points. Their final analytical sample consisted of 7,892 children, and Morgan, Farkas and Wu claim that this sample was representative of the full sample. I followed the same exclusion strategy and came up with a sample of 7,935. The small difference (43 cases) might be due to rounding in cut off procedures. However, when the Mplus program estimated the LGM with random time points, the robust maximum likelihood (MLR) estimation procedure deleted cases with missing values on time scores but non missing values on the corresponding dependent variables. Overall my final analytical sample included 7,809 cases. Tables 2-

1 and 2-2 show descriptive statistics for the full sample, Morgan, Farkas and Wu’s (2009) sample and my replication sample.

In the work of Morgan, Farkas and Wu (2009) one of the main purposes was to estimate the growth curves of children with learning difficulties in mathematics (MD) during kindergarten. They created four different dummy codes to categorize children with learning difficulties. The same strategy was followed in this study, and consistent with Morgan's work, a 10% cut off was used. Fall and spring kindergarten math test scores were used and students who scored in the lowest 10% were labeled as 'Difficulties' or 'D'. Non-MD group represented students who were above the 10% cut off in both fall and spring kindergarten tests. D10 was set to 1 for a student who had MD in the fall semester, but not in the spring semester. D01 was set to 1 if a child had MD in the spring semester but not in the fall. D11 was set to 1 if a child had MD in both semesters.

Table 2-1. Demographic Characteristic for the ECLS-K Full and Analytical Samples

Characteristic	Full Sample (N=12,385)	Replication Study Sample (N=7809)	Morgan's Sample (N=7892)
Gender			
Male	51.2%	50.9%	50.9%
Female	48.8%	49.1%	49.1%
Race			
White	79.9%	82.9%	82.7%
Black or African American	20.1%	17.1%	17.3%
IEP			
Yes	7.0%	6.3%	6.4%
No	93.0%	93.7%	93.6%

IEP= Individualized Education Plan.

Table 2-2. SES and Math IRT scores for the ECLS-K Full and Analytical Samples

Characteristic	Full Sample (N=12,385)	Replication Study Sample (N=7809)	Morgan's Sample (N=7892)
SES*	0.12(.78)	0.14(.77)	0.14(.77)
Fall Kindergarten Mathematics IRT scores	23.88(8.98)	24.33(9.02)	24.31(9.00)

Note: Standard deviations are in parenthesis; IRT = item response theory.

*Using the WKSESL variable

Statistical Models

Hierarchical Linear Modeling

In the work of Morgan, Farkas and Wu (2009) HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004) was used to analyze the data. They used a slopes and intercepts as outcomes model. In my study I tried to replicate their most complex model. Following are the level-1 and level-2 equations for their most complex model:

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_{1i}t + \pi_{2i}t^2 + e_{ti} \quad (2\text{-}17)$$

where i=1,2,3,......n is the index for subjects, t=0,2,4 are the time points, π_{0i} is the initial status of the student at time zero, π_{1i} is the linear slope, π_{2i} is the quadratic term. The growth rate of person i at any specific time point is the first derivate of the growth model at that point (Raudenbush & Bryk, 2002). If π_{2i} is positive, it can be said that the student is growing at an accelerating rate, if it is negative, growing at a decelerating rate. The e_{ti} is the measurement error and it is assumed to have a mean of 0, a constant variance and be distributed normally.

Level 2:

$$\begin{aligned}\pi_{0i} &= \beta_{00} + \beta_{01}(Age) + \beta_{02}(SES) + \beta_{03}(Race) + \beta_{04}(Gender) + \\ &\beta_{05}(repeating\ kindergarten) + \beta_{06}(reading) + \beta_{07}(approaches) + \\ &\beta_{08}(IEP) + \beta_{09}(D00) + \beta_{010}(D01) + \beta_{011}(D11) + r_{0i}\end{aligned} \tag{2-18}$$

$$\begin{aligned}\pi_{1i} &= \beta_{10} + \beta_{11}(Age) + \beta_{12}(SES) + \beta_{13}(Race) + \beta_{14}(Gender) + \\ &\beta_{15}(repeating\ kindergarten) + \beta_{16}(reading) + \beta_{17}(approaches) + \\ &\beta_{18}(IEP) + \beta_{19}(D00) + \beta_{110}(D01) + \beta_{111}(D11) + r_{1i}\end{aligned} \tag{2-19}$$

$$\pi_{2i} = \beta_{20} \tag{2-20}$$

where β is the coefficient for a particular predictor on initial math status and math growth r_{0i} and r_{1i} are random error. Due to lack of enough degrees of freedom the variance of random error for the quadratic term was not estimated.

Latent Growth Curve Model

Based on Morgan's HLM model, we can write the Equation 2-15 for the unconditional model as;

$$y_i = \Lambda \eta_i + \varepsilon_i \tag{2-21}$$

$$y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix}, \Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}, \eta_i = \begin{bmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix} \tag{2-22}$$

The level 2 matrices for a conditional LGM model based on Morgan`s HLM can be written as;

$$\eta_i = \begin{bmatrix} \alpha_i \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \mu_\eta = \begin{bmatrix} \mu_\alpha \\ \mu_{\beta_1} \\ \mu_{\beta_2} \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_{\alpha 1} & \gamma_{\alpha 2} & \cdots & \gamma_{\alpha 11} \\ \gamma_{\beta_1 1} & \gamma_{\beta_1 2} & \cdots & \gamma_{\beta_1 11} \\ \gamma_{\beta_2 1} & \gamma_{\beta_2 2} & \cdots & \gamma_{\beta_2 11} \end{bmatrix}, \zeta_i = \begin{bmatrix} \zeta_{\alpha i} \\ \zeta_{\beta_1 i} \\ \zeta_{\beta_2 i} \end{bmatrix}$$

$$x_i = \begin{bmatrix} Age \\ SES \\ Race \\ Gender \\ rep.\,kinder. \\ reading \\ approaches \\ IEP \\ D00 \\ D01 \\ D11 \end{bmatrix} \tag{2-23}$$

Time points

The ECLS-K provides exact measurement dates for each subject. Given this information, it is possible to calculate the number of days between assessments for each individual. The first mathematics assessment was conducted at the end of the fall kindergarten semester. The date of this assessment was assigned as the starting point and given the value of zero. For example student with the id 0001002C took the first assessment on 30 November 1998 (fall kindergarten) and second assessment on 26 May 1999(spring kindergarten). Using the compute elapse = (date2-date1)/86400 option on SPSS, it has been found that there are 177 days between these two assessment points. The same process was employed for all students and all other time points. Table 2-3 shows descriptive statistics for number of days after the first assessment.

Morgan, Farkas and Wu (2009) used fall and spring of kindergarten math IRT grades to identify children with MDs. In order to analyze mathematics skills growth, they used spring of first grade, spring of third grade and spring of fifth grade math IRT scores. Using fixed time points, they assigned time points of 0, 2 and 4 for the grades.

Table 2-3. Descriptive Statistics for Number of Days between the Fall Kindergarten Assessment and the Other Assessments

Semester	Number of Students	Minimum	Maximum	Mean	SD
Spring KG	7809	120	261	185.6	20.4
Spring1st	7809	477	644	550.6	23.0
Spring3rd	6548	1199	1366	1275.3	25.8
Spring5th	5174	1904	2091	1983.3	29.0

Note: SD= Standard Deviation

Four different time coding strategies

In order to generate parallel results with Morgan's study, the first model was estimated using the fixed time points of 0, 2 and 4 to represent spring of first grade, spring of third grade and spring of fifth grade, respectively. This procedure assigns 1 point for each education year.

The second model was also estimated with fixed time points, but the average of number of days between assessments was used. Because growth analysis started from spring of first grade, I treated the spring first grade assessment dates as the beginning point. Thus, I was able to find how many days passed before the third grade and fifth grade assessments. Calculations showed that on average there were 725 days between first grade and third grade assessments; and, there were 1344 days between first grade and fifth grade assessments. For ease of interpretation each average divided by constant number of 100. Time points were 0, 7.25, and 13.44 represents spring of first grade, spring of third grade and spring of fifth grade, respectively.

The third model was estimated without fixed time points. Adding the command of 'type=random' to analysis line in the Mplus program, inter-individually varying time points of assessments were taken into account. The averages of time points were 0, 7.25 and 13.44. Table 2-4 shows descriptive statistics for time points used.

Table 2-4. Number of days between assessments divided by 100

Term	N	Min	Max	Mean	SD_
Spring 1st grade	7809	0	0	0	0
Spring 3rd grade	6548	6.42	8.08	7.25	0.23
Spring 5th grade	5174	13.48	15.13	14.33	0.29

Note: N= Sample size, SD= Standard deviation

The fourth and the last model also estimated with heterogeneous time points. However, the number of days was divided by 358 instead of 100 in order to set the mean of time points to 0, 2 and 4. This time coding scheme provides estimates directly comparable to Model 1 results. Table 2-5 shows descriptive statistics for the time points used in the fourth model.

Using unconditional and conditional LGM models, I analyzed first, third and fifth grade math scores. Average math IRT scores were increased toward upper grades, starting with 59.98 at the first grade, 95.28 for the third grade and 116.79 for the fifth grade. Table 2-6 shows the descriptive statistics for IRT scores. Figure 2-2 shows the trajectory of IRT scores and indicates that increase in scores is not linear. In order to get parallel results to Morgan's analyses, I also mean centered continuous variables in level 2 at conditional model (i.e. SES and age).

LGM results can be obtained with different estimation procedures. The maximum likelihood (ML) procedure is the most widely used among these. All analyses were performed with Mplus (Muthen & Muthen 2008). The ML estimation method could not be employed for all four models. Instead robust maximum likelihood (MLR) was used for all models. It is known that parameter estimates are the same in MLR and ML, but standard errors might be different.

Table 2-5. Number of days between assessments divided by 358

Term	N	Min	Max	Mean	SD
Spring 1st grade	7809	0	0	0	0
Spring 3rd grade	6548	1.79	2.26	2.02	0.63
Spring 5th grade	5174	3.77	4.23	4.00	0.81

Note: N= Sample size, SD= Standard deviation (Please see Appendix A for the SPSS syntax that executes this scaling)

Table 2-6. Descriptive Statistics for Mathematics IRT Scores First through Fifth Grade

Assessment	N	Min	Max	Mean	SD
Spring 1st grade	7809	11.17	120.50	59.98	17.03
Spring 3rd grade	6548	33.60	146.59	95.28	20.96
Spring 5th grade	5174	47.08	150.94	116.79	20.26

Note: N: Sample size, SD: Standard deviation

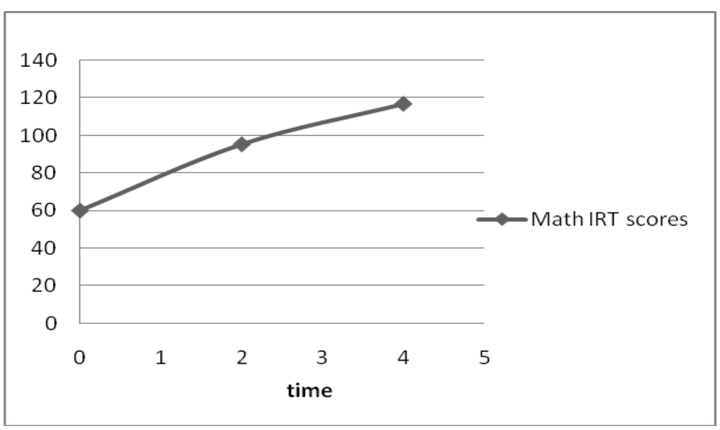

Figure 2-2 Mathematics Growth Trajectory

Results

In this chapter analyses of results of applying four different LGM models and their comparisons to Morgan’s work are presented. For the first step, unconditional LGM results and Morgan’s random regression coefficients (RRC) model results are reported. These two models are statistically equivalent in terms of having the same level1 and level2 equations. Moreover, using an analytical sample of 7,809 cases, a RRC model was estimated and the exact same parameter estimates and standard errors were obtained with a LGM estimated using Mplus.

Intercept values for all five models were roughly same. In terms of slope and quadratic term means, Model 1 and Model 4 produced essentially identical estimations with Morgan's model. However, Model 2 and Model 3 produced smaller estimated slopes. These differences occurred due to scale differences in time coding. The effect of scaling on parameter estimates is explained mathematically below.

Table 2-7. Parameter estimates of RRC model and the unconditional LGM model with different time coding combinations.

Means	RRC	Model 1	Model 2	Model 3	Model 4
Intercept	59.63*	59.99*	59.99*	60.00*	60.01*
Linear Slope	20.82*	20.82*	5.72*	5.70*	20.40*
Curvature	-1.77*	-1.71*	-0.127*	-0.125*	-1.60*

$$\Lambda_{mdel1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}, \Lambda_{mdel2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 7.25 & 52.56 \\ 1 & 14.33 & 205.35 \end{bmatrix}, \Lambda^{\#}{}_{mdel3} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 7.25 & 52.56 \\ 1 & 14.33 & 205.35 \end{bmatrix}, \Lambda^{\#}{}_{mdel4} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}$$

= average of individually varied time points
* p<.05

As shown earlier in Equation 2-1, the model for the score for person *i* at time *t* is

$$y_{it} = \alpha_i + \lambda_t \beta_{1i} + \lambda_t^2 \beta_{2i} + \varepsilon_{it}$$

where λ_t consisted of T values as time indicators. If we set $T = T^{'}$ as an alternative scale for time, then;

$$y_{it} = \alpha_i^{'} + \lambda_t^{'} \beta_{1i}^{'} + \lambda_t^{2'} \beta_{2i}^{'} + \varepsilon_{it} \qquad (2\text{-}24)$$

Because $T = \dfrac{T^{'}}{K}$

$$y_{it} = \alpha_i^{'} + \frac{\lambda_t^{'}}{K} \beta_{1i}^{'} + \frac{\lambda_t^{2'}}{K} \beta_{2i}^{'} + \varepsilon_{it} \qquad (2\text{-}25)$$

and $\alpha_i = \alpha_i^{'}$, $\beta_{1i}^{'} = \beta_{1i} / K$ and $\beta_{2i}^{'} = \beta_{2i} / K^2$. In LGM the means and the variances of the α_i *and* $\alpha_i^{'}$, are estimated. Let the means be $\mu_j^{'}$ *and* μ_j $(j = 0,1,2)$. From the expressions $\alpha_i = \alpha_i^{'}$, $\beta_{1i}^{'} = \beta_{1i} / K$ and , $\beta_{2i}^{'} = \beta_{2i} / K^2$, $\mu_0^{'} = \mu_0$, $\mu_1^{'} = \mu_1 / K$ and $\mu_2^{'} = \mu_2 / K^2$ when the index for occasions 2 and 3 are fixed at 7.25 and 14.33, these values approximately 3.6 times as large as 2 and 4 respectively. Thus from the preceding development we would expect $\mu_1^{'} \cong \mu_1 / 3.6$ and $\mu_2^{'} \cong \mu_2 / 3.6^2$ where the $\mu_j^{'}$ are the coefficients for the scale 0, 7.25 and 14.33. The results in table 2-7 agree with the development. For example the mean slope for Model 2, which was based on the scale points 0, 7.25 and 14.33, is 5.72 whereas the mean slope for Model 1 is 20.82. The ratio is 3.6. Let the variances be $\theta_j^{'}$ and θ_j from the expressions $\alpha_i = \alpha_i^{'}$, $\beta_{1i}^{'} = \beta_{1i} / K$ and, $\beta_{2i}^{'} = \beta_{2i} / K^2$ $\theta^{'} = \theta_0$, $\theta_1^{'} = \theta_1^{'} / K^2$ *and* $\theta_2^{'} = \theta_2^{'} / K^4$. The results in Table 2-8 agree with this development. For example the variance of the slope for Model 2 is 0.306 and the variance for the slope for model 1 is 4.091 the ratio is 13.37 which is approximately 3.6^2.

For each method of coding time, the model implied achievement mean was computed for each occasion. In the 3rd and 4th methods individually varying times of measurement were used. To calculate model implied means under these methods means of time scores used (i.e 7.25 and 14.33, for the 3d and 4th method respectively). These results are reported in Table 2-9 and indicates that unconditional LGM model with four different time coding strategies produced appreciably same model implied means at the three time points.

I further investigated differences in model fit information. MLR estimation procedure produces only Akaike (AIC) and Bayesian (BIC) fit indices. Loglikelihood, AIC and BIC values for all four models presented in Table 2-10.

The Akaike Information Criterion (AIC) is $AIC = 2k - 2\ln(L)$ where *k* represents the number of parameters in the model and *L* is the maximized value of the likelihood function for the estimated model. The value of AIC provided by model is not directly interpretable. Instead of focusing on the magnitude of AIC, the model with the smallest AIC can be selected as the most adequate model among compared models. In my study, Model 4 provided the smallest AIC value.

The Bayesian Information Criterion (BIC) is $BIC = -2\ln L + k\ln(N)$ BIC measures are used to compare the fit of models estimated from the same sample. Results showed that Model 4 has slightly smaller BIC values. However, all four models can be assumed essentially equivalent due to very small differences among the loglikelihood, AIC and BIC for the four coding methods.

Table 2-8. Variances of Parameter Estimates

Parameter	Model 1	Model 2	Model 3	Model 4
Intercept	233.020 (17.21)	232.360 (17.19)	232.810 (17.26)	233.07(17.24)
Linear Slope	4.091 (0.87)	0.306 (0.07)	0.306 (0.7)	4.15(0.86)
Residual	63.500 (3.80)	63.370 (3.80)	62.850 (3.77)	62.70(3.72)

Note: Standard errors are in parenthesis.

Table 2-9. Estimated Mean of IRT Based Math Scores ($y_{it} = \alpha_i + \lambda_t \beta_{1i} + \lambda_t^2 \beta_{2i} + \varepsilon_{it}$)

Term	Observed	Model1	Model2	Model3	Model4
Spr. 1^{st}	59.98	59.99	59.99	60.00	60.01
Spr. 3^{rd}	95.28	94.79	94.79	94.75	94.82
Spr. 5^{th}	116.79	115.91	115.88	116.01	116.01

Table 2-10. Model Fit Information for Unconditional LGM

Estimations	Model 1	Model 2	Model 3	Model 4
Loglikelihood	-93567.73	-93574.20	-93574.20	-93501.92
AIC	187149.45	187162.40	187051.71	187017.83
BIC	187198.19	187211.14	187100.45	187066.57
Adjusted BIC	187175.95	187188.90	187078.20	187044.33

For the second step, I examined conditional LGM models and Morgan`s intercepts and slopes as outcomes (ISAO) model. Table 2-11 shows the parameter estimates for the equation for the intercept term. Significance levels of predictors at $\alpha = .05$ were the same for all five models. With two exceptions, *repeat kindergarten and reading difficulty,* estimates obtained by using ISAO and LGM were essentially the same. Differences in these two parameter estimates might be due to slight difference between samples. Note that, because intercept associated with time point 0 with all four time coding strategies, parameter estimates for LGM were roughly identical.

Table 2-12 shows parameter estimates for the slope term. There are some differences between ISAO and its equivalent LGM Model 1. In order to see if these differences occurred due to different software programs or estimation procedures, the same sample used for Model 1 was implemented in HLM6 software. HLM6 produced exactly the same results with LGM Model 1. Again I suspect that differences between Morgan`s model and LGM might be due to different samples.

Table 2-11. Comparison of parameter estimates for the intercept between Morgan's ISAO model and the conditional LGM model with different time coding combinations.

Intercept	Morgan's	Model1	Model2	Model3	Model4
Mean	60.80*	59.72*	59.71*	59.71*	59.71*
D10	-11.74*	-9.50*	-9.48*	-9.46*	-9.47*
D01	-17.51*	-18.07*	-18.05*	-18.04*	-18.05*
D11	-19.69*	-18.14*	-17.98 *	-17.84*	-17.85*
Age in months	0.72*	0.41*	0.42*	0.42*	0.42*
SES	5.46*	5.06*	5.05*	5.08*	5.09*
White	6.24*	6.87*	6.86*	6.86*	6.86*
Female	-4.19*	-4.01*	-4.08*	-4.09 *	-4.09*
Repeat kindergarten	-3.81	4.63	4.66	4.65	4.64
Reading difficulty	-0.12	-1.68	-1.68	-1.73	-1.73
Approaches difficulty	-6.44*	-7.01*	-7.06*	-7.01 *	-7.01 *
IEP	-6.70*	-7.70*	-7.69*	-7.72*	-7.71*

*p<.05

The main purpose of the study is to examine effect of different time coding strategies in LGM estimations. Similar to the unconditional LGM results, pairs of Model 1/Model 4 and Model 2/Model3 produced roughly same estimations. Differences in these two pairs can also be explained based on Equation 2-23 and 2-24. Coefficients for Model 1 and 4 approximately 3.6 times larger than coefficients for Model 2 and 3.

Fit information for conditional LGMs is presented in table 2-13. Consistent with the unconditional fit information, Model 4 has slightly smaller values which indicate better fit. However, all four models can be assumed essentially same due to small differences.

Table 2-12. Comparison of parameter estimates for the linear slope between Morgan's ISAO model and the conditional LGM model with different time coding combinations.

Linear Slope	Morgan's	Model 1	Model 2	Model 3	Model 4
Mean	20.55*	21.21*	5.83 *	5.81*	20.80*
D10	-1.30*	-1.30*	-0.36*	-0.35*	-1.26*
D01	-1.29*	-0.62	-0.17	-0.19	-0.67
D11	-1.96*	-1.96*	-0.55*	-0.56*	-1.99*
Age in months	-0.20*	-0.17*	-0.05*	-0.05*	-0.20*
SES	0.53*	0.46*	0.13*	0.14*	0.48*
White	1.35*	0.77*	0.21*	0.21*	0.73*
Female	-0.75*	-1.34*	-0.37 *	-0.37*	-1.33*
Repeat kindergarten	-0.62	-2.15*	-0.60*	-0.60*	-2.13*
Reading difficulty	-0.50	-0.48	-0.13	-0.14	-0.48
Approaches difficulty	-0.55	-0.06	-0.02	-0.03	-0.08
IEP	-0.06	-0.84	-0.23	-0.23	-0.82
Mean of Curvature	-1.77*	-1.71*	-0.13*	-0.13*	-1.60*

*p<.05

Table 2-13 Model Fit Indices for Conditional LGM

Estimations	Model 1	Model 2	Model 3	Model 4
Loglikelihood	-112158.20	-112147.77	-112093.66	-112077.20
AIC	224392.39	224371.54	224263.33	224230.40
BIC	224656.99	224636.14	224527.92	224495.00
Adjusted BIC	224536.23	224515.38	224407.17	224374.25

Discussion

This comparison based study examined the impact of omitting the inter-individual time heterogeneity on measurement occasions in LGM. All models resulted in the same fit, which indicates that using heterogeneous time points (i.e., exact distance between assessment dates for each subject) instead of fixed time points could not provide a more adequate fit for the data. Differences in parameter estimates were explained by a mathematical expression, which indicated that these differences occurred due to time point scales. However it is still a question if using inter-individually varying time points would change the model fit and parameter estimates for datasets which have relatively more varied measurement dates. In order to see if increased heterogeneity in time points would change the estimates in an unconditional linear LGM, a simulation study was conducted.

CHAPTER 3
SIMULATION STUDY

In this chapter, the design of the simulation study and the followed procedures are described. The factors of the interest are range of assessment time points, distribution of assessment time points and sample size. A Monte Carlo simulation technique was used to investigate of the effects of the factors on the parameter estimates and model fit indices. With this simulation study, I aimed to understand under what conditions researchers should consider to take into account the exact number of days between assessments for each individual rather than using fixed time points to represent assessment waves when they analyze a LGM. A total of nine different conditions were created to represent assessment time points. However the mean assessment time did not differ and each condition has mean values of 0, 2 and 4. Hence, in a real life situation, a researcher could decide to ignore the variation in measurement occasion and fit a LGM using the mean assessment times.

Design

The design of the simulation had three between-subject factors. These factors were range of assessment time points, distribution of assessment time points and sample size. The range of the assessment time points had three levels, narrow, moderate and wide. Narrow ranged assessment time points were created based on ECLS-K calendar dates of each assessment. This condition included zeros for all cases as the first assessment wave time points, values between 1.79 and 2.26 for all cases as the second assessment wave time points and values between 3.77 and 4.23 for all cases as the third assessment wave time points. As explained in chapter 2, these values were obtained by dividing the number of days between assessments by the constant number 358.The moderate range time point condition also included zeros for all cases as the first assessment wave, but twice the range as in the narrow range condition for other assessment points; so, the second assessment period ranged between 1.58 and 2.52 and the third period ranged between 3.54 and 4.46. The wide range time points included zeros for initial status, values between 1 and 3 for the second assessment and values between 3.01 and 5 for the third assessment period. Increasing the range allowed me to increase the heterogeneity in time points assigned for each assessment wave. Standard deviations of the time scores for both second and third assessment waves were 0.13, 0.26 and 0.58 for narrow, moderate and wide range respectively. Table 3-1 shows the characteristics of the distributions of measurement occasions.

Table 3-1.Characteristics of assessment time points distributions

Range	Measurement	Min.	Max.	Mean	SD
Narrow	Occasion 2	1.79	2.26	2	0.13
	Occasion 3	3.77	4.23	4	0.13
Moderate	Occasion 2	1.58	2.42	2	0.26
	Occasion 3	3.54	4.46	4	0.26
Wide	Occasion 2	1	3	2	0.58
	Occasion 3	3.01	5	4	0.58

SD=Standard deviation

The second between subject factor was the distribution differences of the time points. Three levels were created for this factor; uniform, moderately skewed and extremely skewed. Range, mean and standard deviations of time points were the same across all distributions. A uniform distribution is the simplest continuous distribution in probability. It has constant probability density on an interval (a, b) and zero probability density elsewhere. The distribution is specified by a lower limit (a) and an upper limit (b). Its probability density function is:

$$P(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases} \tag{3-1}$$

In a real life situation, this distribution would be created for example, if an assessment team collects data for approximately the same number of children in each week during an assessment occasion. Moderately skewed and extremely skewed distributions would represent the attempt to collect as many measurements as possible at once, with only a few measurements being collected later. In order to create skewed distributions, the Fleishman (1978) power transformation was applied to the normal distribution. Using the polynomial Equation 3-2, a normally distributed X variable can be transformed into a skewed Y variable.

$$\begin{aligned} &[X \sim N(0,1)] \\ &Y = a + bX + cX^2 + dX^3 \end{aligned} \tag{3-2}$$

The mean and variance of the X variable is known priori, but in order to have the desired levels of skewness and kurtosis mean and variance for the Y variable, one needs to solve the equation with different set of constants. Each set of constants creates different distributions, thus different skewness and kurtosis values. This transformation is the standard in simulation studies to simulate skewed data, because it allows precise control of the skew and kurtosis values. In my study, a moderately skewed distribution had skewness value of 1.25 and kurtosis value of 2.75 and an extremely skewed distribution had skewness value of 1.75 and kurtosis value of 3.75 were used. In order to get these specific skewness and kurtosis values, the set of constants in Equation 3-2 were obtained from the table provided by Fleishman (1978). Histograms for the distributions are shown in Figure 3.

The third between-subject factor was sample size. This factor also included three levels, 200, 2000 and 8000. A sample size of 100 has been recommended as a minimum sample size for LGM by Hamilton, Gagne and Hancock (2003). I decided to use a sample size of 200 to represent small sample sizes. Larger sample sizes of 2000 and 8000 were included in my simulation study because there is greater availability of large longitudinal data sets in the past decades (e.g ECLS-K, NLSY). My replication study was also built on a large sample size (N=7809). Overall, the simulation study contained 3x3x3=27 conditions and one thousand datasets were generated for each condition. Population parameters for the simulated datasets were obtained from the LGM study published by Biesanz, Deeb-Sossa, Papadakis, Bollen&Curran (2004). The population parameters are shown in Table 3-2.

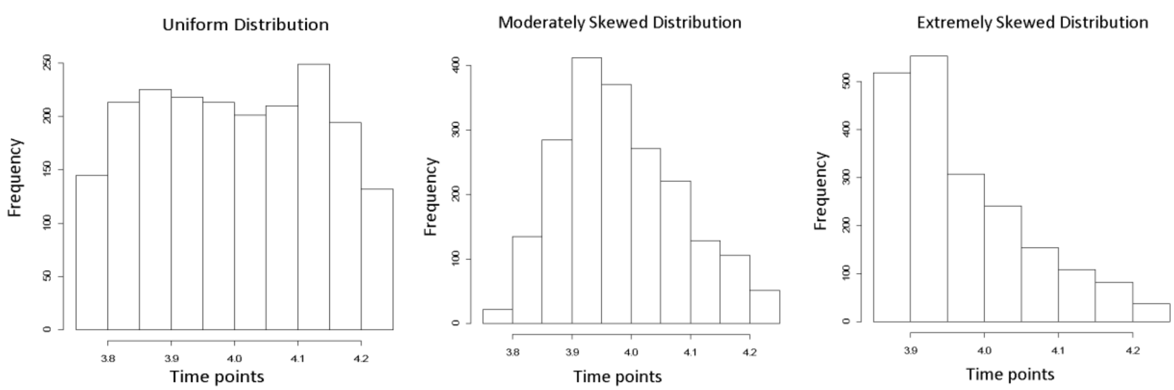

Figure 3-1. Distributions for the third assessment wave time points in narrow range.

Table 3-2. Population Parameters

Parameter Name	Parameter Value
Intercept	39.46
Slope	8.06
Intercept Variance	28.78
Slope Variance	8.20
Intercept-Slope Covariance	12.44
Residual Variance 1	8.32
Residual Variance 2	12.01
Residual Variance 3	57.17

Data Generation

The model used to generate the data was

$$y_i = \Lambda_i \eta_i + \varepsilon_i \tag{3-3}$$

with

$$i = 1, 2, \ldots, n$$

Or in matrix terms

$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & T_1 \\ 1 & T_2 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_{1i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix} \tag{3-4}$$

where *y* is a 3 x 1 vector containing scores of three assessment waves for individual *i*, Λ_i is the 2 x 3 matrix of factor loadings with T's representing individual time points, η is the 2 x 1 vector containing an intercept and linear slope and ε_i is a 3 x 1 vector of disturbances.

All data simulations were conducted using the R statistical software (R Development Core Team, 2009). A thousand datasets were simulated for each combination of conditions, resulting in 27000 datasets. The following steps were used to simulate the data:

- Simulate A, a *n* x 3 matrix of measurement times. The A matrix contained zeros in the first column. The second and third columns contained different time point values for the various conditions but the mean values of 2 and 4 respectively were constant across conditions.
- Create B, a 2 x 2 covariance matrix for α and β based on the population parameters.
- Create C, a *n* x 2 matrix based on B containing intercepts in the first column and slopes in the second column which were sampled from a normal distribution with the means, variances and covariances as shown in Table 3-2 .
- Create D, a *n* x 3 matrix where each column contains the intercepts (i.e., the first column of C)
- Calculate E, a *n* x 3 matrix of growth values created by pre-multiplying A by the transpose of the second column of C. The E matrix included zeros in the first column and growth values in the second and third columns.
- Create F, a 3 x 3 diagonal covariance matrix of disturbances based on population parameters.
- Create G, a *n* x 3 matrix of disturbances for each measurement occasion. Each column of G had multivariate normal distribution with a mean of zero and variance based on F matrix.
- Calculate H = D+E+G, where H is a *n* x 3 matrix including individual *y* scores in each row, and first assessment scores in the first column, second assessment scores in the second column and third assessment values in the third column.

Data Analysis

A LGM with fixed measurement occasions was fit to the simulated datasets using the Mplus 5.2 (Muthen & Muthen, 2008) software to determine how much difference occurred between estimates and population parameters. Table 3-3 shows an example of the Mplus code. Mplus produced 1000 sets of results for each condition. Each set included estimates of residual variances, intercept mean and variance, slope mean and variance, and standard errors for each estimate. The results obtained also included a variety of model fit information: the chi-square fit statistics, the Akaike information criterion (AIC), the fit indices Bentler's (1990) comparative fit index (CFI), Tucker-Lewis index (Tucker and Lewis,1973) TLI, the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) . The mean of the parameter estimates from 1000 iterations of each condition was calculated using SPSS software. Relative parameter biases for the 27,000 estimates of mean of intercepts, mean of slopes, residual variances, variance of intercept, variance of slope, and intercept-slope covariance were calculated by using the following formula (Hoogland & Boomsma, 1998):

$$B(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta}, \qquad (3\text{-}5)$$

where $\bar{\hat{\theta}}$ is the mean of parameter estimates across replications of a condition and θ is the population parameter.

The relative standard error bias was calculated using the following formula:

$$B(S_{\hat{\theta}}) = \frac{\bar{S}_{\hat{\theta}} - SD(\hat{\theta})}{SD(\hat{\theta})}, \qquad (3\text{-}6)$$

where $\bar{S}_{\hat{\theta}}$ is the mean of estimated standard errors for $\hat{\theta}$ and $SD(\hat{\theta})$ is the empirical standard error, calculated as the standard deviation of the estimates of $\hat{\theta}$. Hoogland and Boomsma (1998), stated that relative biases between ± 0.05 and relative standard error biases between ± 0.1 are considered acceptable.

I calculated the relative parameter bias for each parameter under each condition. Some of the relative bias estimations were unacceptable. For these parameters I conducted analyses of variance (ANOVAs) to figure out which factors affected the relative bias. In the ANOVA analyses, factors were range, distribution and sample size. The dependent variables were 27,000 deviations of the estimates from the population parameters, divided by the population parameters (i.e. relative deviations).In addition η^2 (eta squared) were calculated to compare the effects of factors.

Table 3-3. Example of Mplus LGM Program

```
Variable:
Names are y1-y3;
Analysis: type=general;
Model:
Int slope | y1@0 y2@2 y3@4;
```

Results

Convergence Rates and Improper Solutions

Convergence rates were at least 99.9% for all conditions. Improper solutions occurred due to non-positive definite residual covariance (theta) and latent factor covariance (PSI) matrices. The percentage of non positive definite solutions was near zero for the sample size of 8,000. For the sample size of 2000, only the conditions with a wide range of measurement occasions were associated with approximately 10% improper solutions. The conditions with a sample size of 200 had the highest percentage of improper solutions. Especially with the conditions that had a wide range of measurement occasions, the improper solution rate reached 40%, in other range conditions the rate was approximately 25%. Table 3-4 shows the percentages of convergence and improper solutions for each condition.

Relative Bias of Parameter Estimates and Standard Errors

Table 3-5 presents the relative bias of the estimates of the means of the intercept and slope, the covariance between intercept and slope, the residual variances, the slope variance and the intercept variance for all 27 conditions. The relative bias of the estimates of the means of the intercept and slope and the covariance between the intercept and slope were acceptable in all conditions. Except for the residual variances for the first assessment, all other relative bias for the residual variances were positive, which means they were over estimated. Substantial relative biases for the slope and intercept variances were all negative, which means they were underestimated. Because these parameters showed unacceptable bias with some conditions, I conducted analyses of variance (ANOVAs). Effect sizes η^2 were calculated based on ANOVA results, but values below .02 are not presented. The residual variance bias for the first assessment and intercept variance were all negative, however ANOVA results showed that none of the η^2 for factors was larger than .02. These results imply that the substantial relative biases for the intercept variance and residual variance of the first assessment cannot be explained clearly with range, distribution or sample size differences.

Residual variances for the second assessment were affected by range (η^2=.603), by distribution (η^2=.147) and also by range x distribution interaction (η^2=.092). Residual variance for the third assessment were affected by range (η^2=.412) and range x distribution interaction (η^2=.053). From the ANOVA results, it is noticeable that larger differences in residual variance estimates for the second and third assessments were associated mostly with the range factor. Results showed that, increasing the range of the measurement occasions caused an increase in bias. In other words, the level of overestimation was increased when the range was wider. Distribution was the second effective factor and results showed that the uniform distribution of the measurement occasions was associated with an increase in bias of the estimates. Extremely skewed distribution of the measurement occasions moderated the effect of range. The interaction between range and distribution also affected the relative bias. Results showed that, residual variance estimates for the uniformly distributed conditions that had a wide range of measurement occasions had the largest bias.

ANOVA for the bias in slope variance estimates indicated that range (η^2=.061), distribution (η^2=.088), were associated with differences in relative bias estimates. Also there was a significant interaction between the range and distribution factor (η^2=.001). The extremely skewed distribution of measurement occasions caused an increase in absolute value of the bias estimates and the effect was larger when the range was wide. The range factor also affected the relative bias for the slope variance. Results showed that absolute value of the bias increased with an increase in range of the measurement occasions. The sample size factor was not associated with relative bias estimations. Relative bias for standard errors for all parameter estimates for the conditions simulated were acceptable and they presented in Table 3-6.

Model Fit

Mplus provided several model fit information including chi-square statistic (χ^2). In my simulation study, LGMs were constructed on nine known parameters, and eight unknown parameters were estimated with the model. With one degree of freedom LGMs were over identified. The chi square statistic provides fit information for over identified models and a statistically significant test statistic implies that the model specification does not exactly generate the means or the covariance matrix of the observed variables. In my study, consistent with the literature, p≤.05 for the chi-square test statistic was used as the significance criteria. The calculation of the estimated chi-square value for testing the simultaneous null hypotheses is

$$\chi^2 = (N-1)F_{ML}(\Sigma, \mu, \Sigma(\theta), \mu(\theta)) \tag{3-7}$$

$$F_{ML} = \log|\Sigma(\theta)| + tr(\Sigma(\theta)^{-1}\Sigma) - \log|\Sigma| - q - \left[\mu - \mu(\theta)\right]' \Sigma^{-1}(\theta)\left[\mu - \mu(\theta)\right] \tag{3-8}$$

where N is the sample size, F_{ML} is the maximized value of the likelihood function, Σ is the observed covariance matrix, $\Sigma(\theta)$ is the model implied covariance matrix, μ represents the mean structure of the observed values, $\mu(\theta)$ represents the model implied mean structure, and *q* is the number of observed values. The simultaneous null hypotheses being tested are;

$$H_0 : \mu = \mu(\theta) \ and \ \Sigma = \Sigma(\theta) \tag{3-9}$$

The average chi square statistic across conditions and percentage of significant chi-square tests for each condition are presented in Table 3-7. Results showed that the percentage of significant chi-square statistics was below .05 in 4 of the 27 conditions, indicating that the models fit well under these 4 conditions. Results showed that, on average, 8.9% of the chi square test statistics were significant under the conditions with sample size of 200, whereas 41% were significant with sample size of 2,000 and 59.6% were significant with sample size of 8,000. These results indicate that an increase in sample size was associated with an increase in percentage of significant chi-square test statistics. Results also showed that on average 21.2% of the chi-square test statistics were significant under the conditions with narrow range measurement occasions, whereas 39.9% was significant for the moderate range conditions and 48.4% for the wide range conditions, indicating that an increase in range of measurement occasions was associated with an increase in percentage of significant chi-square test statistics. The percentage of significant chi-square test statistics was not affected by the distribution. However, because chi-square test statistic is very sensitive to small misspecifications as the sample size increases, researchers do not rely solely on this statistic to assess the model fit. Sample sizes in the simulation study were 200, 2000, 8000. The larger the sample size, the more likely the rejection of the model with chi-square test statistic. With very large samples, even small differences between the observed means and covariance matrix and the implied mean and covariance matrix may be found significant.

Other fit information, including TLI, CFI, RMSEA, SRMR and AIC values, presented in Table 3-8. The TLI and CFI are two of the fit indices commonly reported in SEM studies (Bollen & Curran,2006). Equations for TLI and CFI, which includes the test statistic for a baseline model, are

$$TLI = \frac{T_b / df_b - T_h / df_h}{T_b / df_b - 1} \tag{3-10}$$

$$CFI = \frac{(T_b - df_b) - (T_h - df_h)}{T_b - df_b} \tag{3-11}$$

where T_b is the test statistic for the baseline model, df_b is the degrees of freedom for the baseline model, T_h is the test statistic for the hypothesized model and df_h is the degrees of freedom for the hypothesized model. The TLI and CFI values generally range between zero to 1 and value of 1 suggests an ideal fit. Values lower than 0.95 raise concerns about the adequacy of a model (Hu and Bentler, 1999). For the simulated conditions TLI values varied between .991 and 1, indicating that models fit well. CFI values varied between .997 and 1, which also indicate that models fit well

Other fit indices provided by Mplus are RMSEA and SRMR. Hu and Bentler (1999) suggest that values smaller than .06 for RMSEA and smaller than .08 for SRMR indicate a good fit. The formula for the RMSEA and SRMR are;

$$RMSEA = \sqrt{\frac{T_h - df_h}{(N-1)df_h}}. \tag{3.12}$$

where N is the sample size, T_h is the test statistic for the hypothesized model and df_h is the degrees of freedom for the hypothesized model and

$$SRMR = \sqrt{\left\{ 2\sum_{i=1}^{p}\sum_{j=1}^{i}\left[(s_{ij} - \hat{\sigma}_{ij} / (s_{ii}s_{jj}) \right]^2 \right\} / p(p+1)} \tag{3.13}$$

where p is the number of observed variables, s_{ij} is the observed covariances, $\hat{\sigma}_{ij}$ donates the reproduced covariances, and s_{ii} and s_{jj} are the observed standard deviations.

For the simulated conditions SRMR values were all smaller than .016, indicating that models fit well. RMSEA is more conservative than CFI and TLI (Leite, 2007). Results showed that under 4 out of 27 conditions, the models did not fit well according to RMSEA. Conditions with 200, 2000 and 8000 sample sizes, wide range and extremely skewed distribution had RMSEA values of .069, .063, and .076 respectively. Also the condition with 2000 sample size, wide range and moderately skewed distribution had a RMSEA value of .068.

The Akaike Information Criterion (AIC) provided by model is not directly interpretable. Instead of focus on magnitude of AIC, the model with the smallest AIC can be selected as the most adequate model among compared models. AIC values estimated from the same sample sizes indicated that a decrease in the range of the measurement occasions was associated with a decrease in AIC values. For example, from the sample size of 200, average AIC values were 4091, 4114, and 4026, for the conditions with narrow, moderate, and wide range, respectively. It is also noticeable that within a sample size and range level, an increase in skewness was associated with a decrease in AIC values, indicating that conditions with extremely skewed distributions had smaller AIC values than conditions with moderately skewed and uniform distributions.

Discussion

My simulation study focused on whether an unconditional LGM can produce accurate results when conducted with fixed time points for measurement waves when there are inter individually varying differences in assessment dates. Using known population parameters which were taken from a published study, three waves of dependent scores were generated based on three different factors. The first of these three factors was sample sizes with three levels. Small, moderate and relatively large sample sizes of 200, 2000 and 8000 were evaluated to determine if they produced differences with respect to overall fit of the model or bias in parameter estimates and standard errors. Results showed that sample size differences did not appreciably affect model fit, parameter estimates or standard errors. However, a small sample size of 200 was associated with non-positive definite solutions. Almost 35% of the Mplus analyses of simulated data sets produced improper solutions with a small sample size. When the sample size is not large enough, improper solutions may occur due to mere sampling fluctuation. Anderson and Gerbing (1984) explained how parameter matrices (Theta-

Delta, Theta-Epsilon, PSI and PHI) may be non-positive definite through mere sampling fluctuation.

The second factor of interest was range in individual assessment dates. This factor had three levels, narrow moderate and wide. Even though range was varied over levels, each range condition had the same mean values for time points, 0,2 and 4 for the first, second and third assessments respectively. These values were chosen to represent assessment waves for every two years. Narrowly ranging time points were simulated based on the ECLS-K. Widely ranging time points captured the assessment dates right after first year and just before the third year for the second assessment dates. For the third assessment, dates between third year and fifth year were captured. Wider ranges for the assessment dates than the ones considered in this study would be unrealistic. Even though there was a significant interaction between the range and distribution factor, the large eta squared values for the range effect deserves discussion. The range factor substantially affected the residual variance estimates. Increased heterogeneity in individual assessment dates caused larger residual variances for the second and third assessment. Large residual variances indicate an increase in unexplained variation of the dependent variable scores. These results showed that a larger portion of the dependent variable variation will remain unexplained if fixed time points are used when heterogeneity exists in assessment dates. In other words there will be a larger amount of error variance when heterogeneity in measurement occasions is omitted in the analyses. The interaction between the range and distribution factor affected the residual variance estimates, and results showed that the absolute value of relative bias were larger for the conditions with widely ranging assessment dates with a uniform distribution. However the bias of the residuals with a uniform distribution decreased as range changed from wide to narrow. In other words, skewness moderated the range effect. The smallest absolute value of relative bias for the second residual variance was obtained under the condition that had narrowly ranging assessment dates with an extremely skewed distribution. Moreover, the smallest absolute value of relative bias for the third residual variance was estimated under the condition that also had narrowly ranging assessment dates with a moderately skewed distribution. Under these two conditions, residual variances estimated with a high degree of accuracy.

Negatively biased slope variance estimates were found in the simulation study. The interaction between the range and distribution factor affected the slope variance estimates. The largest negative bias was estimated for the condition with a sample size of 200, a wide range of assessment dates and extremely skewed distribution of assessment dates. The second largest negative bias was estimated also for a condition with a wide range of assessment dates and extremely skewed distribution. The absolute value of bias for the slope variance with a wide range condition decreased as distribution change from extremely skewed to uniform. Relative bias estimates were not negative for the uniform distribution but with a sample size of 8000. This finding is consistent with an empirical example results provided by Singer and Willet (2003). They did not use individually varying assessment dates as the time indicator but used participant's month based age in an unconditional LGM. Participants' ages at the beginning of the study varied between 72 and 84 months and caused heterogeneity in time points. They concluded that estimated variance components would be larger if heterogeneity in time points was ignored and fixed time points were used. The reason is that LGM with fixed time points fits worse, and fixed time points introduce error into the analysis. There is more unexplained variation in individual slopes. However the distribution of age was not reported in their study. In my study the distribution factor had three levels, uniform, moderately skewed and extremely skewed. Moderately skewed and extremely skewed distributions were weakly associated with slope variance estimates, and associated with underestimation of these values. Results showed that bias of the slope variance was larger when there was extreme skewness. In other words, estimated slope variances were smaller than the population parameter when there was skewness in the distribution of measurement occasion. For the simulated conditions, these findings are reasonable, because with fixed time points the mean of the slope will be to the right side of most of the skewed distribution, and individuals to the left of the distribution will have similar slopes. This similarity might reduce the variance of the slope.

As reported in the results section, a few of the relative bias estimates for the intercept variances were unacceptable; however, calculated eta squared values were smaller than .02 indicating that effect of the simulated factors were very small. The mean intercept values were estimated without bias under all 27 conditions. These results were expected because in each condition intercepts were estimated independently from the manipulations for the measurement occasions. Mean slope values were also estimated without bias indicating that omitting inter individually varying assessment points did not affect the mean slope estimations in this simulation study. Correct estimates for the mean slope might have occurred due to same mean values of time points. Even with the range and distribution manipulations, mean values of the measurement time points were same across all conditions; 0, 2 and 4 for the first, second and third measurement occasions respectively. The covariance between slope and intercept also estimated without bias across all conditions. In all conditions, data sets were generated with correlation coefficient of .81 between intercept and slope.

Chi-square tests for the model fit showed that only four of the simulated conditions produced an adequate fit. Results showed that conditions with a small sample size and a narrow range for measurement occasions are more likely to provide acceptable model fit based on chi-square test statistic. It is known that chi-square statistic is conservative and researchers do not solely report this statistic to decide the model fit in SEM studies. RMSEA fit indices for the models showed that conditions with the extremely skewed measurement occasions distribution did not provide an acceptable model fit. However these unacceptable RMSEA indices were close to cut off criteria of .06. Furthermore, there was an inconsistency between RMSEA and AIC values. Results showed that AIC values were smaller for the extremely skewed distributions for all sample size and range conditions. CFI, TLI and SRMR values were all acceptable. Researchers generally make their final decision about the model fit based on multiple sources. My overall conclusion about the model fits for the simulated conditions is that omitting the inter individual heterogeneity in measurement occasions does not cause serious problems.

Table 3-4. Percentages of convergence and improper solutions for each condition

Sample Size	Range	Distribution	Convergence	Imp. Solutions
200	Narrow	Uniform	100%	24.2%
		M. Skewed	100%	24.3%
		E. Skewed	100%	27.8%
	Moderate	Uniform	100%	24.7%
		M. Skewed	100%	33.5%
		E. Skewed	100%	28.6%
	Wide	Uniform	99.9%	43.1%
		M. Skewed	100%	36.2%
		E. Skewed	100%	44.5%
2000	Narrow	Uniform	100%	.003%
		M. Skewed	100%	.011%
		E. Skewed	100%	.006%
	Moderate	Uniform	100%	.003%
		M. Skewed	100%	.013%
		E. Skewed	100%	.020%
	Wide	Uniform	100%	.070%
		M. Skewed	100%	.127%
		E. Skewed	99.9%	.094%
8000	Narrow	Uniform	100%	0%
		M. Skewed	100%	0%
		E. Skewed	100%	0%
	Moderate	Uniform	100%	0%
		M. Skewed	100%	0%
		E. Skewed	100%	0%
	Wide	Uniform	100%	0%
		M. Skewed	100%	0%
		E. Skewed	100%	16%

Table 3-5. Comparison of relative parameter bias estimates across conditions.

Sample Size	Range	Distribution	$VAR(\varepsilon_1)$	$VAR(\varepsilon_2)$	$VAR(\varepsilon_3)$	$\psi_{\beta\beta}$	$\psi_{\alpha\alpha}$	μ_α	μ_β	$\psi_{\alpha\beta}$
200	Narrow	Uniform	**-.065**	**.153**	-.025	.035	.016	.001	.011	-.004
		M. Skewed	-.018	**.083**	.001	-.021	-.005	.001	-.008	-.012
		E. Skewed	.038	.028	.027	-.021	-.010	-.001	-.008	.005
	Moderate	Uniform	**-.187**	**.600**	-.015	.017	.044	.001	.001	.044
		M. Skewed	**.111**	**.157**	**.129**	**-.062**	-.039	-.002	-.016	.005
		E. Skewed	.029	**.165**	**.066**	**-.046**	-.015	-.001	-.015	-.017
	Wide	Uniform	.023	**2.088**	**.589**	-.017	-.013	.001	.018	.019
		M. Skewed	.021	**1.080**	**.251**	**-.077**	-.019	-.001	-.027	-.031
		E. Skewed	**.236**	**.711**	**.305**	**-.131**	**-.070**	-.002	-.042	.011
2000	Narrow	Uniform	**-.072**	**.170**	-.018	.022	.021	.001	.005	-.011
		M. Skewed	.034	.036	.027	-.021	-.011	-.001	-.007	-.001
		E. Skewed	.032	.031	.026	-.021	-.010	-.002	-.008	.001
	Moderate	Uniform	**-.106**	**.543**	.028	.039	.031	.001	.009	-.014
		M. Skewed	**.068**	**.210**	**.086**	-.040	-.021	-.001	-.014	.002
		E. Skewed	**.048**	**.178**	**.070**	-.045	-.017	-.001	-.018	-.001
	Wide	Uniform	**.095**	**2.011**	**.454**	-.015	-.027	-.001	.001	.023
		M. Skewed	**.176**	**.981**	**.347**	**-.096**	**-.052**	-.001	-.033	.007
		E. Skewed	**.163**	**.828**	**.292**	**-.106**	**-.047**	-.001	-.039	-.002
8000	Narrow	Uniform	**-.079**	**.179**	-.018	.024	.023	.001	.005	-.012
		M. Skewed	.015	**.050**	.023	-.015	-.004	-.001	-.006	-.002
		E. Skewed	.019	.036	.022	-.020	-.005	-.001	-.008	-.004
	Moderate	Uniform	**-.124**	**.554**	.024	.036	.036	.001	.008	-.020
		M. Skewed	.026	**.250**	**.066**	-.029	-.008	-.001	-.012	-.007
		E. Skewed	**.075**	**.170**	**.080**	**-.047**	-.022	-.001	-.018	-.001
	Wide	Uniform	.011	**1.997**	**.436**	.002	-.004	-.001	.003	.005
		M. Skewed	**.115**	**1.114**	**.296**	**-.078**	-.033	-.001	-.031	-.003
		E. Skewed	**.191**	**.799**	**.301**	**-.116**	**-.055**	-.002	-.041	.003

$VAR(\varepsilon_1)$: Residual-variance-1, $\psi_{\beta\beta}$:Slope-variance, $\psi_{\alpha\alpha}$:Intercept-variance, μ_α: Intercept-mean, μ_β: Slope-mean, $\psi_{\alpha\beta}$:Covariance between intercept and slope

Table 3-6. Relative bias estimates for standard errors.

Sample Size	Range	Distribution	$VAR(\varepsilon_1)$	$VAR(\varepsilon_2)$	$VAR(\varepsilon_3)$	$\psi_{\beta\beta}$	$\psi_{\alpha\alpha}$	μ_α	μ_β	$\psi_{\alpha\beta}$
200	Narrow	Uniform	-.027	-.009	.009	-.022	-.034	-.007	-.012	-.004
		M. Skewed	.025	.006	.000	-.003	.020	-.034	.032	.004
		E. Skewed	.003	-.008	-.007	.007	.000	.041	.014	-.003
	Moderate	Uniform	-.030	-.027	-.014	-.022	-.028	-.028	.030	.010
		M. Skewed	-.034	-.026	-.020	-.004	-.029	-.033	.011	-.002
		E. Skewed	-.023	-.020	-.037	.031	-.026	-.032	-.033	-.025
	Wide	Uniform	-.016	-.004	.001	-.011	-.020	-.007	.042	-.022
		M. Skewed	.007	-.018	-.030	-.045	.019	-.011	.026	-.008
		E. Skewed	-.011	-.056	-.042	-.017	-.052	.019	.017	-.001
2000	Narrow	Uniform	.009	.046	.020	.046	.000	-.033	.009	.009
		M. Skewed	-.009	.007	.007	-.006	-.002	-.036	.056	-.006
		E. Skewed	.010	-.002	-.014	-.053	-.002	-.035	-.078	.022
	Moderate	Uniform	-.021	.023	.001	.007	-.010	-.031	.025	.000
		M. Skewed	-.006	.012	.023	.017	-.014	.039	.057	.014
		E. Skewed	.002	-.006	-.031	.006	-.052	.038	.052	.007
	Wide	Uniform	.004	.019	-.031	-.020	.003	-.032	.079	.030
		M. Skewed	-.039	-.017	-.035	.003	-.014	-.032	.076	-.049
		E. Skewed	.000	-.053	-.007	-.004	.012	-.032	.064	-.022
8000	Narrow	Uniform	.018	.057	.046	-.002	.032	-.033	.039	-.042
		M. Skewed	.033	.041	.018	.021	-.011	-.035	-.076	.008
		E. Skewed	.045	.035	.018	.017	.022	-.035	-.078	.006
	Moderate	Uniform	-.033	-.038	-.024	-.039	-.030	-.031	.053	-.014
		M. Skewed	-.036	.001	-.020	-.003	-.031	-.035	-.074	-.002
		E. Skewed	-.020	.010	-.001	-.016	-.042	-.035	-.078	.024
	Wide	Uniform	-.013	.014	-.021	-.023	-.015	-.032	.037	.026
		M. Skewed	-.008	.003	-.026	-.027	-.018	-.032	-.058	.005
		E. Skewed	-.053	-.063	-.042	-.030	-.016	-.032	-.070	-.055

$VAR(\varepsilon_1)$: Residual-variance-1, $\psi_{\beta\beta}$:Slope-variance, $\psi_{\alpha\alpha}$:Intercept-variance,
μ_α: Intercept-mean, μ_β: Slope-mean, $\psi_{\alpha\beta}$:Covariance between intercept and slope

Table 3-7. Chi-square statistic results across conditions

Sample Size	Range	Distribution	Average Chi Sq.	P(Chi.Sq<).05
200	Narrow	Uniform	1.345	9.2%
		M. Skewed	1.005	**3.9%**
		E. Skewed	1.059	5.5%
	Moderate	Uniform	2.051	17.0%
		M. Skewed	1.409	10.0%
		E. Skewed	1.115	6.1%
	Wide	Uniform	.515	**.5%**
		M. Skewed	.698	**2.0%**
		E. Skewed	2.575	25.6%
2000	Narrow	Uniform	3.370	33.7%
		M. Skewed	1.291	7.8%
		E. Skewed	1.411	9.7%
	Moderate	Uniform	6.462	67.9%
		M. Skewed	2.596	26.0%
		E. Skewed	2.054	17.2%
	Wide	Uniform	2.419	21.9%
		M. Skewed	10.908	94.4%
		E. Skewed	9.737	90.1%
8000	Narrow	Uniform	13.113	94.5%
		M. Skewed	1.398	9.8%
		E. Skewed	1.962	16.6%
	Moderate	Uniform	29.195	100%
		M. Skewed	2.554	24.6%
		E. Skewed	10.914	90.2%
	Wide	Uniform	.751	**2.5%**
		M. Skewed	14.462	98.5%
		E. Skewed	47.293	100%

Table 3-8. Comparison of model fit information across conditions.

Sample Size	Range	Distribution	CFI	TLI	RMSEA	SRMR	AIC
200	Narrow	Uniform	.998	.998	.034	.009.	4098
		M. Skewed	.999	.999	.025	.008.	4087
		E. Skewed	.999	.999	.026	.008	4088
	Moderate	Uniform	.997	.993	.056	.012	4130
		M. Skewed	.998	.997	.036	.010	4108
		E. Skewed	.999	.999	.027	.009	4103
	Wide	Uniform	.999	1.000	.010	.007	4272
		M. Skewed	.999	1.000	.016	.007	4183
		E. Skewed	.995	.989	**.069**	.016	4164
2000	Narrow	Uniform	.999	.998	.027	.005	40899
		M. Skewed	.999	.999	.011	.003	40806
		E. Skewed	.999	.999	.011	.003	40799
	Moderate	Uniform	.999	.996	.047	.007	41278
		M. Skewed	.999	.999	.022	.004	41017
		E. Skewed	1.000	.999	.017	.004	40956
	Wide	Uniform	.999	.999	.022	.005	42523
		M. Skewed	.997	.992	**.068**	.011	41805
		E. Skewed	.998	.993	**.063**	.010	41641
8000	Narrow	Uniform	.999	.998	.037	.005	163602
		M. Skewed	.999	.999	.006	.001	163231
		E. Skewed	.999	.999	.008	.002	163153
	Moderate	Uniform	.998	.995	.058	.008	165073
		M. Skewed	.999	.999	.011	.002	164099
		E. Skewed	.999	.999	.033	.005	163852
	Wide	Uniform	.999	1.000	.003	.001	170004
		M. Skewed	.999	.997	.040	.006	167363
		E. Skewed	.997	.991	**.076**	.012	166474

CHAPTER 4
CONCLUSION

In the educational research, it is important to accurately assess individual changes in a behavior or skill. It is also valuable to understand which factors are related to these changes. In order to examine individual growth in detail, large and complex longitudinal studies have been conducted in last decades. One of the popular methods to analyze longitudinal data sets is LGM. Individual observations are defined as a function of time in a LGM. Hence, to accurately define the time indicators in a LGM is important. There are two commonly used approaches to create time indicators; using the ages of the participants or the measurement waves. The literature indicates that time coding strategies for a LGM were examined generally with time points based on participants' ages at the measurement occasions. However, in educational research it is common to use measurement waves as the time points, (i.e. spring assessment, fall assessment). When using the measurement waves as the time indicators, it is ideal to assess each participant at the same time; unfortunately; this is impractical with large studies. The literature indicates that in the applied LGM studies, researchers tend to ignore heterogeneity in measurement dates and assume each child assessed at the same time. In this thesis, I examined the effects of omitting the inter individually varying measurement dates in LGM.

The ECLS-K dataset provides exact dates of measurement for each student. Morgan, Farkas and Wu (2009) published a study in which they examined mathematical growth for a subsample taken from ECLS-K. They made the assumption that each individual was assessed at the same time within a measurement wave. Based on their study, a conditional quadratic growth model was constructed and examined with different time coding strategies in chapter 2. I used the exact number of days between assessments to create time points for each child in order to take into account the heterogeneity in measurement dates. With the particular dataset, results showed that there were no appreciable differences in compared models. Model fit was essentially the same across models and differences in the parameter estimates did occur due to different time scales but not due to heterogeneity in assessment dates. A mathematical explanation was provided for the differences in parameter estimates. These results indicate that researchers can omit the heterogeneity in assessment dates when working on ECLS-K data set, because heterogeneity in these dates is not large enough to affect estimates.

In Chapter 3, a simulation study was conducted to see if different range of the measurement occasions would cause biased estimates in an unconditional linear growth model. The results of this thesis support the conclusion that, there will be a larger amount of error variance when heterogeneity in measurement occasions is omitted. In other words, a larger portion of the dependent variable variation will remain unexplained especially if measurement occasions have a wide range. Based on these results, I recommend that researchers should avoid using the fixed time points approach if there is a large amount of heterogeneity in assessment dates.

In Chapter 3, effects of distribution differences of the measurement occasions were also examined. Uniform distributions of measurement occasions represented a situation in which the same number of participants is assessed in each attempt within an assessment wave. Skewed distributions represented a situation in where most data is collected during a short interval of time. These are realistic situations, and results showed that omitting the heterogeneity might cause incorrect estimates for the slope variance if occasions have a skewed distribution. Another conclusion that can be drawn from this thesis is that a small sample size resulted in approximately 35% improper solutions in simulated conditions.

It is also noticeable that, even though biased estimates were produced with fixed time points approach, all models provided acceptable model fit indices for the simulated data sets. However the chi square statistic indicated lack of exact fit in most conditions. Results showed that having acceptable model fit values do not justify using fixed time points approach. Moreover, consistent with the literature, my empirical study based on Morgan, Farkas and Wu (2009), support the conclusion that models constructed on person-specific time points of measurement yield more adequate fit to the data. The overall conclusion of this thesis is that, researchers should estimate their growth models with person-specific time points when the variability of time points is large. If the heterogeneity in assessment dates is not large, there should be no substantial differences in estimates and model fit information between the models estimated with fixed time points and person-specific time points. At this point, fixed time points approach could be preferred when using Mplus software, because Mplus does not provide some of the fit information for the models estimated with person-specific time points.

One limitation of this thesis is that my simulation study focused only on unconditional linear growth models. Future research is needed to investigate the effects of omitting the heterogeneity in assessment dates in conditional growth models. Both time invariant and time-varying predictors should be included in these conditional models. It is also necessary to investigate non-linear growth models. Another limitation is that, I used only 3 waves of measurement for the dependent variable. This is the minimum number that a LGM requires. Future research might focus on four or more waves.

APPENDIX

CALCULATING THE NUMBER OF DAYS WITH SPSS

- compute date1=DATE.DMY(c3asmtdd,c3asmtmm,c3asmtyy).
- formats date1(DATE11).
- variable widht date1(11).
- compute date2=DATE.DMY(c4asmtdd,c4asmtmm,c4asmtyy).
- formats date2(DATE12).
- variable widht date2(12).
- compute elapse12=(date2-date1)/86400.
- execute.
-
- *To create four time points.*
-
- compute fix = 0.
- compute fix1 = 2
- compute fix2 = 4.
- compute fixnday = 0.
- compute fixnday1 = 7.25.
- compute fixnday2 = 13.44.
- compute randomnday = 0.
- compute randomnday1 = elapse12/100.
- compute randomnday2 = elapse14/100.
- compute random= 0.
- compute random1=elapse12/358.
- compute random2= elapse14/358.
- execute.

LIST OF REFERENCES

Baer, J., & Schmitz, M. (2000). Latent growth curve modeling with a cohort sequential design. Social Work Research, 24, 243.

Biesanz, J., Deeb-Sossa, N., Papadakis, A., Bollen, K., & Curran, P. (2004). The role of coding time in estimating and interpreting growth curve models. Psychological Methods, 9, 30-52. doi:10.1037/1082-989X.9.1.30.

Blozis, S., & Cho, Y. (2008). Coding and centering of time in latent curve models in the presence of interindividual time heterogeneity. Structural Equation Modeling, 15, 413-433. doi:10.1080/10705510802154299.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. The Elementary School Journal, 108, 115-130. doi:10.1086/525550.

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. Sociological Methods & Research, 32, 336-383. doi:10.1177/0049124103260222.

Bollen, K.A., & Curran, P.J. (2006). Latent curve models*: A structural equation approach*. Hoboken, NJ: Wiley. ISBN: 047145592X

Raudenbush, S. W., & Bryk, A. S. (2002).Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Cheadle, J. (2008). Educational investment, family context, and children's math and reading growth from kindergarten through the third grade. *Sociology of Education*, 1-31. Retrieved from PsycINFO database.

Collins, L., & Sayer, A. (2001). *New methods for the analysis of change*. Washington, DC US: American Psychological Association. doi:10.1037/10409-000.

Davison, M. (2008). Hypothesis generation in latent growth curve modeling using principal components. *Educational Research & Evaluation*, *14*, 321-334. doi:10.1080/13803610802249498.

DiPerna, J., Pui-Wa, L., & Reid, E. (2007). Kindergarten predictors of mathematical growth in the primary grades: an investigation using the early childhood longitudinal study kindergarten cohort. *Journal of Educational Psychology*, *99*, 369-379.

Duncan, T., & Duncan, S. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, *35*, 333-363

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

Dunn,G., Everitt, B., & Pickles, A.(1993). *Modeling covariances and latent variables using EQS*. London: Chapman & Hall.

Farrell, A., Sullivan, T., Esposito, L., Meyer, A., & Valois, R. (2005). A latent growth curve analysis of the structure of aggression, drug use, and delinquent behaviors and their interrelations over time in urban and rural adolescents. *Journal of Research on Adolescence (Blackwell Publishing Limited)*, *15*, 179-203. doi:10.1111/j.1532-7795.2005.00091.x.

Gottman, J., & Rushe, R. (1993). The analysis of change: Issues, fallacies, and new ideas. *Journal of Consulting & Clinical Psychology*, *61*, 907.

Gresham, F., & Elliott, S. (1990). Social skills rating system. Circle Pines , MN: American Guidance Service.

Hancock, G., Kuo, W., & Lawrence, F. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, *8*, 470-489.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course*. Greenwood, CT: Information Age Publishing, Inc.

Hertzog, C., & Nesselroade, J. (2003). Assessing psychological change in adulthood: an overview of methodological issues. *Psychology & Aging*, *18*, 639-657. doi:10.1037/0882-7974.18.4.639.

Hong, G., & Raudenbush, S. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*, 205-224. doi:10.3102/01623737027003205.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *PJE. Peabody Journal of Education*, *77*, 189-215. http://search.ebscohost.com.lp.hscl.ufl.edu, doi:NO_DOI

Kim, H. (2008). An analysis of developmentally appropriate and culturally responsive practices and the learning trajectories of kindergarten, first-grade, and third-grade children from ECLS-k: Teachers' beliefs and practices as mediators. *Dissertation Abstracts International Section A*, 68, 4965.

Lawrence, F., & Hancock, G. (1998). Assessing change over time using latent growth modeling. *Measurement & Evaluation in Counseling & Development (American Counseling Association)*, 30, 211

Lee, E. (2008). A latent growth curve analysis of the impact of school mobility on the reading scores of poor and non-poor children in the U.S. *Dissertation Abstracts International Section A*, 69, 874.

Leite, W. (2007). A comparison of latent growth models for constructs measured by multiple Items. *Structural Equation Modeling*, *14*, 581-610. doi:10.1080/10705510701575438.

Luo, Z. (2003). Children's mathematical development from kindergarten to first grade: Identifying early predictors for success. *Dissertation Abstracts International*, 64, 443.

Marks, A., & Coil, C. (2007). Psychological and demographic correlates of early academic skill development among American Indian and Alaska native youth: A growth modeling study. *Developmental Psychology*, *43*, 663-674. doi:10.1037/0012-1649.43.3.663.

McArdle, J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110. doi:10.1111/1467-8624.ep7264163.

Mckinnon, M. (2002, January). Latent growth curve and multilevel analysis of reading acquisition. *Dissertation Abstracts International Section A*, 63, 2133.

Mehta, P., & West, S. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*, 23-43. doi:10.1037/1082-989X.5.1.23.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122. doi:10.1007/BF02294746

Morgan, P., Farkas, G., & Qiong, W. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*, 306-321.

Muthen, B., & Siek-toon, K. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning & Individual Differences*, *10*, 73.

Perez-Johnson, I. (2008). Parsing Hispanic-White achievement gaps: The influence of individual, family, and school factors on mathematics achievement differences in the elementary grades. *Dissertation Abstracts International Section A*, 69, 1281.

Preacher, K. (2009). A review of “An introduction to latent variable growth curve modeling: Concepts, issues, and applications” (2nd ed.). *Structural Equation Modeling*, *16*, 186-190.

Raudenbush, S.W. (2001). Toward a coherent framework for comparing trajectories of change. In Collins, L. M. & Sayer, A. G. (Eds.) *New methods for the analysis of change* (pp. 33-64). Washington, DC: APA.

Rovine, M., & Molenaar, P. (2001). A structural equations modeling approach to the general linear mixed model. *New methods for the analysis of change* (pp. 67-98). Washington, DC US: American Psychological Association. doi:10.1037/10409-003

Singer J. & Willett. J (2003) Applied Longitudinal Data Analysis: *Modeling Change and Event Occurrence.* New York: Oxford University Press. DM000623770UO

Sukkyung, Y., & Sharkey, J. (2009). Testing a developmental-ecological model of student engagement: a multilevel latent growth curve analysis. *Educational Psychology*, *29*, 659-684. doi:10.1080/01443410903206815.

Wen-Jui, H. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology*, *44*, 1572-1590. http://search.ebscohost.com.lp.hscl.ufl.edu, doi:10.1037/a0013886

Willett, J. (1997). Measuring change: What individual growth modeling buys you. *Change and development: Issues of theory, method, and application* (pp. 213-243). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers. http://search.ebscohost.com.lp.hscl.ufl.edu

Willett, J. (2004). Investigating individual change and development: The multilevel model for change and the method of latent growth modeling. *Research in Human Development*, *1*, 31-57.

Willett, J., & Keiley, M. (2000). Using covariance structure analysis to model change over time. *Handbook of applied multivariate statistics and mathematical modeling* (pp. 665-694). San Diego, CA US: Academic Press. doi:10.1016/B978-012691360-6/50024-0.

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

Woodcock, R., McGrew, K., & Werder, J. (1994). Woodcock-McGrew-Werder Mini-Battery of Achievement

Xitao, F., & Xiaotao, F. (2005). Power of latent growth modeling for detecting linear growth: Number of measurements and comparison with other analytic approaches. *Journal of Experimental Education*, *73*, 121-139

Vagi, S. (2008). Socioeconomic status and achievement in math and reading in kindergarten through elementary school: The role of social capital. *Dissertation Abstracts International*, 68, 7005.

Voelkle, M., Wittmann, W., & Ackerman, P. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning & Individual Differences*, *16*, 303-319. doi:10.1016/j.lindif.2006.01.001.

BIOGRAPHICAL SKETCH

Burak Aydin was born in Istanbul, Turkey. He received his bachelors in arts degree in mathematics education from Kocaeli University, Turkey. He served for the Turkish Government as a mathematics teacher for seven months. He later qualified for a scholarship to study abroad and in the fall of 2008, enrolled for graduate studies in the Department of Educational Psychology at the University of Florida and will receive his master of art in education in research and evaluation methodology form the Department of Educational Psychology in August 2010.