5 Linear and Quadratic Growth Models for Continuous and Dichotomous Outcomes

Ann A. O'Connell, Jessica A. R. Logan, Jill M. Pentimonti, and D. Betsy McCoach

5.1 Introduction

Studies of individual change or growth over time pervade the educational and social science research literature, and recent advances in methodology in tandem with technical improvements in statistical software have greatly increased the options that researchers have at their disposal for tracking and describing patterns of change (Collins, 2006; Fitzmaurice & Molenberghs, 2009). Along with this plethora of options comes a corresponding challenge for applied researchers in choosing an appropriate model to represent their data. Unfortunately, neither simplicity nor complexity of a selected model or analytic approach will guarantee its usefulness. Adding further confusion to this situation is the observation that more than one type of model or approach might reasonably describe the patterns in the data. So, what is a researcher to do?

Hierarchical linear modeling (HLM; or multilevel modeling; Chapter 4, this volume) provides a framework that carries particular appeal for fitting individual growth curves to different kinds of data and growth trajectories. Within a multilevel framework, linear models can readily be adapted for nonlinear growth, and nonlinear models can be built for describing change in dichotomous, ordinal, count or other forms of discrete or limited data. In the hierarchy of a multilevel model, repeated observations—which include the targeted outcome, an indicator for time, and any potential time-varying covariates—constitute level 1 of the data. These are variables that vary *within* individuals over time and are situated at the occasion-level of the model. Variables that vary *between* individuals (but not within) such as gender, socioeconomic status, previous education history, and so on provide information at level 2, or the person-level of the model. Multilevel models or HLM are also referred to as mixed-effects regression models (Hedeker & Gibbons, 2006; Raudenbush & Bryk, 2002), a moniker that refers to the primary characteristic of these models to simultaneously capture and describe each person's trend over time as well as estimate variation in these trends for the population of interest.

Longitudinal data for modeling growth typically take the form of repeated observations on the same individuals and are collected over three or more points in time (see Chapter 10 of this volume for a discussion of two-wave change). In addition to capturing change within persons over time (assessed at level 1), the HLM framework allows for examination of between-person differences in linear and nonlinear patterns of change (assessed at level 2). We focus on the use of polynomial models (e.g., linear, quadratic, or cubic, etc.), which provide a simple yet sophisticated general strategy for capturing the shape of change. Perhaps because of their simplicity, researchers have noted that polynomial models have benefits over piecewise models or other methods (Collins, 2006; Willet, Singer, & Martin, 1998), particularly for outcomes that change in a curvilinear pattern over time or those with ceiling or floor effects (Hedeker & Gibbons, 2006).

Among the advantages of the HLM approach to Individual Growth Curves (IGCs) are that variations in the timing of data collection across individuals can be easily accommodated statistically, as can the existence of randomly missing data points. Both of these features are quite common in longitudinal designs, but not always handled optimally in alternative analysis strategies, such as univariate or multivariate repeated measures analysis of variance (O'Connell & McCoach, 2004). A further advantage of the HLM framework is that competing structures for the matrix of variances and covariances of the repeated measures can be fit and compared. In addition, the multilevel framework for analyzing longitudinal data also provides a stepping-stone for other more advanced methods of analysis, such as structural equation or latent growth curve modeling (Curran, 2003; Duncan, Duncan, & Stryker, 2006; Meredith & Tisak, 1990) or growth mixture models (Muthen & Asparouhov, 2009), which are topics included in later sections of this book.

5.1.1 Overview

Often, the simplest representation of growth at level 1 is a linear trajectory over time. For example, if a researcher wanted to examine height for children ages 6 to 12, she would estimate the number of inches gained per child per year on average, and this number would be the same for children ages 6 and 7 and children ages 11 and 12 (example normative growth charts are presented in Figure 1.1 [Petscher, Logan, and Zhou, Chapter 1, this volume]). However, assuming a linear growth trajectory without considering other options may result in a serious misspecification of the underlying growth process. If the researcher were to assume a linear trajectory and continue measuring height up through age 20 (when people have generally stopped growing), she would likely misrepresent the growth pattern of the participants. When the level-1 model is misspecified, this can lead to incorrect parameter estimates and serious errors of inference in the level-1 model. Misspecification can also result in incorrect parameter estimates and errors of inference for the effects of level-2 variables on level-1 slopes and intercepts (McCoach & Kaniskan, 2010; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Snijders & Berkhof, 2008).

When the effect of time on the response variable tends to level off and become weaker (or accelerate and become stronger) over time, fitting a quadratic growth model may be advantageous (Hedeker & Gibbons, 2006; Willet et al., 1998). Returning to the example of measuring height across age, fitting quadratic growth would allow our researcher to account for the slowing down of growth (also called deceleration) that occurs at later ages. The examples provided in this chapter demonstrate how to fit linear growth models, and how quadratic models can be used to fit common forms of nonlinear change. We also demonstrate how to use model fit indices and residual plots to determine the most appropriate shape of growth to avoid problems of misspecification.

A sufficient number of time points are required for application of a polynomial model. In general, one additional occasion of data collection is needed to compensate for each additional randomly varying parameter representing the growth trajectory. Thus, at least three occasions of data are required for a linear growth model (since we estimate two randomly varying parameters: the intercepts and the slopes). At least four occasions of data are required for a quadratic model in which all three growth parameters (intercept, linear trend component, quadratic trend component) are allowed to randomly vary across people, and at least five are required for a cubic growth model in which all four growth parameters (intercept, linear, quadratic, and cubic components) are allowed to randomly vary across people. Although it is possible to estimate nonlinear models with only three time points (e.g., Christ, Silberglitt, Yeo, & Cormier, 2010; Crowe, Connor, & Petscher, 2009), these require specific model constraints.

In this chapter, we focus on polynomial models for change and demonstrate a stepped approach to fitting polynomial growth models. In the next sections, we describe the data used for illustration of the model-building process for linear and nonlinear growth for continuous outcomes. Using HLMv7 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) as a template for analysis, we detail a process of model building for linear growth, reviewing alternatives for coding the passage of time, and then expanding on the model-building process for nonlinear growth. Our approach emphasizes examination of residual plots as an aid to detection and representation of an optimal statistical model. We then provide a discussion and demonstration of a nonlinear model for change in a dichotomous outcome over time: the multilevel logistic model for change. We end by reviewing and summarizing critical measurement, design and statistical power issues for the optimal construction of linear and nonlinear growth models.

5.2 Modeling Change in Continuous Outcomes

5.2.1 Demonstration Data Set: Children of the NLSY79

Data for the analyses presented here were drawn from the children of the National Longitudinal Survey of Youth 1979 (Children of the NLSY79), one of a set of surveys included in the National Longitudinal Surveys (NLS) program (NLS, 2005). The Children of the NLSY79 survey is sponsored and directed by the U.S. Bureau of Labor Statistics and the National Institute for Child Health and Human Development. The survey is managed by the Center for Human Resource Research at The Ohio State University, and interviews were conducted by the National Opinion Research Center at the University of Chicago. The data set used for analysis in the current illustration is an SPSS file, named "NLSY_Growth.sav."

The Children of NLSY79 data set consists of a nationally representative sample of approximately 8,000 children, born to mothers who were included in the original NLSY79. The children were given multiple assessments biannually, beginning in 1986. For the purposes of this chapter, children were selected for inclusion in our demonstration sample if they were born in 1992 or 1993, making them approximately 6 years old when assessed in 1998. Children were only included if they were assessed on the Peabody Individual Achievement Test (PIAT) Math Assessment (Dunn & Markwardt, 1970) on five different occasions between 1998 and 2008, when their ages were approximately 6, 8, 10, 12, and 14 years. For simplicity, only children with complete data on the selected predictors were included. In total, 245 students met the inclusion criteria.

5.2.1.1 Variables

For our demonstrations, *TIME* was coded as 0, 2, 4, 6, or 8, referencing time in years since baseline PIAT assessment in 1998. We selected one continuous child-level predictor, mother's education level, and one dichotomous child-level predictor, sex (0 = male, 1 = female). From the total of 245 children, 48.2% were boys and 51.8% were girls. The mean education level of the mothers for the children in our sample was 13.6 years. Table 5.1 provides a summary of the variables used, and Table 5.2 contains PIAT means across the five measurement occasions, by gender.

5.2.2 Patterns of change

Prior to fitting individual growth curves, simple change plots should be examined to ascertain the shape of observed patterns in the data. With 245 children, 245 plots could be

Table 5.1 Variables for the children of the NLSY79 longitudinal models.

Variable	*Description*
Level 1	
MATH	PIAT raw scores
TIME	Assessment collection point (coded 0,2,4,6,8)
TIMESQ	Squared "Time" variable
Level 2	
SEX01	Recoded gender (0 = male; 1 = female)
MOMEDGRA	Number of years of schooling (continuous, 0 represents 0 years)

Note: NLSY79 = National Longitudinal Survey of Youth 1979; PIAT = Peabody Individual Achievement Test.

Table 5.2 Means (standard deviations) and sample sizes on the Peabody Individual Achievement Test (PIAT) MATH across five waves, gender, and combined sample.

Group	*Time 1*	*Time 2*	*Time 3*	*Time 4*	*Time 5*	*Overall*
Male (n = 118)	14.30 (4.98)	32.58 (9.59)	47.35 (10.25)	55 (10.36)	61.38 (11.07)	42.12 (7.69)
Female (n = 127)	15.91 (6.40)	31.24 (10.22)	46.72 (9.91)	53.45 (11.24)	57.46 (11.47)	40.96 (8.33)
Combined Sample (n = 245)	15.13 (5.80)	31.89 (9.92)	47.02 (10.06)	54.20 (10.83)	59.35 (11.43)	41.52 (8.04)

formed; we present a sample of only the first six children in separate plots in Figure 5.1. These simple plots indicate a general positive trend over time, with rapid growth in mathematics over the initial three to four year period that begins to taper off towards the end of the study. Thus, while a linear (straight-line) model might provide a parsimonious summary of the pattern of change over time, a quadratic model may capture the curvilinear trend evident in the plots as well as provide better fit to the data. These are issues that we examine in a principled way in the sections to follow. Before we demonstrate our stepped approach, we set some groundwork for interpreting growth models through a simple linear model.

5.2.3 The General Linear Model for Growth and the Metric for Time

Within a multilevel framework, one of the simplest polynomial models for growth is the unconditional linear growth model, which we briefly review here and through demonstration in a later section. In this model at level 1, an individual i' s score at time t can be predicted by an intercept, π_{0i}, and a linear growth slope, π_{1i}. The subscript i indicates that the model estimates a separate intercept and a separate linear growth slope for each person in the sample. Therefore, each individual in the sample can have a unique linear growth rate and a unique intercept. A set of two level-two equations are used to predict the intercepts and the linear growth slopes, π_{0i} and π_{1i}, respectively. Thus, the unconditional linear growth model for Y_{ti}, representing the i^{th} person's response at the t^{th} occasion, is:

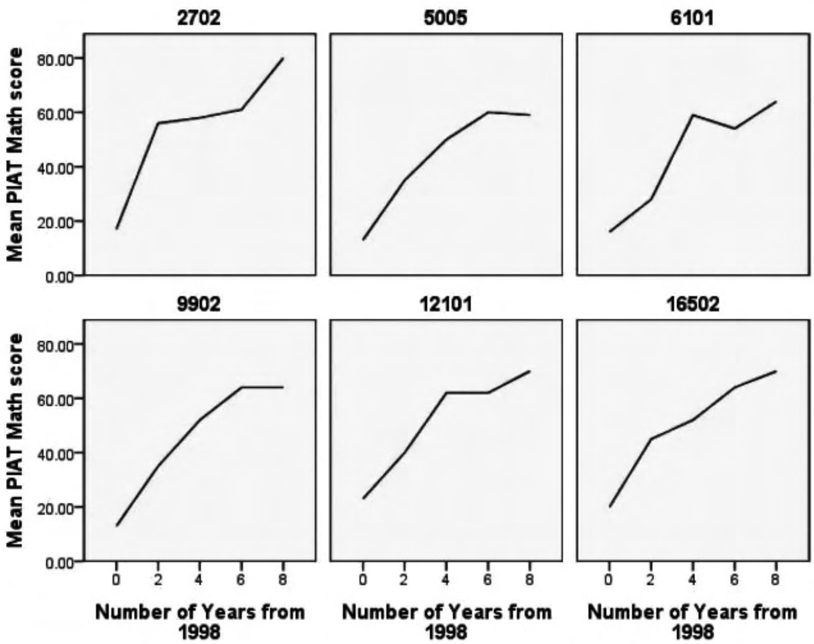


Figure 5.1 Growth plots for sample of first 6 National Longitudinal Survey of Youth 1979 (NLSY79) students.

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + e_{ti};$$

Level 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}.$$

In this model e_{ti} represents the occasion-level residuals, that is, the difference between person i's score at time t from his or her model predicted score $\pi_{0i} + \pi_{1i}$ $(TIME)_{ti}$) at that time point. The person-level residuals, r_{0i} and r_{1i}, are random effects that represent the deviation of person i's intercept and slope from an overall intercept (β_{00}) and overall slope (β_{10}). In the unconditional linear growth model, the variance/covariance matrix for the person-level residuals r_{0i} and r_{1i} provides estimates of the between-person variability in the intercepts and the slopes, τ_{00} and τ_{11}, respectively, and the covariance between the slopes and the intercepts, τ_{01}. That is, var(r_{0i}) = τ_{00}, var(r_{1i}) = τ_{11}, and $cov(r_{0i}, r_{1i}) = \tau_{01}$.

Because the intercept π_{0i} represents the value of Y_{ti} when time = 0, the interpretation of the intercept depends upon the way in which time is coded. Commonly, analysts code time so that the initial time point equals 0 and subsequent values for time represent the increase in minutes, days, months, years, etc., for the other time points. When time is coded in this manner, the intercept represents person i's initial status or his/her score at the start of the study and the between-person variance in these intercepts τ_{00} can be interpreted as the between-person variability in initial status. In other words, the tau parameter captures how much variation there is between people in terms of their predicted initial values on the dependent variable. The between-person variance in the time slopes τ_{11} represents the variability between people in terms of their estimated linear growth rates. When standardized, the covariance parameter τ_{01} in the unconditional linear growth model represents the correlation between people's initial scores (or intercepts) and their growth rates.

Although the most common way to center time is at initial status (which is how we coded time for the NLSY data, measured in years beginning at 0 years for the baseline assessment), there are a variety of other ways to center time. For example, time can be centered at any specific wave, age, or time-point. Centering time provides a meaningful and standard intercept across participants. In the NLSY example, time could be centered at the end of the study period, which is 8 years beyond the baseline assessment. To do this, we would subtract 8 from the variable TIME, and the intercept then becomes the expected value for math score when time = 0, or 8 years after the baseline assessment. Therefore, each individual's intercept is centered at the same time-point. Although many options exist for coding and centering time, researchers must make their choice carefully. It is important not to center the time variable outside the range of data for the participants in the sample. Thus, if a researcher collected students' math scores at the beginning of first, second, and third grades, he or she would not want to center time at the beginning of kindergarten.

Before turning to our primary demonstrations, we caution our readers to carefully consider the metric of time not only for coding purposes, but in terms of study design as well. The data collection schedule should be consistent with the degree of change expected over the duration of the study period. For example, annual assessments of mathematics skill embodies a reasonable frequency of data collection, and the estimated linear trend (slope) across five annual occasions would represent the expected change in math skill score for each increase of 1 year from baseline. This change would likely be a meaningful quantity, more so than if data were collected each month during the same 5 years. In that instance, the expected change per month, with time coded as 0, 1, 2, . . . 60 months, may be prohibitively small or possibly close to zero (more so for even finer-grained data collection occasions), creating estimation problems within the model. Thus, the time metric must serve two purposes: First, the distance between data collection points must be sufficient to measure a desired and meaningful amount of change, and second, the coding scheme utilized uses zero to represent a point at which the researcher is interested in examining the variability of the sample.

5.2.4 Demonstration of Linear and Quadratic Growth Models—Continuous Outcome

Following the recommendations of Francis, Schatschneider, and Carlson (2000), we conduct the analyses for individual growth curves in two phases: a within-participants (unconditional) phase and a between-participants (conditional) phase. In general, the research question addressed in the within-participants (or unconditional) phase is, What is the shape of growth (in math achievement)? Or, what does change over time look like? Decisions on shape can be aided by a comparison of competing growth models (i.e., linear versus quadratic) to determine the pattern that best balances goodness of fit and parsimony. Regardless of how growth is defined, model misspecification in the unconditional phase can bias the results of all future conditional models. Thus, in our example, the goal of the unconditional phase is to select the best-fitting model of student growth in math skills.

In the second phase, referred to as the conditional phase, we build upon the best-fitting unconditional model identified in the first phase, such that predictors or covariates are included in order to predict differences in within-participant growth trajectories. For example, do boys and girls share the same pattern of change over time? Research questions that can be addressed in the conditional phase vary much more widely than those in the unconditional phase and are discussed individually in later sections. In general, however, the goal of the conditional phase involves the prediction of growth patterns from individual (i.e., gender) or group (i.e., classroom or school) characteristics.

First, we describe the steps necessary to complete the first phase (i.e., within-participants or unconditional phase). In order to determine the model that accurately portrays growth in our example, we investigate and assess the goodness of fit of a series of different models of growth. These forms of growth range from most to least parsimonious, with the more complex models providing the most information about growth. Additionally, as there are five time points in the present example, analyses are possible with both linear and quadratic models.

Though technically we could also estimate a cubic (third power) and quartic (fourth power) function, these require many degrees of freedom, which will subtract from our ability to model covariates and random effects. It is important that the researcher thinks carefully when selecting the form of growth. If the goal is to examine the true form of growth, then either of these (cubic or quartic) is a viable option, but model parsimony should also be considered. For some data sets, a quartic function may yield a significantly better fit to the data than a quadratic model, but the added complexity may be overshadowed by the simpler model and be cumbersome for interpretation and generalization.

To build understanding of models for linear and quadratic change, we will investigate the following increasingly more complex models for growth: (A) random intercept, no slope (empty model); (B) random intercept, fixed slope (linear); (C) random intercept, random slope (linear); (D) random intercept, fixed slope, fixed curvature (quadratic); (E) random intercept, random slope, fixed curvature (quadratic); and (F) random intercept, random slope, random curvature (quadratic). We also provide a demonstration of the steps necessary to model both linear and quadratic growth with predictors (i.e., between-participants or conditional phase).

In the examples to follow, full maximum likelihood estimation (FEML, also sometimes called full information maximum likelihood or FIML) is used to allow for model-fit comparisons using the deviance statistics. Another common choice for estimation of IGCs is restricted maximum likelihood (REML). When sample sizes are large, parameter estimates will be similar for these two methods (Raudenbush & Bryk, 2002). In situations in which sample sizes are small, FEML estimates of a growth model's variance components will tend to be smaller than those estimated through REML; thus, FEML is generally preferred for larger samples. In terms of comparing deviances of nested models, REML is more restrictive than is FEML; models compared based on REML must be nested only through their random effects, whereas FEML models can be nested by either their fixed or random components. Because not all of our models are solely nested in terms of their random effects, and given our sample size of 245 students, we chose FEML in order to statistically compare the difference in deviances among our sets of nested models. This process is demonstrated in the following sections, after briefly demonstrating model building through the HLM program.

5.3 Getting Started with HLM

In this section, we provide a brief description of how data files need to be structured for HLM analyses, the process for reading data into HLM the program, and how to build a basic linear growth model. For this chapter, we modeled growth using HLM version 7 (HLMv7), though much of the instruction here is consistent with earlier versions of the program. We present a cursory overview of the program and encourage the reader to consult the book *Hierarchical Linear Models* (Raudenbush & Bryk, 2002) or the program's technical manual (Raudenbush et al., 2011) for additional information on model fitting.

5.3.1 Data Setup

First, we are assuming that each person has been measured on the same assessment at multiple time points. In the models examined here, the data are organized in univariate or long form. This means that the outcome and predictor(s) will each be structured as a single column (i.e., one column per variable), and each person will have multiple rows, with one row for each time point. At least three variables are required to fit a hierarchical growth model: one representing the person identifier, one representing the occasion or time-point identifier, and one representing the assessment score at that time point. For example, in the NLSY example data set that accompanies this chapter (Growth_NLSY.sav) each child (*CHILDID*) is represented up to 5 times (once for each measurement occasion), time is represented by one index variable (*TIME*), and the outcome score is represented by one variable (*MATH*). We also have a quadratic term representing $TIME^2$, which in the data file is identified as *TIMESQ*.

For assessing individual growth in the HLM framework, variables that change over time such as the time variable and the outcome of interest are considered to be level-1 variables, and those that vary at the person level are considered to be level-2 variables (e.g., race, gender, or other person-specific characteristics). In earlier versions of the software, the HLM program traditionally required two data files, one containing the level-1 (or time-level) data and a second containing the level-2 (or person-level) data; records on each file are distinguished by each person's unique identifier (i.e., *CHILDID*). However, in more recent versions of the software these files can be combined into one larger file, with multiple rows per person containing the person's unique identifier, information for each time point, and all person-level characteristics; this is referred to as a person-period data set. Whether you choose to use one combined file or two separate files for the data, these files should be saved in an accessible folder on your computer. In our examples, we use a combined dataset.

5.3.2 Reading in Data

Once you have finalized the dataset, you can begin analyzing the data. Open the HLM software by selecting

Programs > SSI, Inc > HLM

Next, we need to import the data file. To begin importing data, select

File >Make New MDM file >Stat Package Input

This will open a dialogue box titled "Select MDM type," which asks about the analysis you intend to run. There are several options listed, and you can learn what each indicates in the HLM technical manual (Raudenbush & Bryk et al., 2011), but for the purposes of this chapter we exclusively model *time* nested within *students* (a two-level model). Select this option by toggling the button next to "HLM2," and clicking "OK."

This will open a dialogue box titled "Make MDM—HLM2," pictured in Figure 5.2. First, identify the input data type; our data is in SPSS which is the default data type shown, but other forms are permissible. Within the dialogue box, there are five steps that should be followed to read in the data, which are briefly indicated in the corresponding text boxes within Figure 5.2. For Step 1, use the "Structure of Data" box to identify that these data are "longitudinal (occasions nested within persons)." In Step 2, from the "Level-1 Specification" box, select "Browse," which will prompt you to select the data file that contains your level-1 data. In this case, we use Growth_NLSY.sav. Continuing in this step, you should next specify your missing data options. In the data file, we have no missing data, and so "No" can remain checked. If you do have missing data on one or more variables at level 1, we recommend that you select the cases to be deleted "when making MDMs" as this creates a working data set for which the samples included in each model are nested, a requirement for model comparison purposes (for more information on missing data options, see Raudenbush et al., 2011, section 2.6). Finally, after the file has been selected, click on the (now-available) "Choose Variables" button. This will open a screen listing all variables in the data set, where you will be prompted to select an ID and any level-1 variables you wish to analyze. Note that only one ID variable can be selected from this menu, and it should represent the unique person-level cases (e.g., *CHILDID*). We also need to select the outcome variable (*MATH*), and the variable representing time (*TIME*). With these selected, click "OK," which will return you to the "Make MDM" screen (Figure 5.2).

The third step is similar to the second, but the data file containing the level-2 variables is selected; in our case this is the same as above since our data is combined into one file (Growth_NLSY.sav). After selecting the data file, the person-level ID must be indicated (*CHILDID*). Any person-level covariates that you expect to examine in the model building process should also be selected here. For example, in later sections, we include gender (*SEX01*) as a predictor of children's performance, and so this variable should be selected and added to the MDM file here as well.

Fourth, save the MDM file by, first, giving it a name in the textbox on the right (complete with .mdm extension) and, second, clicking the button that reads "Save mdmt file." With the file saved, click "Make MDM file"; this will cause an output window to pop up containing descriptive information about the data files. After reviewing this window to

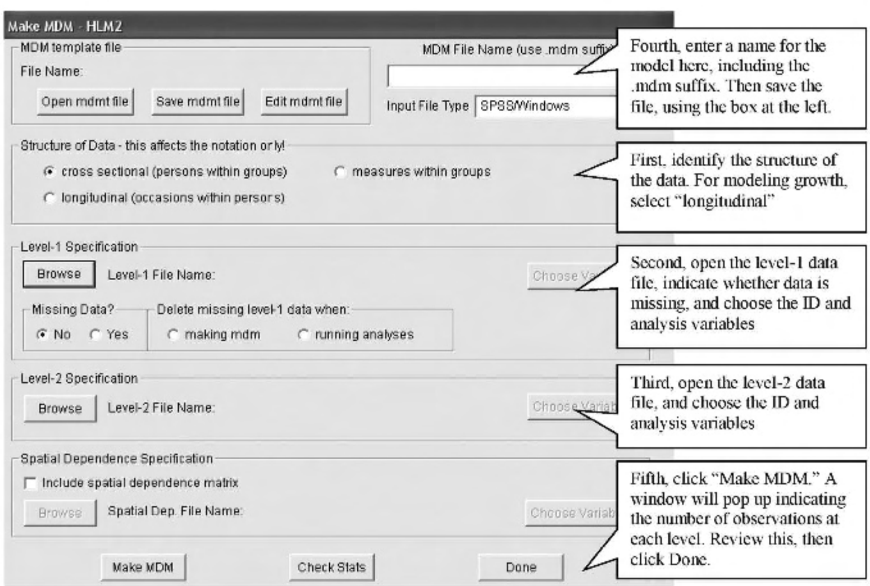


Figure 5.2 Reading in data.

ensure it is as expected (i.e., descriptive statistics and the correct number of level-1 and level-2 units), click on "Done" to move to model building.

5.3.3 Model Building

The model building dialogue box is pictured in Figure 5.3. Initially, the model screen is blank, with potential variables listed on the left side (Step 1, Figure 5.3). Before moving to the model building phase, we suggest changing the HLM default preferences by selecting "Preferences" from the File menu heading, and checking the boxes for "Show Mixed Model," "Use level subscripts," and "Create Graph Files." To begin building the model, ensure first that the tab labeled "Level-1" on the left side of the screen is highlighted (with >> <<). Then select the level-1 outcome by clicking on it. In our example, the outcome is *MATH*. Selecting *MATH* opens the small dialogue box presented in the first panel of Figure 5.3. The only option available is "Outcome variable." Once this is selected, the program automatically fills in a simple two-level model (Step 2; Figure 5.3). This model is referred to as the empty model and is Model A in our series of models demonstrated for the unconditional phase.

The next step is to build a model that accounts for growth over time. To do so, click on *TIME* on the far left, and select one of the presented options. The same menu appears as it did in Step 1 of Figure 5.3, but now more options are available. For the analyses that follow, we add *TIME* uncentered (as described above, our coding already centers *TIME* at baseline assessment). For a thorough discussion on the different ways of entering predictors into an analysis, see Raudenbush and Bryk (2002, p. 31). The model with *TIME* included is presented in Step 3; see Figure 5.3. Each of these parameters is described in detail in subsequent sections. For now, we note that the default is to constrain the person-specific slopes associated with time (π_{1i}) to be constant across children (i.e., the level-2 units). In other words, the default is to exclude the random error term associated with *TIME*, r_{1i}. Throughout this chapter, we discuss the inclusion and exclusion of these residuals. In HLMv7, random errors in a model can be added or removed simply by clicking on them at this screen (Figure 5.3; Step 4). Once you have built the model you wish to test, you can run it by clicking on "Run Analysis" from the menu bar (recall that FEML will be used for our analyses, which can be selected from the "Estimation Settings" option under the main menu heading of "Other Settings").

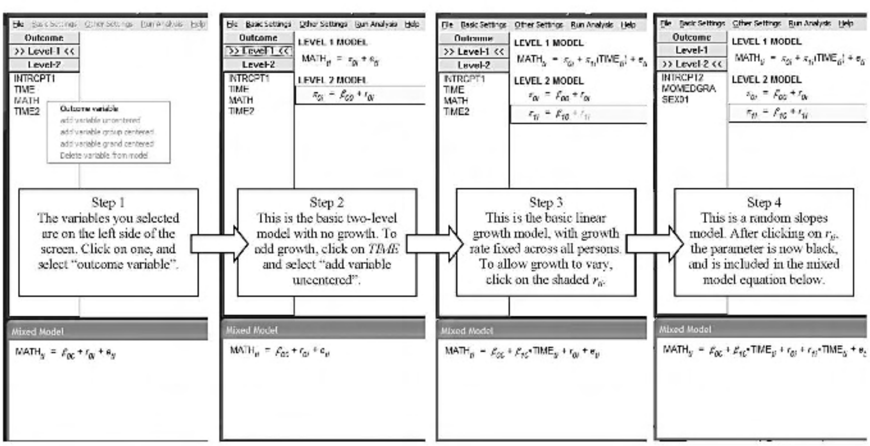


Figure 5.3 Basic model building in HLM7.

The goal of this section was to provide a brief introduction to building individual growth models using the HLM software. In the sections that follow, we begin with interpretation and discussion of the empty model and then build and discuss increasingly more complex linear and quadratic models for growth.

5.3.4 Unconditional Phase: Identifying the Model at Level 1

Recall that in the unconditional phase, the goal is to identify the best model to capture the shape of growth; this is typically accomplished through a series of model comparisons. Results and interpretations for each of the six models mentioned previously (Models A through F) and used to help identify the nature of growth for our data are presented in Table 5.3 and will be discussed in order in the following sections. From a pedagogical perspective, we believe that this intentional strategy helps to clarify differences across these models in a clear and organized fashion, particularly in terms of interpreting parameter estimates for linear and quadratic models.

5.3.4.1 Model A

The first step in identifying the most appropriate model of growth across a series of occasions involves first understanding how much variation in math scores exists *within* students and how much variation exists *between* students—ignoring time. This is accomplished by fitting a baseline model with no growth, by which we mean a model that contains random intercepts for all persons at level 1, and no slope terms (Table 5.3, model A). Note that this is the same model presented in Step 2 of Figure 5.3. The baseline model can be written for each level separately or combined as a single model. In fact, there are always two possible ways to represent hierarchical models, either with separate equations representing level 1 (time) and level 2 (students), as in our introduction, or with a mixed model, which contains coefficients from both levels in one combined equation. For the purposes of this chapter, we primarily present mixed-model equations. In the HLM program, the default is separate equations, but the mixed-model equation function can easily be activated by selecting

File > Preferences > Show Mixed Model

Doing so provides an additional window in which to view the mixed model, although the separate models are visible as well. All combined equations presented here match what HLM presents in the Mixed Model window.

The baseline model is expressed as follows:

Level 1: $MATH_{ti} = \pi_{0i} + e_{ti}$

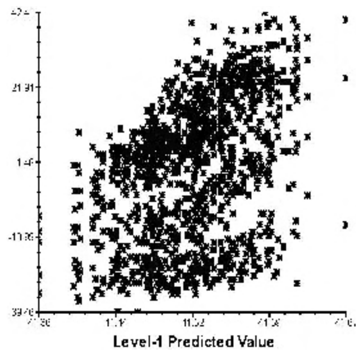
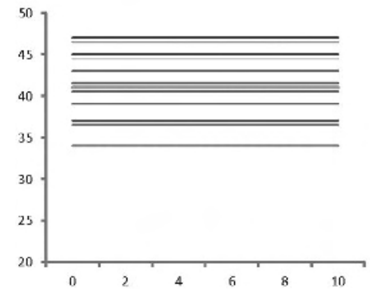
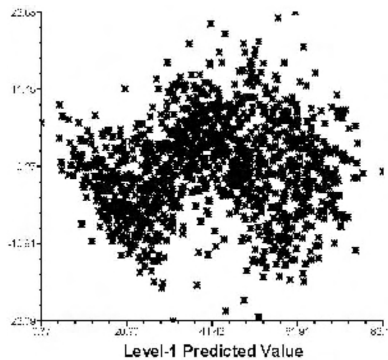
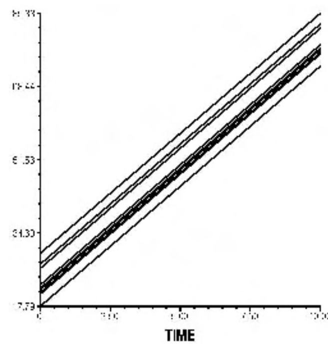
Level 2: $\pi_{0i} = \beta_{00} + r_{0i}$

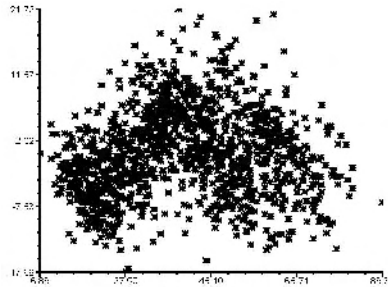
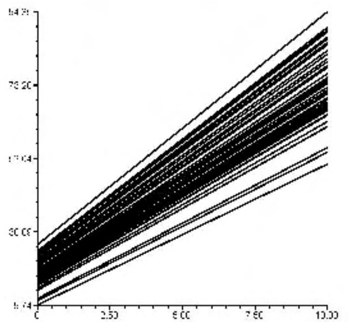
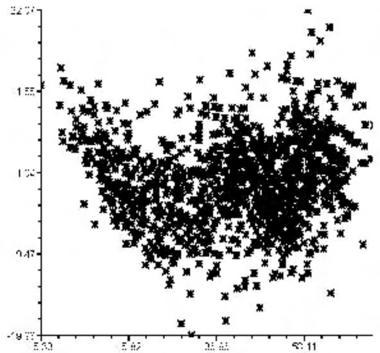
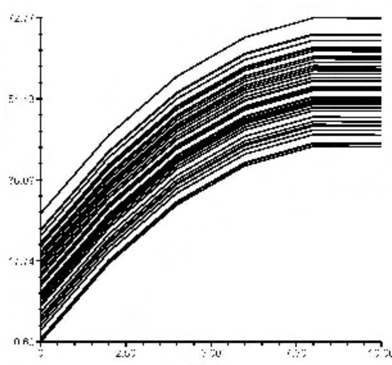
Mixed-Model or Combined Form:

$$MATH_{ti} = \beta_{00} + r_{0i} + e_{ti}$$

In this model, $MATH_{ti}$ represents the dependent variable (i.e., math score) at any time t for student i. Because there is no change over time being estimated, each person's intercept π_{0i}

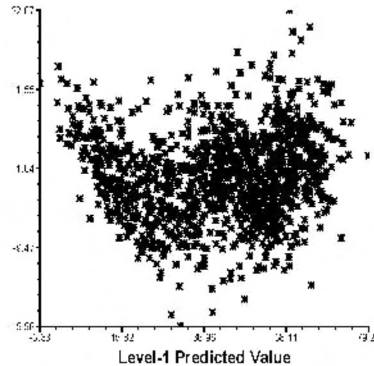
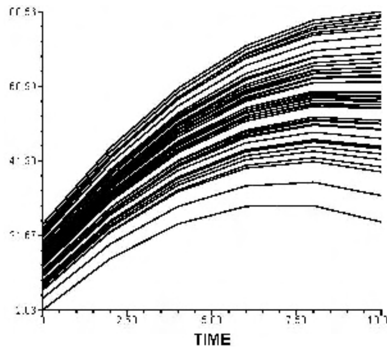
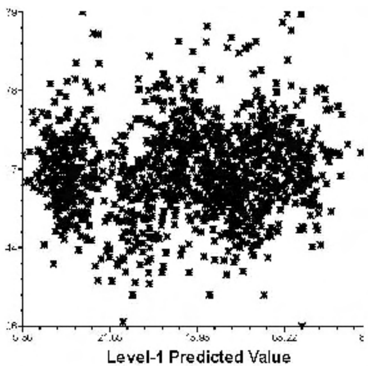
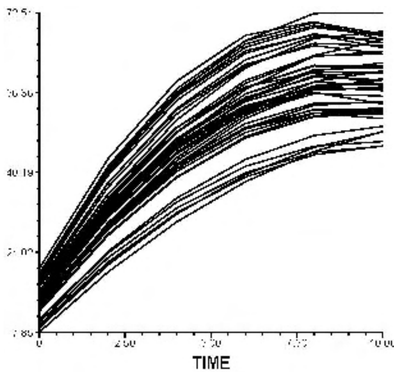
Table 5.3 Models A through F: Explanations for the unconditional phase.

Equations	*Coefficients/ Deviance*	*Residual Plots*	*Graphical Representation*	*Interpretation*
Model A Random intercept No slope $MATH_{ti} = \beta_{00} + r_{0i} + e_{ti}$	$\beta_{00} = 41.52^{*}$ **Variance** $\tau_{00} = 0.47^{ns}$ $\sigma^2 = 355.08^{*}$ **Deviance** = 10,669.65 **Parameters** = 3			The estimated mean math score for students is 41.52, which is significantly different from zero. The variance component (0.47) indicates that students' average math score ignoring differences by time varies, although not significantly.
Model B Random intercept Fixed slope $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + r_{0i} + e_{ti}$	$\beta_{00} = 19.37^{*}$ $\beta_{10} = 5.54^{*}$ **Variance** $\tau_{00} = 52.85^{*}$ $\sigma^2 = 57.33^{*}$ **Deviance** = 8,856.85 **Parameters** = 4			The average math score for students at the first time point (baseline) was 19.37. Students increase by a significant 5.54 points every year. The variance component (52.85) signifies that students significantly vary in math achievement at the first time point (baseline).

Equations	*Coefficients/Deviance*	*Residual Plots*	*Graphical Representation*	*Interpretation*
Model C Random intercept Random slope $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + r_{0i} + r_{1i}$ $*TIME_{ti} + e_{ti}$	$\beta_{00} = 19.37*$ $\beta_{10} = 5.54*$ **Variance** $\tau_{00} = 20.93*$ $\tau_{11} = 0.49^{ns}$ $\sigma^2 = 52.23*$ **Deviance** = 8,780.98 **Parameters** = 6			The average math score for students at the first time point was 19.37. Students increase 5.54 points every year. The variance component for intercepts (20.93) signifies that students significantly vary at the first time point. The variance component for slopes (0.49) demonstrates that differences in students' rate of change is small and does not significantly vary.
Model D Random intercept Fixed slope Fixed curvature $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + \beta_{20} *TIMESQ_{ti}$ $+ r_{0i} + e_{ti}$	$\beta_{00} = 14.91*$ $\beta_{10} = 9.99*$ $\beta_{20} = -0.56*$ **Variance** $\tau_{00} = 56.32*$ $\sigma^2 = 39.97*$ **Deviance** = 8,503.47 **Parameters** = 5			The average math score for students at the first time point was 14.91, with instantaneous rate of change at baseline of 9.99 points. However, change in the rate of change decreased by 0.56 each year. The variance component (56.32) signifies that students significantly vary at the first time point.

(Continued)

Table 5.3 (Continued)

Equations	*Coefficients/Deviance*	*Residual Plots*	*Graphical Representation*	*Interpretation*
Model E Random intercept Random slope Fixed curvature $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + \beta_{20}$ $*TIMESQ_{ti} + r_{0i} + r_{1i}$ $*TIME_{ti} + e_{ti}$	$\beta_{00} = 14.91*$ $\beta_{10} = 9.99*$ $\beta_{20} = -0.56*$ **Variance** $\tau_{00} = 24.28*$ $\tau_{11} = 0.61*$ $\sigma^2 = 33.90*$ **Deviance** = 8,383.99 **Parameters** = 7			The average math score for students at the first time point was 14.91, with instantaneous rate of change of 9.99 at baseline. Change in the rate of change decreased by 0.56 points each year. The variance component (24.28) signifies that students significantly vary at baseline. The variance component for linear component (0.61) demonstrates that students' instantaneous rate of change at baseline significantly varies.
Model F Random intercept Random slope Random curvature $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + \beta_{20} *TIMESQ_{ti}$ $+ r_{0i} + r_{1i} *TIME_{ti} + r_{2i}$ $*TIMESQ_{ti} + e_{ti}$	$\beta_{00} = 14.91*$ $\beta_{10} = 9.99*$ $\beta_{20} = -0.56*$ **Variance** $\tau_{00} = 14.76*$ $\tau_{11} = 3.56*$ $\tau_{22} = 0.03^{ns}$ $\sigma^2 = 31.33*$ **Deviance** = 8,348.70 **Parameters** = 10			The average math score for students at baseline was 14.91, with instantaneous rate of change at baseline of 9.99 points. Change in the rate of change decreased by 0.56 each year. The variance component (14.76) signifies that students significantly vary at baseline. The variance for the linear component (3.56) demonstrates that students' instantaneous rate of change at baseline does significantly vary. However, the variance component for the quadratic term (0.03) signifies that the change in the rate of change between students does not significantly vary.

reflects their estimated mean math score across the five time points, and β_{00} represents the estimated *overall* mean math score across all students. In terms of variability, r_{0i} represents random error between students on the overall intercept (i.e., around β_{00}), and e_{ti} represents random error within students from their own mean score. This baseline model is often referred to as an empty model or an unconditional means model, because it contains no predictors.

Estimates for the mixed-model coefficients, level-1 residual plots, graphical representation of the model, and general interpretation of the results for this empty model are presented in Table 5.3 as Model A. Mixed-model coefficients are drawn directly from the HLM output, as we show in the following. Residual plots and graphical representations of the model were created for comparison across all models. The residual plots provide a visual indication of how well the data fit each model. In the residual plots, the *X*-axis is the predicted value of math score, and the *Y*-axis is the residual score. A residual plot with no discernible pattern suggests homoscedasticity and can be used to form part of the evidence in favor of good model fit. Model A's residual plot shows poor model fit: the general trend in the plot is increasing, suggesting a systematic relationship between the predicted values and the residuals that is not adequately captured by this empty model.

In the graphical representations of the model presented in Table 5.3, the *X*-axis represents time (in years since 1998), and the *Y*-axis represents the estimated math score. Each line indicates the model predictions for an individual student (a random 10% of the sample are plotted). Note that in the graphical representation of Model A, all of the lines have no gradation; they do not change across time. This is because our empty model has not allowed for change over time; instead, one predicted score (i.e., their estimated mean) is generated across all available data for an individual, ignoring potential patterns due to time. This model allows for intercepts to randomly vary between students, but the predictions over time do not vary within individuals; the prediction is the same regardless of time point. We note that these level-1 estimates are not determined through ordinary least squares methods, but through a weighted estimation process called empirical Bayes (EB) estimation, which has important implications for examining and interpreting results from this empty model. EB estimates are also called "shrinkage" estimators, such that each student's individual prediction tends to be "shrunken"—or weighted—towards the overall mean of the sample (Raudenbush & Bryk, 2002).

Portions of the HLM output from the model are provided in the following:

```
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                    Standard             Approx.
   Fixed Effect         Coefficient  Error     T-ratio   d.f.      P-value
 ----------------------------------------------------------------------------
For      INTRCPT1, P0
   INTRCPT2, B00          41.516735   0.540168 76.859      244       0.000
 ----------------------------------------------------------------------------


Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect  Standard            Variance    df    Chi-square   P-value
               Deviation           Component
-----------------------------------------------------------------------------
INTRCPT1,      r0        0.68522      0.46952   244    221.88437    >.500
 level-1,      e        18.84366    355.08363
-----------------------------------------------------------------------------
```

```
Statistics for the current model
 -------------------------------------------------------------------
Deviance                          = 10669.648371
Number of estimated parameters = 3
```

In the Final Estimation of Fixed Effects table, the coefficient for the intercept can be found ($\beta_{00} = 41.52$, $p < .001$ in this model). This coefficient signifies that the estimated overall or grand mean math score for all students is 41.52 and, according to its corresponding t test is significantly different from 0. These values are also presented in the Coefficients column of Table 5.3 (Model A). Fixed effects results from an additional table labeled the Final Estimation of Fixed Effects with Robust Standard Errors (output not shown) are available in the output but in general are not used for interpretation. Robust standard error values can be compared to the multilevel standard errors above, and large differences in these values signify a misspecified model (see Raudenbush and Bryk [2002] and Maas and Hox [2004], for in-depth discussion of this issue). These issues are beyond the context of our current chapter and thus we present and discuss the multilevel standard errors as in the preceding table.

Next, in the Final Estimation of Variance Components table, the variance for the random error *between* students on the intercept can be found on the line labeled $r0$ which contains the information for the r_{0i} : $var(r_{0i}) = \tau_{00} = 0.47$, $p > .50$. The variance for the random error within students is found on the next line labeled e : $var(e_{ti}) = \sigma^2 = 355.08$. The within-person variance is quite large[1] while between-person differences on the intercepts is quite small, but recall that differences in math achievement due to time are not yet included in this empty model.

Although this model does not describe growth, it is useful because it allows for the partitioning of between to total variation also known in HLM as the intra-class correlation or ICC. The calculation of the ICC is based on the empty model and involves dividing the variance component from level 2 (τ_{00}) by the total variance in the model ($\tau_{00} + \sigma^2$). In Model A, $\sigma^2 = 355.36$, and $\tau_{00} = 0.47$. This calculation finds that the ICC is very small (.001), suggesting that—for this model—most of the variance in math scores is due to variation within students across occasions (99.9%) but not between students. Since most of the variance lies within persons, occasion-level predictors (such as time) are needed to account for these intra-individual differences. Without accounting for time, there is so much variability within children that very little variance seems to be due to child differences. This attribution of variance between versus within students can be re-examined by comparing the ICC for this unconditional model and the residual ICC (Snijders & Bosker, 2012) for a model in which time is included as a fixed effect (Model B).

The graph of the residuals for this model (Table 5.3) reveals a positive linear trend, suggesting that this model is not the best fit for the data. With a well-fitting model, we would expect to see no pattern to the residuals. When a discernible pattern is observed, there is an indication that a more appropriate model should be investigated, and it is quite likely that the positive trend in these residuals may be eliminated once the effect of time is incorporated into the model. We examine that possibility in our next model.

Finally, the deviance and corresponding number of estimated parameters can be found at the end of the output (deviance = 10,669.65, number of parameters = 3 in this model). The number of parameters includes the coefficients at level 2 plus all variances/covariances: β_{00}, τ_{00}, and σ^2. These values can be used for model comparisons and we describe this process after presenting and interpreting our next unconditional model.

5.3.4.2 Model B

Our next step is to search for a model that adequately describes the patterns observed in Figure 5.1. We begin by considering the simplest form of a linear growth model:

$$MATH_{ti} = \beta_{00} + \beta_{10} * TIME_{ti} + r_{0i} + e_{ti}$$

where $MATH_{ti}$ is the value of the dependent variable (i.e., math score) at any time t for student i, β_{00} represents the estimated mean student score when *time* = 0, β_{10} represents the estimated linear rate of growth, or the amount of change expected as *time* increases by one year, r_{0i} represents the random error between students on the intercept, and e_{ti} represents random error within students after accounting for time. Note that this is the same model presented in Step 3 of Figure 5.3.

Recall that in this data set *time* is coded as the number of years that have elapsed since 1998, such that 0 represents 1998, when students entered the study at age 6. This first linear model includes a random intercept (i.e., allowing children to vary in math scores at baseline, i.e., age 6 when *time* = 0) but a fixed slope (i.e., in this model children are assumed to grow in a linear fashion and at the same rate). Examining the graphical representation in Table 5.3 (Model B), note that each student's predicted trajectory is allowed to vary in terms of their intercepts or predicted values at baseline (age 6), but students' growth lines are parallel. The parallel lines are a result of the growth rate being fixed rather than varying across students. Thus, this model imposes an assumption of equal slopes (equal growth rates) on the data.

Relevant output for this simplest linear model is presented in the following:

```
Final estimation of fixed effects:
-------------------------------------------------------------------------------
                                          Standard                 Approx.
   Fixed Effect             Coefficient   Error       T-ratio      d.f.      P-value
-------------------------------------------------------------------------------
For        INTRCPT1,    P0
    INTRCPT2, B00             19.366163   0.596754    32.456       244       0.000
For       TIME  slope,  P1
    INTRCPT2, B10              5.537143   0.076485    72.395       979       0.000
-------------------------------------------------------------------------------


Final estimation of variance components:
-------------------------------------------------------------------------------
Random Effect               Standard       Variance      df      Chi-square     P-value
                            Deviation      Component
-------------------------------------------------------------------------------
INTRCPT1,         r0          7.26983       52.85044     244     1374.29332    0.000
 level-1,         e           7.57162       57.32947
-------------------------------------------------------------------------------

Statistics for the current model
-------------------------------------------------------------------------------
Deviance                        = 8856.846644
Number of estimated parameters = 4
```

In the Final Estimation of Fixed Effects table, the coefficient for the intercept is β_{00} = 19.37 (p <.001), and the coefficient for the slope or linear rate of change is β_{10} = 5.54

(p <.001). Therefore, in this model the mean math score for students at baseline is estimated to be 19.37 points and students are expected to increase by a statistically significant 5.54 points every year. The random error between students on the intercept (τ_{00} = 52.85, p < .001) indicates significant variability between students on their initial math score. The graphical representation of the model indicates this variation in baseline math achievement and a positive linear trend over time with all lines parallel, corresponding to the fixed slope in this model. The residual plot shows evidence, however, of a clear quadratic trend (e.g., the shape is curved), suggesting that a model using linear growth is not the most appropriate fit for this data.

Next, with time included as a fixed effect we can calculate the residual ICC after accounting for linear growth. Following the formula in the previous section, the residual ICC = (52.85 / (52.85 + 57.33)) = .48. This indicates that 48% of the variance is due to between-student differences, with the remainder due to variation within students across time. Recall that the ICC for Model A suggested that .1% of the variance was attributable to differences between students. By accounting for growth over time, we have a better estimate of the proportion of variance in math scores that is due to students.

Finally, the deviance and corresponding number of estimated parameters for Model B can be found at the end of the output (deviance = 8,856.85, number of parameters = 4). Note that this deviance is much smaller than that obtained in Model A. In general, the smaller the deviance, the better the fit, but this reduction must be balanced against the number of parameters used to obtain the smaller deviance. In the following, we describe how the deviances can be used to statistically compare the fit of two nested models.

5.3.4.3 Model Comparisons

As we have noted, models can be compared in the following ways: (a) through examination of residual plots and (b) through the deviance test which is a comparative index of goodness of fit versus parsimony (Singer & Willett, 2003). Additional model comparison strategies are outlined in McCoach and Black (2008).

First, as residuals represent the difference between actual and predicted scores, residual plots visually demonstrate how well a model represents the data. Thus, a model that fits well will tend to have residuals close to zero and an absence of systematic patterns in the residuals.

Second, for a given set of competing nested models of the data, the selection of a final model may be determined by balancing differences in model fit indices with model complexity as measured through the number of fixed and random effects. This is typically accomplished by comparing deviances. In order to do so, the models must be nested, and the deviance from the less complex model (fewer parameters) is compared to the more complex model. There are two ways models may be considered nested: in terms of solely their random effects or in terms of both random and fixed effects. Two models in which a particular predictor included at level 1 is treated as fixed in one model and allowed to vary at random in the second model are nested in their random effects; alternatively, models may also be nested in terms of their random and fixed effects (i.e., a new variable may be added [fixed effect] and allowed to vary at random across individuals [random effect]). This distinction is important, because models estimated using REML yield a deviance statistic that should only be used to compare two models nested solely in terms of their random effects. Two models estimated through FEML can be compared via their deviance statistics if they are nested in terms of their random and/or fixed effects. The critical criterion is nestedness.

For their appropriate form of nestedness and estimation strategy, the difference in deviances for two nested models follows a chi-square distribution with degrees of freedom equal to the difference in number of parameters between the two models. This comparison assesses the reduction in deviance attributable to the additional parameters in the more complex model. If the reduction in deviance is not statistically significant relative to the chi-square critical value and associated degrees of freedom, then the two models yield similar fit and the simpler model would be preferred. If the reduction in deviance is statistically significant, then there is support for the more complex model providing better fit to the data (i.e., significant reduction in deviance). Note that the preference for either model based on this chi-square test does not confirm either model as a good-fitting model; the comparison is always and only relative to the two particular models being compared.

We used FEML for all models here to facilitate comparisons. In comparing Model A (deviance = 10669.65, 3 parameters) to Model B (deviance = 8,856.85, 4 parameters), the chi-square difference test yields the following reduction in deviance: χ^2 (1) = 1,812.8, which represents a significant reduction in deviance given the addition of one parameter, with $p < .0001$. Thus, Model B, which includes time as a fixed effect at level 1 is preferred over Model A. Review of the residual plots supports this choice as well.

5.3.4.4 Model C

Although we expected that Model B would be preferred over Model A, the residual plot for Model B is indicative of a clear quadratic trend. Before jumping to a model that incorporates this suspected quadratic trend, we examine a model with random—rather than fixed—slopes, in keeping with our stepped approach to understanding the patterns of change at level 1. This next linear model allows the intercept to vary between students as well as the linear slope (rate of change). The model is written as

$$MATH_{ti} = \beta_{00} + \beta_{10} {}^{*}TIME_{ti} + r_{0i} + r_{1i} {}^{*}TIME_{ti} + e_{ti}.$$

As before, $MATH_{ti}$ is the dependent variable (i.e., math score) at any time t for student i, β_{00} represents average student score at baseline (time = 0, or age 6), β_{10} represents the average linear rate of growth, or the amount of change expected as *time* increases by one year, r_{0i} represents the random effects (error) between students on the intercept, r_{1i} represents the random effects (error) between students on the slope, and e_{ti} represents random error within students after accounting for time. Note that we demonstrated how to fit this model in HLMv7 in previous sections; it is the same model presented in Step 4 of Figure 5.3.

The results for this more complex linear model are summarized in Table 5.3 as Model C; corresponding to the previous sections, these are obtained from the fixed and random effects tables of the output (the output is not included here to conserve space). Results show that the average math score for students at the first time point was estimated as 19.37 ($p < .001$) points and students increased an average of 5.54 ($p < .001$) points each year. Thus, the general pattern of change for these particular data is the same as in Model B, but here the slope of 5.54 is an *average* across students; some students may have a stronger slope, and some may have a weaker slope. The variance component for the intercept ($\tau_{01} = 20.93$, $p < .001$) signifies that students significantly vary at the first time point. Note that this is considerably smaller than in Model B. The variance component for the slope ($\tau_{11} = 0.49$, $p = .098$) was not statistically significant, and indicates that students' linear rate of change does not differ substantially across students. However, the chi-square difference test comparing Model B to Model C suggests that including time as a random effect does yield a significant reduction in deviance, with $\chi^2(2) = 75.87$, $p < .0001$. The chi-square difference test is known to be superior to the direct assessment of the variance component noted above (Singer & Willet, 2003), despite the fact that those tests are supplied as the default in HLM and other statistical packages. Note that the number of parameters estimated in Model C is six, which includes the estimated covariance between the level-two residuals for the intercept and slope.

The graphical representation of Model C shows variability in initial status (baseline math achievement) with a positive linear slope for each student, but note that all lines have slightly different slopes, because slopes were allowed to vary in this model. Similar to Model B, however, the graph of the residuals for Model C shows a clear quadratic trend, suggesting once more that a model based on linear growth may not provide the most appropriate depiction of these data, and instead, we should consider fitting a quadratic term to the model. We turn to that approach next.

5.3.4.5 Models D and E

Continuing in the unconditional phase, our next set of models include quadratic forms of growth. Therefore, it will be necessary to add a quadratic growth parameter to the model. To create this quadratic parameter, the *time* variable is squared. In the example data set, this variable is called *timeSQ*. The squared variable is then added to the model already containing the linear component, *time*. Including the *timeSQ* parameter allows for curvature in the growth model capturing the potential curvilinear change in math growth. Another way to think of the inclusion of *timeSQ* is that the effect of time now *depends on* or *varies by* time. Thus, the coefficient for the quadratic term represents either positive or increasing change in the effect of change over time, or negative and decreasing change in the effect of change over time, across the overall growth trajectory.

When a quadratic term is first included in the model, it is typically treated as a fixed effect (i.e., each person is forced to have the same rate of curvature—either increasing or decreasing). Given this constraint, there are two possible ways to treat the linear term. If the previous step of the model-building process (the linear phase) indicated that the linear term should be fixed, then the linear model term would continue to be fixed in the initial quadratic modeling phase. However, because the previous step in our model-building process indicated that slopes should be random, we added the quadratic function to a random slopes model. To illustrate differences in interpretation, both models are presented as Models D (fixed linear term) and E (random linear term) in Table 5.3.

Model E thus includes both a linear and a quadratic growth term, allowing intercepts to vary between students, as well as the slope (linear term), while the curvature (quadratic term) is fixed (held constant) across students. This model can be written as

$$MATH_{ti} = \beta_{00} + \beta_{10} * TIME_{ti} + \beta_{20} * TIMESQ_{ti} + r_{0i} + r_{1i} * TIME_{ti} + e_{ti}$$

with $MATH_{ti}$ representing the math score at any time t for student i and β_{00} representing average student score at baseline (*time* = 0). The linear and quadratic components are interpreted much differently than in the previous linear models. To understand these interpretations, we need to know a little bit about the nature of curves in general.

All quadratic functions have either a maximum (peak) or a minimum (trough); the slope of a line tangent to this turning point—when the curve changes from increasing to decreasing or vice versa—is flat and will always be zero. However, for growth models, the time frame of our study may not necessarily cover this turning point. For example, our graphs of Models D and E capture only one portion of a quadratic function. Quadratic models can be concave down, opening downwards with a maximum peak, or concave up, opening upwards with a minimum trough. Different sections of either shape could be captured within the range of time represented by the data. In addition to shape, curves can also be understood in terms of lines drawn tangent to the curve, across the study period. Looking at the graphs in Table 5.3 for Models D and E, the slopes of the tangent lines to these curves when $TIME = 0$ are always positive. Over time, the slopes for these tangent lines are becoming gradually weaker, even becoming negative for some students. The parameter estimates for our models describe these patterns.

At any time point, we refer to the slope of the line tangent to the curve as the *instantaneous rate of change* at that time point. Principles of calculus tell us how to find the slope of the line tangent to a curve: we take the derivative of the curvature function with respect to time. For a quadratic function of growth over time, this becomes: $\frac{\partial y}{\partial t} = \beta_{10} + 2 * \beta_{20} * time_{ti}$. We can refer to this as the simple slope of the tangent line for any given value for $time_{ti}$. When $time = 0$, the simple slope of the tangent line is β_{10}. Thus, in the quadratic model, β_{10} represents the overall linear component for the growth trajectory and estimates the average instantaneous rate of growth when $time = 0$ (or for wherever *time* is centered). If β_{10} is positive, the tangent lines are increasing right at the point where $time = 0$; if β_{10} is negative, the individual growth curves tend to be decreasing when $time = 0$.

Now, how do we interpret β_{20}, the coefficient for *timeSQ* in our quadratic model? Looking back at our simple slope expression, β_{20} represents half of the change in the simple slopes as *time* increases by one unit (it is half of the change, because $2 * \beta_{20}$ represents the entire expected change in the simple slopes via our formula). The quadratic parameter informs us as to how fast or slow the simple slopes are changing over time and is often referred to as the *curvature parameter* (Singer & Willett, 2003).

If β_{20} is negative, the simple slopes are always decreasing or becoming more negative; if β_{20} is positive, the simple slopes are always increasing or becoming more positive. If β_{20} is large and statistically significant, the effect of time on rate of change over time is strongly and significantly increasing or decreasing, and the curvature parameter adds to our understanding of the data. If the curvature parameter is positive, the overall trend is concave up (with a minimum at the bottom of the curve); if the curvature parameter is negative, the overall trend is concave down (with a maximum at the top of the curve). Within the time range of the sample data, any portion of the curve might be represented; model graphs as we have provided in Table 5.3 are useful in understanding the interpretation of the parameters in polynomial models.

For Model E with the curvature parameter assumed to be fixed across all students, the random effects r_{0i} represents the random error between students on the intercept, r_{1i} represents the random error between students on the linear term, and e_{ti} represents random error within students after accounting for quadratic change.

The results for this model are presented in Table 5.3 as Model E. The average math score for students at the first time point was 14.91 ($p < .001$) points, with an instantaneous rate of change at baseline on average of 9.99 ($p < .001$) points; this means that at initial status, the curves are trending upward (positive simple slopes). The negative coefficient associated with the quadratic term indicates that the effect of time is slowing somewhat with each passing year, and the model graphs should be concave down in shape. The variance component for the intercept ($\tau_{00} = 24.28$) signifies that students significantly vary at baseline. The variance component for the linear trend ($\tau_{11} = 0.61$) indicates that students' instantaneous rate of change at baseline also varies significantly; some may be stronger, and some may be weaker.

The curves in the graphical representation of Model E each have a different overall trajectory (as the linear component was allowed to vary across students) but a fixed rate of curvature—in this case, deceleration; that is, the change in the rate of change slows down at the same rate for all students. Examining the residual plots for Model E, the graph of the residuals shows that the clear curvilinear pattern observed in previous models is beginning to disappear and the graph of residuals appears more random. Finally, for this model, the deviance is 8,383.99 and the number of parameters is seven. We compare this to the deviance for our next model.

5.3.4.6 Model F

As a final step in our unconditional phase, we fit the most complex form of quadratic model, which allows both the linear and the quadratic components to vary. This model can be written as

$$MATH_{ti} = \beta_{00} + \beta_{10} *TIME_{ti} + \beta_{20} *TIMESQ_{ti} + r_{0i} + r_{1i} *TIME_{ti} + r_{2i} *TIMESQ_{ti} + e_{ti}.$$

Parameter estimates are interpreted similar to Model E, with the exception that the curvature parameter—representing how dramatically the curve changes over time—is allowed to vary across people.

The results are presented in Table 5.3 as Model F, and show that the average math score for students at baseline was 14.91 ($p < .001$) points, with an instantaneous rate of change at baseline of 9.99 ($p < .001$) points. The negative coefficient associated with the quadratic term once again indicates that the effect of time is slowing somewhat with each passing year, and the model graphs should be concave down in shape (the simple slopes are leveling off over time). The variance component for the intercept ($\tau_{00} = 14.76$, $p < .001$) indicates significant variation in baseline math scores. The variance component for the linear component ($\tau_{11} = 3.56$, $p < .05$) also indicates significant variability between students in their instantaneous rate of change at baseline. However, the variance component for the quadratic term ($\tau_{22} = .03$, $p = .327$) indicates that differences in the rate of curvature are small and thus do not significantly vary between students. Nonetheless, it is instructive to compare the deviances between Model F and Model E and to examine the residual plots and graphical representations for a model with a varying curvature parameter. The residual plot illustrates that the previously observed quadratic pattern in the residuals has disappeared, which is what we expect to observe with a model that appropriately captures a quadratic trend. The graphical representation of the model still includes lines with a slight curvature, but notice that the curve of some lines is flatter than the curve of others; this is a result of allowing the slopes and quadratic terms for each of the students to individually vary.

Finally, the deviance for this model is 8348.70, with ten parameters. Comparing Model E to Model F, χ^2 (3) = 35.29, $p < .01$. Thus, Model F is retained as our preferred model of unconditional growth for these data.

5.3.5 Conditional Phase—Adding Person-Level Predictors

After the form of growth has been identified in the unconditional phase of model building, the researcher can then focus on the conditional phase. Of fundamental importance to the process of modeling growth is the understanding that the conditional phase should only be started after the model determined to best represent the data has been identified. In other words, covariates should only be added to level 2 of the model after defining the form of growth through the unconditional phase.

In the conditional phase, predictors[2] of the randomly varying intercepts or other individual growth parameters from level 1 of the model can be included. For example, a researcher may hypothesize that where growth begins (initial status) is related to the gender of the person being tested, or that growth is faster for students from families with higher socioeconomic status (SES). The inclusion of such level-2 or person-level predictors is driven in part by the research question and in part by their impact on model fit. In two-level growth modeling, where multiple time points are nested within each child or person, most predictors of interest will be included at the person-level (an exception is the use of time-varying covariates at level 1; see McCoach & Kaniskan [2010] for examples of these kinds of predictors). Level 2 predictors are used to investigate whether person-level characteristics are related to variations in the patterns of growth identified through the unconditional modeling phase.

Predictors can be either categorical or continuous, and just as with standard ordinary least squares regression, the inclusion of a predictor changes the interpretation of the other terms within the model. For models that contain interaction terms or powers of predictors, such as the quadratic models considered in the previous section, or when the interpretation of the intercept of a model is an important goal of the research, it is often reasonable to include covariates that are coded in such a way as to contain a meaningful zero. Similar to our method for coding time, this process is called centering and is typically done for continuous covariates. In some situations (e.g., dichotomous data), centering can be used for categorical predictors, although it is more common to simply use dummy codes or another scheme in which 0 retains meaning. For example, in the NLSY data set, gender was originally coded such that 1 represented male and 2 represented female. We recoded this gender variable so that 0 represents male and 1 represents female and named this new variable *sex01*. Because this variable now has an inherently meaningful interpretation for the value 0, we can include it as an uncentered predictor. In HLMv7, this can easily be done by selecting "Add variable uncentered" when clicking to include it in the equation.

An example of a continuous predictor from the NLSY is mother's education level (momedgra). Just as with the categorical predictor, it is important that continuous predictors be coded in such a way as to have a meaningful zero point and that participants in the study have data at that zero point. In the NLSY data set, for example, mother's education level is coded such that 0 represents no schooling at all (1 represents first grade only, 2 represents second grade only, etc.). A check of the frequency distribution for momedgra reveals that 0.4% of mothers (n = 6) reported having zero years of education, with the next lowest reported years of education being sixth grade (n = 12, 0.8%). As limited data exist at the lowest end of this distribution, the zero, though meaningful, is not very representative of the overall sample.

If a variable does not have a meaningful or reliable zero, researchers can consider using grand mean centering when including the variable as a level-2 predictor in multilevel growth models. Under grand mean centering, the mean of that variable for all level-2 units (i.e., persons or children) is subtracted from every corresponding level-2 score. This process facilitates interpretation of model parameters and guards against extrapolation bias. Grand mean centering within HLM7 can be done by selecting "Add variable grand-mean centered" when clicking to include it in any equation.

For the purposes of demonstration, we include a comprehensive account of potential conditional models in Table 5.4, which includes the equations, estimates, graphical depictions, and interpretations of a conditional phase for each model presented in Table 5.3.

Table 5.4 Models A1 through CI (linear) and D1 through F2 (quadratic): Explanations with student-level predictors.

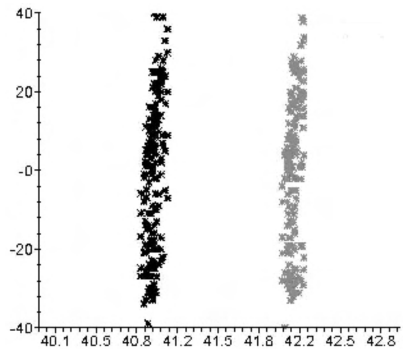
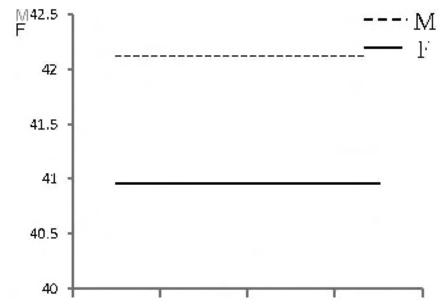
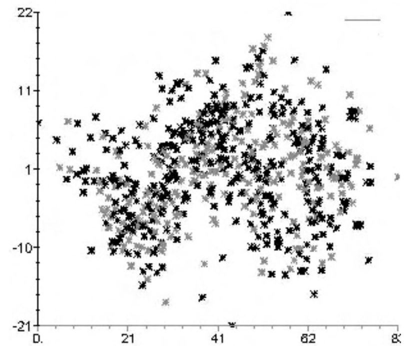
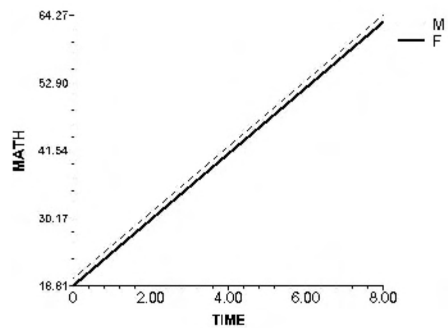
Equation	*Coefficients/Deviance*	*Residual Plots*	*Graphical representation*	*Interpretation*
Model A1 Random Intercept with interaction No Slope $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + r_{0i} + e_{ti}$	$\beta_{00} = 42.12^*$ $\beta_{01} = -1.16^{ns}$ **Variance** $\tau_{00} = 0.45^{ns}$ $\sigma^2 = 354.76^*$ **Deviance** = 10,668.48 **Parameters** = 4			The predicted mean math score for boys, ignoring time of assessment, is 42.12. The score for girls is lower by 1.16 points, at about 41, but the difference is not statistically significant.
Model B1 Random Intercept with interaction Fixed Slope $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + r_{0i} + e_{ti}$	$\beta_{00} = 19.97^*$ $\beta_{01} = -1.16^{ns}$ $\beta_{10} = 5.54^*$ **Variance** $\tau_{00} = 52.51^*$ $\sigma^2 = 57.33^*$ **Deviance** = 8,855.55 **Parameters** = 5			The estimated average math score for boys at the first time point was 19.97. The average score for girls at this same time point was not significantly different (1.16 points below = 18.81 points). Students, regardless of gender, increased 5.54 points each year.

Table 5.3 Models A through F: Explanations for the unconditional phase.

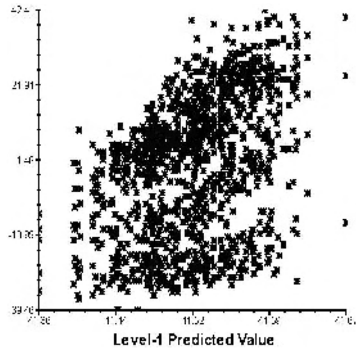
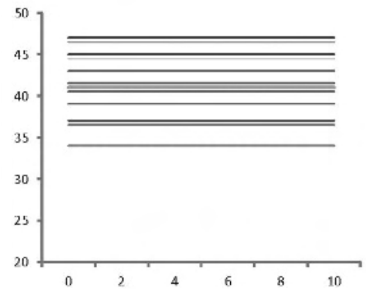
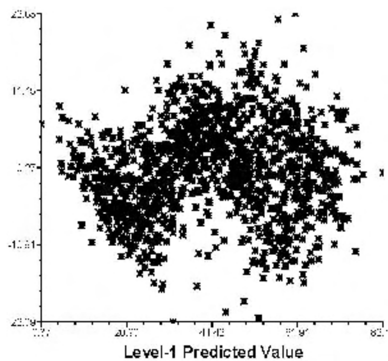
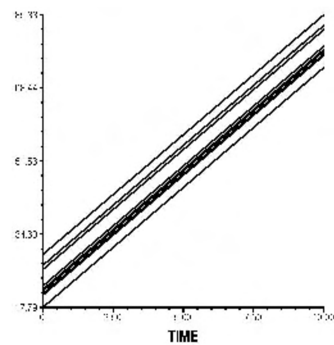
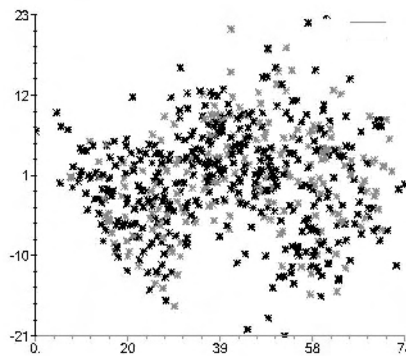
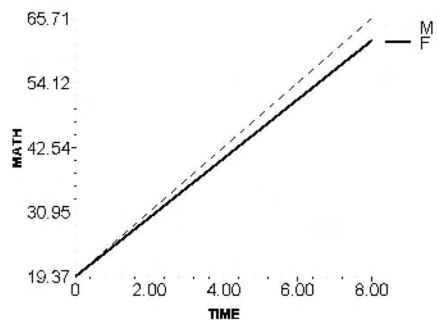
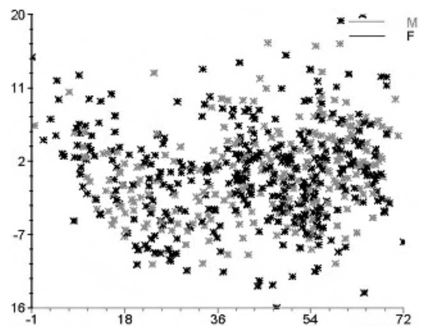
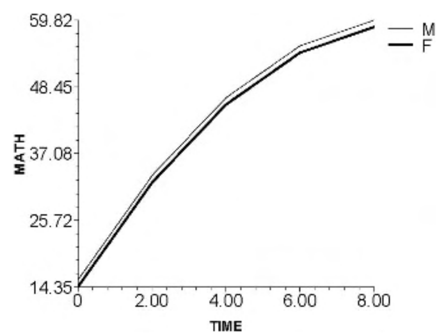
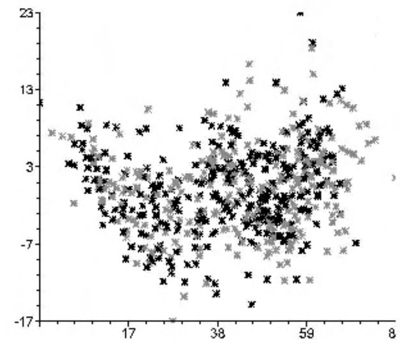
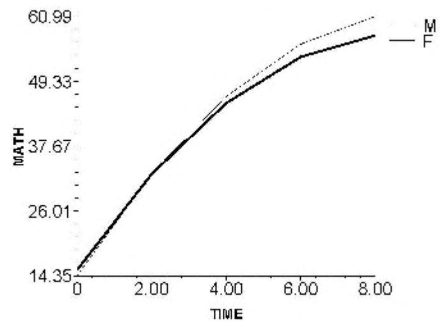
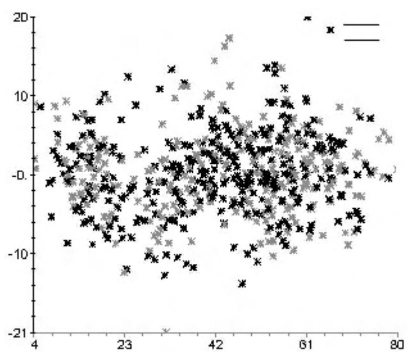
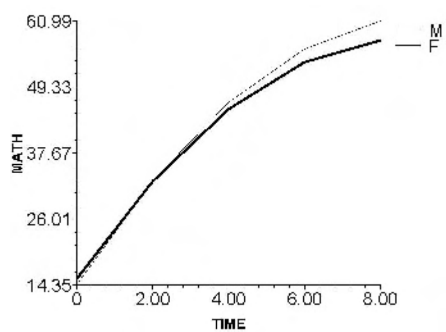
Equations	*Coefficients/ Deviance*	*Residual Plots*	*Graphical Representation*	*Interpretation*
Model A Random intercept No slope $MATH_{ti} = \beta_{00} + r_{0i} + e_{ti}$	$\beta_{00} = 41.52$* **Variance** $\tau_{00} = 0.47^{ns}$ $\sigma^2 = 355.08$* **Deviance** = 10,669.65 **Parameters** = 3			The estimated mean math score for students is 41.52, which is significantly different from zero. The variance component (0.47) indicates that students' average math score ignoring differences by time varies, although not significantly.
Model B Random intercept Fixed slope $MATH_{ti} = \beta_{00} + \beta_{10}$ $*TIME_{ti} + r_{0i} + e_{ti}$	$\beta_{00} = 19.37$* $\beta_{10} = 5.54$* **Variance** $\tau_{00} = 52.85$* $\sigma^2 = 57.33$* **Deviance** = 8,856.85 **Parameters** = 4			The average math score for students at the first time point (baseline) was 19.37. Students increase by a significant 5.54 points every year. The variance component (52.85) signifies that students significantly vary in math achievement at the first time point (baseline).

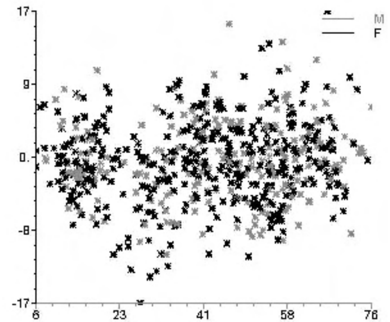
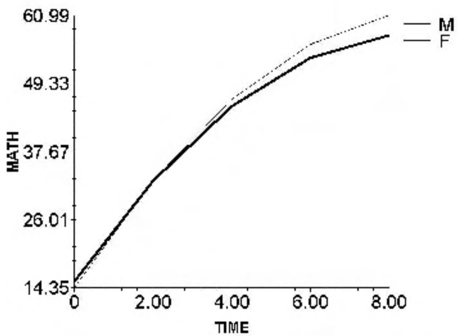
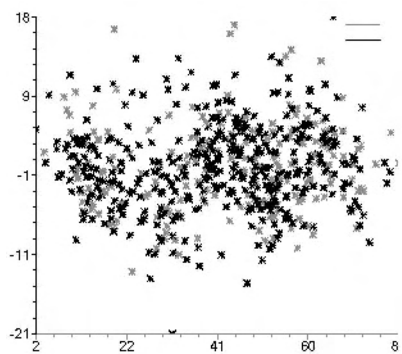
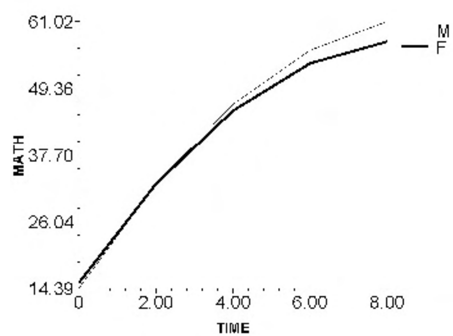
Table 5.4 (continued)

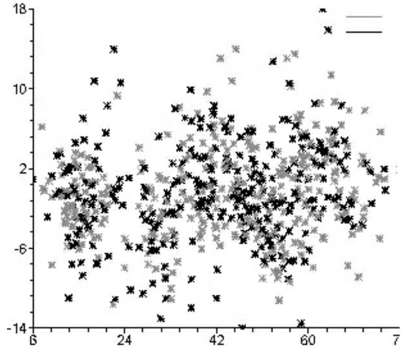
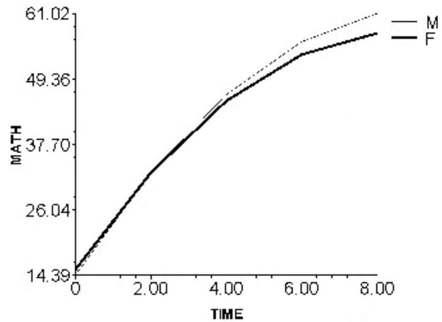
Equation	*Coefficients/Deviance*	*Residual Plots*	*Graphical representation*	*Interpretation*
Model C1 Random Intercept with interaction Random Slope with interaction $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} * (time)_{ti} *(sex01)_i + r_{0i} + r_{1i} *(time)t_i + e_{ti}$	$\beta_{00} = 18.80*$ $\beta_{01} = 1.09^{ns}$ $\beta_{10} = 5.83*$ $\beta_{11} = -0.56*$ **Variance** $\tau_{00} = 21.63*$ $\tau_{11} = 0.46^{ns}$ $\sigma^2 = 51.75*$ **Deviance** = 8,769.73 **Parameters** = 8			Average score for boys at study start is 18.80. Girls' start is nonsignificantly different by 1.09 points. Boys gain on average 5.83 points each year; this differs significantly from girls, who gain 5.27 (–.56 points less) points per year. In this model, there is significant variation in baseline scores, but not in time slopes.
Model D1 Random Intercept with interaction Fixed Slope Fixed Curvature $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{20} *(timeSQ)_{ti} * + r_{0i} + e_{ti}$	$\beta_{00} = 15.52*$ $\beta_{01} = -1.16^{ns}$ $\beta_{10} = 9.99*$ $\beta_{11} = -0.56*$ **Variance** $\tau_{00} = 55.98*$ $\sigma^2 = 39.97*$ **Deviance** = 8,502.18 **Parameters** = 6			Average score for boys at study start is 15.52. Girls' start is nonsignificantly different, at 14.36 points. Instantaneous rate of change at baseline is 9.99 points regardless of gender. Change in the rate of change decreases by .56 points each year. After accounting for gender, there is still significant variation in baseline scores.

Equation	*Coefficients/Deviance*	*Residual Plots*	*Graphical representation*	*Interpretation*
Model D2 Random Intercept with interaction Fixed Slope with interaction Fixed Curvature $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} *(time)_{ti} *(sex_{01})_i + \beta_{20} *(timeSQ)_{ti} + r_{0i} + e_{ti}$	$\beta_{00} = 14.35^*$ $\beta_{01} = 1.09^{ns}$ $\beta_{10} = 10.28^*$ $\beta_{11} = -0.56^*$ $\beta_{20} = -0.56^*$ **Variance** $\tau_{00} = 56.14^*$ $\sigma^2 = 39.18^*$ **Deviance** = 8,482.49 **Parameters** = 7			Average score for boys at study start is 14.35. Girls' start is nonsignificantly different. At baseline, the instantaneous rate of change for boys is 10.28 points; girls have a significantly lower instantaneous rate of change. The change in rate of change over time becomes smaller by .56 points per year regardless of gender. Significant variance in the baseline scores remains.
Model E1 Random Intercept with interaction Random Slope with interaction Fixed Curvature $MATH_{ti} = \beta_{00} + \beta_{01}* (sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} *(time)_{ti} *(sex_{01}) + \beta_{20} *(timeSQ)_{ti} + r_{0i} + r_{1i}(time)_{ti} + e_{ti}$	$\beta_{00} = 14.35^*$ $\beta_{01} = 1.09^{ns}$ $\beta_{10} = 10.28^*$ $\beta_{11} = -0.56^*$ $\beta_{20} = -0.56^*$ **Variance** $\tau_{00} = 23.98^*$ $\tau_{11} = .53^*$ $\sigma^2 = 33.90^*$ **Deviance** = 8,368.99 **Parameters** = 9			Fixed effects retain same interpretation as in Model D2. After accounting for gender, significant variation in baseline scores (23.98) and in the instantaneous rate of change at baseline (.53) still remain.

(Continued)

Table 5.4 (Continued)

Equation	*Coefficients/Deviance*	*Residual Plots*	*Graphical representation*	*Interpretation*
Model F1 Random Intercept with interaction Random Slope with interaction Random Curvature $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} *(time)_{ti} *(sex01)_i + \beta_{20} *(timeSQ)_{ti} + r_{0i} + r_{1i}(time)_{ti} + r_{2i}(timeSQ)_{ti} + e_{ti}$	$\beta_{00} = 14.37^*$ $\beta_{01} = 1.05^{ns}$ $\beta_{10} = 10.29^*$ $\beta_{11} = -0.57^*$ $\beta_{20} = -0.56^*$ **Variance** $\tau_{00} = 14.75^*$ $\tau_{11} = 3.52^*$ $\tau_{22} = 0.03^{ns}$ $\sigma^2 = 31.29^*$ **Deviance** = 8,334.03 **Parameters** = 12			With only slight variation, interpretation of fixed effects remains the same as in Models D2 and El. After accounting for gender, significant differences in baseline scores and in the instantaneous rate of change at baseline still remains, but no variation in the curvature parameter is evident.
Model E2 Random Intercept with interaction Random Slope with interaction Fixed Curvature with interaction $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} *(time)_{ti} *(sex01)_i + \beta_{20} *(timeSQ)_{ti} + \beta_{21} *(timeSQ)_{ti} *(sex01)_i + r_{0i} + r_{1i}(time)_{ti} + e_{ti}$	$\beta_{00} = 14.39^*$ $\beta_{01} = 1.02^{ns}$ $\beta_{10} = 10.25^*$ $\beta_{11} = -0.49^{ns}$ $\beta_{20} = -0.55^*$ $\beta_{21} = -0.01^{ns}$ **Variance** $\tau_{00} = 23.98^*$ $\tau_{11} = 0.53^*$ $\sigma^2 = 33.90^*$ **Deviance** = 8,368.96 **Parameters** = 10			Average score for boys at study start is 14.39. Girls' start is nonsignificantly different by 1.02 points. Instantaneous rate of change at baseline is 10.25 for boys and is lower by .49 points for girls, but this difference is not significant. The change in the rate of change over time is negative for boys and slightly more negative for girls, but this difference is also not significant. After accounting for gender, significant variability in baseline scores and in the instantaneous rate of change at baseline still remains.

Equation	Coefficients/Deviance	Residual Plots	Graphical representation	Interpretation
Model F2 Random Intercept with interaction Random Slope with interaction Random Curvature with interaction $MATH_{ti} = \beta_{00} + \beta_{01} *(sex01)_i + \beta_{10} *(time)_{ti} + \beta_{11} *(time)_{ti} *(sex01)_i + \beta_{20} *(timeSQ)_{ti} + \beta_{21} *(timeSQ)_{ti} *(sex01)_i + r_{0i} + r_{1i}(time)_{ti} + r_{2i}(timeSQ)_{ti} + e_{ti}$	$\beta_{00} = 14.39^*$ $\beta_{01} = 1.02^{ns}$ $\beta_{10} = 10.25^*$ $\beta_{11} = -0.49^{ns}$ $\beta_{20} = -0.55^*$ $\beta_{2l} = -0.01^{ns}$ **Variance** $\tau_{00} = 14.75^*$ $\tau_{11} = 3.52^*$ $\tau_{22} = 0.03^{ns}$ $\sigma^2 = 31.29^*$ **Deviance** = 8,334.00 **Parameters** = 13			Fixed effects retain same interpretation as Model E2. After accounting for gender, significant differences in baseline scores and in the instantaneous rate of change at baseline still remain. No significant variation in the change in the rate of change is evident after accounting for gender differences.

In the following, we concentrate on illustrating the conditional phase based on Model F, which we determined as the best representation of growth for the models we examined.

There are two primary conditional models that were augmented from the unconditional Model F. First, we examine a model in which gender is used to predict differences in the intercept and the linear term within the random intercept, the random slope, and the random quadratic model (Model F1; Table 5.4). This analysis posits that there would be gender differences in students' initial baseline scores (initial status) and in their instantaneous rate of change at baseline (linear component). Next, we examine a model that also includes gender as a predictor of the curvature parameter, the random quadratic term (Model F2; Table 5.4). In an effort to focus more completely on interpretation of conditional quadratic models, we utilize gender as the only predictor here; the extension is straightforward for a properly centered or coded continuous covariate.

5.3.5.1 Model F1

We included gender (*sex01*) as a predictor of the intercept and the linear term for Model F1:

$$Math_{ti} = \beta_{00} + \beta_{01} * (sex\,01)_i + \beta_{10} * (time)_{ti} + \beta_{11} * (time)_{ti} * (sex\,01)_i + \\ \beta_{20} * (timeSQ)_{ti} + r_{0i} + r_{1i}\,(time)_{ti} + r_{2i}\,(timeSQ)_{ti} + e_{ti}$$

Note that the inclusion of *sex01* appears as a main effect, but also as part of an interaction term with *time*, which is a level-one variable. This interaction represents a *cross-level interaction*. When a variable is included in a model to investigate differences in a slope parameter, that resulting parameter—here, β_{11}—is referred to as a cross-level interaction effect. Interpretation of these relevant parameters is provided below.

5.3.5.1.1 INTERCEPT DIFFERENCES

The first type of question we can ask with a dichotomous predictor is whether the dependent variable differs between the two categories at the intercept. In the NLSY application, this is equivalent to asking whether math achievement at age 6 (*time* = 0) is different based on the student's gender. The inclusion of the gender variable means that β_{00} now estimates the average performance at baseline when gender is equal to zero (which in this data set indicates a male student), and the differential between males and females at baseline is captured by $\beta_{01.}$ For males at baseline, the estimated overall initial math achievement score is β_{00} = 14.37; and for girls, the estimated overall math achievement would be $\beta_{00} + \beta_{01}$ = 14.37 + 1.05 = 15.42; however, the difference between boys and girls at baseline is not significantly different from zero.

5.3.5.1.2 INSTANTANEOUS GROWTH-RATE DIFFERENCES

There was a significant difference in the instantaneous rate of change at baseline between boys and girls. This instantaneous rate of change for boys is estimated to be β_{10} = 10.29 ($p < .05$) and for girls we would estimate this as about .57 points lower, $\beta_{11} = -.57$ ($p <$ .05). Thus, overall the simple slopes at baseline are positive (increasing) for both boys are girls, but for girls the slopes of these tangent lines are somewhat smaller. Both of these parameter estimates exhibited significant variability between students, with variance components that are statistically greater than 0 (Table 5.4).

5.3.5.1.3 CURVATURE PARAMETER DIFFERENCES

The curvature parameter in Model F1 is significant and negative, $\beta_{20} = -.56$, showing deceleration in the simple slopes or a downward curvature over time; in this model, no variation in the curvature over time was observed; however, the more reliable deviance comparison between Model E1 (Table 5.4) and Model F1 suggests retaining the quadratic term as a random effect, $\chi^2\ (3) = 34.96$, $p < .01$.

5.3.5.1.4 MODEL ASSESSMENT

The graphical representation of the results and interpretation of this model are presented in Table 5.4, Model F1. In the graphical representation of Model F1, there are two lines presented, one for males and one for females. As noted above, the size of the coefficient associated with gender at initial status shows only a small difference (1.05 points) between boys and girls that is not significantly different from zero. Thus, although girls on average tended to score lower in math at baseline, results in average math performance at age 6 do not significantly differ by gender.

We can use deviance comparisons to assess whether Model F1 (with gender as a predictor of the intercept and linear component) provides a better fit to the data than the unconditional model, Model F. This comparison is statistically significant $\chi^2\ (2) = 14.73$, $p < .001$; thus Model F1 is preferred.

5.3.5.2 Model F2

In Model F2, we allow the curvature parameter to vary across students and showcase the inclusion of a categorical predictor on the quadratic term. This leads to an additional cross-level interaction in our model, and each student is allowed to accelerate (or decelerate) at his or her own rate. Note that the inclusion of an interaction between *sex01* and the quadratic term alters the interpretation of the variance component; this variability represents the amount of variation in the estimated curvature parameter after accounting for gender differences. Model F2 can be represented as

$$Math_{ti} = \beta_{00} + \beta_{01} * (sex\,01)_i + \beta_{10} * (time)_{ti} + \beta_{11} * (time)_{ti} * (sex\,01)_i + \beta_{20} * (timeSQ)_{ti} + \\ \beta_{21} * (timeSQ)_{ti} * (sex\,01)_i + r_{0i} + r_{1i}\,(time)_{ti}\,(sex\,01)_i + r_{2i}\,(timeSQ)_{ti} + e_{ti}$$

In this model, *sex01* appears as a main effect, as part of an interaction term with *time*, and as part of an interaction with *timeSQ*.

5.3.5.2.1 INTERCEPT DIFFERENCES

The interpretation of the intercept differences in model F2 is nearly identical to that of model F1. By including gender, the coefficient β_{00} now represents the average performance of males at baseline (when age = 6). From Table 5.4, we see that the estimated mean score for boys at baseline is 14.39. The estimated score for girls can be derived by adding the value of the coefficient β_{01}, as it is representative of the differences in baseline score between males and females. From Table 5.4, this value is 1.02, and is nonsignificant, indicating that males and females have essentially the same score at baseline.

5.3.5.2.2 INSTANTANEOUS GROWTH-RATE DIFFERENCES

Examining Table 5.4, there was a significant difference in the linear rate of change at baseline between boys and girls. This instantaneous rate of change for males is estimated to be a significant 10.25 points per year. In this model, the instantaneous rate of growth was not found to be significantly different for girls than for boys $\beta_{11} = -.49$ $p > .05$. However, significant between-student variability remained in the linear trend, even after accounting for gender (τ_{11}, Table 5.4).

5.3.5.2.3 CURVATURE PARAMETER DIFFERENCES

In Model F2, potential differences between boys and girls on the curvature parameter were assessed via a cross-level interaction. Just as with previous parameter estimates, the inclusion of gender on this parameter means that the coefficient associated with curvature (β_{20}) represents the estimated change in simple slopes over time for boys. The value suggests that the change in the growth curvature over time for males significantly decreases at a rate of .55 points per year. Thus, achievement in mathematics is decelerating or slowing down over time. There was no significant difference between males and females in the overall change in the growth rate over time ($\beta_{21} = -.01$ $p > .05$). In addition, the variance component for the curvature effect indicated no between-student variability to be explained after accounting for gender differences (τ_{22}; Table 5.4).

5.3.5.2.4 MODEL ASSESSMENT

The interpretation of the graphical representation of Model F2 follows that of Model F1. There is one line presented for boys and one for girls. Recall that the differences between the two groups at baseline were minimal and nonsignificant ($\beta_{01} = 1.02$), which is demonstrated by the close distance between the two depicted lines. Over time, differences in math scores diverge between boys and girls, but not significantly so. Yet variability in instantaneous rate of change at baseline between students is evident, even after controlling for gender.

Finally, just as was established with the linear models, a chi-square comparison test between the final unconditional model (Model F) and the corresponding conditional model (Model F2) can be used to determine whether the inclusion of the predictor is warranted. Thus deviance comparisons between Model F (deviance = 8,348.70, parameters = 10) and Model F2 (deviance = 8,334.00, parameters = 13) indicate that Model F2 was a better fit and should be retained ($\chi^2(3) = 14.70$, $p < .001$).

Because the gender difference on the curvature parameter was found to be nonsignificant, the model-building process warrants the next step be to determine whether to retain this model (Model F2), or the model without this predictor (Model F1). This can be done with the same chi-square comparison test used previously. In this case, the deviance comparisons between Model F1 (deviance = 8,334.03, parameters = 12) and Model F2 (deviance = 8,334.00, parameters = 13) was nonsignificant, ($\chi^2(1) = .03$, $p = 1$). Thus, by parsimony, Model F1 would be chosen as the final model.

5.4 Modeling Change in Dichotomous Outcomes

Our preceding examples demonstrate a useful process for modeling linear or quadratic change in outcomes that are continuous and reasonably assumed to follow a normal distribution. Often, however, outcome variables that interest us are dichotomous—having only two values—such as whether or not a student attains proficiency in reading or mathematics over several years of school, or whether or not an adolescent has used alcohol or other drugs at distinct waves during a multiwave intervention study. Binary or dichotomous outcomes are those in which "success" may indicate that a child achieves proficiency or that an adolescent did not use alcohol or other drugs; success is often coded 1 and its absence (or "failure") is coded 0. A widely used method for describing the relationship between binary or dichotomous outcomes and a collection of explanatory variables is logistic regression. When binary data are clustered, as in the preceding longitudinal examples in which the multiple observations or outcomes are nested within individuals, the analysis is referred to as either a multilevel logistic regression or a mixed-effects logistic regression model.

Multilevel logistic regression is one type of a class of models called generalized linear mixed models (GLMM). In fact, all of the linear and quadratic models presented above are also GLMMs, but they represent the simplest and special case of GLMM in which the residuals from level 1 are normally distributed. The GLMM can also be used to model multilevel nominal, ordinal, or time-to-event data; these are often best understood as extensions of the multilevel logistic model.

There are many effective reviews of single-level logistic regression in the literature (e.g., O'Connell, 2006), and extensive work can be found on the multilevel logistic model (e.g., O'Connell, Goldstein, Rogers, & Peng, 2008) and its use in longitudinal studies (e.g., Hedeker & Gibbons, 2006). Here, we present a data application that follows the approach demonstrated earlier: first describing the pattern of level 1 change and then modeling variation in this pattern across individuals at level 2.

5.4.1 Demonstration Sample Data: ECLS-K

The data were drawn from the Early Childhood Longitudinal Study (ECLS), a national database developed and managed through the National Center for Education Statistics (NCES, 2002). The ECLS-K kindergarten cohort follows nearly 20,000 students assessed during fall and spring of kindergarten and spring of first grade, with additional follow-ups in the spring of third, fifth, and eighth grade. The ECLS sampling plan used a 30% subsample of the kindergarten cohort for inclusion in a fall first-grade wave of data collection. We focus here on the arithmetic proficiency data for a sample of n = 3,440 kindergarten children selected from the ECLS-K. Because our primary focus is to illustrate the application of the multilevel logistic model to longitudinal data, we limited our selected sample to those children who were included in the ECLS database for all four of the kindergarten and first-grade waves, who did not change schools between kindergarten and first grade, were not repeating kindergarten, and had no missing observations on the level-2 predictors utilized for this demonstration.

The dependent variable was formed from whether or not a student had achieved proficiency in early numeracy, originally assessed on a six-point scale, 0 to 5. We identified proficiency as mastering either level 4 or 5, which would reflect ability to solve simple addition and subtraction problems (level 4), and/or simple multiplication and division problems (level 5). ECLS's proficiency scale was constructed so that mastery at level 5, for example, could only be achieved after mastery of all previous levels (West, Denton, & Germino-Hausken, 2000). Thus, our dichotomous variable, *mathprof*, represents "proficiency in simple computation" (coded as 1), versus "not proficient" (coded as 0). Time is measured in months from the baseline kindergarten assessment: 0, 8, 12, and 18 months. We coded *male* as 1 for boys (49.5%) and 0 for girls (50.5%). We used the number of family risk factors, *numrisks*, such as having parents whose primary language is not English or

Table 5.5 Observed proportions attaining proficiency for the ECLS-K mathematics data.

Months	*Girls*	*Boys*	*Risk_0*	*Risk_1*	*Risk_2*	*Risk_3*	*Risk_4*	*Overall*
0	.04	.04	.06	.02	.00	.00	.00	.04
8	.18	.21	.25	.15	.06	.04	.00	.20
12	.33	.36	.42	.26	.23.	.10	.00	.36
20	.75	.74	.82	.70	.65	.56	.50	.77

living in a family receiving welfare or food stamps, as an additional continuous explanatory variable. This count variable ranged from 0 to 4, with 4 indicating more risks for poor performance in school (Zill & West, 2001). In our sample, 35.2% of children had one or more family risks.

Table 5.5 contains the proportion proficient in months since the baseline fall kindergarten assessment for boys versus girls, and based on the number of family risks. From this table, we see that overall proficiency increases over the kindergarten through first-grade period, with no evident differences between boys and girls, but with fewer children reaching proficiency when their number of family risks is larger.

5.4.2 *Review: Single-level Logistic Regression*

Before demonstrating the multilevel logistic regression model and its application to the study of change, we review a few basics for the single-level logistic model. Predicting a dichotomous (0,1) outcome through ordinary least squares regression is often referred to as a linear probability model, because the predicted value $E(Y_i \mid x)$, which we write as $\pi_i(x)$, represents the probability of the success outcome for the i^{th} person, conditional on the given set of predictors (Hosmer & Lemeshow, 2000; Long, 1997). However, probability is a limited quantity, and is constrained to lie between 0 and 1. Standard regression approaches are thus not appropriate for dichotomous outcomes, since predictions could fall outside the allowable 0,1 range. In fact, standard regression assumes that the residuals from this model would be independent and normally distributed as well as homoscedastic (following a common variance, σ^2). However, the variance for a binary sample follows a binomial distribution, and more specifically, a Bernoulli distribution, which is a special case of the binomial where each independent outcome (or "trial") results in either a success (1) or a failure (0). The variance of a Bernoulli distribution is directly related to its mean: $Var(Y \mid x) = \pi_i(x)^*(1-\pi_i(x))$. This makes the sampling distribution for binary variables clearly heteroscedastic (i.e., variable across people), so options other than the standard regression model should be employed.

5.4.2.1 *Linear and Non-linear Models*

Despite problems with the standard regression model for dichotomous data, one of its benefits is the fact that it represents a model that is *linear in the parameters*, often simply referred to as a *linear model*. Regression models that contain predictors that are themselves products or functions of other variables, such as interaction terms or the quadratic models in the previous sections that included a time*time (i.e., time^2) predictor, are still considered linear models (Fox, 2008). However, regression models that include a product or function of other regression coefficients in that model are termed *nonlinear models*.

Estimated coefficients for these kinds of regression models cannot be solved through ordinary least squares methods, but this is often easily overcome through maximum likelihood estimation (MLE) strategies or approximations to MLE. However, models that are linear in their parameters have advantages, primarily because we can easily add products or powers of predictors to capture non-linear patterns of change, and we can systematically attribute changes in the model to the addition of these new regressors, similar to our approach in the previous sections. Thus, it would be advantageous to derive a regression model for dichotomous data that is linear in the parameters.

The logistic regression model takes advantage of the fact that the relationship between the probability of success, $P(Y_i = 1) = \pi_i$, and a covariate x_i tends to follow an S-shaped curve, that is, a nonlinear pattern that can conveniently be modeled through a logistic function. Rather than attempt to predict probabilities directly, however, logistic regression reframes the model in terms of odds and logits (defined in the next paragraph)—quantities that can easily be manipulated back and forth and that are formed from the original success probabilities.

Although the desired goal in a logistic model is to understand factors related to the probability of success, the model is inherently nonlinear in terms of probability, but it is linear in terms of the logits. This defining feature of the logistic regression model makes it an attractive option for analysis of dichotomous data. By definition, the odds of success is a simple quotient comparing the probability of success to the probability of failure, and a logit is defined as the natural log of the odds of success. The logit is the quantity that is modeled in logistic regression. That is, with q predictors,

$$Odds = \frac{P(Y_i = 1 \mid \underline{x})}{P(Y_i = 0 \mid \underline{x})} = \frac{P(Y_i = 1 \mid \underline{x})}{1 - P(Y_i = 1 \mid \underline{x})} = \frac{\pi_i(\underline{x})}{1 - \pi_i(\underline{x})}$$

$$\text{logit}(\pi_i(\underline{x})) = \ln\left(\frac{\pi_i(\underline{x})}{1 - \pi_i(\underline{x})}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots \beta_q X_{iq} = x^T \beta.$$

Note that this model for the logits contains no residual term; these residuals are heteroscedastic, so no single representation exists, and the expected pattern of variation in the outcomes is predefined by use of the binomial sampling model. Note also that exponentiating the logit yields a direct expression for the odds in terms of the regression coefficients, and working backward to write the success probability in terms of the odds yields an expression for the logistic regression model that displays its nonlinear nature:

$$\exp(\text{logit}) = \exp(x^T \beta)$$

$$\pi_i(\underline{x}) = \frac{odds}{1 + odds} = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}.$$

At first glance, logistic regression may seem like a roundabout process to construct a model that is linear in the parameters. However, using the logit as the "new" outcome yields a model that is easy to work with and for which all of the familiar modeling strategies of standard regression will apply, such as including dummy-coded predictors, interaction terms, or powers of predictors. Interpretations of slopes and intercepts remain the same, with the exception that the outcome is a logit. Thus, interpretation of slopes refers to incremental changes in the logit due to a one-unit increase in a predictor, rather than incremental change in probability or in the dichotomous outcome itself. Our demonstration that follows makes this clear.

5.4.3 Longitudinal (Multilevel) Models for Dichotomous Outcomes

A natural extension of the single-level logistic regression model, the multilevel logistic regression model can be used to model dichotomous outcomes that represent repeated observations obtained on the same individuals over time. Individual growth models can be fit to dichotomous data following the principles outlined in the previous sections. That is, in the unconditional phase, we examine and describe the patterns of change occurring at level 1or the occasion level. If necessary, polynomial terms for time can be included at this level to capture curvilinear change. Once we feel that the level-1 pattern of change is adequately incorporated into the model, the conditional phase is used to investigate differences in patterns of growth due to predictors at the person level. In an educational context, these might include variables such as gender or race/ethnicity, background variables such as mother's education level, or psychosocial variables such as general persistence in school. All of the methods utilized in the previous section can be put to use within the multilevel logistic model.

The general form of the unconditional individual growth model for dichotomous outcomes, assuming that time is measured in months since baseline and allowing the effect of time to vary across individuals, would look like

$$\ln\left(\frac{\pi_{\mathrm{ti}}(\underline{x})}{1-\pi_{\mathrm{ti}}(\underline{x})}\right) = \beta_{00} + \beta_{10} MONTHS_{ti} + r_{0i} + r_{1i} MONTHS_{ti}.$$

All modifications to this model, such as including or eliminating random effects, building in a quadratic term for months to capture potential for nonlinear growth, or including level-2 predictors for a conditional model of growth will follow the same processes as established in our earlier sections.

5.4.4 Estimation

Logistic regression for clustered or longitudinal data introduces several complexities to model estimation that were not encountered in the multilevel model for continuous outcomes. The literature discusses adaptations to estimation for clustered data in general (Brown & Prescott, 2006; Raudenbush & Bryk, 2002), and maximum likelihood strategies are typically used to obtain estimates for both the fixed and random components of the model. The solution to an integral necessary for the estimation process, however, becomes intractable when the outcomes in a mixed-effects model are dichotomous. Several approximation techniques have been utilized to overcome this problem, and fall into two forms: approximate the likelihood through pseudo-likelihood methods, or approximate the integral through numerical methods (Breslow, 2003; Diaz, 2007; Pinheiro & Bates, 2006). One numerical strategy that seems to work well in sufficiently large samples is adaptive quadrature (AQ). This option has recently been included in the HLMv7 software, and it is the one we utilize here. A benefit of this choice is that the method yields a deviance statistic that can be appropriately compared across competing nested models, as was described previously for the continuous outcomes. This feature is not available in all estimation methods for generalized linear mixed models (see O'Connell, Reed, Ren, & Li, 2010, for a brief review).

We also report only the unit-specific (or subject-specific) estimation results. For generalized linear mixed models, it is possible to obtain population-averaged results (sometimes called marginal results). These provide estimates for the fixed-effect coefficients that are "averaged" across all random effects. Using the unit-specific results is more consistent with how we have used and interpreted results from all models for the continuous outcomes presented above. See Raudenbush and Bryk (2002) and Brown and Prescott (2006) for more detail on this issue. Finally, graphical methods for plotting model results for dichotomous outcomes are not available to the same degree as for continuous outcomes, so we emphasize interpretation of the basic results.

5.4.5 Demonstration: Modeling Longitudinal Dichotomous Outcomes: ECLS-K Data

We begin by examining a plot of the logits over time for our sample of n = 3440 kindergarteners assessed on mathematics proficiency four times through first grade (Figure 5.4). Because variability of dichotomous outcomes is already incorporated through use of the binomial sampling model and is heteroscedastic, a summary across all students in the sample is the best way to examine the general pattern of change.

Although there seems to be a slight departure from a perfect straight-line pattern to proficiency during summer between kindergarten and first grade, with only four waves of data a linear growth model may be the best degree polynomial to fit to the data. It would be possible to fit a quadratic, but such an approach is not strongly supported by this simple plot, and a linear trend appears reasonable. Further, models with both random and fixed effects for the quadratic trend failed to converge, which can occur if the variability in that random effect is very small; thus, we report the results for the linear models. We also decided to code time in months in reverse, so that our conditional model could examine whether there are differences *at the end of first grade* (rather than at baseline as examined in all NLSY79 models in our earlier sections) in mathematics proficiency by gender, or by the number of family risks. Thus, time was coded –20 at Wave 1 (kindergarten entry), –12 (Wave 2, spring kindergarten), –8 (Wave 3, fall first grade), and 0 (Wave 4, spring first grade). Table 5.6 provides the logit results from two multilevel logistic models, G1 and G2; Model G1 is the unconditional growth model with a random intercept and random effect for time. Model G2 is a conditional growth model with *male* and *numrisks* as predictors of the intercepts (status at end of first grade) and of the linear slope for time in months prior to spring first grade.

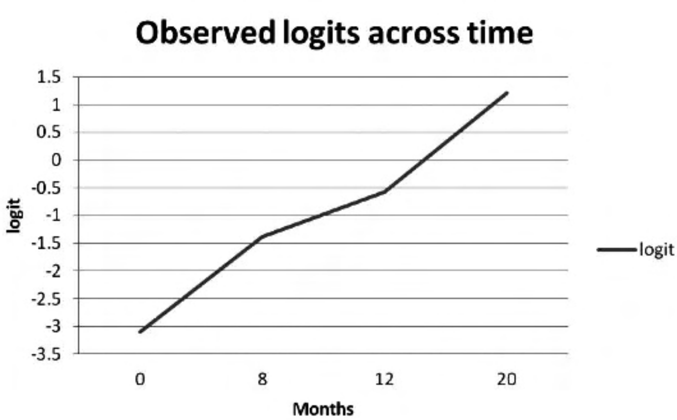


Figure 5.4 Unconditional logits for proficiency over time for the Early Childhood Longitudinal Study–Kindergarten (ECLS-K) data.

Table 5.6 Unconditional (G1) and Conditional (G2) growth models for dichotomous assessment of proficiency in early mumeracy (mathprof): ECLS-K.

Parameters	*G1 Unconditional*	*G2 conditional*
Intercept (Spring FG)		
β_{00}	1.41(.06)	1.68(.08)
β_{01}(male)		.02(.09)
β_{02}(numrisks)		–.52(.06)**
τ_{00} (variance(intercepts))	1.64(.18)**	1.55(.20)**
Months		
β_{10}	.29(.01)**	.28(.01)**
β_{11}(male)		–.001(.01)
β_{12} (numrisks)		.03(.007)**
τ_{11} (variance(slopes))	.002(.0002)	.002(.0002)**
Months-squared		
β_{20}	—	—
–2LnL	35,307.81	35,081.10
Number parameters	5	9

5.4.5.1 Unconditional Model G1

First, we see that at final status (spring first grade), the expected logit of success for children in this sample was 1.41. To interpret this intercept, we can exponentiate: *exp*(1.41) = 4.10 = *odds* of proficiency at baseline, and the estimated *probability* of proficiency at baseline is then 4.10 / (1+ 4.10) = .804, which can be compared with the observed value of .77 at wave four presented in Table 5.5. The random effects information implies that there is a lot of variability in the baseline logits for proficiency, τ_{00} = 1.64 (*s.e.* = .18). A simple Wald's test, using 1.64 / .18, yields a large *Z*-value, so the amount of variation at the end of first grade is statistically significant between children.

For each month from the final assessment, the expected logit for proficiency in mathematics changes by $\beta_{10} = .29$ points. Because time is reverse coded, this amount represents an increase in the logit over time across the growth trajectory; correspondingly, the probability of being proficient also increases over time. To see this, we can create a predicted logit for a child at the beginning of kindergarten (months = –20), $\hat{\eta}_{1i} = 1.41 + .29(-20) = -4.39$. Exponentiating the logit prediction yields the prediction in odds, *exp* (–4.39) = .012; and solving for probability yields .012/(1+ .012) = 1.2% (with small odds, odds and probability are similar). Continuing this process, at Wave 2 (months = –12), the predicted logit is $\hat{\eta}_{1i} = 1.41 + .29(-12) = -2.07$, with odds = .126 and corresponding probability of proficiency estimated at 11.2%. At Wave 3 (months = –8), the predicted logit = –.91, odds = .403, and prediction is 28.7%. These do not seem to match well with the raw proportions in Table 5.5, so our next strategy is to investigate predictors that might help better explain the observed proportions.

Finally, we note that the random effect for months is quite small, but is statistically significant. This suggests that children increase in math proficiency in a very similar pattern between kindergarten and first grade. It would be reasonable to compare the previous model to one without a random effect for months (in fact, that model has a statistically smaller deviance). However, we keep this random effect in the model simply to facilitate a more interesting discussion.

5.4.5.1.2 ODDS RATIOS

Before moving to the conditional phase, we offer an alternative approach to interpreting logistic results. For predictors in a logistic (or multilevel logistic) model, exponentiating their respective coefficient yields the *odds ratio*, which can be directly interpreted as the factor by which the odds of proficiency changes as that predictor increases by one unit. For the months variable, $\beta_{10} = .29$ and $exp(.29) = 1.335$. We interpret this as follows: For each one-unit increase in months (i.e., from –20 months to –19 months), the estimated odds of proficiency increases 1.335 times. Note that for months = –20, the odds = .01265; and for months = –19, the odds = .01689. Their ratio is the odds ratio: .01689/ 0.1265 = 1.335. Thus, the odds of proficiency increases by a factor of 1.335 each month. Laws of exponents can be used to find expected increase in odds for continuous predictors over a larger range, such as expected change in odds over an 8-month period, say, from –20 (*wave* 1) to –12 (*wave* 2) months. To do this, multiply the parameter estimate by the period of interest, for example, 8 months, prior to exponentiating; thus, $exp(.29*8) = 10.176$; so the odds of proficiency at –12 months should be 10.176 times the odds of proficiency at –20 months (an 8-month change). This is the expected change across an 8-month range. At –20 months, the odds of proficiency are .012, so 8 months later, we expect the odds would be 10.176*.012 = .122, which corresponds (within rounding) to the value found using the regression equation in the preceding section. For continuous predictors, this method is particularly appealing. A 1.335 increase in the odds each month may not seem substantively important, but across 8 months, a 10.176 increase in the odds of proficiency clearly is relevant.

5.4.5.2 Conditional Model G2

Two student-level predictors of the intercept (final status in logits) and the slope are included at level 2: *male* and *numrisks*. Results are provided in Table 5.6. No significant differences due to gender were observed at the end of first grade (intercepts) or for the change over time; this is consistent with the raw data summarized in Table 5.5. However, a

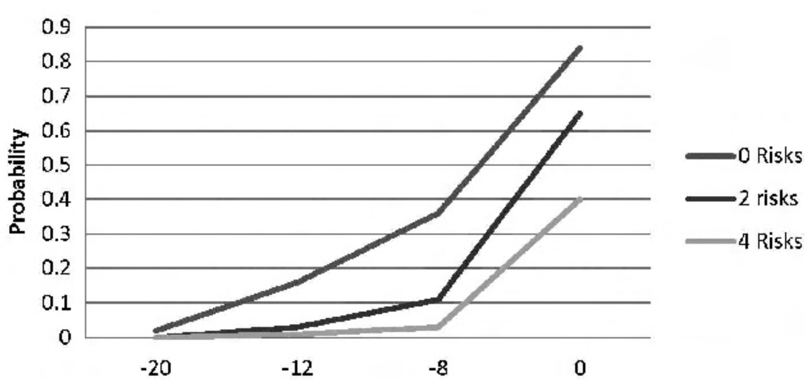


Figure 5.5 Predicted growth in proficiency over time by number of family risks.

significant contribution of the number of family risks was observed, and this effect was negative for the intercept, $\beta_{02} = -.52$. At the end of first grade, children with more family risks were *less likely* to attain proficiency in early numeracy; as the number of family risks increases, the odds of proficiency at the end of first-grade decreases by a factor of ($exp(-.52 = .59)$, all else held constant. This means that for each additional family risk, the odds of proficiency decreases by slightly more than half. This is a statistically significant finding.

The effect of the number of family risks on the linear trend is positive, with $\beta_{12} = .03$. However, it is important not to be misled by the positive slope in this context. Recall that time is reverse coded to clock time of assessments prior to the end of first grade: *months* = −20, −12, −8, 0. Although there are clear benefits to centering time at the end of first grade to focus attention on the potential gender differences or the potential impact of family risks at that time point, caution is required in interpreting the direction of the logit predictions for the predictors on the slope. In general, reviewing the mixed-model format can help in interpreting the direction of effects regardless of how time may be coded

$$\begin{aligned}\eta_{ti} = {} & \beta_{00} + \beta_{01} * male_i + \beta_{02} * numrisks_i \\ & + \beta_{10} * months_{ti} + \beta_{11} * months_{ti} * male_i + \beta_{12} * months_{ti} * numrisks_i \\ & + r_{0i} + r_{1i} * months\end{aligned}$$

In our example, the cross-level interaction between *months* and *numrisks*, $\beta_{12} = .03$, and is statistically significant. Due to the reverse coding of time, this implies that as the number of family risks *increases*, the overall effect of time on likelihood of proficiency becomes *more negative*, all else held constant; thus, children with more family risks are *less likely* to attain proficiency relative to their lower risk peers by the end of first grade. Figure 5.5 contains a graphical depiction for the expected proficiency probabilities for children with 0, 2 and 4 family risks. Note that the graph is nonlinear; as discussed previously, the relationship between probability of an event (such as attaining proficiency) and a covariate (in this case, *months*) is nonlinear, and is the primary justification for application of the multilevel logistic model for growth.

Finally, we can use the deviances from the two competing models, G1 and G2, provided in Table 5.6 as −2lnL, to determine their relative fit. The conditional model provides a statistically significant improvement in fit related to the unconditional model, $\chi^2(4) = 226.71$, $p < 001$.

5.4.6 Summary

Although the logistic regression model is common in the social sciences, other options are available for identifying a function that adequately describes the nonlinear pattern between probability and the predictors, such as the probit model, but in practice, these two approaches tend to yield similar findings. Regardless of the choice for the structural part of the model, the motivation for model building remains the same whether outcomes are dichotomous or continuous: First, researchers must identify the most appropriate pattern for growth in an unconditional phase; second, predictors of variability in these growth patterns are examined in the conditional growth phase.

5.5 Design, Measurement, and Statistical Power Issues Affecting Quality of Growth Models

Longitudinal researchers often struggle with the optimal design of longitudinal studies. Is it better to have more people collected over fewer waves of data collection or fewer people with more observations across time? What is the optimal spacing of data collection points?

There are trade-offs in terms of the study duration, the frequency of observation, and the sample size. If there is a great deal of within-person variance in the growth trajectories, then adding additional observations may provide more insight into the process under examination than adding people to the sample does. In addition, having larger numbers of observations per person is more important for nonlinear growth models than for linear growth models. Adding additional observations can aid in understanding the functional form for complex nonlinear trajectories. In addition, adding more time points bolsters statistical power most for models that contain higher order polynomial terms (Raudenbush & Liu, 2001). On the other hand, if there is a great deal of between-person variance, then increasing the sample size (i.e., persons) may be more informative and provide greater statistical power. Finally, there is a level of diminishing returns in terms of adding time points. Whereas increasing from 3 to 4 time points is likely to result in a large increase in power, increasing from 9 to 10 time points will have much less of an effect on power. In general, increasing sample size increases the power to detect group differences in the growth slope more than increasing time points does (Raudenbush & Liu, 2001).

In addition to power, there are several measurement and statistical issues that should also be considered when conducting growth analyses. We discuss several of these issues below.

5.5.1 Regression to the mean, measurement error, and reliability

Regression to the mean is an important, but commonly misunderstood statistical phenomenon. When using an independent variable (such as a test score at Year 1) to predict scores on a dependent variable (such as a test score at Year 2), errors in prediction will occur whenever the correlation between the two variables is less than perfect (Campbell & Kenny, 1999). These errors in prediction will make it appear that people with initially extreme scores have scores closer to the mean on the posttest. There are two primary sources for this imperfection: lack of congruence between the two assessments (in format or in content tested) and measurement error. Reliability is a related concept in that it describes the consistency of scores across time, test forms, or internally within the test itself. Measurement error and reliability are inversely related: The greater the measurement error, the lower the reliability of the scores. The goal, of course, is to minimize the degree of measurement error in scores, but it is impossible to completely eliminate. Because hierarchical linear modeling is a regression-based technique, the assumptions of linear regression models continue to apply. One commonly overlooked assumption of linear regression is that the independent variables are measured reliably. When one or more predictor variables are measured with error, the results of the analysis might be misleading. The likelihood of Type II error increases for measures that exhibit low reliability, while the likelihood of Type I error increases for the other variables in the model (e.g., Osborne & Waters, 2002). Therefore, it is especially important to provide evidence of reliability of scores for all of the continuous independent variables in the model.

5.5.2 Attrition and Missing Data

One strength of individual growth curve modeling is that it can easily handle missing data at level 1 if the data are missing at random. When data are missing completely at random or missing at random, individual growth modeling should still produce valid results (Singer & Willett, 2003) because one can assume "that the probability of missingness is unrelated to unobserved concurrent outcomes (conditional on all observed outcomes)"

(p. 159). However, when attrition is systematically related to scores on the outcome variable of interest (after controlling for the independent variables in the model), the estimates of the growth parameters are likely biased, potentially invalidating the inferences about the phenomenon of interest. Therefore, researchers must examine the nature of the missingness within the sample prior to conducting growth analyses. The interested reader should consult Enders (2010) for an excellent introduction to the issue of missing data.

5.5.3 Transforming the Dependent Variable

If the dependent variable is not normally distributed, one option is to transform the dependent variable. This may be a reasonable option; however, a few words of caution are in order. First, any transformation that is made to the dependent variable at one time point should be made to the dependent variable across all time points. Otherwise, the dependent variable is no longer on the same metric throughout the study. Because participants are growing across time, it is not uncommon to observe a variable that is positively skewed at the beginning of the study, fairly normally distributed in the middle of the study, and negatively skewed at the end of the study. Such pattern of results is indicative of floor effects during the early time points and ceiling effects during the latter time points. However, transforming the outcome variable cannot solve this problem, as any transformation that improves Time Point 1 will make the last time point even more non-normal and vice versa. Also, researchers should resist the temptation to standardize the outcome variable (by making z-scores at each time point). This forces the variance of the outcome variable to be equal (and to equal 1.0) across every time point, which may not be useful, given that researchers are usually interested in understanding how the variance in the outcome changes across time.

5.6 Summary

The goal of our chapter was to present a principled and stepped approach for building and interpreting polynomial individual growth models. This process involves identifying the pattern of change observed over time at level 1, prior to including level 2 predictors of variation in these patterns. Typically, it is not necessary to consider all possible variations on including or excluding random effects in the models during the unconditional phase; we did so here to emphasize understanding of model results arising from slight variations in how parameters are treated and how growth may be represented. However, the approach we advocate—beginning with unconditional models at level 1 and building toward conditional models at level 2—can be appropriately utilized for modeling growth in continuous outcomes as well as outcomes that are discrete, limited, or non-normally distributed, such as the dichotomous example included earlier. As others have argued, identifying the most appropriate functional form of growth is essential to deriving an adequate representation of developmental change; this includes design of the temporal sequencing of assessments in order to be able to recognize complex curvilinear patterns if they truly exist in the phenomena under study (Collins, 2006; Grimm, Ram, & Hamagami, 2011; Willett et al., 1998).

Through this chapter, we have also clarified differences in terminology between linear and nonlinear growth and between linear and nonlinear models. Polynomial models can be used effectively to capture many forms of nonlinear growth and can be incorporated into models where the underlying distribution is non-Gaussian, such as the logistic distribution. This makes polynomial models for growth attractive and straightforward to apply for many types of outcome variables.

Finally, our demonstrations indicate the power of using residual plots as an aid in identifying the functional form of change. Examination of residuals is often ignored throughout the model-fitting process, but as we have shown here in Tables 5.3 and 5.4, residual plots can bolster one's understanding of the modeling process, and inform in the selection of appropriate models of the data.

Notes

1. Under FEML in HLMv7, a Wald test can be conducted with $Z = \sigma^2/(s.e.(\sigma^2))$; these values can be found in an earlier section of the printout. However, since variation over time is expected in longitudinal designs, we focus our interpretation on between-person differences.
2. Whether a variable is a predictor or a covariate within a growth model is determined by the question rather than the modeling, thus we use the word *predictor* throughout, but these can also be interpreted as covariates.

References

Breslow, N. (2003). *Whither PQL?* (University of Washington Biostatistics Working Paper Series, No. 192). Retrieved from http://www.bepress.com/uwbiostat/paper192

Brown, H., & Prescott, R. (2006). *Applied mixed models in medicine* (2nd ed.). West Sussex, England: Wiley.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York, NY: Guilford Press.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39*, 447–462.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Reviews in Psychology*, *57*, 505–528.

Crowe, E. C., Connor, C. M., & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of Educational Psychology*, *47*, 187–214.

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

Diaz, R. E. (2007). Comparison of PQL and Laplace 6 estimates of hierarchical linear models when comparing groups of small incident rates in cluster randomized trials. *Computational Statistics and Data Analysis*, *51*, 2871–2888.

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

Dunn, L. M., & Markwardt, F. C. Jr. (1970). *Peabody Individual Achievement Test: Manual*. Circle Pines, MN: American Guidance Service.

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

Fitzmaurice, G., & Molenberghs, G. (2009). Advances in longitudinal data analysis: An historical perspective. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 3–27). Boca Raton, FL: Chapman & Hall/CRC Press.

Fox, J. (2008). *Applied regression analysis and generalized linear models*. Thousand Oaks, CA: Sage.

Francis, D. J., Schatschneider, C., & Carlson, C. D. (2000). Introduction to individual growth curve analysis. In D. Drotar (Ed.), *Handbook of research in pediatric and clinical child psychology* (pp. 51–73). Norwell, MA: Kluwer Academic/Plenum Publishers.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357–1371.

Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Maas, C. M., & Hox, J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*(2), 127–137.

McCoach, D. B., & Black, A. C. (2008). Evaluation of model fit and adequacy. In A.A. O'Connnell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 245–271). Charlotte, NC: Information Age Publishing.

McCoach, D. B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in Psychology*, *1*: 17. doi: 10.3389/fpsyg.2010.00017 http://www.frontiersin.org/Psychology.

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

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

National Center for Education Statistics. (2002). *User's manual for the ECLS-K first grade public-use data files and electronic code book: NCES 2002-135*. Washington, DC: U.S. Department of Education.

National Longitudinal Surveys. (2005). *NLS handbook*. Washington, DC: Bureau of Labor Statistics. Available online [http://www.bls.gov/nls/nlsdoc.htm].

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks, CA: Sage.

O'Connell, A. A., Goldstein, J., Rogers, H. J., & Peng, C. Y. J. (2008). Multilevel logistic models for dichotomous and ordinal data. In A. A. O'Connell and D. B. McCoach (Eds.), *Multilevel Modeling of Educational Data* (pp. 199–242). Charlotte, NC: Information Age Publishing.

O'Connell, A. A., & McCoach, D. B. (2004). Applications of hierarchical linear models for evaluations of health interventions. *Evaluation & the Health Professions*, *27*(2), 119–151.

O'Connell, A. A., Reed, S., Ren, W., & Li, J. (2010, April 29–May 4). *Estimation methods and software comparison for hierarchical generalized linear models*.Paper presented at the 91st Annual American Educational Research Association Annual Meeting, Denver, CO.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, *8*(2). Retrieved from http://pareonline.net/getvn.asp?v=8&n=2

Pinheiro, J. C., & Bates, D. M. (2006). Approximations to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*, *4*, 12–35.

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

Raudenbush, S. W., Bryk, A., Cheong, Y.F., Congdon R., & du Toit, M. (2011). *HLM7: Hierarchical Linear and Nonlinear Models*. Lincolnwood, IL: SSI.

Raudenbush, S. W., & Liu, X. F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*(4), 387–401.

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

Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In H. Goldstein, J. De Leeuw, & W. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–176). New York, NY: Springer.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's Kindergartners: Findings from the early childhood longitudinal study, kindergarten class of 1998–99: Fall 1998* (NCES 2000–070). Washington, DC: USDE, National Center for Education Statistics.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, *10*(2), 395–426.

Zill, N., & West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school* (NCES 2001–035). Washington, DC: USDE, National Center for Education Statistics.