Technical Report # 1320

A Primer on Longitudinal Data Analysis In Education

Joseph F. T. Nese

Cheng-Fei Lai

Daniel Anderson

University of Oregon

behavioral research & teaching

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Note: Funds for this dataset were provided by the Oregon Department of Education Contract No. 8777 as part of Project OFAR (Oregon Formative Assessment Resources) Statewide Longitudinal Data System (SLDS), CFDA 84.372A Grant Program that is authorized by the Educational Technical Assistance Act of 2002.

Abstract

Longitudinal data analysis in education is the study growth over time. A longitudinal study is one in which repeated observations of the same variables are recorded for the same individuals over a period of time. This type of research is known by many names (e.g., time series analysis or repeated measures design), each of which can imply subtle differences in the data or analysis, but generally follows the same definition. The purpose of this paper is to provide an overview of longitudinal data analysis in education for practitioners, administrators, and other consumers of educational research, focusing on: the purposes of longitudinal data analysis in education, some of its benefits and limitations, and the various analyses used to model student growth trajectories.

A Primer on Longitudinal Data Analysis in Education

Longitudinal data analysis in education is the study of student growth over time. A longitudinal study is one in which repeated observations of the same variable(s) are recorded for the same individuals over a period of time. This type of research is known by many names (e.g., time series analysis or repeated measures design), each of which can imply subtle differences in the data or analysis, but generally follows the same definition. The purpose of this paper is to provide an overview of longitudinal data analysis in education for practitioners, administrators, and other consumers of educational research, focusing on: the purposes of longitudinal data analysis in education, some of its benefits and limitations, and the various analyses used to model student growth trajectories.

1. Purposes

Longitudinal data analysis, also known as growth modeling and growth curve analysis, has as its primary purpose the measurement of change, or trajectories. Growth trajectories refer to both the intercept (initial or starting point) and the slope (growth, or change over time). There are two general objectives that are addressed by longitudinal data analysis: (a) how the outcome variable changes over time, and (b) predicting or explaining differences in these changes (Singer & Willett, 2003). The first purpose is more narrow, and looks at the description of the *functional form* of growth; that is, is growth linear, or non-linear. It is important to note here that growth can increase and/or decrease, accelerate and/or decelerate, and that an important part of longitudinal data analysis is modeling the correct functional form of growth.

The second purpose is much broader than the first, and addresses the relation between the trajectory and independent variables of interest (e.g., instructional program, public vs. private schooling, absences, socioeconomic status). In the coming sections, examples of these two purposes are provided, and different analyses that help answer questions related to these purposes are illustrated. Specific longitudinal educational data, described next, is used to help elucidate these purposes.

1.1 Description of Data

The following longitudinal data are used to help illustrate examples about growth and analyses throughout this paper. These data come from a larger study conducted in 2009-2010 to develop a comprehensive reading and mathematics assessment system. The sample includes 186 students in grade 4 who were administered eight oral reading fluency (ORF) measures over one academic year. Measures were administered in October, November, December, January, February, March, April, and May. Students with ORF results from at least four testing occasions were included in the sample.

For the ORF administration, students were shown a narrative passage (approximately 250 words) and were given 60 seconds to "do their best oral reading." The assessor followed along as the student read, indicating on the test protocol each word the student read incorrectly (producing the wrong word or omitting a word). If a student hesitated for more than three seconds, the assessor provided the correct word, prompted the student to continue, and marked the word as read incorrectly. Student self-corrections were marked as correct responses. After one minute, the assessor marked the last word read and calculated the total number of words read correctly (wcpm), by subtracting the number of incorrect words from the total words read.[1]

1.2 Examples of Applied Longitudinal Data Analysis in Education

Many teachers and special educators use students' work, test scores, and products to monitor skill development over time. Working within a response to intervention (RTI) framework, teachers are also often expected to monitor student progress to identify discrepancies in academic performance levels and trajectories between students and groups. In this context, and in our example data described in section 1.1, data can be students' scores on curriculum based measures (CBM), and student growth over time can be used to evaluate the effectiveness of instruction. These approaches often involve repeated performance sampling, graphic displays of time-series data, and qualitative descriptions of performance, which allow inferences to be made about both inter-individual (between-student) differences and intra-individual (within-student) improvement (Deno, Fuchs, Marston, & Shin, 2001).

1.2.1 Single-subject research

Perhaps the most basic application of longitudinal data analysis in education is single-subject research. In this type of experimental research individuals serve as their own control, meaning that comparisons are made to the individual's previous performance (Gast, 2010). In single-subject research, data for each individual are presented on a separate line graph so that data are collected repeatedly, graphed regularly, and analyzed frequently to make data-based decisions on an on-going basis (Gast, 2010). Single-subject research is considered experimental because the design includes a baseline phase that provides repeated measurement prior to an intervention to establish a pattern that can be used to compare post-intervention change in performance (Gast, 2010). In general, the researcher is attempting to qualify the effectiveness of the intervention based on a comparison to baseline data, which can be done with one or multiple individuals. It is important to note that single-subject research is a separate type of research from most of those discussed here, largely because there is no estimation of parameters, in other words, it is a nonparametric approach. Nonparametric generally means an approach that does not estimate parameters based on a population. Its counterpart, parametric, describes most statistical analyses that estimate parameters (e.g., regression coefficients, or growth trajectories) based on a larger population.

1.2.2 Describe growth

As mentioned, one purpose of longitudinal data analysis is to describe the functional form of growth. Here, functional forms of growth are placed into three categories: linear, polynomial, and piecewise. The most parsimonious, or simple, form of growth is linear growth. In linear growth models, growth is assumed and constrained to change at a constant rate over time, either increasing (a positive slope parameter) or decreasing (a negative slope parameter).

The second category of functional form is polynomial growth models, those that include exponential growth rates. Although these models encompass all possible orders of polynomial growth, longitudinal data analysis in education typically only includes quadratic and cubic growth. (Note that a polynomial growth model must include all growth terms prior to the final order, so that quadratic models include linear and quadratic terms, and cubic models include

---

[1] Note that ORF measures were developed as part of the easyCBM© progress monitoring and assessment system (Alonzo, Tindal, Ulmer, & Glasgow, 2006).

linear, quadratic, and cubic terms) Using the data described in section 1.1, time is modeled in months, from 0-7. To include a quadratic growth term, each unit of time is squared when added to the equation (i.e., 0, 1, 4, 9, 16, 25, 36, 49), and for a cubic growth term each unit of time is cubed when added to the equation (i.e., 0, 1, 8, 27, 64, 125, 216, 343). This allows the modeled growth to accelerate and/or decelerate as a function of time.

The last category of functional form is piecewise growth models, those that include different slopes for different time periods. An example of piecewise growth in education is data across two consecutive years, where separate estimated slopes for year 1, for summer, and for year 2 are desired. This model would have three slope parameters, each representing a different period of time (with a theoretical rationale about why one would expect different slopes for each time period).

In education, it is often assumed that growth is linear, but this assumption should always be supported by empirical evidence and statistical tests. The remaining purposes of longitudinal growth analysis in education discussed here relate to the second purpose of exploring the relation between the growth trajectory and independent variables of interest.

1.2.3 Predict and Model Variance of Trajectories

Perhaps the most primary purpose of longitudinal growth analysis in education is to explore the heterogeneity (difference) in change between students, and moreover, to determine the relation between predictors and the shape of each student's growth trajectory (Singer & Willett, 2003, p. 8). In other words, are there differences in where students begin (the intercept) and how students grow (slope), and if yes, what variables explain these differences? For those more familiar with some principles of statistics, these can also be analyzed in terms of the variances of the intercept and slopes, and if there is significant variance in these, what variables account for, or explain, these variances? For example, differences in ORF intercept and within-year ORF growth between students in general education and students receiving special education services can be explored (in which case the predictor is a dichotomous variable that indicates special education status or not).

1.2.4 Trajectories to Predict an Outcome

Using advanced statistical analyses, it is also possible to use growth trajectory parameters to predict distal outcomes. Following the example, intercept and slope estimates can be used to predict year-end reading achievement as measured by scores on the year-end state reading test. The relation between fall ORF skill (intercept) and year-end reading, and the relation between within-year ORF growth and year-end reading can also be estimated. The two relations can be compared to determine which is a better predictor of year-end reading: where one starts or how one grows throughout the year?

1.2.5 Accountability

One last example of a purpose of longitudinal data analysis in education is accountability. In the last example (section 1.2.4), the year-end state reading test scores were used as an outcome variable. These tests, in reading and math in grades 3-8 and content specific subjects in high school, are administered as part of the No Child Left Behind Act (NCLB, 2002). NCLB legislation requires states to implement accountability systems based on student test scores to track Adequate Yearly Progress (AYP); see section 3.2.1 for further discussion. States have used cross-sectional design to track AYP, a design that involves the observations of a population at one specific point in time, for example, observing grade 3 over several years in which each year a different group of student performance is analyzed. Currently, it is becoming more popular for states to use longitudinal data analysis, specifically, value-added approaches to analyze accountability. Value-added approaches consider all students' initial skill level in addition to their growth over time in order to more fairly account for progress. In other words, value-added approaches attempt to separate the effects of teachers and schools from those effects beyond the control of the education system (e.g., family background or SES), and hold states (or districts, schools, teachers) accountable only for the variables related to education. Please see section 3.2.2 for further discussion of value-added models.

2. Data & Assumptions

In this section some of the principles and assumptions of both longitudinal data *and* analysis are discussed. There are a number of data considerations when conducting or reviewing a longitudinal analysis in education, including the form of the observed data, the functional form of growth, and the number and schedule of occasions.

2.1 What Does the Observed Data Look Like?

As mentioned, one purpose of longitudinal data analysis is to describe the functional form of growth, or to determine which of the three categories of form (linear, polynomial, piecewise) best fit the data (section 1.2.2). This can be done in several ways, including an "eye-ball" inspection of the *observed* data. Note that the "observed" data are those that can be calculated directly from the data for each occasion (e.g., means, or averages, at each occasion), and the "estimated" or "predicted" line represents the intercept and slope as estimated by a statistical model (more on statistical models in section 3).

In an eye-ball inspection of the observed data, the observed data are those means that can be calculated directly from the data for each occasion. In the grade 4 ORF example there are eight testing occasions, and the sample ORF means of each occasion can be graphed. Figure 1 displays these observed means for each occasion. This graphic representation of the observed data helps supports the next step, deciding on the functional form of the data. Eye-balling these data, it appears growth could be quadratic, decreasing over time, or even cubic, with decreasing growth then increasing at the end of the year.

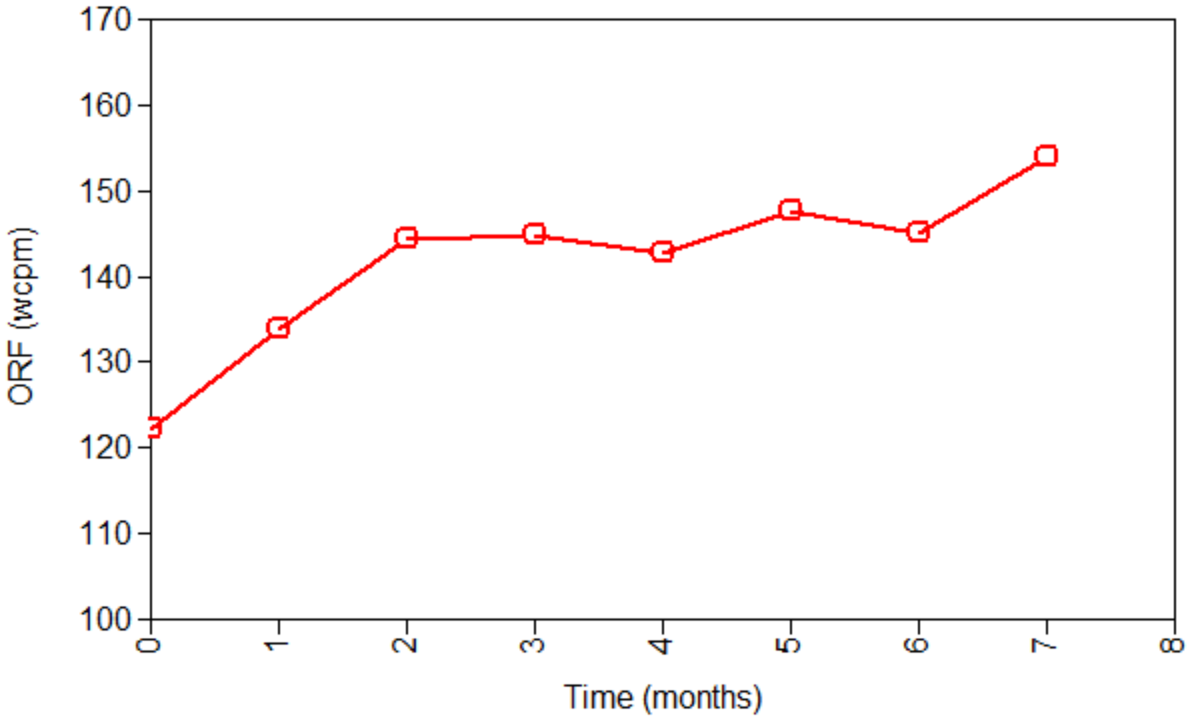

Figure 1. Observed (sample) means of ORF scores across each occasion.

2.2 What is the Functional Form of Growth?

Rigorous statistical tests are more often used to determine the functional form of the data when statistical analyses are involved. These tests are used to determine (a) which functional form best fits the data, based on a statistically significant or meaningful result, (b) whether the parameter associated with a growth term (e.g., quadratic, cubic) are statistically significantly different from zero, suggesting the parameter is a good addition to the model, and (c) whether the variance associated with a growth term is statistically significantly different from zero, suggesting the researcher can add predictors to explain that variance.

Figure 2a shows the predicted *linear* mean ORF growth (i.e., estimated) across time. Here, you can see that the growth is constrained to change at a constant rate over time. Figure 2b shows the predicted *quadratic* mean ORF growth across time. In this graph, you can see the predicted growth rate increases initially and then decelerates over the course of the school year. This is an example of quadratic growth in which change decelerates over time (it can also accelerate, in which case growth would exponentially increase over time). Figure 2c shows the predicted *cubic* mean ORF growth, and here you can see predicted growth rate increases to begin, decelerates around mid-year, and then increases at the end of the year. This is an example of cubic growth, in which there are two bends in the growth; in this case, decelerating and then accelerating. The opposite can also be modeled, accelerating growth followed by decelerating growth.

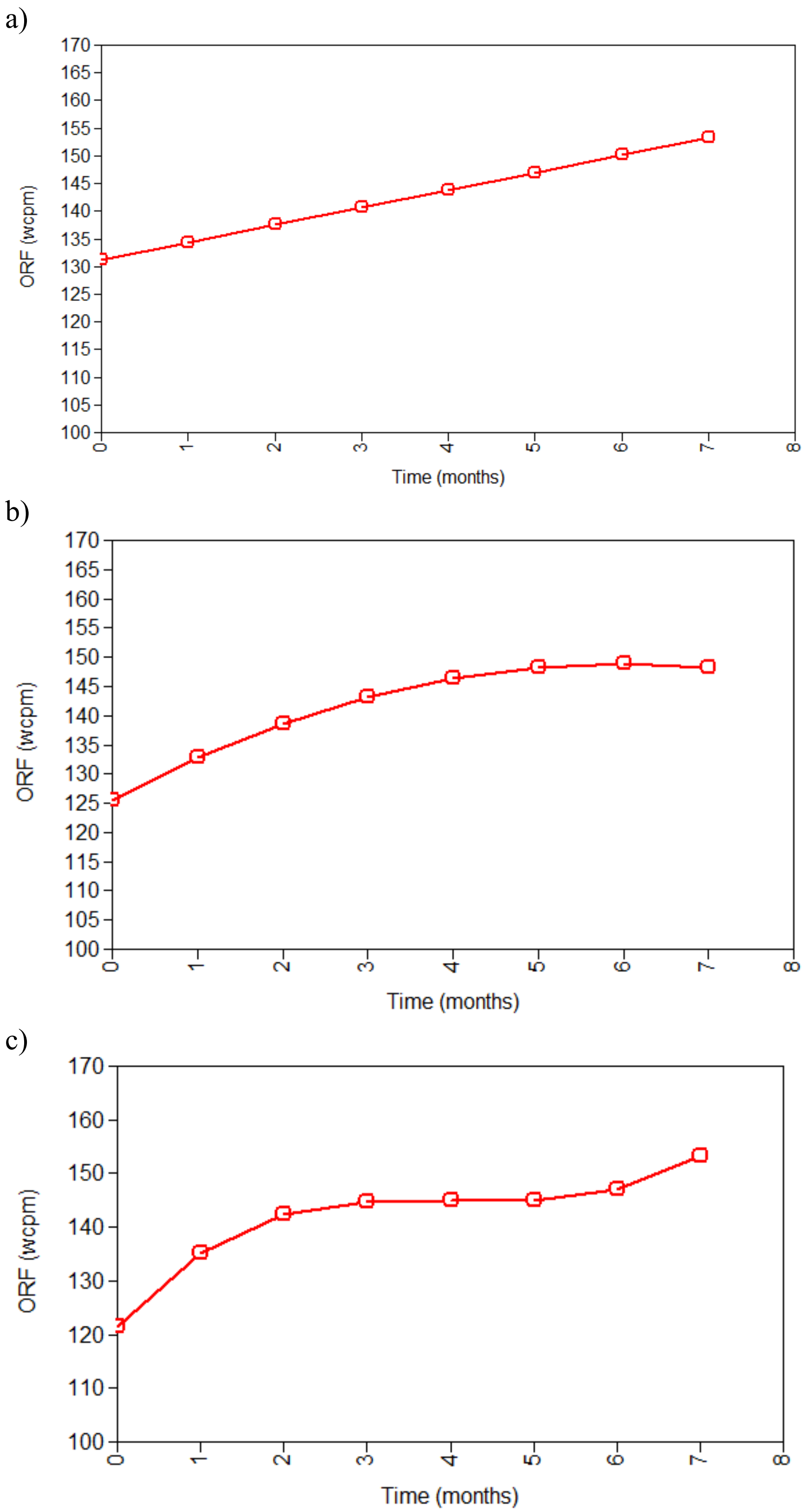

Figure 2. (a) Predicted (estimated) linear mean ORF growth across time. (b) Predicted quadratic mean ORF growth across time. (c) Predicted cubic mean ORF growth across time.

The three graphs in Figure 2 can be compared to the mean observed growth of the sample displayed in Figure 1, and statistical analysis can help determine which model best fits the data. Simply by eye-balling the grade 4 ORF trajectories, one might speculate that the cubic model in Figure 2c would best fit the observed data in Figure 1. Once the functional form of growth is selected, the variance in ORF of the intercept and slope can be explored; then, meaningful predictors can be added to explain these variances.

2.3 Testing Occasions

The importance of exploring the functional form of the longitudinal data, and a warning about the assumption of linear growth without empirical analysis has been emphasized. Given this context, it is not always true that one can model all three categories of functional form. Table 1 provides a guide to the exponential form, points of inflection, and minimum number of occasions needed for specific growth models. The exponential form refers to the exponent for the highest order polynomial in the equation. (Remember that a polynomial growth model must include all polynomial terms prior to the final order.) The points of inflection refer to the number of curves or bends in the predicted growth slopes. Finally, the minimum number of occasions needed specify how many occasions (i.e., observations, time points, or waves of data) are needed to statistically model a specific functional form. Note that for a longitudinal data analysis (linear, polynomial, or piecewise), one needs at least 3 occasions to model growth; having two occasions allows one to look only at *gain*, not *growth* as defined in this paper.

Table 1 only lists functions up to a cubic growth model, however, one could include as many exponents and points of inflection as desired, as long the minimum number of occasions is sufficient; the numbers in the columns simply continue in sequential order. Linear growth can be modeled with 4 or 5 occasions, or quadratic growth with 8 or whatever occasions, but there is a minimum for each form category.

Note that piecewise growth is not listed in Table 1, but for each piece of growth, one needs at least 3 occasions. Referring back to the earlier example of growth over two school years including the summer between (section 1.2.2), you would need at least 9 occasions (3 for year 1, 3 for summer, and 3 for year 2) to fully estimate a linear slope for each piece separately, and more occasions to fully estimate polynomial growth for each piece.

Table 1. Number of occasions needed for different functional form growth models.

Function	Exponential Form	Points of Inflection	Minimum Occasions Needed
Linear	^1	0	3
Quadratic	^2	1	4
Cubic	^3	2	5

2.4 Timing of Occasions

The metric for time is an important consideration, and must be sensible (Singer & Willett, 2003). Think of the scale of time as the x-axis of a graph showing growth, where the y-axis is the outcome variable. In Figures 1 and 2, months are used as the time metric, but this can be changed to suit the purpose and data structure. Time can be measured in terms of age or calendar, for example years, months, weeks, days, or hours. But units other than time can be represented on the x-axis. Take, for example, a car warranty, where a car is guaranteed based on the number of months *or* on the number of miles. In this example, the x-axis can either be time in months, or miles driven, which in a sense is a proxy for time.

In addition, data can be collected on a fixed schedule, in which all individuals are observed at the same time and occasions, or on a flexible schedule, in which individuals are observed at different times on different occasions (Singer & Willett, 2003). Individually-varying occasions demands more complex statistical models than does a fixed schedule, but can still be explored. The timing of the occasions relates to the degree to which the data is missing or incomplete (missingness), which is discussed in section 2.6.

2.5 Same Measure Over Time

A final assumption about longitudinal data analysis concerns the outcome variable, which must be a continuous, psychometrically robust variable whose values change systematically over time (Singer & Willett, 2003, p. 13). A psychometrically robust variable has strong precision of measurement, meaning strong reliability and small error of measurement. In general, the outcome must represent the same intended construct and maintain the same scale at every occasion. Note that one test may not necessarily represent the same construct at every age, and the amount of the outcome refers to the distance between scores being constant across time in order to measure growth at all.

2.6 Missing Data

Similar to other types of data analysis, missing data is a ubiquitous problem for longitudinal data analysis. Missing data is problematic for many reasons, including: (a) decreasing the representativeness of a sample (e.g., dropouts could be systematically different from non-dropouts in a study), (b) loss of statistical power to detect meaningful effects, (c) producing biased or inaccurate results, and (d) negatively affecting both internal and external validity of the study.

Missing data occurs in various patterns, such as participants refusing to participate, dropping out in the middle of a study (i.e., attrition), participating on selected occasions (i.e., participants are involved in some occasions but not others), and providing partial response by either omitting items, or answering some parts/types of items and not others.

Despite the challenges of missing data, there are statistical methods to control for missing data. For example, one can determine if the missing data problem can be ignored because the missingness is random and unrelated to other variables (for more information see Little & Rubin, 1987). Other ways of handling missing data include predicting, deleting, or imputing the missing values. In addition, some statistical software uses an estimation technique (i.e., maximum likelihood) that allows the inclusion of all students who have been observed on at least one occasion. It is important to note, however, that some of these methods are more complex and advantageous than others, and hold caveats.

2.7 Advantage over Gain Scores

Research has often addressed student change to understand how each student's learning or knowledge changes as an increment. That is, the difference between pre- and post-test or before and after an intervention; in other words, observing a student's initial score and subtracting it from the student's final score to obtain a measure of change from beginning to end. This method does not account for change as a continuous process, and there are limitations to this measurement of change (Willett, 1994). Analysis involving two occasions can result in a misleading estimate of change, because there is insufficient data to measure important details of students' learning trajectory over time. Changes may be occurring over time with a meaningful trajectory that can be explored by researchers, but two occasions do not provide an adequate method for studying growth (Willett, 1994).

3. Purposes / Analytic Methods

In this section two general purposes of longitudinal data analysis in education are discussed: describing or modeling growth and accountability. Several ways to represent growth graphically, including methods of exploratory descriptive growth are provided, as are several advanced statistical techniques for modeling growth involving Hierarchical Linear Modeling (HLM) and Structural Equation Modeling (SEM).

3.1 Exploratory Descriptive Growth

The graphic representation of growth using individual empirical growth plots and individual empirical growth plots inter-individual differences in growth is provided here.

3.1.1 Individual Empirical Growth Plots

According to Singer and Willett (2003), one of the simplest ways to observe change in individuals over time (i.e. growth) is to visually inspect individual empirical (or observed) growth plots. These plots are temporally sequenced graphs of individual empirical growth records (i.e., recorded data), and can be created using many major statistical packages, including SPSS (SPSS Inc., 2010), HLM (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2010), and Mplus (Muthén & Muthén, 1998-2007). Because viewing individual plots may be difficult to detect differences and similarities in growth, it is recommended to view sets of plots in a small number of panels. Each individual's empirical growth can be summarized using a trajectory applying either a nonparametric or a parametric approach. Here, nonparametric refers to smoothing trajectories without imposing a specific functional form, and parametric refers to trajectories that are summarized using a functional form such as linear, quadratic, or some other form of growth (for more information, see Chapter 2, Singer and Willett, 2003). It is important to note that these techniques are exploratory approaches to examining growth. In other words, these approaches do not offer statistical tests of significance, or rigorous methods by which to make predictions about future performance or to explain why trajectories occur as they do for different students.

Figure 3 shows empirical growth plots of eight grade 4 students in our data. Exploring individual plots can provide initial growth information. For example, these eight plots suggest that the students generally have an initial ORF score between 70 and 160. There are some students that start with a much higher initial ORF score, and some with a much lower score. Across the 8 time points, some students show more gradual positive growth in the middle of the year and then a slight decrease, some have growth patterns that are relatively stable, and others have growth that goes up and down throughout the year.

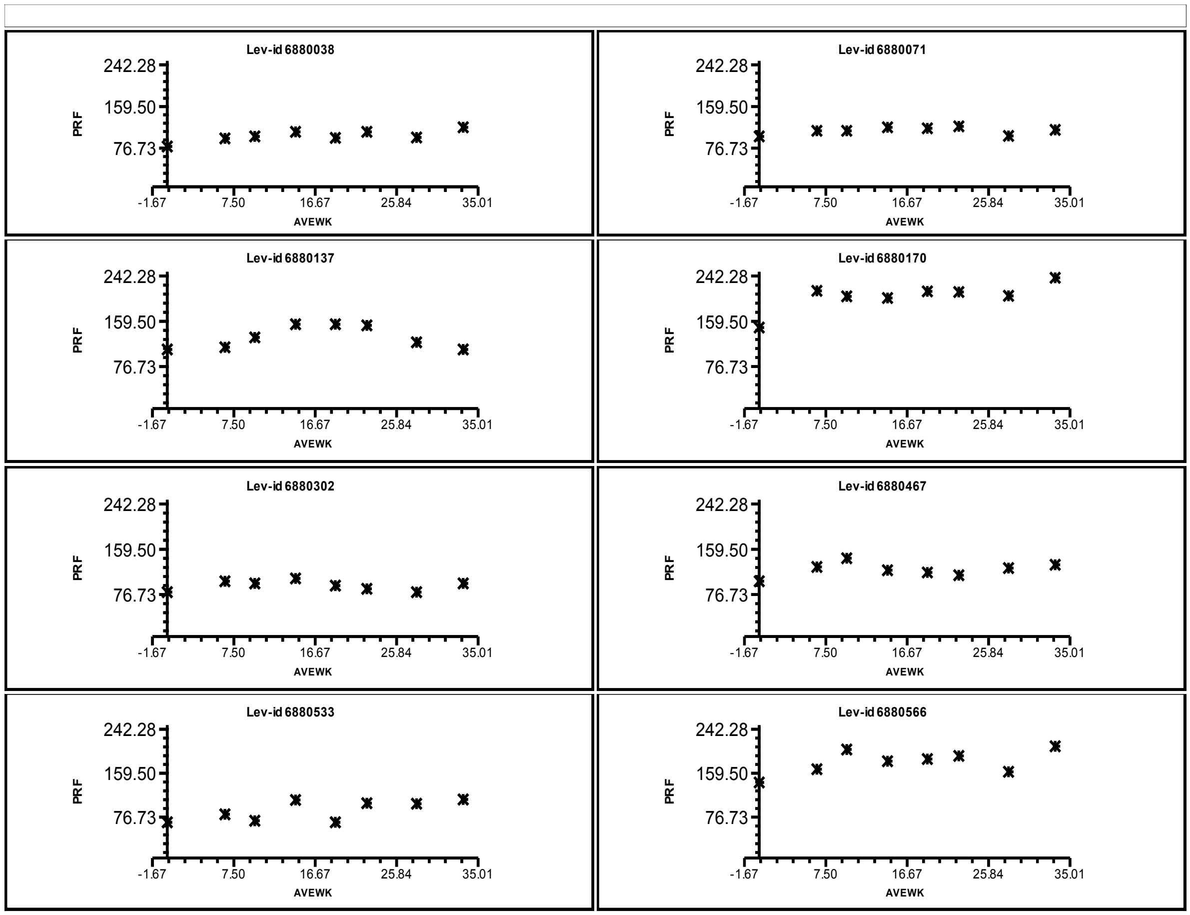

Figure 3. Empirical growth scatter plots of eight grade 4 students on the ORF measures in one year.

Figure 4 shows the nonparametric, smoothed growth trajectories (i.e. no specific functional form was imposed) of the same eight students. In other words, a smooth line was used to connect the eight time points for all eight students. When examining plots like this, it is important to consider the elevation or decline, shape, and slope of each curve.

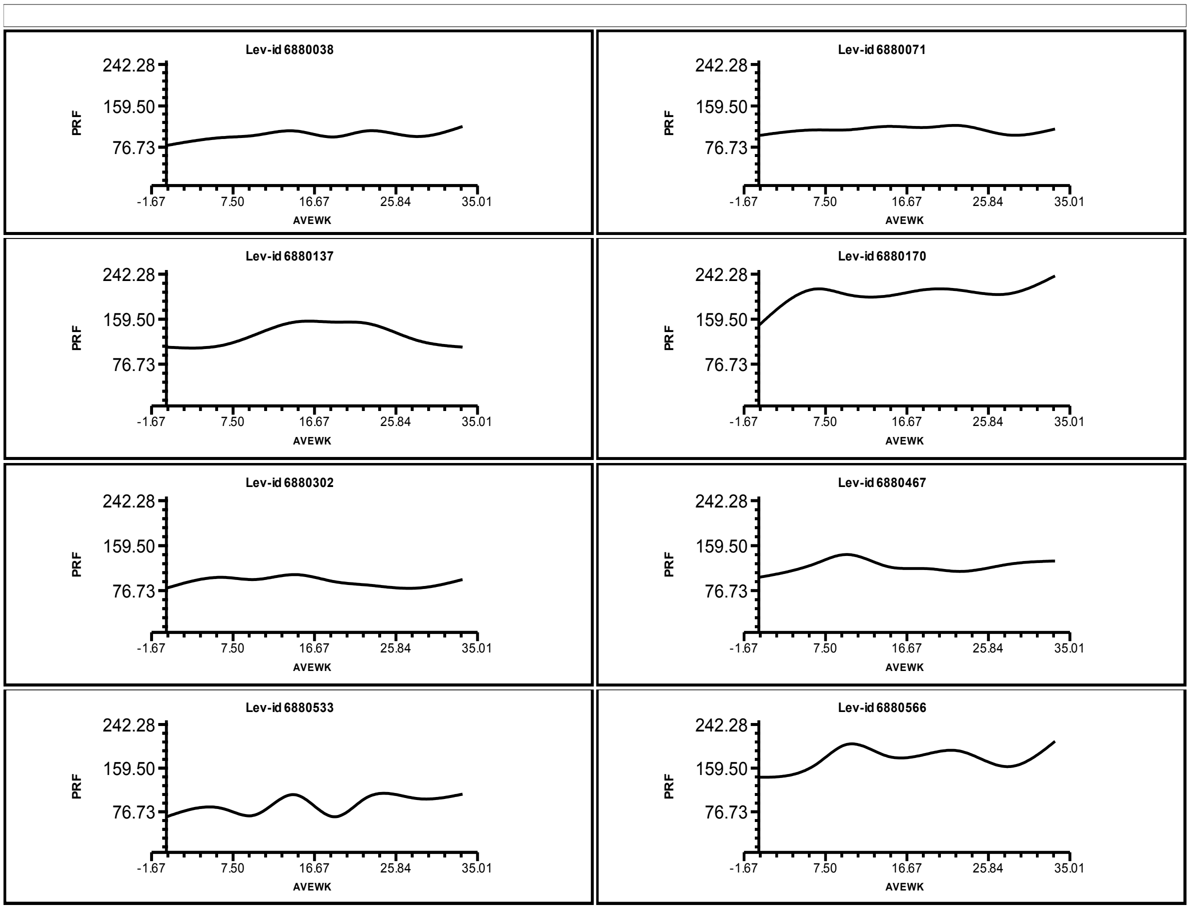

Figure 4. Smooth nonparametric individual growth plots of eight grade 4 students on the ORF measures in one year.

3.1.2 Inter-Individual Differences in Growth

To examine whether all individuals grow similarly or differently, inter-individual growth trajectories must be examined (i.e., differences in growth between students, or how trajectories vary across students). One way to examine this is to plot the set of smoothed individual trajectories onto a single graph. Figure 5 shows the observed graph plot of 50 randomly selected students in the grade 4 ORF data and the fitted average linear growth trajectory for the group. Note that the average growth trajectory in this plot is primarily used as a comparison with the observed individual trajectories, and the slope was constrained (or forced) to be linear, increasing at a constant rate over time. The “average” trajectory in red suggests that students’ ORF scores increase gradually across the academic year, with an average beginning ORF score of 131 words correct per minute (wcpm) and an average growth of 3.15 wcpm per month. However, the graphed individual observed trajectories in black suggest that there is substantial inter-individual difference in growth across the year, both at the intercept and slope. More specifically, some students displayed fluctuating growth, some fairly positive linear growth, and some quadratic growth.

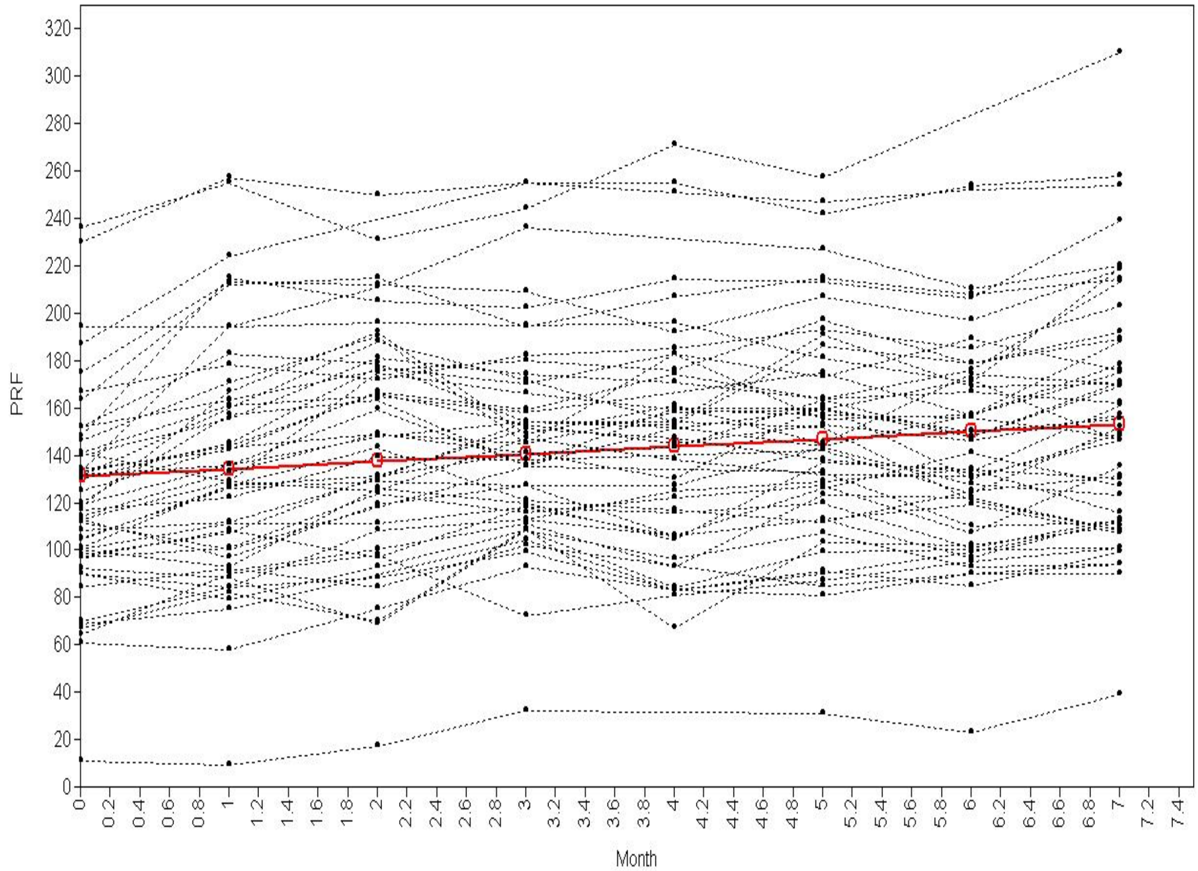

Figure 5. A collection of observed trajectories of 50 random grade 4 students on the ORF measures in one year, with an OLS average growth trajectory in red.

It may be also useful to explore the relation between growth and student characteristics that can be time-invariant (i.e., constant over time), such as ethnicity, or whether a student receives special education services (SPED).[2] Figure 6 shows the observed graph plot of 40% (73 students) randomly-selected students in the grade 4 ORF data separated by SPED status (blue lines represent students receiving general education instruction (GenED), and red lines represent students receiving SPED services). The observed trajectories for SPED students are generally lower than GenED students in both the intercept and growth across the year. Generally, plots like this are used to examine systematic differences in pattern.

[2] Special education status is considered to be constant over time in this example, as no student in the sample stopped receiving special education services during the year. Quite often, however, students demonstrate many different patterns of entering and exiting special education, which is an example of a time-variant variable.

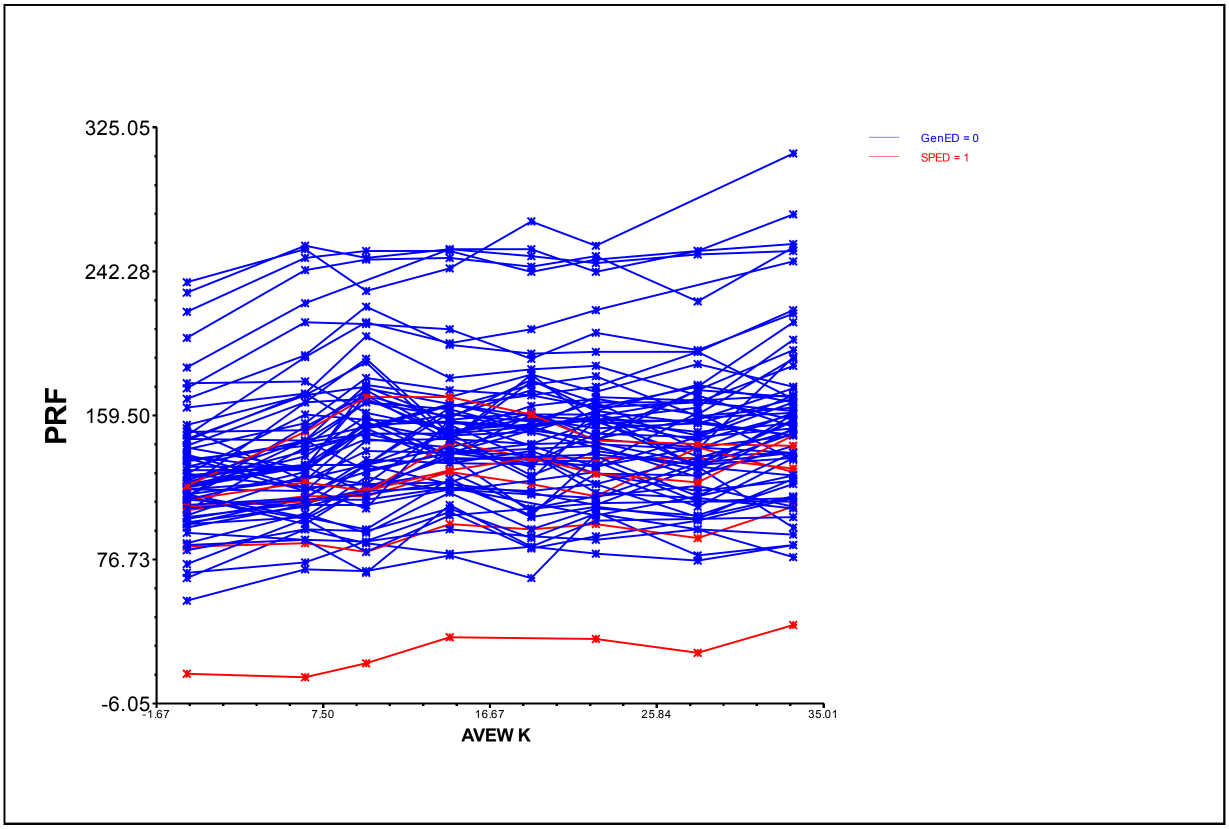

Figure 6. A collection of observed trajectories of 73 randomly selected grade 4 students on the ORF measures in one year, separated by the predictor SPED. GenED = 0 (blue) represents students receiving general education and SPED = 1 (red) represents students receiving special education services.

3.2 Modeling Growth

Several advanced statistical techniques involving Hierarchical Linear Modeling (HLM) and Structural Equation Modeling (SEM) are discussed here.

3.2.1 Regression

Regression is a common parametric model used to summarize individual empirical growth trajectories. Regression is used to determine the relation between independent and dependent variables for the purpose of either (a) trying to best *predict* of the outcome variable, or (b) identifying variables, and their relative importance, in *explaining* the outcome variable. Using the grade 4 ORF data, students' data are plotted in a graph, with time on the x-axis and ORF on the y-axis. A line is then applied to the data, as in Figure 5, that represents the *average* intercept and slope for all students, following the equation: $y = i + b\mathrm{X} + e$, where y is the outcome variable, i is the average intercept, b is the slope, or average growth over time, X is the independent time variable, and e is the residual, or the difference between the line and the observed data (different for each student). In general, regression allows the researcher to explore hypotheses about predicting or explaining initial status and growth, and estimate the relation of predictors to the outcome while controlling for the effects of other predictors. For example, one could explore the effect of receiving special education services on students' ORF trajectories while controlling for students' sex, socioeconomic status, and ethnicity.

3.2.2 Variance at Different levels

An advantage of longitudinal data analysis using more advanced statistical techniques (e.g., HLM and SEM, see below) is that differences in growth within and between students can be explored. One would expect there to be a significant amount of variance within student growth trajectories, because the scores are changing across time. Once one has selected the most appropriate functional form, that is, the differences "within-students," one can explore different time-varying variables that may influence growth. These are variables that are not constant over time, rather change as a function of time. For example, students' sex is constant over time, but students' school absences do change over time. In the example data provided in section 1.1, if a researcher records the number of absences at the same time as she records ORF scores, then she can use this variable to try to explain some of the within-student variance.

Once the most appropriate functional form is selected and any time-varying variables of interest are modeled, then the differences between students can be explored. This means that the students are assumed to have the same general functional form of growth, but individual students can have different values of their individual growth parameters. Thus, individual students can differ based on the intercept (initial status) and slope(s) (e.g., linear, quadratic, cubic). One can also explore if differences in intercept and slope are related to variables of interest, such as socioeconomic status or special education status. For example, is there a relation between ORF initial status and special education status, and is there a relation between rate of ORF growth and special education status? What about SES, or type of reading program, or access to reading materials at home? In addition, variance can be modeled at a school, district, or state level as well. This means that school variables that may have an influence on ORF status and growth can be analyzed. For example, does being in a Title 1 school have an effect on ORF status or growth? What about public versus private schools, or schools with teachers with more years of experience? All of these research questions can be introduced into statistical models of longitudinal data.

3.2.3 Univariate (e.g., HLM) and Multivariate (e.g., SEM)

One of the most prominent analytic methods for modeling growth is Hierarchical Linear Modeling (HLM). The term HLM applies to the analytic technique, as well as the popular software program that executes these analyses (Raudenbush & Bryk, 2002). The term *hierarchical* refers to the structure of the data, in which *levels* of data are nested within other levels. For example, students (level-1) are nested in schools (level-2), meaning that there are separate groups of students who are nested within separate schools. This can be expanded to schools (level-2) that are nested in different districts, or states. This kind of data structure can also be applied to longitudinal data analysis, in which case testing occasions (level-1) are nested within students (level-2), meaning that each student has taken any number of repeated tests over time.

In longitudinal data analysis in education this hierarchical, or nested, data structure is important because it separates the variance component into a within-student component (level-1) and a between-student component (level-2). OLS regression assumes that the residuals across students are independent of one another; that is, there is assumed to be no correlation, or similarities, between students that is not accounted for by the model. But this assumption may not always be accurate, as there are often similarities among students who attend the same school. Since schools draw students from the same neighborhood, students often share similar background contextual characteristics such as SES, and since these students are taught by the same teachers, they also share similar educational contextual characteristics. Thus, an advantage of HLM is that it can partition the variance at these different levels to account for these shared contextual factors.

HLM also has the flexibility to test the effects of predictor variables on the initial status and the change in slope, and offers a statistical test for both individual effects and group variation in growth. Another advantage is that HLM (and other software) can include all students who have been observed on at least one occasion, and as long as the data are missing at random then results can be interpreted as if there were no missing data.

3.2.4 Latent Growth Modeling

Another prominent analytic method for modeling growth is latent growth modeling, a special case of the more general approach known as structural equation modeling (SEM). Latent growth modeling is also known as latent growth curve modeling or latent growth curve analysis. SEM is a statistical technique used to test and estimate causal relations among observed and unobserved variables. For the review of SEM in this series, please see Anderson, Patarapichayatham, and Nese (2010).

In general, compared to HLM, SEM offers an alternative approach to model specification and estimation (Singer & Willett, 2003). In terms of model specification, HLM takes a univariate approach to longitudinal data analysis, in which the outcome is a single variable so that each student has a separate row of data for each occasion. SEM takes a multivariate approach, in which there multiple outcomes so that each student has a separate column for each testing occasion. Figure 7 shows an example of the data structure for each approach. Figure 7a represents the univariate approach, where you can see the ORF (outcome) data for two students, across months, and Figure 7b represents the multivariate approach, where you can see the data for nine students, and each of the variables with the ORF prefix represents a separate testing occasion. Notice that the structure of "stuid = 1" is the same across the data, reading down in Figure 3a and across in Figure 5b.

a)

stuid	month	prf
1	0	135
1	1	134
1	2	154
1	3	158
1	4	153
1	5	154
1	6	178
1	8	172
2	0	160
2	1	-999
2	2	-999
2	3	154
2	4	155
2	5	143
2	6	-999
2	8	160

b)

stuid	prfsb	prf.pm3s	prf.pm4s	prfsm	prf.pm7s	prf.pm8s	prf.pm9s	prfs
1	135	134	154	158	153	154	178	
2	160	-999	-999	154	155	143	-999	
3	149	153	147	141	156	142	143	
4	223	222	222	227	223	201	230	[illegible]
5	156	170	180	170	182	165	175	
6	128	121	138	133	138	117	131	[illegible]
7	185	170	163	170	162	141	169	[illegible]
8	103	115	130	129	133	-999	-999	[illegible]
9	180	192	197	178	256	176	212	[illegible]

Figure 7. (a) Univariate data structure used in HLM, and (b) multivariate data structure used in SEM.

With the multivariate data structure, what is a two-level model in HLM (occasions within students) is a one-level model in SEM. And thus a three-level model in HLM is then a two-level model in SEM, and so on.

Perhaps the most fundamental and important difference between the two approaches, is that latent growth modeling using an SEM approach offers more flexibility in its model specification based on its approach to estimation. SEM is also known as covariance structure analysis, given its model estimation procedures. SEM uses an estimation procedure based on the probability distributions of the variance-covariance and mean structures of the data. In an SEM framework, the intercept and slope(s) are treated as unobserved variables, or random effects. HLM, on the other hand, generally uses an estimation procedure based on the probability distribution of the outcome variable based on the random effects and parameters. In general, SEM offers more flexibility with access to the variance-covariance and mean structures. SEM provides an opportunity to model: (a) differences in residual variances over time (as opposed to the HLM default that assumes equal residual variance across all occasions); (b) correlated residuals over time that can be specified to the researchers' needs; (c) regressions among the outcomes over time; (d) growth modeling as part of a larger latent variable model; (e) growth modeling of factors measured by multiple indicators; (f) regressions among growth factors and random effects; (g) estimated time score models; and (h) the general flexibility to fix, constrain, and/or correlate variances and means.

3.2.5 Modeling Heterogeneity in Growth Patterns

As introduced in the previous section, random coefficient modeling (e.g., Raudenbush & Byrk, 2002) and latent growth curve modeling (Meredith & Tisak, 1990) are common growth models used to study growth in education research. These modeling techniques, however, assume that individuals come from one population with a single average intercept (i.e. starting point) and one average growth trajectory, which may not always be reasonable. Alternative approaches such as growth mixture modeling (GMM) and latent class growth analysis (LCGA) relax the single population assumption and allow different classes of individuals to vary around different mean growth curves. GMM assumes class variant parameters on intercept (starting point) and slope (growth), including their fixed and random effects, and error variances. LCGA is more restrictive, and assumes that there is zero within-class variance on the intercept and slope. These models may be more reasonable when describing growth, especially when there are theoretical reasons to believe that growth patterns may be heterogeneous.

There are educational implications when subgroups of students exist in the population with different growth trajectories. For example, a Head Start program study by Kreisman (2003) found that there were two distinct growth patterns, with a *majority* class that displayed below-average initial reading and math scores and exhibited negative growth over time in both subjects, and a *minority* class that displayed below-average initial reading and math scores, but exhibited steady positive growth. These results would have important practical implications, such as the use of different intervention strategies to address the needs of the two distinct groups of students.

3.2.6 An Alternative Approach to Modeling Non-linear Growth using Benchmark Data

Growth may not always be linear. Quadratic and cubic growth models are common methods, but involve complicated modeling techniques used to examine non-linearity in the growth curves. However, such polynomial growth models present challenges especially for benchmarking data that uses three data points. Kamata, Nese, Patarapichayatham, and Lai (in preparation) suggested a combination of estimating the time score and a growth mixture analyses approach when only three occasions are available for studying growth. The estimated time score model aims to examine non-linearity in the data. Although this approach is typically used in a linear growth model setting (with at least two time scores fixed), one or more time scores can be set as free parameters. For example, using a data set with nine time points, Heck and Takahashi (2006) found a non-linear declining growth curve by freely estimating factor loadings for the last two time points. Thus, this procedure allows for the estimation of non-linear growth, and is an approach to help determine the functional form of growth.

In the case of benchmarking data with fall, winter, and spring scores available, the factor loadings can be specified as $(0, 1, \lambda)$, where λ is the parameter to be estimated. If λ is estimated to be 2, the growth would be linear because of the equal time intervals. If λ is not 2, growth is then non-linear. If $\lambda < 2$ in our example, it would indicate that growth during the winter-spring terms is steeper than the growth during fall-winter terms. On the other hand, if $\lambda > 2$, it would mean that growth for the winter-spring term is flatter than the fall-winter term. In addition to modeling non-linearity in growth, Kamata et al. (in preparation) were able to examine for differences in growth patterns using growth mixture modeling, which could provide additional information about the relation between the two growth patterns of fall-winter and winter-spring terms.

3.3 Accountability

The purpose of accountability in education is to provide information on students, schools, and school systems to improve learning and assist policy making (Sanders & Horn, 1994). Accountability models generally focus on the product and not the process by which it is achieved. Different accountability models are discussed in the following sections.

3.3.1 Models for Determining Adequate Yearly Progress (AYP)

Status and improvement models have been primary federal accountability approaches for determining if schools have met the adequate yearly progress (AYP) status, a requirement of the No Child Left Behind Act of 2001. The status model compares the percentage of students who are proficient based on the current year's test scores to the state's annual targets. Schools that do not meet the AYP status can use the *Improvement* model, which allows them to receive credits if there is at least a 10% decrease of students in a particular subgroup who are not proficient from the previous year to the next. Finally, the *Index* model is an alternative for schools to receive partial credit for student subgroups scoring below proficient if the current year students were to make gains (e.g., below basic to basic) over the previous year students in the same grade. It is important to note, however, that these models do not necessarily compare the same cohort of students. Both the Status and Improvement models have been criticized for not giving credit to schools that may have shown academic improvement for students at different levels of proficiency. Finally, neither model considers that teachers, schools, or districts may have no control over students who are low performing at a given time point. To address these critiques, the Growth Model Pilot program in 2005 and the Differentiated Accountability pilot program (2008) were initiated. These programs suggest a trend toward valuing academic growth and less emphasis on average performance of a school, district, or state.

3.3.2 Value-Added Models

Value-added models (VAM) are used to evaluate the effectiveness of teachers and schools using information from students' academic growth (Braun, 2005). At least two years of data (student and classroom or school) that are matched to either the teacher or the school data are required for VAMs in order to estimate the contributions of schools or teachers to student academic growth. VAMs yield a number associated with each teacher or school, which is then used to compare how different a teacher's performance is from the performance of an average teacher, given the average growth of the students in the class.

VAMs assume that the effect of a teacher is constant on all of the teacher's students in a given subject and year, with the teacher's effect remaining constant or diminishing throughout the year. It is also assumed that teachers have access to equal resources and share similar academic goals in their classes. Although these assumptions could be plausible, they should be tested.

Despite VAMs' popularity as a statistical tool to evaluate teacher and school effectiveness for accountability purposes, there are many practical and statistical concerns. Questions on the fairness of using VAMs arise because students' academic growth may be influenced by many factors, including classroom placement with a teacher, as well as school context and practice (i.e. school resources and implementation of school and district policies) that are not within the control of teachers. Additionally, VAM models, just like many statistical models, cannot offer causal explanations due to lack of random pairings among students and teachers. Finally, longitudinal data that VAMs model require generally cannot avoid missing data, which could subsequently bias the estimated effectiveness of teachers (i.e. teacher effects).

VAMs are a family of models that estimate the contributions of schools or teachers to student academic growth. Some VAMs that are being used currently include the Educational Value-Added Assessment System (EVAAS), the Dallas Value-Added Accountability System (DVAAS) and the Rate of Expected Academic Change (REACH) (for more information about these VAMs, see Sanders, Saxton, & Horn, 1997; Webster & Mendro, 1997; and Doran & Izumi, 2004).

4. Conclusion

The intent of this paper is to provide an overview of longitudinal data analysis in education for practitioners, administrators, and other consumers of educational research. The examples and descriptions presented here are meant as a primer regarding the some of the purposes, benefits and limitations, and analyses used in longitudinal data analysis in education. The first purpose of longitudinal data analysis is to determine the functional form of growth. The second purpose, which often follows the first in a study of growth, is to examine the relation between the trajectory and variables of interest. There are several approaches to this purpose, including but limited to, regression, HLM, and SEM. Accountability models are also used to assist policy making by focusing on the product of the system as parceled by schools or teachers. This paper is intended as a primer on longitudinal growth modeling in education for consumes of such research such as school administrators. For a more thorough discussion of any area introduced in this paper, please see the books by Singer and Willett (2003), Braun (2005), Raudenbush and Bryk (2002), and articles by Meredith and Tisak (1990), Daniel McCaffrey and William Lockwood for information on value-added models.

References

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM© online progress monitoring assessment system.* Eugene, OR: Center for Educational Assessment Accountability. Available at http://easyCBM.com.

Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Princeton, NJ: ETS.

Council of Chief State School Officers Accountability Systems and Reporting State Collaborative. (2009). Guide to United States Department of Education growth model pilot program 2005-2008. Retrieved from http://www.ccsso.org/Documents/2009/Guide_to_United_States_2009.pdf

Deno, S. L., Fuchs, L. S., Marston, D. B., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507–524.

Differentiated Accountability States (2008) http://www2.ed.gov/admins/lead/account/differentiatedaccountability/dapstates.html

Doran, H. C., & Izumi, L. (2004). *Putting education to the test: A value added model for the state of California (Technical Report).* San Francisco, CA: Pacific Research Institute.

Gast, D. L. (2010). *Single-subject research methodology in behavioral sciences*. New York, NY: Routledge.

Heck, R. H., & Takahashi, R. (2006). Examining the impact of proposition 48 on graduation rates in division 1A football and program recruiting behavior: testing a policy change model. *Educational Policy*, *20,* 587-614.

HLM Version 7 (2010). Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T. & du Toit, M. Scientific Software International, Inc.

Kamata, A., Nese, J. F. T., Patarapichayatham, C., & Lai, C. F. (in preperation). Methodological note on modeling nonlinear growth for benchmarking data.

Kreisman, M. B. (2003). Evaluating academic outcomes of Head Start: An application of general growth mixture modeling. *Early Chidlhood Research Quarterly, 18,* 238-254.

Little, R. J. A., & Rubins, D. (1987). *Statistical analysis with missing data*. New York: Wiley.

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

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus User's Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén.Nese, J. F. T., Biancarosa, G., Anderson, D., Lai, C. F., & Tindal, G. (in press). Within-year oral reading fluency with CBM: A comparison of models. *Reading and Writing*. doi: 10.1007/s11145-011-9304-0

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical linear models* (Second Edition). Thousand Oaks: Sage Publications.

Sanders, W. L., & Horn, B. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8,* 299-31.

Sanders, W. L., Saxton, A. & Horn, B. (1997). The Tennessee value-added assessment system: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press, Inc.

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

SPSS version 19 (2010). SPSS Inc.

Webster, W. & Mendro, R. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press, Inc.

Willet, J. B. (1994). Measuring change more effectively by modeling individual change over time. In T. Husen & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education,* (second ed.). Elmsford, NY: Pergamon Press.