Dissertations and Theses in Statistics   Statistics, Department of

8-2010

Estimating Teacher Effects Using Value-Added Models

Jennifer L. Green
University of Nebraska-Lincoln, jgreen@montana.edu

ESTIMATING TEACHER EFFECTS USING VALUE-ADDED MODELS

by

Jennifer L. Green

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professor Erin E. Blankenship

Lincoln, Nebraska

August, 2010

ESTIMATING TEACHER EFFECTS USING VALUE-ADDED MODELS

Jennifer L. Green, Ph.D.

University of Nebraska, 2010

Advisor: Erin E. Blankenship

Value-added modeling is an alternative approach to test-based accountability systems based on the proportions of students scoring at or above pre-determined proficiency levels. Value-added modeling techniques provide opportunities to estimate an individual teacher's effect on student learning, while allowing for the possibility to control for the effect of non-educational factors beyond a school system's control, such as socioeconomic status. However, numerous considerations exist when using value-added models to estimate teacher effects and defining what the *teacher effects* really describe.

Chapter 2 provides an introduction to value-added methodology by describing several value-added models available for estimating teacher effects and their respective advantages and disadvantages. Modeling variations and their impact on estimated teacher effects are also discussed in addition to the various statistical and psychometric issues associated with estimating value-added teacher effects.

Because value-added analyses require high-quality longitudinal data that are often not available, Chapters 3 and 4 propose methodology for analyzing less-than-ideal assessment data. Chapter 3 proposes value-added methodology for analyzing longitudinal student achievement data not on a single developmental scale and addresses issues arising

when using a layered, longitudinal mixed model to analyze gains in standardized scores. The chapter also discusses methods for estimating teacher effects on student learning before and after entering professional development programs and applies these methods of analysis to achievement data.

Chapter 4 describes the use of curve-of-factors methodology to analyze longitudinal achievement data collected from two differently scaled assessments in a single year and subject, such as mathematics. Assuming data come from a curve-of-factors model structure, a simulation study evaluates the performance of the proposed curve-of-factors model in its ability to accurately rank teachers in the presence of either complete or missing test data and compares it to the performance of the Z-score methodology proposed in Chapter 3.

ACKNOWLEDGEMENTS

I am grateful for all of the guidance and encouragement my advisor, Dr. Erin Blankenship, has provided me throughout this whole process. Without her support and unending patience, this dissertation would not have reached its full potential. I also thank my other committee members, Dr. Walt Stroup, Dr. Steve Kachman, Dr. Ruth Heaton and Dr. Jim Bovaird, for their help and support throughout my academic program. I am eternally grateful for my committee's willingness to spend countless hours answering questions and helping me solve the seemingly unsolvable. In addition, I thank Dr. Jim Lewis for all of the support and guidance he has provided over the past three years. I appreciate the faculty's sincere enthusiasm for my research and their investment in my education and my future. I especially thank Joe, Brianna, my parents, my family and my friends for their patience, support, understanding and encouragement during this time. Without you, this dissertation could not have been written. To all: thank you for believing in me when I needed it the most.

GRANT INFORMATION

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Over the past several years, there has been a national effort to hold students to higher academic standards. This effort includes holding states accountable for assessing measurable student outcomes. Value-added modeling is an alternative approach to test-based accountability systems interested in the proportions of students scoring at or above pre-determined proficiency levels. Value-added modeling techniques estimate the contribution of educational factors, such as teachers, to growth in student achievement, while allowing for the possibility to control for the effect of non-educational factors beyond a school system's control, such as socioeconomic status. Value-added modeling methods provide opportunities to estimate the proportion of variability in achievement or student growth attributable to teachers, as well as estimate an individual teacher's effect on student learning.

School districts and policymakers desire to use teacher effect estimates for a variety of purposes, from informing educational systems how students are affected by current practices and conditions to making high-stakes decisions regarding teacher salary and/or employment. These estimates are also desired to evaluate the effectiveness of professional development programs. However, even though value-added modeling methods infer causal effects of teachers on student growth, the assessment data are not obtained from randomized, experimental studies. Consequently, several obstacles need to be addressed before value-added modeling should be used in these ways.

1.1 Value-Added Models for Estimating Teacher Effects

Chapter 2 serves as a background and introduction to value-added methodology. Several value-added models available for estimating teacher effects are described, as are the models' respective advantages and disadvantages. Modeling variations, such as the use of layered versus non-layered design matrices and the specification of teacher effects as fixed or random, are also discussed, and the impact of such considerations on estimated teacher effects is explained in detail using an example provided by Wright and Sanders (2008). The various statistical and psychometric issues associated with estimating value-added teacher effects are highlighted, providing a summary of the current state of value-added modeling research and recommendations for future work.

1.2 Estimating the Impact of a Professional Development Program on Student Learning

Professional development programs focus on preparing teachers to meet the recent initiatives on improving the quality of student instruction, but rigorous evaluations are needed to determine whether these programs are actually effective. Value-added modeling techniques provide opportunities to estimate the relationship between teacher development and student learning, but most require student achievement data to be on a single developmental scale over time (McCaffrey, Lockwood, Koretz, & Hamilton, 2003). Typically, available assessment data do not meet such requirements, limiting analyses that can be conducted. Chapter 3 proposes alternative value-added methodology, specifically the use of Z-scores, for analyzing less-than-ideal longitudinal student achievement data collected from a mixture of norm- and criterion-referenced assessments to estimate the impact of a professional development program on student learning. The

chapter discusses methodology for estimating teacher effects on student learning before and after entering professional development and addresses issues arising when using a layered, longitudinal linear mixed model to analyze gains in standardized scores. The methodology is applied to data collected from a mathematics professional development program in mathematics education, the Math in the Middle Institute Partnership (M^2), and the results are discussed.

1.3 Using Parallel Processing Methodology to Estimate Teacher Effects

Few studies have addressed how to use value-added models to analyze achievement data not on a single developmental scale (Green, Smith, Heaton, Jiao, & Stroup, under review; Rivkin, Hanushek, & Kain, 2005), and even fewer, perhaps none, have discussed how to use information from multiple instruments in a single year that are on different scales, potentially both within and between instruments over time. When modeling multiple outcome measures, instead of a single measure across time, parallel process, or multivariate, growth curve models can estimate the relationship between the growth trajectories for each of the parallel measures and allow researchers to investigate changes in latent factors over time instead of changes in observed scores. Chapter 4 describes the use of parallel processing, specifically curve-of-factors, methodology to analyze longitudinal student achievement data collected from two different assessments in a single subject, such as mathematics, and estimate teachers' effects on student learning. Assuming data come from a curve-of-factors model structure, a simulation study evaluates the performance of the proposed curve-of-factors model in its ability to

accurately rank teachers in the presence of either complete or missing test data and compares it to the performance of the Z-score methodology proposed in Chapter 3.

Chapter 2

Value-Added Models for Estimating Teacher Effects

2.1 Introduction

Since the enactment of No Child Left Behind (NCLB) (2001), education systems, in theory, have held students to higher academic standards, and states are accountable for assessing measurable student outcomes. States receiving Title I funds for improving the academic achievement of disadvantaged students must require schools to make adequate yearly progress (AYP). While states are given latitude with regards to what is meant by "adequate," in general, this means the proportion of students achieving pre-determined proficiency levels on state assessments is expected to increase annually until all students in particular grades are deemed proficient or higher.

Value-added modeling is an alternative approach to test-based accountability systems interested in the proportions of students scoring at or above pre-determined proficiency levels. Value-added modeling techniques estimate the contribution of educational factors, such as teachers, to growth in student achievement, while allowing for the possibility to control for the effect of non-educational factors beyond a school system's control, such as socioeconomic status. Value-added modeling methods provide opportunities to estimate the proportion of variability in achievement or student growth attributable to teachers, as well as estimate an individual teacher's effect on student learning. When these methods identify large differences in teacher effectiveness, they also have the potential to help researchers identify what characteristics highly effective

teachers possess and motivate informed improvements in education (McCaffrey, Lockwood, Koretz, & Hamilton, 2003).

Teacher effect estimates can be used for a variety of purposes, from informing educational systems how students are affected by current practices and conditions to making high-stakes decisions regarding teacher salary and/or employment. However, even though value-added modeling methods infer causal effects of teachers on student growth, the assessment data are not obtained from randomized, experimental studies. Consequently, several limitations exist when defining what *teacher effects* really describe. Defining teacher effects requires identifying to what a particular teacher's impact on a student's growth in achievement will be compared, such as other teachers in the school, district, or entire state. The definition also depends on the outcomes used to measure achievement; the scope and purpose of the instruments can limit what is measured and, consequently, restrict the part of a teacher's total impact on a student that can be estimated (McCaffrey et al., 2003). Other factors affecting students' growth in achievement, such as characteristics of classrooms and schools, can be confounded with teacher effect estimates, so the purpose for obtaining such estimates needs to be clearly defined and should dictate how precisely the effects need to be estimated. Typically, teacher effects merely account for unexplained classroom-level heterogeneity (Lockwood, McCaffrey, Mariano, & Setodji, 2007).

Studies investigating value-added teacher effects provide evidence teachers have differing effects on student learning (Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti, & Miller, 2002; Wright, Horn, & Sanders, 1997) that persist over time (Sanders & Rivers, 1996), but these studies have shortcomings (McCaffrey et al., 2003). Section 2.2

describes proposed value-added models for estimating teacher effects and discusses their respective advantages and disadvantages. Section 2.3 covers the impact of different modeling decisions on the estimation of teacher effects, and Section 2.4 highlights various statistical and psychometric issues associated with estimating such effects. The chapter concludes with a summary of the current state of value-added modeling research and recommendations for future work.

2.2 Value-Added Models

Multiple authors have championed the use of value-added models to analyze longitudinal student achievement data (Doran, 2003; Drury & Doran, 2003; Hershberg, Simon, & Lea-Kruger, 2004; Lissitz, 2005; Sanders, Saxton, & Horn, 1997). These methods fall into three categories: covariate adjustment models, gain score models and multivariate models (McCaffrey et al., 2003).

2.2.1 Covariate Adjustment Models

Covariate adjustment models, for example,

$$y_{ig} = \mu_{ig} + \beta y_{i,g-1} + T_g + e_{ig}, \tag{2.1}$$

regress each student's current achievement score, y_{ig}, on his or her prior score, $y_{i,g-1}$, for the year of data collection, $g = 1, 2, 3, \ldots, m$. The student-specific mean, μ_{ig}, adjusts for factors affecting a student's level of achievement, such as free-and-reduced lunch and English Language Learner (ELL) identifiers. It can also account for many other factors, including characteristics of schools. The teacher effect, T_g, estimates the current year teacher's impact on a student's outcome. The residual errors, e_{ig}, are assumed to be

normally distributed with mean zero and variance σ_{eg}^2 and independent of the prior year

scores and teacher effects.

Covariate adjustment models are easy to specify and fit, and they do not require

performance on measures used in successive years to be placed on a single

developmental scale so growth can be measured across grades or ages. This is

particularly beneficial for school systems using a mixture of norm-referenced and/or

criterion-referenced tests, where reported student scores from the two types of

instruments reflect different measurements: either relative academic performance or

proficiency on predetermined criteria, respectively. Teacher effects from prior years are

embedded in the previous year's score, so the effects persist indefinitely even though they

are not explicitly specified in subsequent years' models. However, information is lost

about student performance by estimating models separately for each year, so critics argue

these methods are not really measuring student growth. Covariate adjustment methods

also require complete student records, so missing student outcomes must either be

imputed or removed from the analysis. When data are not missing completely at random,

list-wise deletion can lead to biased estimates of all effects. In general, covariate

adjustment models are easy to work with, but potentially over-simplify the complexity of

student growth over time.

2.2.2 Gain Score Models

Gain score models,

$$d_{ig} = y_{ig} - y_{i,g-1} = \mu_{ig} + T_g + e_{ig}, \tag{2.2}$$

treat the difference between two successive scores, d_{ig}, for student i as the response for

the gth year of data collection. The student-specific mean, μ_{ig}, adjusts for factors

affecting a student's growth from one year to the next. It can account for many factors, including characteristics of schools. The teacher effect, T_g, estimates the current year teacher's impact on a student's growth. The residual errors, e_{ig}, are assumed to be normally distributed with mean zero and variance σ^2_{eg} and independent of the teacher effects.

Gain score models are also easy to specify and fit. These methods model students' gains in scores, so time-invariant factors, such as gender, race and poverty level, affecting a student's level of achievement need not be estimated. Prior year teacher effects are not typically specified in the model, which assumes they persist undiminished over time. Although "covariate" methods do not require tests to be on a single developmental scale, "gain" methods do, so changes in performance are not confounded with changes in tests (McCaffrey et al., 2003). In addition, gain score models require complete student records and lose information about student growth by assuming pairs of gains for the same student are independent. Overall, gain score models are easy to work with and explicitly model student growth in achievement, but they have stringent scale requirements and potentially over-simplify the complexity of student growth over time.

2.2.3 Multivariate Models

Multivariate methods jointly model all student scores, including relationships between each student's set of outcomes. These approaches also accommodate missing data, making efficient use of all available information. Specifying a multivariate model provides more flexibility, allowing the exploration of several assumptions, such as the persistence of teacher effects and the residual covariance structure of student outcomes. In some instances, these models are robust to omitted covariates, but they are

computationally intensive and require much more in the way of computing resources than either the gain score or covariate adjustment methods. Even though multivariate methods are often recommended, they are not widely adopted because the required resources and high-quality longitudinal data are not readily available. Three common multivariate approaches include the cross-classified model (Raudenbush & Byrk, 2002), the Education Value-Added Assessment System (EVAAS) teacher model (Sanders et al., 1997), and the variable persistence model (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004).

2.2.3.1 Cross-Classified Models

Hierarchical linear models (HLMs) can model students' growth over time, but when assessing teachers' influence on rates of learning, the models require each outcome to belong to only one student, who in turn remains in a single teacher's classroom over the course of the study (Raudenbush & Byrk, 2002). The nested structure required by HLMs is a limitation when studies want to model students' growth over the course of several years and, subsequently, several teachers. Rather, students who share the same teacher(s) in one year do not share the same teacher(s) in the next year, resulting in a cross-classified structure (Rasbash & Browne, 2008; Raudenbush & Byrk, 2002); HLMs do not model this additional complexity.

Cross-classified models, for example,

$$
\begin{aligned}
y_{i0} &= (\mu + m_i) + T_0 + e_{i0} \\
y_{i1} &= (\mu + m_i) + (\beta + b_i) + T_0 + T_1 + e_{i1} \\
y_{i2} &= (\mu + m_i) + 2(\beta + b_i) + T_0 + T_1 + T_2 + e_{i2} \\
y_{i3} &= (\mu + m_i) + 3(\beta + b_i) + T_0 + T_1 + T_2 + T_3 + e_{i3}
\end{aligned}
\tag{2.3}
$$

model scores, $y_{i,(g-1)}$, for student i at the $g = 1, 2, 3, 4$ year of data collection as a function of the average achievement, μ, and the average learning rate, β. The student-

specific intercepts, m_i, and slopes, b_i, are assumed to be independent across students and normally distributed with mean zero, variances τ_0 and τ_1, respectively, and covariance τ_{01}. The random teacher effects, T_k, are the expected deflections to students' growth curves when encountering teacher k. These effects are assumed to be independently, normally distributed with mean zero and constant variance across years. Teacher effects are also assumed to be independent of all other effects in the model. The random errors, e_{ig}, are assumed to be normally distributed and independent of each other, both within and between students, because the individual growth curves are assumed to "capture all the student-related influences on scores" (McCaffrey et al., 2003, p. 58).

More generally, the cross-classified model from above can be specified as,

$$y_{i,(g-1)} = (\mu + m_i) + a_{i,(g-1)}(\beta + b_i) + \sum_{k=1}^{K}\sum_{h=0}^{g-1} D_{hik}T_k + e_{i,(g-1)}, \tag{2.4}$$

where $a_{i,(g-1)}$ assumes the value *(g-1)* at year *g*, and the term $D_{hik} = 1$ if student *i* encounters teacher *k* at time *h*; $D_{hik} = 0$ otherwise. The teacher effects, T_k, are summed over time, so a student's score is attributed to all previous and current teachers the student had for a particular subject. These types of models can also be extended to include other factors, such as student- and teacher-level covariates (Raudenbush & Byrk, 2002), as well as higher-order polynomials in *g* (Raudenbush, 2004).

Cross-classified models explicitly model individual growth curves, often using a linear trend instead of separate means for each year. The linear trend used to model student growth places restrictions on the residual error covariance matrix. Subsequently, whenever the covariance between m_i and b_i, τ_{01}, is greater than zero, the variability of scores increases over time (McCaffrey et al., 2004). Raudenbush (2004) acknowledged

cross-classified approaches have stronger variance assumptions than models with unstructured variance-covariance matrices, but he argued this additional assumption potentially makes more appropriate and efficient use of student achievement data. Because cross-classified models can become complex quickly, other constraints may also need to be imposed to simplify a model.

In the cross-classified model described, teacher effects persist undiminished into the future, so contributions of past as well as current teachers, are accounted for in a student's set of scores. Consequently, the variability due to teacher inherently increases with each additional year of data collection (McCaffrey et al., 2004). Scores must also be on a single developmental scale (McCaffrey et al., 2003).

2.2.3.2 EVAAS Model

One prominent multivariate longitudinal linear mixed model is the Education Value-Added Assessment System (EVAAS) layered model (Sanders et al., 1997). This approach assumes teacher effects are independent and persist undiminished over time and subject. For a single track of students within a school system, a simplified version of the EVAAS model for a particular subject, such as math or reading,

$$
\begin{aligned}
y_{i1} &= \mu_1 + T_1 + e_{i1} \\
y_{i2} &= \mu_2 + T_1 + T_2 + e_{i2} \\
y_{i3} &= \mu_3 + T_1 + T_2 + T_3 + e_{i3}
\end{aligned}
\qquad (2.5)
$$

models scores, y_{ig}, for student i at year $g = 1, 2, 3$ as a function of year-specific means, μ_g. Random teacher effects are included for all teachers a student encounters for the subject during the course of the study. The teacher effects are assumed to be normally distributed with zero mean and year-specific variances; these effects are assumed

independent both within and across year. The random errors, e_{ig}, are assumed to be normally distributed and independent across students. Within-student correlations are assumed to follow an unstructured covariance structure with time-specific variances.

EVAAS jointly models more than one subject per grade for multiple cohorts of students across several school systems. The EVAAS teacher model (Sanders et al., 1997),

$$y_{ijklmn} = \mu_{iklm} + \sum_{p=1}^{N_{mk}} c_{m(ijkl)_p} t_{m(ijkl)_p} + e_{ijklmn} \,, \tag{2.6}$$

is much more complex, where y_{ijklmn} is the measurement on the nth student in the mth subject and the lth grade who encountered the jth teacher in the ith school system during the kth year. Separate fixed means, μ_{iklm}, are estimated for each grade, year, subject and school system combination. The random effect of the jth teacher who taught in the ith school system during the kth year, l^{th} grade and m^{th} subject is t_{ijklm}, and c_{ijklm} is the fractional contribution of teacher j to the student's score, accounting for instances when a student has multiple teachers for a subject in the same year. Finally, e_{ijklmn}, is the random deviation of the nth student's measurement from the fixed mean.

The $c_{m(ijkl)_p} t_{m(ijkl)_p}$ terms are summed so a student's score is attributed to all previous and current teachers the student had for subject m, creating a *layered* model. The teacher effects are summed over the index p, which tracks the student across years and allows for multiple teachers in the same year. The total number of teachers a student had though year k in subject m is N_{mk}.

The random teacher effects are assumed independent across teacher, subject, grade and time, even when the same teacher teaches multiple subjects, grades and/or

years. Separate teacher variance components are estimated for each year, subject and grade combination, creating a heterogeneous, diagonal variance-covariance matrix for the random teacher effects. The EVAAS teacher model also uses an unstructured variance-covariance matrix to account for relationships between each student's set of scores across subjects and grades, but assumes different students' scores are independent. Teachers' impact is analyzed based on at least three years of student data (Sanders et al., 1997).

In both the cross-classified and the EVAAS teacher models, teacher effects persist undiminished into the future, so contributions of both current teachers and past teachers are accounted for in a student's set of scores. Consequently, the total teacher contribution to the variability of scores increases over time, even though the total variance may not, depending on the testing instrument used (McCaffrey et al., 2004). However, the EVAAS model, unlike the cross-classified model, does not place restrictions on the overall grade-specific means or the covariance structure of repeated measurements on the same student (McCaffrey et al., 2004; Wright, Sanders, & Rivers, 2006). The unstructured within-student covariance matrix allows each student to serve as his or her own control, making it unnecessary to account for factors affecting a student's level of achievement (Sanders et al., 1997). Yet, both of these models are computationally intensive and require scores be on a single developmental scale (McCaffrey et al., 2003).

2.2.3.3 Variable Persistence Model

McCaffrey et al. (2004) proposed the variable persistence model, a generalized version of multivariate longitudinal models for student outcomes. This approach is similar to the EVAAS teacher model, but it allows prior teachers to have variable contributions to current scores rather than assuming complete persistence of these effects.

For a single track of students within a school system, a simplified version of the variable

persistence model for a particular subject,

$$
\begin{aligned}
y_{i1} &= \mu_1 + T_1 + e_{i1} \\
y_{i2} &= \mu_2 + \alpha_{21}T_1 + T_2 + e_{i2} \\
y_{i3} &= \mu_3 + \alpha_{31}T_1 + \alpha_{32}T_2 + T_3 + e_{i3}
\end{aligned}
\qquad , \qquad (2.7)
$$

models scores, y_{ig}, for student i at year $g = 1, 2, 3$ as a function of year-specific means,

μ_g. Random teacher effects are included for all teachers a student encounters for the

subject during the course of the study. The teacher effects are assumed to be normally

distributed with zero mean and year-specific variances; these effects are assumed

independent both within and across year. The persistence of prior teacher t on subsequent

scores at year g is α_{gt}, which is estimated. The random errors, e_{ig}, are assumed to be

normally distributed and independent across students. Within-student correlations are

estimated using an unstructured covariance structure with time-specific variances.

The variable persistence model can be extended to include student- and school-

level covariates (McCaffrey et al., 2003). A special case of the variable persistence model

(Lockwood, McCaffrey, Mariano, et al., 2007),

$$
y_{ig} = \mu_g + \boldsymbol{\beta}'_g \mathbf{x}_{ig} + \sum_{g^* \le g} \alpha_{gg^*} \boldsymbol{\varphi}'_{ig^*} \boldsymbol{\theta}_{g^*} + e_{ig} , \qquad (2.8)
$$

includes covariates, but not school effects, where y_{ig} is student i's score in year g, $g = 1$,

2, …, m, and μ_g is the year-specific mean. Time-invariant and time-varying covariates

for student i are included in the vector, $\mathbf{x}_{ig}$. The random teacher effects for each year,

$\boldsymbol{\theta}_g$, are linked to students by $\boldsymbol{\varphi}_{ig}$, which allows fractional contributions of teachers to

student i's score during year g. The persistency parameters, α_{gg^*}, model the strength of

past teacher effects on current student measurements, where $\alpha_{gg^*} = 1$ when $g^* = g$. The term e_{ig} represents random error. Similar to the EVAAS teacher model, the variable persistence model can be extended to account for multiple cohorts of students, subjects per grade and school systems, but the model quickly grows in complexity (Lockwood, McCaffrey, Mariano, et al., 2007).

Although the variable persistence model allows teacher effects to persist into the future, it does not assume these effects persist undiminished. Instead, the persistency parameters are freely estimated, so the total teacher contribution to the variability of scores does not inherently increase over time, unlike the cross-classified and EVAAS teacher models (McCaffrey et al., 2004). However, as the persistency parameters approach zero, correlations between future scores of students who shared a teacher in the past are no longer accounted for, eliminating one of the benefits of using the layered modeling approach (Wright & Sanders, 2008). In fact, McCaffrey et al. (2004) and Lockwood, McCaffrey, Mariano, et al. (2007) found estimated teacher effect persistency parameters to be relatively small.

McCaffrey et al. (2004) showed, under certain restrictions, each of the alternative value-added modeling approaches discussed is a special case of the general, variable persistence model (Figure 2.1). Although the multivariate methods allow efficient use of all available data and are typically recommended, they are computationally intensive and require much more in the way of computing resources than either the gain score or covariate adjustment methods. While similar in many regards, each of the multivariate approaches has unique features (Figure 2.2). For instance, unlike the EVAAS teacher and
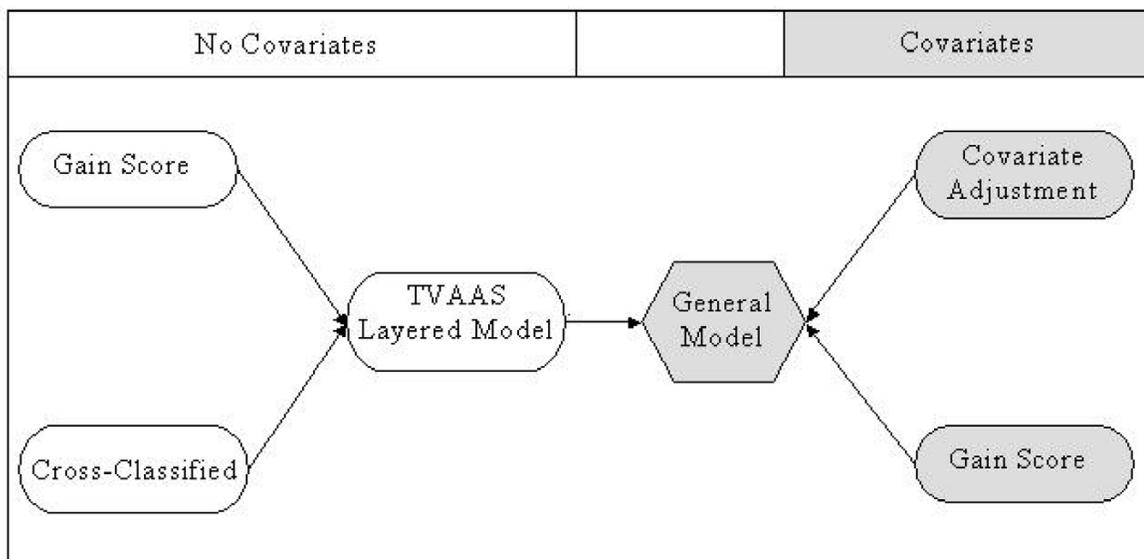
Without covariates, gain scores and the cross-classified model are special cases of the layered model with restrictions to the overall time trend and/or the distribution of residual errors. The layered model is a special case of the general model with restrictions to the αs and without covariates. The covariate adjustment and gain-score model with covariates are special cases of the general model with restrictions to the distribution of residual errors and the αs.

Note. From "Models for Value-Added Modeling of Teacher Effects," by D. F. McCaffrey, J. R. Lockwood, D. Koretz, T.A. Louis, and L. Hamilton, 2004, *Journal of Educational and Behavioral Statistics, 29*(1), p. 78. Copyright 2004 by Sage Publications, Inc. Adapted with permission. McCaffrey et al. (2004) refer to the EVAAS layered model as the TVAAS layered model.

Figure 2.1: Relationship among Models

the variable persistence models, the cross-classified framework models individual growth curves for students, so its within-student covariance matrix is restricted. Also, while both the cross-classified and EVAAS teacher models assume teacher effects persist undiminished in the future, the variable persistence model allows the strength of these effects on future scores to vary. Consequently, the variable persistence model only requires scores be linearly related instead of on a single developmental scale (McCaffrey et al., 2004): that is, the cumulative distribution functions for different scales should not cross (McCaffrey et al., 2003). While the additional complexities of these modeling

approaches can be beneficial, they also come with the price of additional computational

burdens.
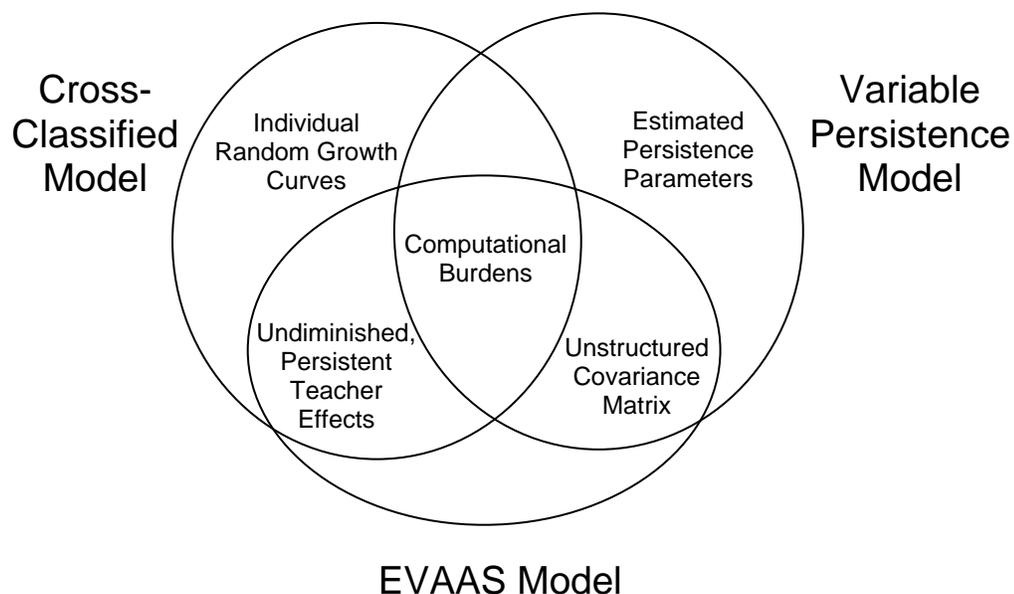
Figure 2.2: Comparison of Multivariate Models for Modeling Longitudinal Student
Achievement Data

2.3 Estimating Teacher Effects

Studies investigating value-added teacher effects provide evidence teachers have

differing effects on student learning (Rivkin et al., 2005; Rowan et al., 2002; Wright et

al., 1997) that persist over time (Sanders & Rivers, 1996), but these studies have

shortcomings. Statistical and psychometric issues arise when estimating teacher effects

using longitudinal student achievement data (McCaffrey et al., 2003). Different modeling

variations, such as the use of layered versus non-layered design matrices and the

estimation of random versus fixed teacher effects, are described, and the impact of such

considerations on estimated teacher effects follows.

2.3.1 Layered Models

Wright and Sanders (2008) distinguish between the layered and non-layered model in the construction of the coefficient matrix for teacher effects (Table 2.1). In the non-layered model, each student's outcome in a given year is linked only to the current teacher. In contrast, the layered model links a student's achievement to current and previous teachers within a given time span. Therefore, the coefficient matrix for the layered model can have several "1"s in a row, connecting past teachers with subsequent student outcomes. This approach accounts for the "correlation of future scores for students who [have] shared a past teacher" (Lockwood, McCaffrey, Mariano, et al., 2007, p. 126).

Student	Year	Teacher	Matrix for non-layered model			Matrix for layered model		
			A	B	C	A	B	C
01	1	A	1	0	0	1	0	0
01	2	B	0	1	0	1	1	0
01	3	C	0	0	1	1	1	1
02	1	B	0	1	0	0	1	0
02	2	C	0	0	1	0	1	1
02	3	A	1	0	0	1	1	1

Note. The example used estimates one overall teacher effect for each teacher instead of a separate teacher effect for each year.

Table 2.1: Comparison of Coefficient Matrices for Teacher Effects in Non-Layered and Layered Models

2.3.2 Best Linear Unbiased Predictors

Teacher effects may be specified as either fixed or random effects depending on the intended scope of inference. If teacher effects are treated as fixed, the underlying assumption is the observed teachers are the only units of interest; conclusions drawn apply only to the teachers for whom data were collected. If teacher effects are specified as random, it is assumed the observed teachers represent a random sample from a population of teachers to which conclusions can be applied. Random teacher effects are estimated by best linear unbiased predictors (BLUPs), which are also referred to as shrinkage estimators (Raudenbush & Byrk, 2002; Robinson, 1991). This estimation procedure weights the average deviations of each teacher's students' scores from the overall average score. The weighting takes into account sample size and variability within, as well as across, teachers' classrooms to "shrink" teacher effect estimates to the overall mean of the distribution of teacher effects, assumed to be zero. Each teacher's random effect is estimated relative to all other teachers in the sample, and the variability of the estimated teacher effects is assumed to represent variability present in the population of teachers. Such shrinkage and relative estimation does not occur in the estimation of fixed effects.

2.3.3 Impact of Model Specification on Teacher Effect Estimates

Wright and Sanders (2008) explain the EVAAS model in further detail by comparing it to three other sub-models. The researchers use simple examples to illustrate the effect layered design matrices and/or within-student correlations have on estimated teacher effects. In these examples, Wright and Sanders (2008) use a special case of the linear mixed model (Laird & Ware, 1982),

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}, \qquad\qquad (2.9)$$

where $\mathbf{y}$ is a vector of test scores, $\mathbf{X}$ tracks the year of each test score, $\mathbf{\beta}$ is a vector of overall mean scores for each year and $\mathbf{Z}$ is the coefficient matrix for $\mathbf{u}$, the vector of random teacher effects, assumed to be normally distributed with $E(\mathbf{u}) = \mathbf{0}$ and $Var(\mathbf{u}) = \underset{t \times t}{\mathbf{D}} = \tau^2 \mathbf{I}_t$, where t is the total number of teachers. Random errors, $\mathbf{e}$, are also assumed to be normally distributed with $E(\mathbf{e}) = \mathbf{0}$,

$$Var(\mathbf{e}) = \underset{ns \times ns}{\mathbf{R}} = \mathbf{I}_n \otimes \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho\sigma^2 \\ \rho\sigma^2 & \cdots & \rho\sigma^2 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{R}_0 \text{ and } Cov(\mathbf{u},\mathbf{e}) = \mathbf{0}, \text{ where } n \text{ is}$$

the number of students with test scores and s is the number of test scores for each student; residuals from different students are assumed to be independent, but measures on the same student are assumed to be correlated. The overall variance of test scores is

$$\mathbf{V} = Var(\mathbf{y}) = \mathbf{ZDZ}^T + \mathbf{R} = \sigma^2 \left(\theta\mathbf{ZZ}^T + \mathbf{R}_0 \right) = \sigma^2 \mathbf{V}_0, \quad \text{where} \quad \theta = \tau^2 / \sigma^2. \quad \text{Variance}$$

components are not estimated from the data, because the focus is on the estimation of fixed means and random teacher effects. In each of the examples, the ratio of teacher variance to residual variance, θ, is fixed at 1000 to reduce the amount of shrinkage in the teacher effect estimates and, subsequently, make the estimates easier to interpret; the within-student correlation, ρ, is specified as either 0 or 0.7. Substituting $\mathbf{V}_0$ for $\mathbf{V}$, $\mathbf{R}_0$ for $\mathbf{R}$ and $\theta\mathbf{I}_t$ for $\mathbf{D}$ in the estimating equations for fixed and random effects does not change the estimates produced.

One of the examples involves three years of scores for nine different students (Figure 2.3). These data can also be found in Table 10 of Wright and Sanders (2008). In

each year there are three different teachers: teachers A, B and C in year one; teachers D, E and F in year two; and teachers G, H and J in year three. As shown in Figure 2.3, the teachers fall into two different tracks: students who have teacher A or B in year one do not share any teachers over the course of the three years with the student who has teacher C in year one; students who have teacher A or B in year one either have teacher D or E in year two and teacher G or H in year three, whereas the student who has teacher C in year one has teachers F and J in years two and three, respectively. Subsequently, the estimated random effects of teachers depend on deviations for students in the teachers' respective tracks, where deviations are the differences between students' scores and the corresponding yearly mean. The students' scores are also assumed to be scaled so changes in scores from year to year are meaningful.
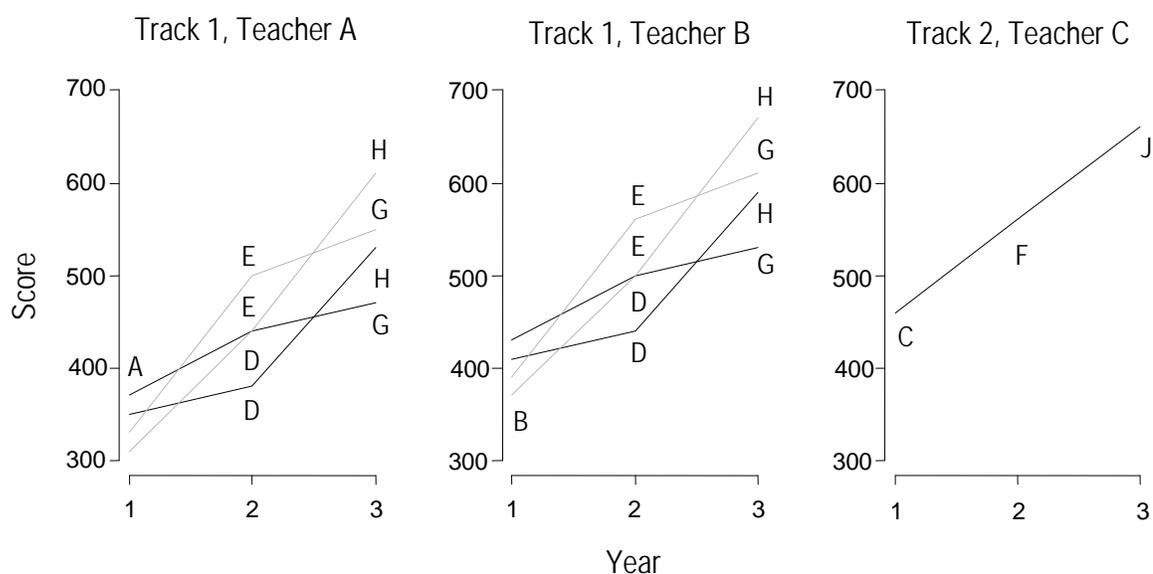
Figure 2.3: Student Scores and Teachers over Time

Estimates of random teacher effects are compared for four different models: 1) the non-layered model (NLM) with within-student correlations set to zero ($\rho = 0$), abbreviated NLM(0); 2) the non-layered model with $\rho = 0.7$, abbreviated NLM(0.7); 3)

the layered model (LM) with $\rho = 0$, abbreviated LM(0); and 4) the layered model with

$\rho = 0.7$, abbreviated LM(0.7). For the previously described data, Wright and Sanders

(2008) provide formulas for calculating the effects of the second and third year teachers

using each of the four models. Estimated teacher effects during the first year are not

value-added and, while estimated, are not discussed; students do not have previous scores

from which to determine the value a year one teacher has added. In the formulas, "M"

represents the mean of the deviation scores $(\mathbf{y} - \mathbf{X\beta})$ for a group of students, the

subsequent integer indicates the year for which the deviations were calculated and the

subscript denotes the students' corresponding teacher(s). For example, $M2_D$ represents

the average year two deviation score of students taught by teacher D, and $M1_{DE}$

represents the average year one deviation score of students taught by teachers D and E in

year two. Differences in teacher effect estimating equations for the four models are

discussed using this notation.

     The NLM(0) teacher effect estimates are merely unadjusted means of students'

deviation scores. For instance, the estimated random effect of teacher D in year two is the

mean deviation score of students taught by teacher D during year two,

$$M2_D = \Sigma \, (\text{Score} - \text{Year Two Fixed Effect Mean}) \, / \, \text{Number of Students}$$

$$= \{(440 - 500) + (380 - 500) + (500 - 500) + (440 - 500)\} \, / \, 4$$

$$= -60.$$

Similarly, the estimated random effect of teacher G is the average deviation score of the

students he or she taught in year three,

$$M3_G = \Sigma \, (\text{Score} - \text{Year Three Fixed Effect Mean}) \, / \, \text{Number of Students}$$

$$= \{(470 - 600) + (550 - 600) + (530 - 600) + (610 - 600)\} / 4$$

$$= -60.$$

The NLM(0) estimates do not incorporate expected deviation scores, because within-student correlations are assumed to be zero; no basis exists for which to obtain such expected deviations. Consequently, teachers with generally higher achieving students will appear to be more effective than teachers with generally lower achieving students, even when the latter have huge impacts on their students' learning.

The NLM(0.7) teacher effect estimates adjust the means of students' deviation scores in a given year to account for students' deviation scores in other years, as in analysis of covariance (ANCOVA). For example, the estimated random effect of teacher D in year two is the mean deviation score of students taught by that teacher during that time adjusted by the students' expected deviation scores. The formula,

$$M2_D - b(M1_D - M1_{DE}) - b(M3_D - M3_{DE}),$$

can be interpreted in three parts: $M2_D$, $b(M1_D - M1_{DE})$ and $b(M3_D - M3_{DE})$. The mean deviation score of students taught by teacher D in year two, $M2_D$, is -60; this is the NLM(0) estimate for teacher D.

However, $M2_D$ is adjusted by the expected year one deviation score for students taught by teacher D in year two, $b(M1_D - M1_{DE})$. The coefficient, b, is the pooled within-teacher multiple regression coefficient for predicting students' year two deviation scores from their corresponding year one and year three deviation scores. In this example, $b = \rho / (1 + \rho) = 0.7 / (1 + 0.7) = 0.411765$. The mean year one deviation score of students taught by teacher D in year two is:

$$M1_D = \Sigma\,(\text{Year One Score for Students Taught by Teacher D} - \text{Year One Fixed}$$

$$\text{Effect Mean}) / \text{Number of Students}$$

$$= \{(370 - 400) + (350 - 400) + (430 - 400) + (410 - 400)\} / 4$$

$$= -10.$$

The expected year one deviation score also incorporates the mean year one deviation score of students taught by teacher D or E in year two,

$$M1_{DE} = \{\; \Sigma\,(\text{Year One Score for Students Taught by Teacher D} - \text{Year One Fixed}$$

$$\text{Effect Mean}) + \Sigma\,(\text{Year One Score for Students Taught by Teacher E} - \text{Year}$$

$$\text{One Fixed Effect Mean})\;\} / \text{Number of Students}$$

$$= \{(370 - 400) + \ldots + (410 - 400) + (330 - 400) + \ldots. + (370 - 400)\} / 8$$

$$= -30.$$

Hence, the students taught by teacher D in year two had an above-average deviation score in year one, $b(M1_D - M1_{DE}) = 0.411765\{(-10) - (-30)\} = 8.235$.

The mean, $M2_D$, is also adjusted by the expected year three deviation score for students taught by teacher D in year two, $b(M3_D - M3_{DE})$. The mean year three deviation score of students taught by teacher D in year two is:

$$M3_D = \Sigma\,(\text{Year Three Score for Students Taught by Teacher D} - \text{Year Three}$$

$$\text{Fixed Effect Mean}) / \text{Number of Students}$$

$$= \{(470 - 600) + (530 - 600) + (530 - 600) + (590 - 600)\} / 4$$

$$= -70.$$

This mean is then compared to the mean year three deviation score of students taught by teacher D or E in year two,

$M3_{DE} = \{ \Sigma$ (Year Three Score for Students Taught by Teacher D – Year Three

Fixed Effect Mean) + Σ (Year Three Score for Students Taught by Teacher E

– Year Three Fixed Effect Mean) $\}$ / Number of Students

$= \{(470 - 600) + \ldots + (590 - 600) + (550 - 600) + \ldots + (670 - 600)\} / 8$

$= -30,$

to produce an expected year three deviation score of $0.411765\{(-70) - (-30)\} = -16.471$.

Together, these three parts comprise the NLM(0.7) estimate for the random effect

of teacher D in year two:

$M2_D - b(M1_D - M1_{DE}) - b(M3_D - M3_{DE})$

$= -60 - 0.\,411765\{(-10) - (-30)\} - 0.411765\{(-70) - (-30)\}$

$= -60 - 0.\,411765\{20\} - 0.411765\{-40\}$

$= -51.8.$

Although slightly higher than teacher D's NLM(0) estimate of -60, the estimate of -51.8

still indicates teacher D had a lower than average teacher effect. The NLM(0.7) estimate

is simply an adjusted mean of the students' deviation scores. Although the adjustments

based on expected deviations are conventional, some properties of these adjustments are

undesirable. For instance, future scores are used to adjust for earlier scores, even though

the covariate is potentially affected by the prior year teacher; this is typically

inappropriate for ANCOVA and may also be inappropriate in this context. The NLM(0.7)

estimate for the effect of teacher D uses the students' third year deviation scores to adjust

for their second year deviation scores, even though teacher D's impact may also affect

student performance in year three. Additionally, the NLM(0.7) estimates reduce a

teacher's effect if his or her students' subsequent deviation scores are above-average, i.e., a teacher is penalized if students go on to have higher than expected gains.

The NLM(0.7) estimates are also susceptible to within-track centering of the covariates. The expected deviation scores are found by comparing a teacher's mean student deviation score to the mean deviation scores of students in the same track rather than the mean deviation scores of all students for whom scores are available. For example, the NLM(0.7) estimate for teacher D uses $M1_{DE}$ and $M3_{DE}$ to center the year one and year three covariates, respectively. Teacher D and E's students are not in the same track as teacher F's student. Consequently, the year one and year three deviation scores of teacher F's student do not impact the centering of the covariates for estimating teacher D and teacher E's NLM(0.7) effects. Although these scores are included in the estimation of the year one and year three fixed effect means, the fixed effect means used in the calculation of $M1_D$, $M1_{DE}$, $M3_D$ and $M3_{DE}$ cancel when finding the differences $(M1_D - M1_{DE})$ and $(M3_D - M3_{DE})$. Therefore, teacher F's student deviation scores do not affect the NLM(0.7) estimates of teacher D and E's effects. Similarly, when estimating the NLM(0.7) effect for teacher F in year two, the year one and year three mean deviation scores are found for only the students in that same track. This example is particularly problematic, because the within-track centering of the covariates simplifies the formula for this estimate,

$$M2_F - b(M1_F - M1_F) - b(M3_F - M3_F),$$

to be merely the unadjusted mean deviation score for teacher F's student in year two, $M2_F$, the same as the NLM(0) estimate; no adjustments are made, because the expected deviations are zero.

In year three, the NLM(0.7) estimates have similar issues. The estimated effect for teacher G in year three,

$$M3_G - b(M2_G - M2_{GH}) - b(M1_G - M1_{GH}),$$

is the mean deviation score for teacher G's students in year three, $M3_G$, adjusted for the students' performance in years one and two, $b(M1_G - M1_{GH})$ and $b(M2_G - M2_{GH})$, respectively. Although the covariates for years one and two are not affected by the year three deviation scores, the year three NLM(0.7) estimates are still susceptible to within-track centering. Consequently, the NLM(0.7) estimates for teachers F and J do not account for the student's previous performance and, therefore, are not truly value-added effects.

The LM(0) teacher effect estimates are also adjusted means, but the estimates incorporate between-track differences and between-student correlations. For instance, the random effect of teacher D in year two is estimated using

$$M2_D - M1_{DE} + 0.5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})].$$

The mean deviation score of students taught by teacher D in year two, $M2_D = -60$, is adjusted by the mean year one deviation score for students taught by either teacher D or E in year two, $M1_{DE} = -30$, and the performance of teacher D's students in year three compared to year two, relative to students on the same track, $0.5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})] = 0.5[\{(-70) - (-30)\} - \{(-60) - (-30)\}] = -0.5$. Together, these adjustments produce the estimated LM(0) effect for teacher D in year two, -35, which is higher than the NLM(0) estimate of -60, but still below the average teacher effect. The LM(0) effect for teacher F in year two simplifies to be

$$M2_F - M1_F + 0.5[(M3_F - M3_F) - (M2_F - M2_F)] = M2_F - M1_F,$$

because this track consists of only one teacher in each year, nullifying the adjustment for future scores and reducing the teacher effect to simply a mean gain. However, in both of these estimates, the LM(0) effects account for between-track differences, unlike the NLM(0.7) estimates. The mean deviation of the previous year's scores is track specific, but it incorporates the fixed effect mean for that year, which is estimated using data from all students, irrespective of track. This is contrary to the NLM(0.7) estimates, in which each fixed effect mean cancels due to the centering of each covariate.

The LM(0) estimates also account for the between-student correlations induced by using the layered design matrix. Expected deviation scores are found based on deviation scores for students who have shared at least one common teacher. In the adjustment for teacher D's effect, $0.5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$, teacher D is rewarded if his or her students perform better in year three than they did in year two, relative to other students on the same track. Likewise, if teacher D's students have relatively worse future performance, his or her teacher effect estimate decreases; this is contrary to the estimation of the NLM(0.7) teacher effects. Although no future scores are available to obtain this type of adjustment for year three LM(0) teacher effect estimates, between-student correlations are accounted for through the additional prior year adjustment in the teacher effect estimates. For instance, the formula for the LM(0) estimate for teacher G in year three is $M3_G - M2_{GH}$, where $M2_{GH}$ adjusts the year three mean deviation score for teacher G, $M3_G$, by accounting for prior mean score deviations, between-track differences and between-student correlations. This type of adjustment also occurs in the year two LM(0) teacher effect estimates (e.g., $M1_{DE}$ for teacher D's effect and $M1_F$ for teacher F's effect). Yet, none of the LM(0) teacher effect estimates account for within-student

correlations, so expected deviations scores for a student cannot be based on other scores from the same student, only on scores of students who have shared the same teacher(s).

The LM(0.7) estimates account for both between- and within-student correlations. The LM(0.7) random effect of teacher D in year two is estimated using

$$M2_D - M1_{DE} - 0.7(M1_D - M1_{DE}) + 0.5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})],$$

which is similar to the LM(0) estimate. However, the LM(0.7) utilizes within-student correlations to obtain expected deviation scores for students based on other scores from those same students. The inclusion of the additional term, $0.7(M1_D - M1_{DE})$, compares the mean year one deviation score for students who had teacher D in year two to the mean year one deviation score for students in the same track. The LM(0.7) estimate for teacher G in year three,

$$M3_G - M2_{GH} - b(M2_G - M2_{GH}) - b(M1_G - M1_{GH}),$$

also accounts for within-student correlations by incorporating the students' prior mean deviation scores in both year one, $b(M1_G - M1_{GH})$, and year two, $b(M2_G - M2_{GH})$. By making use of both between- and within-student information, the LM(0.7) teacher effect estimates utilize expected deviation scores based on scores of students who have shared the same teacher(s), as well as other scores from the same student.

Table 2.2 displays the year two and year three teacher effect estimates from each of the four models discussed using the student scores graphed in Figure 2.3. Estimates are not included for year one teachers A, B and C, because their effect estimates cannot be described as value-added. However, comparing the year two and year three teacher effect estimates is particularly enlightening. As illustrated in Figure 2.3, student gains from year one to year two and from year two to year three support the notion teachers E and H had

relatively higher than average effects on student learning. Alternatively, the effects of teachers D and G on student gains from one year to the next were below average, while the effects of teachers F and J were average. The effects of teachers G and H are further magnified by considering each student's score trajectory from year one to year two through the inclusion of within-student correlations; the students' trajectories drastically change after teacher G's and teacher H's instruction, whereas the trajectory for teacher J's student does not change.

Year	Teacher	Model			
		NLM(0)	NLM(0.7)	LM(0)	LM(0.7)
2	D	-60	-51.8	-35	-49
	E	0	-8.2	35	49
	F	60	60	0	0
3	G	-60	-76.5	-30	-46.5
	H	0	16.5	30	46.5
	J	60	60	0	0

Table 2.2: Teacher Effect Estimates from NLM(0), NLM(0.7), LM(0) and LM(0.7)

The non-layered models produce estimates reflecting mean deviation scores; NLM(0) estimates are merely unadjusted mean deviation scores, while NLM(0.7) are mean deviation scores adjusted for students' prior and/or future achievement. Consequently, the teacher effect estimates for these models portray teachers F and J as highly effective teachers, simply because their student had higher than average scores in years two and three, not because the student showed higher than average growth.

Alternatively, the layered models produce estimates reflecting mean gains in deviation scores. As a result, teachers F and J had teacher effect estimates of zero for both layered models, because their student had relatively average mean gains and the student's score trajectory did not change over time; the estimate of zero conveys the teacher did not have a value-added effect on the student's learning. Additionally, accounting for within-student correlations provides more information about student growth over time, pulling the other LM(0.7) teacher effect estimates further from zero than the LM(0) estimates for the same teachers.

The variable persistence model (Equation 2.8) proposed by McCaffrey et al. (2004) allows teacher persistency parameters to vary. However, estimated persistency parameters tend to be near zero (Lockwood, McCaffrey, Mariano, et al., 2007; McCaffrey et al., 2004), so the teacher effect estimates from the low-persistency model approach the undesirable behavior of the NLM(0.7) estimates and do not appear to be value-added; estimates fail to account for between-track differences, and teachers are penalized for students who perform unexpectedly well in future years. As discussed by McCaffrey et al. (2004), the non-layered and low-persistency model estimates are also typically more susceptible to bias from omitted variables correlated with the *level* of student achievement than are the layered model estimates. In fact, the risk of overcorrecting teacher effects arises if non-instructional, time-invariant covariates beyond a school system's control, such as race and poverty status, are included in the layered model. However, teacher effect estimates are susceptible to bias if variables correlated with *gains* in student achievement are omitted, irrespective of how teacher effect persistency is or is not specified.

In general, using the EVAAS layered model (Equation 2.6) to estimate teacher effects has advantages over other modeling approaches. The layered model accounts for both between- and within-student correlations to adjust for past and future student achievement scores. The model also uses all available scores, resulting in "less biased, more stable, more efficient estimates" (Wright & Sanders, 2008, p. 14) than either the gain score (Equation 2.2) or the covariate adjustment model (Equation 2.1). The use of multiple years of data allows estimates to be adjusted, thereby accounting for external pulses occurring in a given year and rewarding teachers whose students perform better than expected in the future. Overall, Wright and Sanders (2008) argue the EVAAS layered model is a competitive option for estimating teacher effects, because of its flexibility and adherence to value-added philosophy.

2.4 Issues

Uncertainty remains concerning whether student background variables should be included as covariates. The EVAAS model does not include such covariates, but researchers can inappropriately extend this practice to other, less sophisticated longitudinal value-added models. Instead, decisions to model student-level covariates should be based upon the interaction of several factors: "the distribution across classes and schools of students with different characteristics, the relationship between the characteristics and outcomes, the relationship between the characteristics and true teacher effects, and the type of model used" (McCaffrey et al., 2003, p. 70). Bias of teacher effect estimates is an issue when students are disproportionately assigned to schools and/or classrooms based on background characteristics related to student outcomes and true

teacher effects. When student background characteristics are correlated with teacher effects, the inclusion of student-level covariates can mask the effects of teachers. Fixed effects estimation of the covariates overcorrects for these background characteristics and results in underestimation of true, random teacher effects. For example, if highly effective teachers are assigned to affluent students, socioeconomic status becomes a proxy for teacher quality and true teacher effects are underestimated. Ballou, Sanders, and Wright (2004) propose a strategy to adjust for bias when student characteristics are correlated with true teacher effects, but strategies to address the issue of bias resulting from contextual effects as a result of aggregate-level variables have yet to be developed. Ultimately, when deciding whether to control for covariates, researchers must choose between potentially confounding teacher effects with student-level variables and potentially biasing teacher effects.

Debates also continue about the persistency of teacher effects. The EVAAS (Equation 2.6) and cross-classified models (Equation 2.4) assume teacher effects persist undiminished into the future. As a result, prior teachers affect a student's score in a particular year, but not a student's gain in scores from one year to the next. Gain score models (Equation 2.2) implicitly make the same assumption in that gains do not rely on prior year teacher effects. Consequently, the models assume teachers do not affect students' future growth. In the variable persistence model (Equation 2.8), however, student gains depend on prior teachers' effects, because the differences between the current and previous year's persistency parameters for teacher effects are not necessarily zero, as they are in complete persistence models. Similarly, covariate adjustment models (Equation 2.1) assume prior teacher effects influence student growth at the rates specified

by the coefficients for prior scores. Yet, the potential advantages of freely estimating persistency parameters need to be carefully weighed against the advantages of using a layered model; as the persistency parameters approach zero, correlations between future scores of students who shared a teacher in the past are no longer accounted for, eliminating one of the benefits of using the layered modeling approach (Wright & Sanders, 2008).

Other issues arise when linking teachers to student scores. In some cases, students have incomplete records, missing some or all of their teacher links in a given year. This can occur when students are taught by multiple teachers in a single subject. Specifically, students may transfer to different schools midyear, be team-taught and/or learn a subject, such as reading, in multiple teachers' classes. Such instances result in complex teacher links, and linking a student's outcome to only one (or even no) teacher in a particular year may confound an identified teacher's effect with other, unidentified teachers' effects. However, determining how to accurately reflect the effect of multiple teachers on a student is not straightforward and requires potentially unrealistic assumptions.

Issues associated with the construction and scaling of instruments used to measure student achievement also exist. Typically, measures of student achievement are assumed to be on an interval scale, where any difference in scores has the same meaning at any point on the scale. For example, a student with test scores of 40 and 60 in consecutive years is assumed to have made the same amount of growth as a student with scores of 20 and 40. However, linking scores from different tests to a single scale for comparisons across grades may require nonlinear transformations, where rates of growth no longer have the same meaning across all ability levels. As illustrated in the previous example,

students with different levels of ability may have the same rates of achievement on the original test scales, but after transforming the scores nonlinearly to a single developmental scale, the students could potentially have different rates of change; the change from 20 to 40 may be more or less significant than the change from 40 to 60 and may subsequently impact teacher effect estimates (McCaffrey et al., 2003). Teacher effect estimates may also be sensitive to the alignment of test content with curricula, and the weighting of test content to obtain one overall achievement score (Lockwood, McCaffrey, Hamilton, et al., 2007; Martineau, 2006). Yet, varying curricula and test content across grades amplify the complexity of measuring growth in a multidimensional domain, such as mathematics, and the potential impact of such test construction and scaling issues on teacher effect estimates should not be ignored.

2.5 Summary and Future Work

Value-added modeling techniques estimate the contribution of educational factors, such as teachers, to growth in student achievement, while allowing for the possibility to control for the effect of non-educational factors. Several value-added models for estimating teacher effects have been proposed as alternatives to current test-based accountability procedures, such as AYP, but each has its respective advantages and disadvantages. Although these methods have the potential to identify highly effective teachers, teacher effect estimates are sensitive to different modeling specifications, including the persistency of teacher effects. Several statistical and psychometric issues exist, and sensitivity of teacher effects to such issues needs to be explored. Consequently, considerations should be made when defining what *teacher effects* really describe, and

teacher effect estimates should be linked to other valid measures of teacher effectiveness. Although value-added teacher effect estimates should not be used in isolation to make high-stakes decisions, value-added methodology can help researchers identify what characteristics highly effective teachers possess and motivate informed improvements in education (McCaffrey et al., 2003).

Chapter 3

Estimating the Impact of a Professional Development Program on Student Learning

3.1 Introduction

Professional development programs focus on preparing teachers to meet the recent initiatives on improving the quality of mathematics instruction, but rigorous evaluations are needed to determine whether these programs are actually effective (Carey, 2004; Guskey, 1994; Hill, 2007a; Loucks-Horsley, Stiles, & Hewson, 1996; National Mathematics Advisory Panel [NMAP], 2008; Shaha, Lewis, O'Donnell, & Brown, 2004). According to Hill (2007a),

> Almost no local professional development—and even most efforts offered
> by respected university faculty, nonprofit, and commercial professional
> developers—is rigorously evaluated, in the sense of researchers looking
> for changes in teacher knowledge and instructional practice. Even more
> seldom do researchers investigate the effect on student learning. More
> often, evaluations simply ask participants to report whether and how the
> program affected their own teaching. (p. 122)

However, teacher self-reports tend to be unreliable and subjective measures of program effectiveness (Borko, 2004; Wilson & Berne, 1999).

Past research has examined the relationship between teaching quality and student learning, investigating the effect teachers' degrees (Ackerman, Heafner, & Bartz, 2006; Rowan, Correnti, & Miller, 2002; Wayne & Youngs, 2003), coursework (Hill, Rowan, &

Ball, 2005; Wayne & Youngs, 2003), certification status (Ackerman et al., 2006; Hill et al., 2005; Rowan et al., 2002; Wayne & Youngs, 2003), teaching experience (Ackerman et al., 2006; Hill et al., 2005; Rowan et al., 2002; Wayne & Youngs, 2003), licensure examination scores (Wayne & Youngs, 2003), pedagogical practices (Ackerman et al., 2006; Rowan et al., 2002) and amount of professional development (Ackerman et al., 2006) have on student achievement. Recent efforts measure not only teachers' mathematical content knowledge, but also their mathematical knowledge for teaching (Ball, Hill, & Bass, 2005; Hill, 2007b; Hill, Schilling, & Ball, 2004), examining its relationship to student achievement (Hill et al., 2005). Various covariate adjustment models and gain score models have been used in conjunction with statistical models, such as analysis of variance (ANOVA), analysis of covariance (ANCOVA) (Sanders, 2006) and hierarchical linear models (HLM) (Raudenbush & Bryk, 2002; Rowan et al., 2002; Wright, Sanders, & Rivers, 2006), to estimate teacher effects. However, researchers criticize these procedures for biasing teacher effects and/or modeling students' achievement status instead of changes in achievement (Rowan et al., 2002), as well as for the methods' approaches for handling missing data (Sanders, 2006). Unfortunately, "assertions about the magnitude of teacher effects on student achievement depend…on the methods used to estimate these effects and on how the findings are interpreted" (Rowan et al., 2002, p. 1536).

Cross-classified models (Raudenbush & Bryk, 2002) and the Educational Value-Added Assessment System (EVAAS) model (Sanders, Saxton, & Horn, 1997) are currently recommended over other models to provide estimates of teacher effectiveness. The EVAAS model is a longitudinal linear mixed effects model that has each student

serve as his or her own control, similar to the cross-classified model which models individual growth curves. Using the EVAAS model, Sanders et al. (1997) have been able "to produce estimates of school and teacher effects that are free of socioeconomic confoundings and do not require direct measures of these concomitant variables" (Wright, Horn, & Sanders, 1997, p. 58). In fact, Sanders (2000) claims, "Our research work…clearly indicates that differences in teacher effectiveness is the single largest factor affecting [students'] academic growth" (p. 334); teachers are the dominant factor impacting student progress (Sanders, 2004; Wright et al., 1997). Darling-Hammond (2000) adds, "[E]ffects of well-prepared teachers on student achievement can be stronger than the influences of student background factors, such as poverty, language background, and minority status" (Conclusions and Implications, ¶ 6).

With teacher effectiveness linked to student achievement, questions remain about what factors influence the quality of teaching (Carey, 2004). Professional development is one factor thought to influence teaching quality. Desimone, Porter, Garet, Yoon, and Birman (2002) write, "Professional development is considered an essential mechanism for deepening teachers' content knowledge and developing their teaching practices" (p. 81). Various authors list characteristics of effective professional development programs (Desimone et al., 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Guskey, 1994; Loucks-Horsley et al., 1996), but rigorous evaluations are needed to determine whether these programs actually affect teaching quality (Carey, 2004; Hill, 2007a; Loucks-Horsley et al., 1996; NMAP, 2008; Shaha et al., 2004).

Previous research has examined the relationship between teacher quality and student learning (Ackerman et al., 2006; Darling-Hammond, 2000; Hill et al., 2005;

Presley, White, & Gong, 2005; Rowan et al., 2002; Wayne & Youngs, 2003), while other

research has investigated the value-added effects of teachers on student learning

(McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Rowan et al., 2002; Sanders

& Rivers, 1996; Sanders et al., 1997; Wright et al., 1997). Yet, the relationship between

teacher development and teacher practices, as well as student learning still needs to be

explored (Fishman, Marx, Best, & Tal, 2003; Frome, Lasater, & Cooney, 2005; NMAP,

2008).

Existing research tries to estimate the effect of professional development

programs with a dichotomous variable (Shaha et al., 2004; Stroup, 2007; Stroup & Fang,

2006). Typically, teachers are assigned either a value of one to indicate their participation

or a value of zero to indicate their absence of participation in a program. However, the

discrete nature of this approach neglects the interactive nature of teachers and the

possibility of *creeping excellence*, where program participants share their newly acquired

knowledge and ideas with non-participating teachers. This approach also disregards the

teachers' varying degrees of participation and changes in practice. Hill (2007a) writes,

"Although teachers might be required to engage in professional development, they are

not required to learn from it" (p. 123). A teacher's participation in professional

development opportunities does not necessitate actual learning or changes in teacher

beliefs and practices. It may also be the case a teacher's absence of participation in a

professional development program does not predicate his or her lack of teaching quality.

Instead, estimating the change in a teacher's effect on student achievement after

participating in a professional development program can be an alternative approach for

estimating the impact of such a program; this approach allows each teacher to serve as his

or her own control and helps address the complexities ignored by merely comparing the effects of participating teachers on student achievement to those of non-participating teachers. In this chapter, alternative methodology for using less-than-ideal longitudinal student achievement data to estimate the impact of a professional development program is proposed and applied to data collected from a mathematics professional development program, the Math in the Middle Institute Partnership (M^2). The chapter concludes with a summary of the results and recommendations for future work.

3.2 Methods for Estimating the Impact of a Professional Development Program on Student Learning

In recent years, education systems, in theory, have held students to higher academic standards (No Child Left Behind, 2001) by holding states accountable for assessing measurable student outcomes. Research efforts have addressed issues associated with analyzing student achievement data (McCaffrey, Lockwood, Koretz, & Hamilton, 2003), but many of the recommended approaches have not been widely adopted because the required resources and high-quality longitudinal data are not readily available. Most value-added modeling (VAM) approaches require student achievement data to be vertically scaled, or at least linearly related, over time (McCaffrey et al., 2003). Such requirements limit analyses that can be conducted on available assessment data, which often are not on a single developmental scale. Few studies have addressed how to use value-added models to analyze achievement data not on a single developmental scale (Rivkin, Hanushek, & Kain, 2005), and even fewer have discussed how to use these data to estimate the impact of professional development on students. The purpose of this section is to investigate how to use a value-added model for analyzing longitudinal

student achievement data collected from a mixture of norm- and criterion-referenced assessments to estimate the impact of a professional development program on student learning.

3.2.1 Background

Multiple authors have championed the use of value-added models to analyze longitudinal student achievement data (Doran, 2003; Drury & Doran, 2003; Hershberg, Simon, & Lea-Kruger, 2004; Lissitz, 2005; Sanders et al., 1997). These methods fall into three categories: covariate adjustment models, gain score models and multivariate models (McCaffrey et al., 2003). Covariate adjustment models regress each student's current achievement score on his or her prior year score, while gain score models treat the difference between two successive years' scores as the response. Both methods require complete student records and lose information about a student's performance over time by estimating models separately for each year. Although "covariate" methods do not require tests to be on a single developmental scale, "gain" methods do so changes in performance are not confounded with changes in tests (McCaffrey et al., 2003).

Multivariate models jointly model all student scores, including relationships between each student's set of outcomes. These approaches also accommodate missing data, making efficient use of all available information. One prominent multivariate longitudinal linear mixed model is the Education Value-Added Assessment System (EVAAS) layered model (Sanders et al., 1997). This approach assumes teacher effects are independent and persist undiminished over time and subject. A more general version was proposed by McCaffrey et al. (2004) where prior year teachers have variable contributions to current year scores. The variable persistence model only requires scores

be on linearly related scales, but the EVAAS model requires scores be on a single developmental scale (McCaffrey et al., 2003). Although computationally intensive, the layered modeling approaches have advantages over other methods (Sanders, 2006; Wright & Sanders, 2008).

Studies investigating VAM teacher effects provide evidence teachers have differing effects on student learning (Rivkin et al., 2005; Rowan et al., 2002; Wright et al., 1997) that persist over time (Sanders & Rivers, 1996), but these studies have shortcomings. Statistical and psychometric issues arise when estimating teacher effects using longitudinal student achievement data (McCaffrey et al., 2003). Lockwood, McCaffrey, Hamilton, et al. (2007) showed estimated VAM teacher effects are sensitive to the ways student achievement is measured. This is particularly problematic when scores are not on a single developmental scale. Rivkin et al. (2005) standardized criterion-referenced test scores to have a mean of zero and a standard deviation of one for each cohort of students within a given academic year to address differences between tests. However, students could not be matched to specific teachers, so only subject- and grade-level means were used to model gains. Although a Z-score approach has limitations (McCaffrey et al., 2003), it is more appropriate to use for modeling gains in achievement than raw test scores when instruments are not on the same scale. It is important to investigate how a Z-score approach can be used on less-than-ideal data to not only estimate teacher effects, but also the impact of a professional development program on student learning.

3.2.2 EVAAS Layered Teacher Model

For a single subject, such as mathematics, a simplified version of the EVAAS

model,

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zt} + \mathbf{e}, \qquad\qquad (3.1)$$

is a special case of the linear mixed model (Laird & Ware, 1982) where $\mathbf{y}$ is a vector of test scores, $\boldsymbol{\mu}$ is a vector of means, and $\mathbf{Z}$ is the coefficient matrix for $\mathbf{t}$, the vector of random teacher effects, assumed to be distributed $N(\mathbf{0}, \sigma_t^2 \mathbf{I})$. Random errors, $\mathbf{e}$, are also assumed to be normally distributed with $E(\mathbf{e}) = \mathbf{0}$ and $Var(\mathbf{e}) = \mathbf{R}$. Residuals from different students are assumed to be independent, but residuals on the same student are assumed to be correlated and are modeled using an unstructured within-student covariance structure. This complex covariance structure accounts for variables affecting students' levels of achievement and is used instead of non-instructional student-level covariates (Wright & Sanders, 2008).

Student	Year	Teacher	$\mathbf{Z}$ matrix for non-layered model			$\mathbf{Z}$ matrix for layered model		
			A	B	C	A	B	C
01	1	A	1	0	0	1	0	0
01	2	B	0	1	0	1	1	0
01	3	C	0	0	1	1	1	1
02	1	B	0	1	0	0	1	0
02	2	C	0	0	1	0	1	1
02	3	A	1	0	0	1	1	1

Note. The example used estimates one overall teacher effect for each teacher instead of a separate teacher effect for each year.

Table 3.1: Comparison of $\mathbf{Z}$ Matrix in Non-Layered and Layered Models

Wright and Sanders (2008) distinguish between the layered and non-layered model in the construction of the $\mathbf{Z}$ matrix (Table 3.1). In the non-layered model, each student's outcome in a given year is linked only to the current teacher. In contrast, the layered model links a student's achievement to current and previous teachers within a given time span. Therefore, the $\mathbf{Z}$ matrix for the layered model can have several "1"s in a row, connecting past teachers with subsequent student outcomes.

3.2.3 Modified Layered Model

To estimate the impact of a professional development program on student learning, the earlier model (Equation 3.1) can be modified to

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{Z}_a\mathbf{t}_a + \mathbf{Z}_b\mathbf{t}_b + \mathbf{Z}_c\mathbf{t}_c + \mathbf{e}. \tag{3.2}$$

The random teacher effect is separated into a common teaching effect, $\mathbf{t}_c \sim N(\mathbf{0}, \sigma_c^2\mathbf{I})$, and before and after professional development participation effects, $\mathbf{t}_b \sim N(\mathbf{0}, \sigma_b^2\mathbf{I})$ and $\mathbf{t}_a \sim N(\mathbf{0}, \sigma_a^2\mathbf{I})$. $\mathbf{Z}_b, \mathbf{Z}_a$ and $\mathbf{Z}_c$ are the corresponding coefficient matrices for $\mathbf{t}_b, \mathbf{t}_a$ and $\mathbf{t}_c$, respectively. The covariate, β_a, accounts for the impact of a professional development program, and $\mathbf{X}_a$ is the coefficient matrix to track participation. The intercept, $\boldsymbol{\mu}$, can be estimated to be the same value for all students.

The $\mathbf{Z}_b, \mathbf{Z}_a$ and $\mathbf{Z}_c$ matrices are constructed so that for each teacher, the teacher effect before participating in a professional development program is $p_b = t_b + t_c$, $\mathbf{p}_b \sim N(\mathbf{0}, \sigma_b^2\mathbf{I} + \sigma_c^2\mathbf{I})$; the teacher effect after the onset of participation is $p_a = \beta_a + t_a + t_c$, $\mathbf{p}_a \sim N(\mathbf{X}_a\beta_a, \sigma_a^2\mathbf{I} + \sigma_c^2\mathbf{I})$; and their covariance is $Cov(p_b, p_a) = \sigma_c^2$. In this model, the teacher effects before and after program participation are assumed to have different variances, but not to be independent. This is a reasonable assumption, because teachers'

effects before and after participating in a professional development program should be related.

The difference between each participating teacher's effects, $p_a - p_b = \beta_a + t_a - t_b$, is specified as the impact of the professional development program on his or her students' achievement. While Sanders et al. (1997) and others (McCaffrey et al., 2004) estimate separate teacher effects for each combination of grade and year in a set of longitudinal data, only two overall teacher effects (one before and one after participation) are estimated for each teacher over the time period spanned by the data.

Table 3.2 illustrates how the $\mathbf{Z}$ matrices change in the modified model (Equation 3.2). A new variable, "PD," is added, indicating whether a teacher had participated in the professional development program (1) or not (0). The corresponding $\mathbf{Z}_b$ and $\mathbf{Z}_a$ matrices displayed link each teacher to the student outcomes, while also distinguishing whether or

Student	Year	Teacher	PD	$\mathbf{Z}_b$ matrix			$\mathbf{Z}_a$ matrix		
				A	B	C	A	B	C
01	1	A	0	1	0	0	0	0	0
01	2	B	1	1	0	0	0	1	0
01	3	C	0	1	0	1	0	1	0
02	1	B	1	0	0	0	0	1	0
02	2	C	0	0	0	1	0	1	0
02	3	A	1	0	0	1	1	1	0

Table 3.2: $\mathbf{Z}_b$ and $\mathbf{Z}_a$ Matrices

not the teacher had participated in the program. For example, students 01 and 02 both had teacher A. However, student 01 had teacher A before the teacher had participated in the program, while student 02 had teacher A after participation. Therefore, the $\mathbf{Z}_b$ matrix links student 01 to teacher A, while the $\mathbf{Z}_a$ matrix links student 02 to teacher A. The $\mathbf{Z}_c$ matrix is the sum of the $\mathbf{Z}_b$ and $\mathbf{Z}_a$ matrices, resulting in the same $\mathbf{Z}$ matrix for the layered model (Table 3.1).

3.2.4 Z-scores

With the EVAAS model, changes in raw scores are not meaningful when test scores in successive years are not on a single developmental scale. To compensate for this problem, standard Z-scores can be used. More specifically, for grade i and student j in a given academic year, the original test score is S_{ij}. The corresponding Z-score is

$Z_{ij} = \dfrac{S_{ij} - \overline{S_{i.}}}{\sigma_i}$ where $\overline{S_{i.}}$ is the average score for grade i, and σ_i is the standard deviation of scores in grade i.

In a given academic year, Z_{ij} indicates how many standard deviations the original score S_{ij} is away from the average score for a grade. Changes in Z-scores reflect changes in relative position across years for a group of students, but not necessarily changes in academic achievement, when measures are on different developmental scales (McCaffrey et al., 2003). The standardized scores allow for within-group comparisons across academic years.

The variance of a student outcome in an academic year is $Var(y) = \sum_{k=1}^{g} c_k^2 \sigma_t^2 + \sigma_e^2$, where g is the number of grades for which data have been collected on a student through

that academic year. In the EVAAS model, $c_k = 1$ for all k, so the variance of an outcome

inherently increases with g. When the standard Z-score is used as the response variable in

Equation 3.1, its variance is restricted to one. Thus, the layered $\mathbf{Z}$ matrix was

standardized so $\sum_{k=1}^{g} c_k^2 = 1$ (Table 3.3). As shown in Table 3.3, Sanders et al.'s (1997)

proposed method of equally weighting each previous and current teacher's contribution

to a student's score in a given academic year was adapted to take into account the

constraint imposed by the fact Z-scores, by definition, must have a variance of one.

Student	Year	Teacher	PD	Standardized layered $\mathbf{Z}$ matrix			Layered $\mathbf{Z}$ matrix		
				A	B	C	A	B	C
01	1	A	0	1	0	0	1	0	0
01	2	B	1	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	0	1	1	0
01	3	C	0	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	1	1	1
02	1	B	1	0	1	0	0	1	0
02	2	C	0	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	0	1	1
02	3	A	1	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	1	1	1

Table 3.3: Standardized Layered $\mathbf{Z}$ Matrix

The same method was also applied to the standardization of the $\mathbf{Z}_c$ matrix when

using Z-scores as the response for Equation 3.2. The nonzero elements in the $\mathbf{Z}_a$ and $\mathbf{Z}_b$

matrices were then assigned the corresponding standardized weights from $\mathbf{Z}_c$ (Table 3.4). Elements in the coefficient matrix, $\mathbf{X}_a$, were defined as the sum of the standardized weights in the $\mathbf{Z}_a$ matrix to track the weighted frequency of previous and current teachers participating in the professional development program and attributed to each student outcome.

Student	Year	Teacher	PD	$\mathbf{X}_a$	Standardized $\mathbf{Z}_b$ matrix			Standardized $\mathbf{Z}_a$ matrix		
					A	B	C	A	B	C
01	1	A	0	0	1	0	0	0	0	0
01	2	B	1	$1/\sqrt{2}$	$1/\sqrt{2}$	0	0	0	$1/\sqrt{2}$	0
01	3	C	0	$1/\sqrt{3}$	$1/\sqrt{3}$	0	$1/\sqrt{3}$	0	$1/\sqrt{3}$	0
02	1	B	1	1	0	0	0	0	1	0
02	2	C	0	$1/\sqrt{2}$	0	0	$1/\sqrt{2}$	0	$1/\sqrt{2}$	0
02	3	A	1	$2/\sqrt{3}$	0	0	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$	0

Table 3.4: Standardized $\mathbf{Z}_b$ asnd $\mathbf{Z}_a$ Matrices

3.3 Example: Math in the Middle Institute Partnership

The Math in the Middle Institute Partnership (M^2) is an NSF-funded mathematics professional development program at the University of Nebraska-Lincoln aiming to "create a University/Educational Service Unit (ESU)/Local School District partnership

with the capacity to educate and support teams of outstanding middle level (Grades 5-8) mathematics teachers to become intellectual leaders in their school, districts and ESUs" (Lewis, Heaton, McGowan, & Jacobsen, 2004, p. 1). One major component of M^2 is the M^2 Institute, a multi-year institute offering participants a coherent program of study to deepen their mathematical knowledge for teaching and develop their pedagogical and leadership skills. The second M^2 project component is a research initiative to understand how changes in teachers' mathematics teaching practice translate into measurable improvement in student performance. The program consists of six cohorts, or groups, of mathematics teachers in Nebraska whose entrances were staggered yearly, with the first cohort beginning in October, 2004.

As part of the research initiative, this example focuses only on the analysis of student achievement data from 2003-04 to 2007-08. During this time, each school district in Nebraska was free to choose whatever student achievement measure it deemed appropriate; a variety of criterion- and norm-referenced tests were administered to various grade levels at various points during the school year across districts. Consequently, student achievement scores from different districts are not directly comparable. Considering these issues, attention is restricted to one of the larger participating school districts, Middleview Public Schools[1] (MPS). Typically, MPS middle schools have roughly 27 students in a mathematics class. The district's mobility rate has ranged between 14.29% and 17.82% since the 2003-2004 academic year. Data were collected from 317 MPS 5th-8th grade teachers who taught mathematics between the 2003-04 and 2007-08 academic years, 37 of whom were M^2 participants. Student achievement data were collected for 5th-8th grade students in the district (Table 3.5), as

---

[1] Names are pseudonyms.

were survey data from the teachers. Yet, no existing statistical models were adequate to address the MPS assessment practice of administering a mix of norm- and criterion-referenced tests to various grade levels each year.

	Academic year administered				
Grade	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008
5	MAT	MAT	MAT & CRT	MAT & CRT	MAT & CRT
6	MAT	MAT	MAT & CRT	CRT	CRT
7	MAT	MAT	MAT & CRT	MAT & CRT	MAT & CRT
8	--	--	CRT	CRT	CRT

Note. Dashes indicate no test was administered to that grade level in that year. MAT = Metropolitan Achievement Test. CRT = Criterion Referenced Test.

Table 3.5: MPS Middle School Assessments between 2003-04 and 2007-08

MPS has shared annual middle school student achievement data with M^2 researchers since the 2003-2004 academic year. This example examines the data from 2003-04 to 2007-08. These data include several variables, such as each student's grade level, mathematics teachers, mathematics courses and mathematics achievement scores. Within each grade level and year combination students took a criterion-referenced test (CRT) and/or the Metropolitan Achievement Test (MAT). The MAT is a norm-referenced test whose purpose is to test "a broad range of students with real world content," and contains questions to measure basic skills and knowledge, as well as "critical thinking processes and strategies" (Pearson, 2008, ¶ 1). The CRTs were developed by MPS mathematics teachers. Questions were written according to specific, predetermined MPS mathematics standards, exceeding the state mathematics content

standards. Students' CRT scores reflect proficiency on the criteria, rather than relative academic performance.

The EVAAS model used by Sanders et al. (1997) and the variable persistence model proposed by McCaffrey et al. (2004) require, respectively, scores from year-to-year be on a single developmental scale or linearly related. Each of the two types of tests used by MPS has a different purpose and potentially measures different mathematical abilities. Scores are not provided on a single developmental scale, so changes in student achievement are not necessarily directly comparable from one grade to the next.

3.3.1 Modified Layered Model Implementation

The modified layered model (Equation 3.2) using standardized layered $\mathbf{Z}_b, \mathbf{Z}_a$ and $\mathbf{Z}_c$ matrices was applied to the MPS middle schools mathematics achievement scores, standardized within each grade and academic year combination. When a student had scores available for both tests in a given academic year and grade combination, the CRT score was used in the analysis. This decision was made in conjunction with MPS senior personnel. They reported students tended to take the CRT more seriously than the MAT, because performance on the CRT is a component of students' math course grades, and performance on the MAT is not. Additionally, curricula and assessment analyses performed by Smith (2004) revealed the CRT to be a better match to M^2 goals than the MAT. Thus, CRT scores were used in years when students had both CRT and MAT scores.

The majority of fifth grade scores were missing teacher links, because fifth grade student achievement data are part of the district's elementary, rather than middle, school data set; with the exception of fifth grade scores linked to M^2 participants, none of the

elementary data are linked to specific teachers. This was the lowest grade level for which data were collected, and removal of the scores would eliminate information about student performance in an entire grade level. Consequently, information key to establishing a baseline for student performance would be lost. Hence, missing fifth grade teachers effects were assigned the value zero. In the few other instances where teacher links were missing, the student scores were still included in the analysis to better track each student's changes in achievement across time.

In some cases, a student had more than one math teacher in a given school year. This is due to the MPS practice of assigning students performing below grade level in mathematics to a Mathematics Intervention class in addition to the regular grade level mathematics course; thus, students are enrolled in a second math course, in which they receive additional instruction in an effort to bring their achievement up to meet grade level standards. Prior to eighth grade, such intervention courses meet every other day (while mathematics classes meet daily); in eighth grade students attend both courses daily. While ideally students would have the same teacher for both the regular mathematics course and the intervention course, scheduling difficulties meant some students had two different teachers. When this happened, students were linked to the regular mathematics course teacher rather than the intervention course teacher.

Using standard mixed models methodology (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006) with $\mathbf{Z} = (\mathbf{Z_b} \mid \mathbf{Z}_a \mid \mathbf{Z}_c)$ and $\mathbf{t'} = (\mathbf{t_b'} \mid \mathbf{t_a'} \mid \mathbf{t_c'})$, the effect of a teacher on student learning before participating in M² was estimated as $\hat{p}_b = \hat{t}_b + \hat{t}_c$ and after as $\hat{p}_a = \hat{\beta}_a + \hat{t}_a + \hat{t}_c$. Estimates were obtained using REML implemented in the GLIMMIX procedure in SAS, Version 9.2 (SAS Institute Inc., 2008).

3.3.2 Results

Table 3.6 shows the estimated variance components and fixed effects. The estimated variances of the before and after participation teacher effects are 0.179 and 0.141, respectively, indicating the teacher effects before participating in M^2 are estimated

Parameter	Estimate	SE
σ_b^2	0.161	0.041
σ_a^2	0.124	0.052
σ_c^2	0.018	0.037
σ_{e5}^2	0.945	0.011
σ_{e6}^2	0.753	0.011
σ_{e7}^2	0.710	0.011
σ_{e8}^2	0.734	0.015
$\sigma_{5,6}$	0.656	0.010
$\sigma_{5,7}$	0.601	0.010
$\sigma_{5,8}$	0.591	0.012
$\sigma_{6,7}$	0.586	0.010
$\sigma_{6,8}$	0.568	0.012
$\sigma_{7,8}$	0.596	0.012
μ	-0.065	0.008
β_a	0.019	0.064

Table 3.6: Variance Components and Fixed Effects Estimates

to be slightly more variable than the teacher effects after participation. The covariance between a teacher's before and after participation effects is estimated to be 0.018, and the resulting estimated correlation is 0.112. Consequently, these variance components do not seem reasonable, because a teacher's effect on student learning before participating in a professional development program is expected to be at least moderately correlated with his or her effect after participation. The overwhelming number of fifth grade scores with missing teacher links appears to create a lot of noise, as evidenced by the relatively large estimated fifth grade residual error variance, 0.945. Yet, subsequent residual errors on the same student are strongly related, with estimated correlations ranging between 0.710 and 0.825.

The estimated intercept, -0.065, is the starting point from which to base subsequent teacher impacts. Although the covariate accounting for program participation is not significantly different from zero, it is estimated to be 0.019 and allows the center of the distribution of teacher effects after participating in M^2 to shift from the assumed mean of zero.

Figure 3.1 compares the estimated before participation teacher effects for M^2 participants and non-participants, illustrating the mean teacher effect for program participants (mean = 0.053, standard deviation = 0.372, n = 37) is estimated to be slightly higher but not statistically (p = .1671) different than that of non-participants (mean = -0.038, standard deviation = 0.359, n = 280). From this, there do not appear to be strikingly obvious differences between non-participants and participants, prior to entering the program.
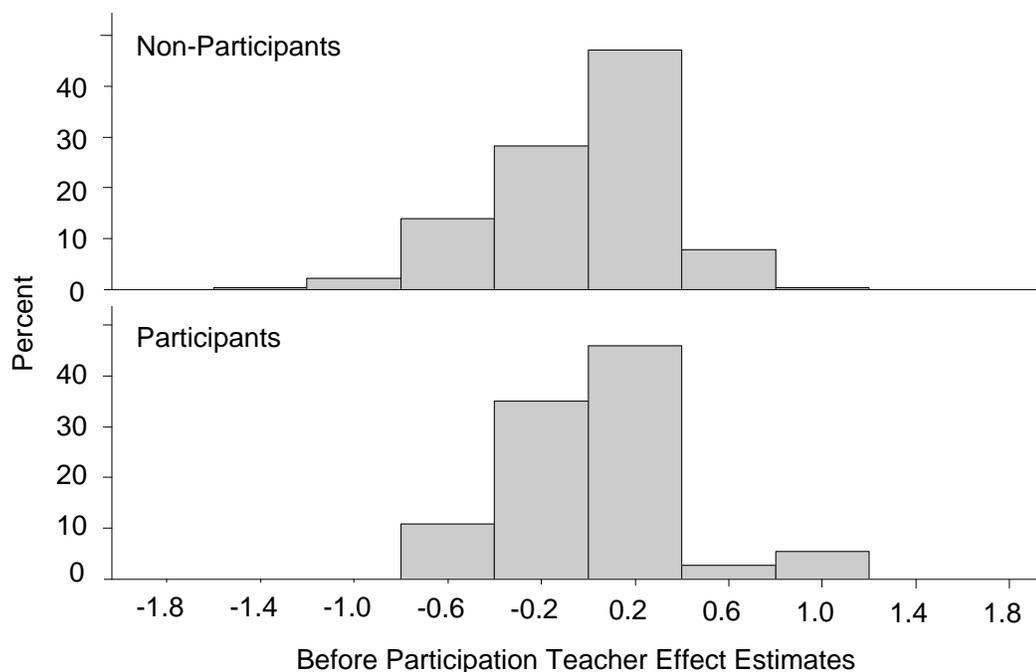
Note. The horizontal axis is in terms of standardized units.

Figure 3.1: Comparison of Before Participation Effects between Teachers Participating (n = 37) and Not Participating (n = 280) in M^2

The estimated impact of participation in M^2 does not indicate a significant change in participants' effects on student learning (Figure 3.2). The average difference between each M^2 participant's estimated after and before participation teacher effects is -0.030 with a standard error of 0.080, while the median difference, -0.007, is slightly higher. This indicates the average change in each teacher's predicted teacher effect after program participation is slightly, though not statistically (p = .7099, df = 36), different from zero.

While these results do not suggest a participation effect, the potential for ceiling effects with criterion-referenced tests exists. The tests are constructed to determine whether students meet pre-determined proficiency levels on specific mathematics criteria, and it is possible for students to consistently answer most to all questions on these assessments correctly. In such instances, the tests are unable to detect changes in a

student's achievement across time and, consequently, limit discernment of value-added teacher effects. Thus, some teacher's value-added effects could be underestimated due to many students in teachers' classrooms reaching the ceiling of the criterion-referenced tests, which is a limiting feature of the instrument, but not necessarily of the model.
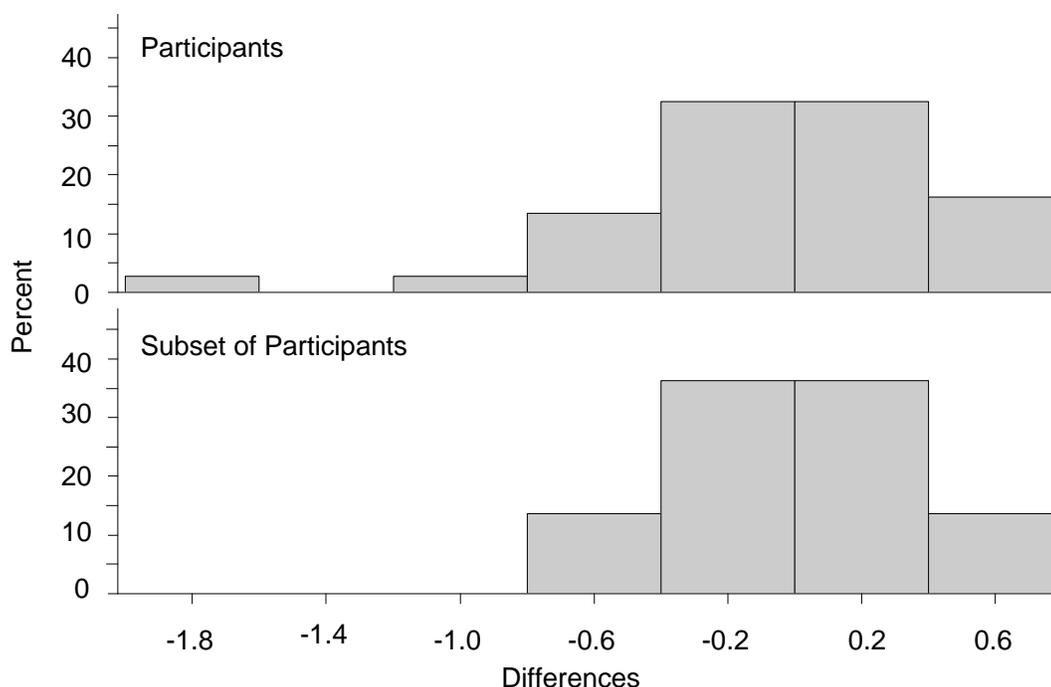
Note. The horizontal axis is in terms of standardized units.

Figure 3.2: Comparison of Differences between Before and After Participation Effects for MPS Teachers (n = 37) and a Subset of MPS Teachers (n = 22) Participating in M^2

Further inspection of the teacher effect estimates and the MPS student achievement data reveals nine teachers participating in M^2 are linked to only fifth grade scores. Consequently, the teacher effect estimates for these fifth grade teachers are not truly value-added, because previous scores on which to base each student's fifth grade achievement are not available. Additionally, six more teachers are not linked to any student scores before their participation in M^2, so these teachers' estimated before participation teacher effects are not meaningful. After removing these 15 teachers'

estimated effects from the comparison, the average difference between each remaining teacher's estimated after and before participation teacher effects is slightly higher (mean = 0.027, SE = 0.073, n = 22), and the median difference remains roughly zero (Figure 3.2). Although the results do not indicate the M^2 Institute had a significant impact on participating teachers' effects on student learning, the instruments may not have been designed to detect such changes. In future research, the *Survey of Mathematical Knowledge for Teaching* (Hill, Schilling, & Ball, 2004) data, collected from teachers before and after M^2 participation, will be used to investigate whether teachers' gains in content knowledge for teaching mathematics during the program can be linked to the changes in teacher effects on student learning.

3.4 Summary and Future Work

Although a Z-score approach has limitations, it is an appropriate alternative to using raw data when analyzing less-than-ideal student achievement data across a mixture of norm- and criterion-referenced tests over time. This chapter addressed issues arising when using a layered, longitudinal linear mixed model to analyze gains in standardized scores, including weighting considerations for variance components. Additional studies should consider other weighting alternatives and investigate the impact of such variance component weighting schemes on the estimation of teacher effects. Because curricula and test content vary across grades, as do mobility rates, future research should also explore whether notable changes in a student's Z-scores from year to year are associated with changes in mobility rates, curricula and/or test content.

Additionally, methods were proposed for estimating teacher effects on student learning before and after teacher participation in professional development programs. Although the specific example used does not indicate the M^2 Institute had a significant impact on participating teachers' effects on student learning, the instruments may have not been designed to detect such changes. When utilizing this methodology, determining whether the goals of the program align with what the instruments assess and acknowledging any existing limitations is essential. Further research should address the potential issue of censoring, and careful consideration needs to be given to what data are needed and how much baseline data should be obtained when estimating the impact of a professional development program. Ideally, these methods can be extended to other VAM approaches, as well as other professional development programs, and could eventually be used to establish potential relationships between changes in a teacher's mathematical knowledge for teaching mathematics and changes in student achievement.

Chapter 4

Using Parallel Processing Methodology to Estimate Teacher Effects

4.1 Introduction

In recent years, education systems, in theory, have held students to higher academic standards (No Child Left Behind, 2001) by holding states accountable for assessing measurable student outcomes. Research efforts have addressed issues associated with analyzing student achievement data (McCaffrey, Lockwood, Koretz, & Hamilton, 2003), but many of the recommended approaches have not been widely adopted because the resources and high-quality longitudinal data required for these approaches are not readily available. Most value-added modeling approaches require student achievement data to be vertically scaled, or at least linearly related, over time (McCaffrey et al., 2003). Such requirements limit the analyses that can be conducted on available assessment data, which often are not on a single developmental scale. Few studies have addressed how to use value-added models to analyze achievement data not on a single developmental scale (Green, Smith, Heaton, Jiao, & Stroup, under review; Rivkin, Hanushek, & Kain, 2005), and even fewer, perhaps none, have discussed how to use information from multiple instruments in a single year that are on different scales, potentially both within and between instruments over time. Section 4.2 describes how to use parallel processing, specifically curve-of-factors, methodology to analyze longitudinal student achievement data collected from two different assessments in a single subject, such as mathematics, and estimate teachers' effects on student learning.

Assuming data come from a curve-of-factors model structure, a simulation study described in Section 4.3 evaluates the performance of the proposed curve-of-factors model in its ability to accurately rank teachers in the presence of either complete or missing test data. The performance of the curve-of-factors model is then compared to that of the Z-score methodology proposed in Chapter 3. The chapter concludes with a summary of the results and recommendations for future work.

4.2 Parallel Processing Methodology

Growth curve models analyze differences in individuals' changes on repeated measurements over time. Growth curve models can be specified either within a multilevel modeling framework as a random coefficients regression model (Raudenbush & Byrk, 2002) or within a structural equation modeling (SEM) framework as a latent growth curve model (Little, Bovaird, & Slegers, 2006). In a random coefficients model, an average growth curve is estimated for all individuals. Because growth curves differ between persons, each individual's random deviations from the average trajectory are also modeled, capturing individual variability around the average trajectory of change (Raudenbush & Byrk, 2002). Although a latent growth curve model can be specified as a random coefficients model in the multilevel framework, modeling growth models within the SEM framework has the advantage of allowing researchers to model changes in a latent construct over time (Little et al., 2006). This is particularly beneficial when modeling multiple outcome measures, instead of a single measure across time. In such instances, parallel process, or multivariate, growth curve models estimate the relationship

between the growth trajectories for each of the parallel measures and allow researchers to investigate changes in latent factors over time instead of changes in observed scores.

4.2.1 Introduction to Parallel Processing

Multivariate, or parallel process, growth curve models "model univariate growth in the context of multiple parallel measures and relate those multiple outcomes to each other" (Little et al., 2006, p. 193). One type of multivariate growth curve model is the associative latent growth curve model (Little et al., 2006). This type of model first establishes the univariate latent growth curve model structures for each of the constructs, or processes, measured over time and then estimates the correlations between the growth factors for all of the constructs. This type of model allows researchers to estimate the relationship between the growth trajectories of multiple constructs measured over time. For example, Cheong, MacKinnon, and Khoo (2003) used an associative latent growth model in a mediation analysis context. In the study, Cheong et al. (2003) explored the impact a drug prevention program had on the growth factors of two different, parallel processes: the mediator process (perceived importance of team leaders) and the outcome process (nutrition behaviors) for high school football players. Structural relations in a mediation model were used to analyze the effect of participation in the prevention program on initial status and change in both the perceived importance of team leaders and the nutrition behaviors over time. The authors also used structural relations, instead of correlations, to explore whether initial status in each of the constructs predicted change in the other construct and whether change in perceived performance of team leaders predicted change in nutrition behaviors over time. In this sense, Cheong et al. (2003) used an associative growth curve model to estimate the relationship between the latent growth

curves of perceived performance of team leaders and the nutrition behaviors over time. Similarly, Roesch et al. (2009) used an associative growth curve model to explore the relationship between the growth trajectories of various mediator processes (psychosocial constructs) and the growth factors of an outcome process (physical activity constructs) in adolescents over time. The authors also used a mediation model to estimate the effect of a health promotion intervention program on those growth factors.

Another type of multivariate growth curve model is the factor-of-curves model (Little et al., 2006). Similar to the associative growth curve model, the factor-of-curves model first establishes the univariate growth trajectories for each of the constructs measured over time. However, instead of estimating the correlations or structural relations between the latent growth factors for the constructs, this type of model uses higher-order growth factors to represent the relationship between the latent growth curves of the processes. This type of model allows researchers to determine whether the growth trajectories for multiple constructs measured over time can be summarized by a common, higher-order latent growth curve. Duncan, Duncan, and Strycker (2000) used a factor-of-curves model to analyze the relationship between the initial status and rate of change in four different behaviors (drug use, marijuana use, deviance and academic failure) of adolescents across time. Instead of correlating the growth factors, the factor-of-curves model allowed them to specify a higher-order latent growth curve of problem behavior with one common initial status and one common rate of growth to describe the relationship among growth trajectories of the lower-order constructs.

The last type of multivariate growth curve model discussed in this chapter is the curve-of-factors model (Little et al., 2006). This type of model is used to estimate a latent

growth curve to describe changes in a latent construct measured by multiple indicators over time. For instance, mathematics achievement can be measured by multiple instruments in a given year. Across time, changes in the latent factor can be described by a higher-order growth curve. This allows researchers to investigate change over time based on a student's common, latent trait, instead of a student's observed scores. The following section describes the use of a modified version of this type of model in a value-added context, where the latent factors are allowed to covary instead of have a specific growth trajectory over time.

4.2.2 Parallel Processing in a Value-Added Context

Parallel processing, specifically curve-of-factors methodology, applied in a value-added context can extend the analysis of student achievement data to situations in which multiple tests with potentially different scales are given each year in a particular subject. Instead of estimating a teacher's effect on changes in a student's scores over time, curve-of-factors models can allow the estimation of a teacher's effect on changes in some common, latent trait measured by the multiple instruments in each year. These models can be extended to account for longitudinal student achievement data cross-classified in
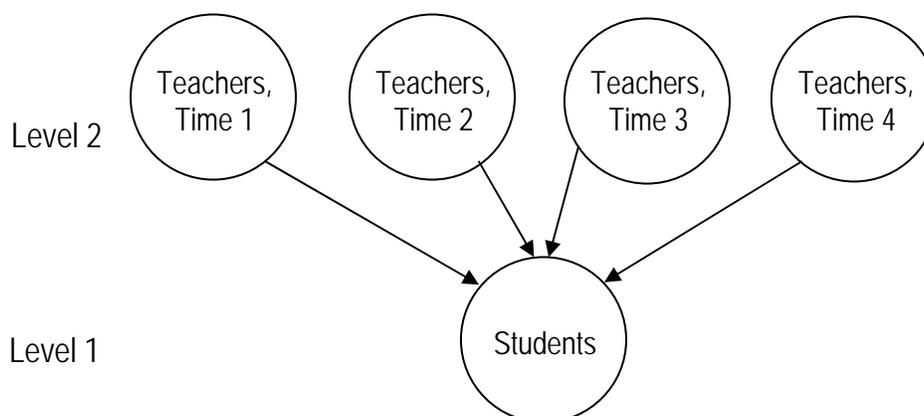
Figure 4.1: Visual of Cross-Classified Data Structure

nature. Figure 4.1 shows an example of such a structure, where students at level one are nested within level two teachers' classrooms at each time. However, not all students who have the same teacher at one time have the same teacher at another time, resulting in a cross-classified structure.

The level one curve-of-factors model shown in Figure 4.2 has two indicators, or observed measurements, e.g., CA1 and MA1, of a latent trait at each of the four time points. The latent factors A_1-A_4 represent the latent trait of interest, e.g., achievement, at each of the times. The set of latent factors for a single individual is assumed to follow a multivariate normal distribution with a mean vector of zero and an unstructured covariance matrix. The covariance between two latent factors is represented by a two-sided arrow connecting the two factors, with the variance of a factor represented by a self-connecting arrow. Latent factors are assumed independent across students. The measurement error variance of an indicator or test, i.e., the portion of an indicator's variance not explained by the underlying construct, is also estimated and represented by a
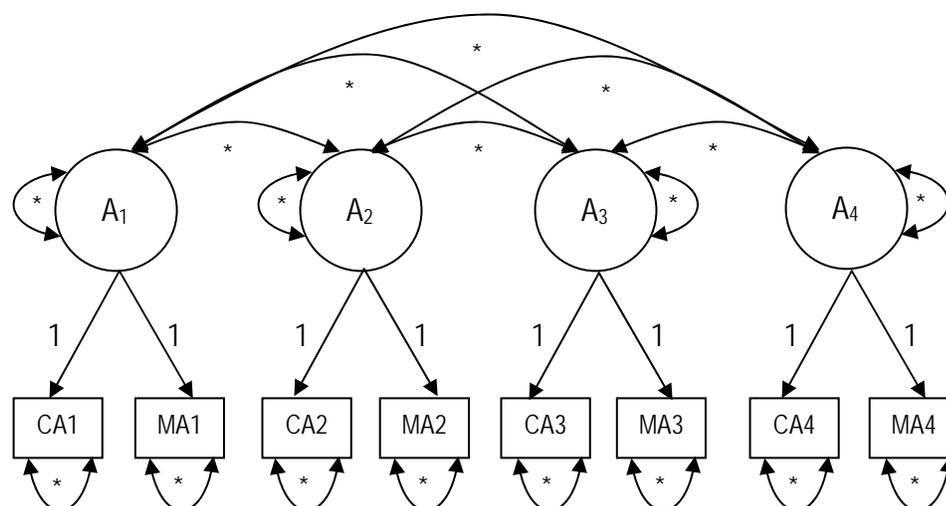
* Freely estimated (i.e., unconstrained) parameter.

Figure 4.2: Level One Curve-of-Factors Model

self-connecting arrow. All measurement errors are assumed to be independent both within and across students, with constant variance across indicators and time. The loadings, or coefficients, regressing the indicators on their corresponding factors, or constructs, are constrained to one for identification purposes.

Including the level two teacher effects in the curve-of-factors model structure (Figure 4.2) cannot be easily depicted in a diagram because of the cross-classified nature of the data. Instead, this relationship can be given by the model equations,

$$\mathbf{A} = \mathbf{Z}_{t_a}\mathbf{t} + \mathbf{d} \tag{4.1}$$

and

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A} \otimes \mathbf{1}_2 + \mathbf{e}, \tag{4.2}$$

which can be combined into one overall equation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_t\mathbf{t} + \mathbf{Z}_d\mathbf{d} + \mathbf{e}, \tag{4.3}$$

where $\mathbf{Z}_t = \mathbf{Z}_{t_a} \otimes \mathbf{1}_2$, $\mathbf{Z}_d = \mathbf{I} \otimes \mathbf{1}_2$ and $\mathbf{A} \otimes \mathbf{1}_2$ adjust for the shift from one response (i.e., the latent factor) at level two to two responses (i.e., the test indicators) at level one in each year. In Equation 4.3, $\mathbf{y}$ is the vector of test scores, $\mathbf{X}$ is the coefficient matrix for $\boldsymbol{\beta}$, the vector of test means, and $\mathbf{Z}_t$ is the coefficient matrix for $\mathbf{t}$, the vector of random teacher effects on latent factors, which can be assumed to be distributed $N\left(\mathbf{0}, \sigma_t^2\mathbf{I}\right)$. As indicated in Equation 4.1, the teacher effects are estimated based on changes in a student's common, latent trait, instead of on changes in a student's scores over time. Including teachers in this manner changes the constructs, A_1-A_4, from exogenous, or independent, factors, as illustrated in Figure 4.2, to endogenous, or dependent, factors. Consequently, the variance of a factor originally represented by a self-connecting arrow

in Figure 4.2 is replaced by a disturbance, or term representing factor-level variability not attributed to teachers. Additionally, the unstructured covariance matrix is no longer for the set of latent factors, but instead for the set of disturbances for the latent trait across time. The vector of random, factor-level disturbances, $\mathbf{d}$, can be assumed to be normally

distributed with $E(\mathbf{d}) = \mathbf{0}$ and $Var(\mathbf{d}) = \left(\mathbf{I} \otimes \begin{bmatrix} \sigma_{d1}^2 & \sigma_{d1,d2} & \sigma_{d1,d3} & \sigma_{d1,d4} \\ \sigma_{d2,d1} & \sigma_{d2}^2 & \sigma_{d2,d3} & \sigma_{d2,d4} \\ \sigma_{d3,d1} & \sigma_{d3,d2} & \sigma_{d3}^2 & \sigma_{d3,d4} \\ \sigma_{d4,d1} & \sigma_{d4,d2} & \sigma_{d4,d3} & \sigma_{d4}^2 \end{bmatrix} \right)$, where

disturbances for the same student are allowed to covary but disturbances for two different students are assumed to be independent; $\mathbf{Z}_d$ is the corresponding coefficient matrix. Random test measurement errors, $\mathbf{e}$, can be assumed to be normally distributed with $E(\mathbf{e}) = \mathbf{0}$ and $Var(\mathbf{e}) = \sigma_e^2 \mathbf{I}$.

Equation 4.3 can be extended to a layered model in which teacher effects are assumed to persist undiminished over time. Wright and Sanders (2008) distinguish between the layered and non-layered model in the construction of the $\mathbf{Z}_t$ matrix (Table 4.1). In the non-layered model, each student's outcome for a given indicator is linked only to the current teacher. In contrast, the layered model links a student's achievement to current and previous teachers within a given time span. Therefore, the $\mathbf{Z}_t$ matrix for the layered model can have several "1"s in a row, connecting past teachers with subsequent student levels of the latent trait. Because multiple indicators are available in a given year, two outcomes in the same year in either the non-layered or the layered model are identically linked to teachers in the respective $\mathbf{Z}_t$ matrix.

Student	Indicator	Teacher	$\mathbf{Z}_t$ matrix for non-layered model				$\mathbf{Z}_t$ matrix for layered model			
			A	B	C	D	A	B	C	D
1	CA1	A	1	0	0	0	1	0	0	0
1	MA1	A	1	0	0	0	1	0	0	0
1	CA2	B	0	1	0	0	1	1	0	0
1	MA2	B	0	1	0	0	1	1	0	0
1	CA3	C	0	0	1	0	1	1	1	0
1	MA3	C	0	0	1	0	1	1	1	0
1	CA4	D	0	0	0	1	1	1	1	1
1	MA4	D	0	0	0	1	1	1	1	1

Note. The example used estimates one overall teacher effect for each teacher instead of a separate teacher effect for each year.

Table 4.1: Comparison of **Z** Matrix in Non-Layered and Layered Models

4.3 Example: Student Achievement Simulation Study

The analysis of student achievement data can be challenging, particularly when high-quality longitudinal data on a single developmental scale are not readily available. Student achievement data from Middleview Public Schools[2] (MPS) exhibit such complexities. Over time, a variety of criterion- and norm-referenced tests have been administered across the middle-level grades within MPS schools. As shown in Table 4.2, student achievement data were collected for 5th-8th grade students in the district between the academic years 2003-2004 and 2007-2008. Green et al. (under review) proposed Z-

[2] Names are pseudonyms.

score methodology to address the MPS assessment practice of administering a mix of norm- and criterion-referenced tests to various grade levels each year. However, the proposed method does not use information from multiple instruments administered in a given year; in each year, scores from only one instrument are used. Thus when multiple assessments are given in a year, the researcher must choose which test to use. A simulation study was conducted to compare the use of a Z-score model to a curve-of-factors model when data were simulated assuming a curve-of-factors model structure. Of particular interest was how estimated teacher effects and percentiles change when analyzing data with this alternative approach instead of the correct model. Model parameters specified for the simulation were based on summary statistics obtained from the MPS student achievement data. Comparisons across models were made for both a complete and a missing tests case to investigate the impact of having only one, instead of two, indicators of a construct in one or more years.

	Academic year administered				
Grade	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008
5	MAT	MAT	MAT & CRT	MAT & CRT	MAT & CRT
6	MAT	MAT	MAT & CRT	CRT	CRT
7	MAT	MAT	MAT & CRT	MAT & CRT	MAT & CRT
8	--	--	CRT	CRT	CRT

Note. Dashes indicate no test was administered to that grade level in that year. MAT = Metropolitan Achievement Test. CRT = Criterion Referenced Test.

Table 4.2: MPS Middle School Assessments between 2003-04 and 2007-08

4.3.1 Simulation Study Description

Student achievement data were simulated for 2,000 students over the course of four years. Each year, data were simulated from two different instrument scales assuming a curve-of-factors model structure (Equation 4.3), for a total of eight observations per student. The MA-indicators and the CA-indicators, representing the MAT and CRT assessments in the MPS data set, were simulated assuming the following mean structures: $\mathbf{1}_{2000} \otimes \begin{bmatrix} 62 & 84 & 54 & 70 \end{bmatrix}^T$ and $\mathbf{1}_{2000} \otimes \begin{bmatrix} 231 & 234 & 236 & 239 \end{bmatrix}^T$, respectively. The total variance for each indicator was 225 units squared. Random measurement errors, $\mathbf{e}$, were assumed to be normally distributed with $E(\mathbf{e}) = \mathbf{0}$ and $Var(\mathbf{e}) = \mathbf{R} = \sigma_e^2 \left(\mathbf{I}_{2000} \otimes \mathbf{I}_8 \right)$. Measurement error variances were simulated to be equal for both instruments across all four years, with 20 percent ($\sigma_e^2 = 45$) of the total indicator variance representing measurement error variance. Factor-level disturbances were simulated assuming an unstructured within-student covariance structure. The vector of random, factor-level disturbances, $\mathbf{d}$, was simulated assuming the covariance structure,

$$Var(\mathbf{d}) = \left(\mathbf{I}_{2000} \otimes \begin{bmatrix} 157.5 & 116.7 & 99.8 & 83.3 \\ 116.7 & 135 & 98.6 & 82.7 \\ 99.8 & 98.6 & 112.5 & 80.5 \\ 83.3 & 82.7 & 80.5 & 90 \end{bmatrix} \right), \text{ where disturbances for the same}$$

student were allowed to covary but disturbances for two different students were assumed to be independent. Because a layered model with complete persistence was assumed, the variances of a student's outcomes in an academic year are

$$Var(CA_g) = Var(MA_g) = \sum_{k=1}^{g} \sigma_t^2 + \sigma_{d_g}^2 + \sigma_e^2, \text{ where } k = 1, 2, \ldots, g \text{ represents the years for}$$

which data were collected on a student through that academic year. Consequently, the

variance of an outcome inherently increases with *g* unless other adjustments are made. For this simulation, the disturbances were simulated to have decreasing variance over time to adjust for the assumed complete persistency of teacher effects, as specified by the layered $\mathbf{Z}_t$ coefficient matrix, and the subsequent increase in variability accounted for by teachers across time. An alternative approach would be to allow indicator variances to increase over time, but the MPS student achievement data motivating the simulation did not support the use of this strategy. Within each year, 20 different teachers were each randomly assigned 100 students. The vector of random teacher effects, $\mathbf{t}$, was assumed to be distributed $N\left(\mathbf{0}, \sigma_t^2 \mathbf{I}_{80}\right)$ with constant variance, $\sigma_t^2 = 22.5$.

Each of the 1000 simulations was analyzed using both the curve-of-factors model from which the data originated (Equation 4.3) and an alternative approach, referenced in this chapter as the Z-score model. The Z-score model,

$$\mathbf{z} = \boldsymbol{\mu} + \mathbf{Z}_t \mathbf{t} + \mathbf{e}, \tag{4.4}$$

is a modified version of the EVAAS model (Sanders, Saxton, & Horn, 1997), where $\mathbf{Z}_t$ is a standardized layered coefficient matrix (Green et al., under review) and $\mathbf{z}$ is a vector of simulated achievement scores, standardized within each year for each of the two different types of instruments. The vector of standardized test scores for an instrument, $\mathbf{z}$, is modeled by an overall intercept, $\boldsymbol{\mu}$, and a vector of random teacher effects, $\mathbf{t}$, assumed to be distributed $N\left(\mathbf{0}, \sigma_t^2 \mathbf{I}_{80}\right)$. Random errors, $\mathbf{e}$, are also assumed to be normally distributed with $E(\mathbf{e}) = \mathbf{0}$ and $Var(\mathbf{e}) = \mathbf{R}$. Residuals from different students are assumed to be independent, but residuals on the same student are assumed to be correlated and are modeled using an unstructured within-student covariance structure.

Because a student had scores available for both tests in all four years, the Z-score method was applied twice: once using MA scores across all four years (MA Z-score model) and once using CA scores (CA Z-score model).

The 1000 simulated data sets were also modified to reflect one pattern of missingness occurring in the MPS data. The modified data sets were identical to those previously described; however, it was assumed the CA1 and the MA4 assessments were not given. Therefore, those simulated test scores were removed from the original data sets. Each of the 1000 modified simulations were analyzed using both a curve-of-factors model (Equation 4.3), where the random measurement errors, $\mathbf{e}$, were assumed to be distributed $N\left(\mathbf{0}, \sigma_e^2\left(\mathbf{I}_{2000} \otimes \mathbf{I}_6\right)\right)$, and the Z-score model (Equation 4.4). As shown in Table 4.3, three different Z-score models were specified in the missing tests case. The CA Z-score model used the standardized CA scores as its response when both instruments were administered in the same year. Alternatively, the MA Z-score model used the standardized MA scores when available. In the third Z-score model, the MA/CA

Model	Year			
	1	2	3	4
Curve-of-factors	MA1	MA2 & CA2	MA3 & CA3	CA4
CA Z-score	MA1	CA2	CA3	CA4
MA Z-score	MA1	MA2	MA3	CA4
MA/CA Z-score	MA1	MA2	CA3	CA4

Table 4.3: Assessments used as Responses for Simulation Analysis with Missing Tests

model, the standardized MA2 assessment scores were specified as the response in year

two, and the standardized CA3 assessment scores were specified as the response in year

three. This third Z-score model differed from the other two because it used scores from

the same type of instrument in two subsequent years rather than three subsequent years.

Data were simulated in SAS, Version 9.2 (SAS Institute Inc., 2008), and teacher

effect estimates were obtained for each set of analyses using REML implemented in

ASReml, Version 3.0 (Gilmour, Gogel, Cullis, & Thompson, 2009). For each simulation

and model combination, the teacher effect estimates were ranked within year. Among the

20 teachers in a given year i, the rank of the kth teacher, $\hat{R}_{k(i)}$, was between 1 and 20,

where $\hat{R}_{k(i)} = 1$ was assigned to the smallest estimated teacher effect and $\hat{R}_{k(i)} = 20$ to the

largest. In instances where multiple teachers had the same predicted teacher effect, the

teachers were assigned the mean of those corresponding ranks. The true teacher effects

for each simulation were similarly ranked and denoted $R_{k(i)}$. The corresponding

percentiles, $\hat{P}_{k(i)}$ and $P_{k(i)}$, are defined as $\hat{P}_{k(i)} = 100\left(\dfrac{\hat{R}_{k(i)}}{20+1}\right)$ and $P_{k(i)} = 100\left(\dfrac{R_{k(i)}}{20+1}\right)$,

respectively. The mean squared error of the estimator, $\hat{P}_{k(i)_p}^{m}$, for estimating the true

percentile, p, of teacher k in year i with model m, $MSE_{i\hat{P}}^{m} = E\left[\left(\hat{P}_{k(i)_p}^{m} - p\right)^2 \middle| p\right]$, was

estimated as $\widehat{MSE}_{i\hat{P}}^{m} = \dfrac{1}{1000}\sum_{q=1}^{1000}\left[\left(\hat{p}_{qk(i)_p}^{m} - p\right)^2\right]$, where $\hat{p}_{qk(i)_p}^{m}$ is the percentile estimate

from model m for teacher k whose true percentile is p in year i and simulation q

(Lockwood, Louis, & McCaffrey, 2002). Taking into account scaling considerations, the

square root of $\widehat{MSE}_{i\hat{P}}^{m}$, $\widehat{RMSE}_{i\hat{P}}^{m}$, was compared across true percentiles and models within

each time. For a given percentile p, the bias and standard error of the estimator, $\hat{P}^m_{k(i)_p}$,

$$Bias^m_{i\hat{P}} = E\left[\left(\hat{P}^m_{k(i)_p} - p\right)\Big| p\right] \text{ and } SE^m_{i\hat{P}} = \sqrt{Var\left[\left(\hat{P}^m_{k(i)_p} - p\right)\Big| p\right]}$$ were estimated as

$$\widehat{Bias}^m_{i\hat{P}} = \frac{1}{1000}\sum_{q=1}^{1000}\left[\left(\hat{p}^m_{qk(i)_p} - p\right)\right] \text{ and } \widehat{SE}^m_{i\hat{P}} = \sqrt{Var\left[\left(\hat{p}^m_{qk(i)_p} - p\right)\Big| p\right]},$$ respectively.

4.3.2 Results

For the Z-score models, the means and standard errors of the estimated teacher effect variance components are almost identical in the complete and missing tests scenarios. The same is true for the curve-of-factors model. In both the complete and the missing tests cases, the means and standard errors of the estimated teacher effect variances for the Z-score models are all 0.27 and 0.04, respectively. This finding is expected, because the only difference between a student's mean-centered MA and CA scores for a given year is the random measurement error, assumed to have constant variance across test and year. The curve-of-factors models in the complete and missing cases have similar means and standard errors of the estimated teacher effect variances, with averages for both cases of 22.59 and standard errors of 3.72 and 3.73 for the complete and missing tests scenarios, respectively. The means of the estimated teacher effect variance components for the curve-of-factors models are close to the true teacher effect variance, 22.5, but this is to be expected because the data are generated assuming a curve-of-factors structure with complete data for all eight tests. The means of the estimated teacher effect variances for the Z-score models, 0.27, indicate teachers are estimated to account for approximately 27% of the total variability present in the data, or 60.75 units squared (as opposed to the true teacher effect variance, 22.5). This result appears contrary to what the curve-of-factors models suggest, but the increasing

proportion of total variability accounted for by teachers each year is modified in the Z-score models through the standardized layered $\mathbf{Z}_t$ coefficient matrix. Consequently, the percent of total variability accounted for by teachers in the Z-score models is a weighted average of the increasing proportion of total variability attributed to teachers in the simulation across years.

Figure 4.3 shows the estimated RMSE for the different models with complete and missing tests for specific percentiles across the four years. The estimated RMSE is largest at the middle percentiles for all models, indicating the difficulty of accurately predicting a teacher's true rank increases as the teacher's true percentile moves toward the average. In year one, the estimated RMSE across percentiles is larger than that of the other three years, irrespective of the availability of all test scores or the model used. Across percentiles, the estimated RMSE does not change drastically from the complete to the missing case for the Z-score models. However, the magnitude of the estimated RMSE across percentiles for the curve-of-factors model that is noticeably smaller in the complete tests case tends toward the magnitude of the estimated RMSE of the Z-score models in the missing tests case.

The estimated bias at specific percentiles for the various models ranges from approximately two percentiles above to three percentiles below the true percentiles across the four years. As depicted in Figure 4.4, the bias at the extreme percentiles is largest in year one for all of the models, irrespective of whether data for all eight tests are available. Similar to the pattern evidenced in the RMSE plots, the estimated bias across percentiles does not change drastically from the complete to the missing case for the Z-score models. Again, the noticeably smaller magnitude of the estimated bias across percentiles for the

Figure 4.3: Comparison of $\widehat{RMSE}_{i\hat{P}}^{m}$ for the Curve-of-Factors (solid), CA Z-score (dash) and MA Z-score (dotted) Models with Complete and Missing Tests, and MA/CA Z-score (dot-dash) Model with Missing Tests for Specific Percentiles and Years
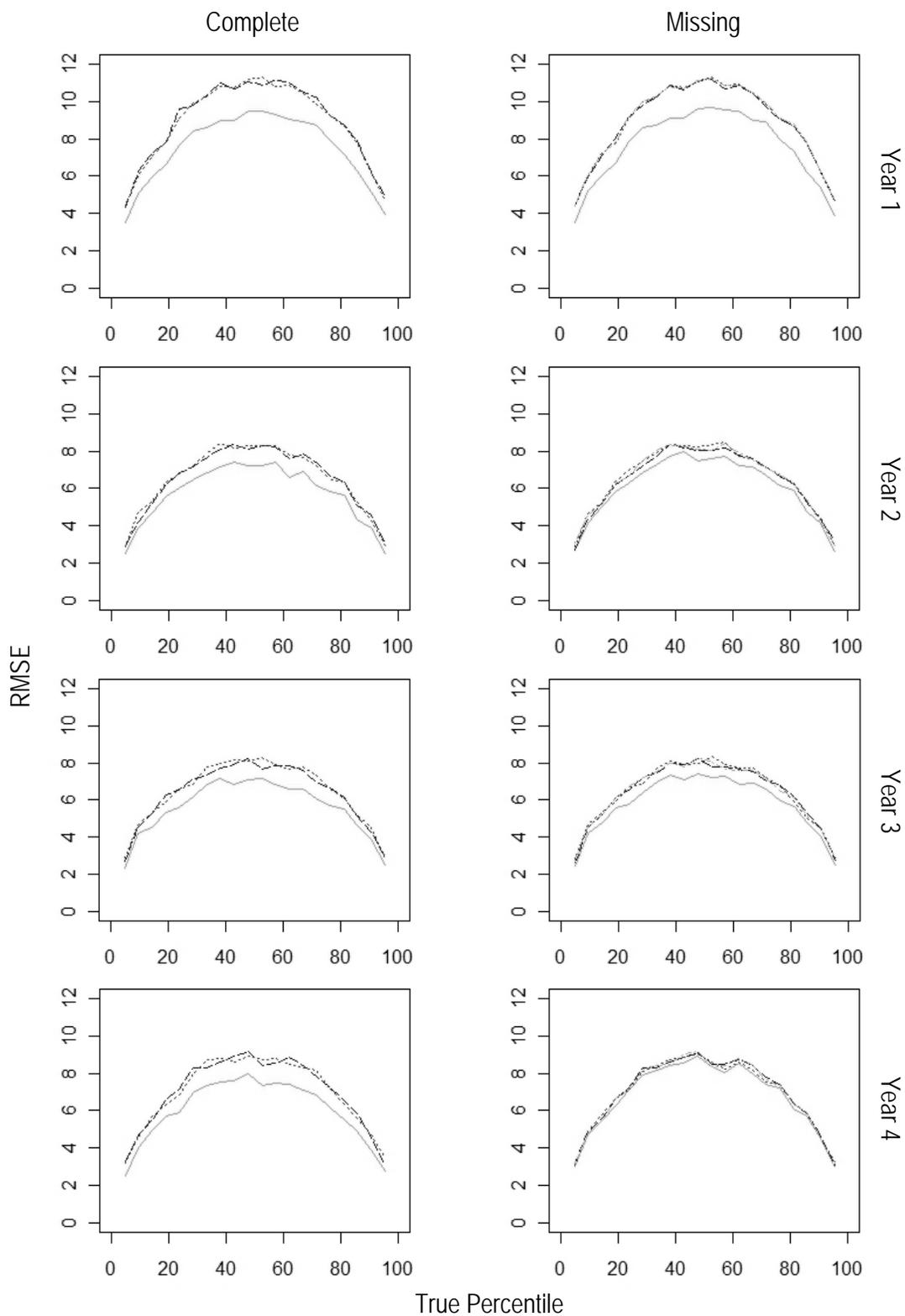
Figure 4.4: Comparison of $\widehat{Bias}_{i\hat{P}}^{m}$ for the Curve-of-Factors (solid), CA Z-score (dash) and MA Z-score (dotted) Models with Complete and Missing Tests, and MA/CA Z-score (dot-dash) Model with Missing Tests for Specific Percentiles and Years
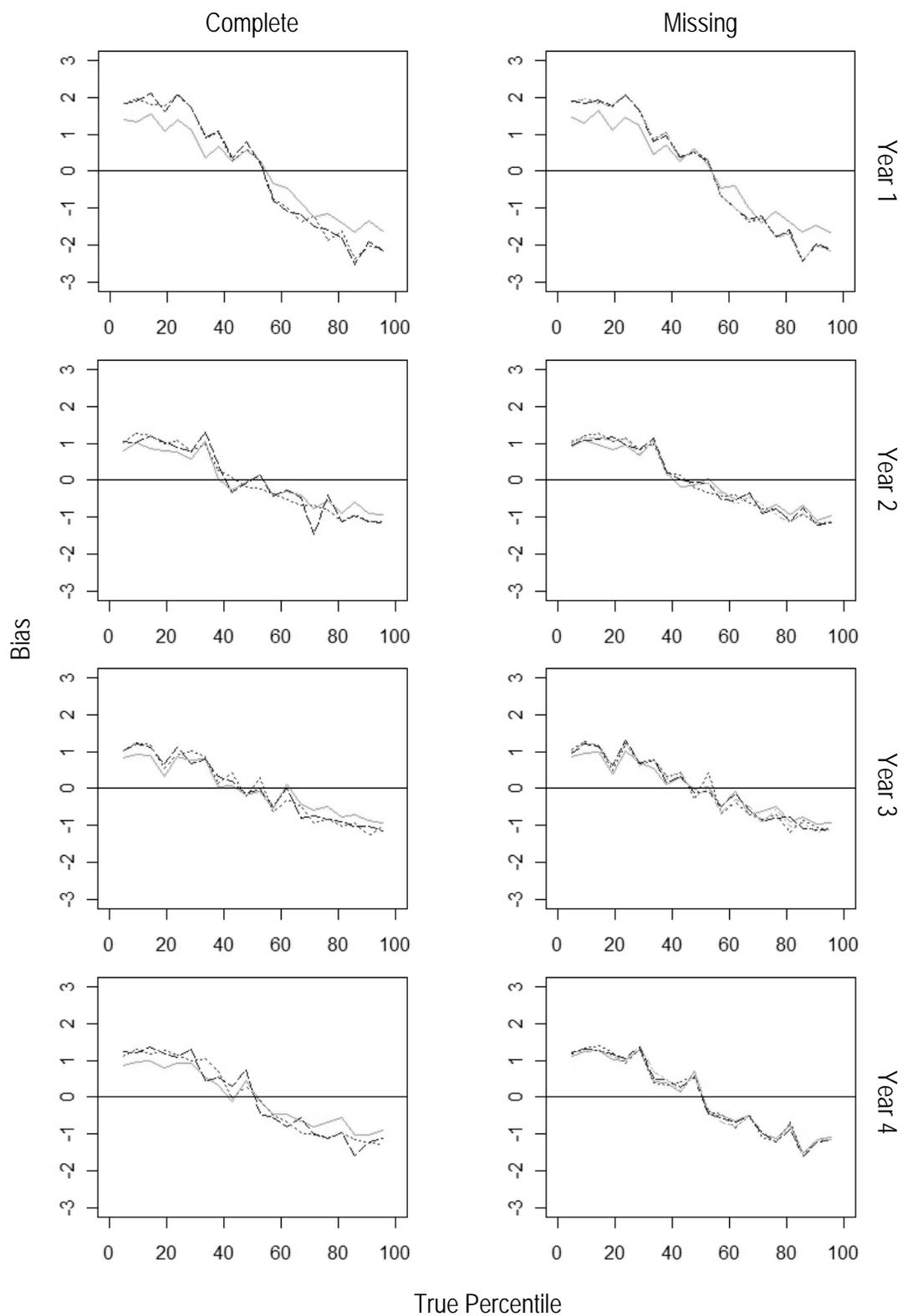
curve-of-factors model in the complete tests case increases toward that of the Z-score models in the missing tests case. Because the magnitude of the bias is relatively small, the patterns present in the plots of the estimated standard errors of the estimator, $\hat{P}^m_{ik_p}$, for specific percentiles are all but identical to those of the RMSE plots in Figure 4.2.

Figure 4.5 depicts the estimated sampling distributions of the estimator, $\hat{P}^m_{ik_p}$, for teachers truly at the 48th percentile and teachers truly at the 76th percentile. Because the simulation has only 20 teachers each year, the 48th and 76th percentiles are closest to the true percentiles of interest, the 50th and 75th percentiles. These distributions are compared across all years for the curve-of-factors and MA Z-score models with complete and missing tests. For clarity, the estimated distributions of other Z-score models are not included due to their similarity to the estimated sampling distributions for the MA Z-score models. In instances where multiple teachers have the same predicted teacher effect, the teachers are assigned the mean of those corresponding ranks. When creating these graphs, teachers originally assigned a non-integer rank are reassigned the next highest integer rank. For example, teachers assigned a rank of 14.5 are reassigned a rank of 15. This strategy errs in favor of the teachers, because it allows the teachers to be ranked higher, rather than lower, than their non-integer rank. Because ties in rankings are infrequent, this strategy only eliminated low relative frequencies representing rare occurrences, rather than meaningful information, in the plots. Comparing the estimated sampling distributions across true percentiles, years and models, there do not appear to be noteworthy differences between those for complete tests and those for missing tests. The estimated distributions at the 48th percentile in years one and four are slightly more variable than those in other years. In all years, the fatter tails illustrate the difficulty
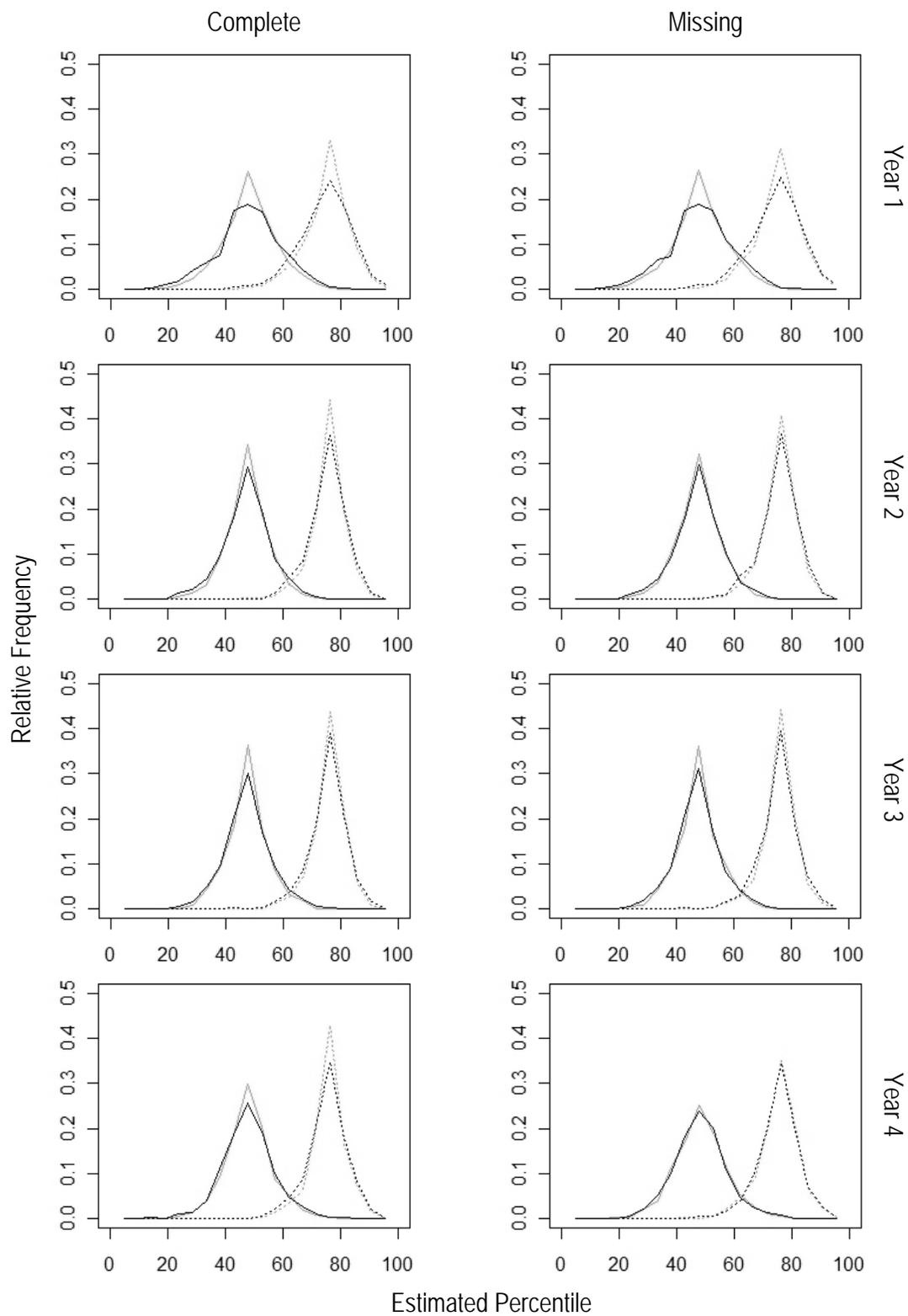
Figure 4.5: Comparison across Years of the Estimated Sampling Distributions of $\hat{P}^m_{ik_p}$ under the Curve-of-Factors (gray) and MA Z-score (black) Models with Complete and Missing Tests for Teachers Truly at the 48th (solid) and 76th (dotted) Percentiles

associated with accurately predicting a teacher's true rank when he or she is close to the average percentile. Although slight differences between the curve-of-factors model and the MA Z-score model appear in the complete tests case, these differences almost disappear in the missing tests case.

Across all four years, the estimated probability of classifying teachers in the upper quartile for each of the models with complete and missing tests increases nonlinearly as the true percentile increases (Figure 4.6). Although there do not appear to be noticeable differences between the estimated probabilities across models and availability of tests, the slopes of the curves in year one differ from those in subsequent years due to the differences in the variability of the sampling distributions for the estimator, $\hat{P}^m_{ik_p}$, at some true percentiles. The range of estimated probabilities of correctly ranking a teacher at the 76th percentile is not centered at .5, as might be expected. Instead, it ranges between approximately .5 and .7, because of the bias associated with the estimator at the 76th percentile.

4.4 Summary and Future Work

Curve-of-factors methodology applied in a value-added context extends the analysis of student achievement data to situations in which multiple tests with potentially different scales are given each year in a particular subject. Instead of estimating a teacher's effect on changes in a student's score over time, the curve-of-factors models allow the estimation of a teacher's effect on changes in some common, latent trait measured by the multiple instruments across years. In the simulation study, the behavior of the curve-of-
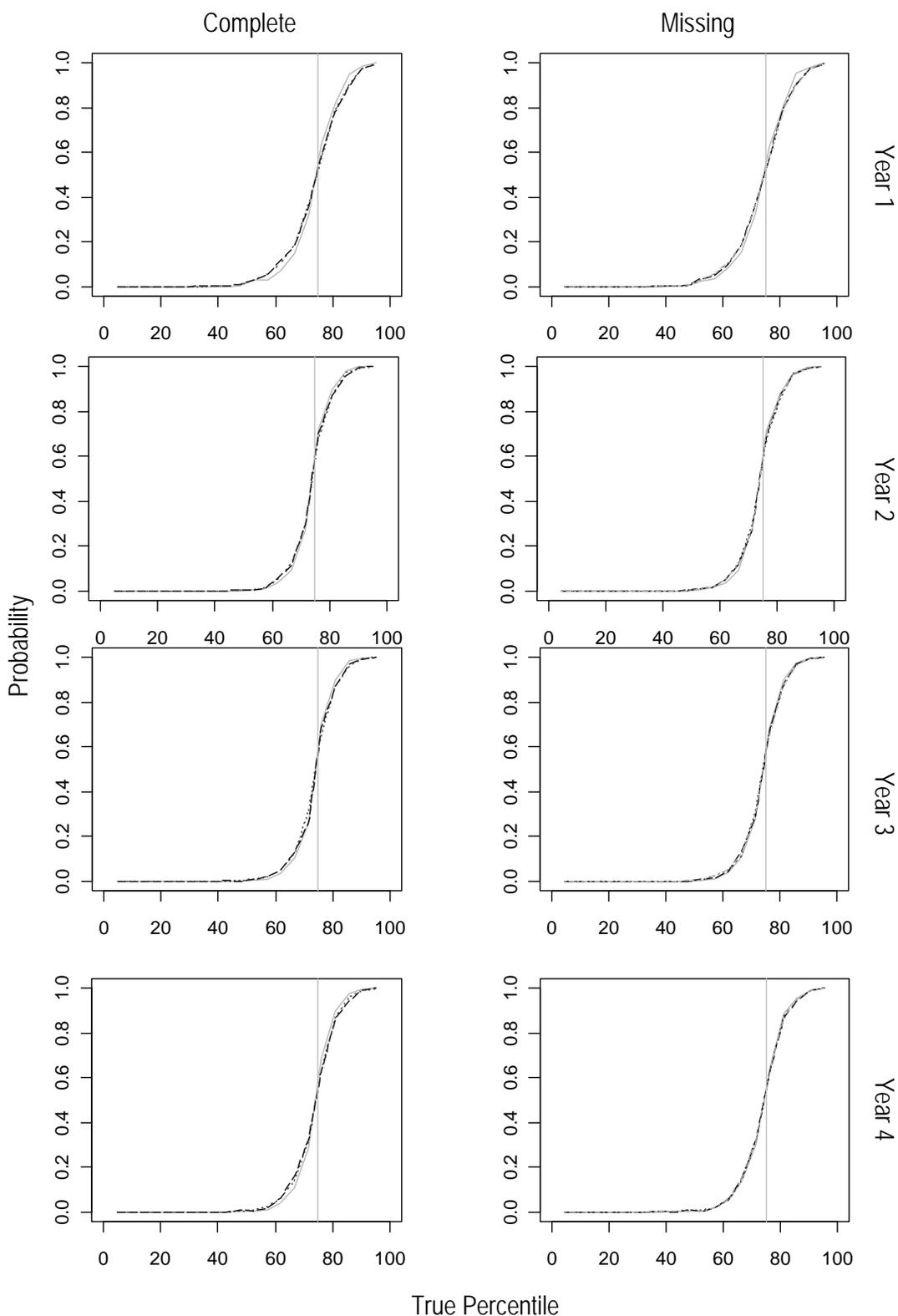
Figure 4.6: Comparison of the Estimated Probability of Classifying Teachers in the Upper Quartile for the Curve-of-Factors (solid), CA Z-score (dash) and MA Z-score (dotted) Models with Complete and Missing Tests, and MA/CA Z-score (dot-dash) Model with Missing Tests for Specific True Percentiles and Years

factors model when all tests are given tends toward that of the alternative Z-score models when a test is missing in both years one and four. With regards to estimating a teacher's true percentile, assuming a curve-of-factors model structure from which the data are originally simulated loses its statistical advantages, such as smaller RMSE, when tests are missing in some of the years. The results from the Z-score models are not noticeably different from one another for both the complete and missing tests cases. This is expected, because the only difference between a student's mean-centered MA and CA scores for a given year is the random measurement error, assumed to have constant variance across test and year. The lack of differences between Z-score models in the complete and missing tests cases is also expected. Because only one test can be used as the response in each year, having only one test in a given year instead of two is not consequential. Overall, the curve-of-factors model does not have drastically better performance relative to the alternative Z-score approach, even though the curve-of-factors model is the structure from which the data are simulated. Additional studies should use simulation to investigate how the estimation of teacher effects and their corresponding percentiles change under other assumed model structures and alternative weighting schemes for teacher effect variance components across years, as well as different ratios of teacher effect variance to error variance, such as those used by Lockwood et al. (2002), in various complete and incomplete test scenarios.

Chapter 5

Conclusions

Value-added modeling techniques estimate the contribution of educational factors, such as teachers, to growth in student achievement, while allowing for the possibility to control for the effect of non-educational factors. Several value-added models for estimating teacher effects have been proposed as alternatives to current test-based accountability procedures, such as adequate yearly progress (AYP), but each has its respective advantages and disadvantages. Although these methods have the potential to identify highly effective teachers, teacher effect estimates are sensitive to different modeling specifications, including the persistency of teacher effects. Furthermore, several statistical and psychometric issues exist, and sensitivity of teacher effects to such issues still needs to be explored.

This dissertation includes three chapters that provide an introduction to value-added methodology and discuss the estimation of teacher effects. Because value-added analyses require high-quality longitudinal data that are often not available, Chapters 3 and 4 proposed methodology for analyzing less-than-ideal student assessment data.

Specifically, Chapter 3 described how to use a value-added model when longitudinal student achievement data are not on a single developmental scale. Although the Z-score approach has limitations, it is an appropriate alternative to using raw data when analyzing less-than-ideal student achievement data across a mixture of norm- and criterion-referenced tests over time. This chapter addressed issues arising when using a layered, longitudinal linear mixed model to analyze gains in standardized scores,

including weighting considerations for variance components. Additionally, this chapter proposed methods for estimating teacher effects on student learning before and after teacher participation in professional development programs. Although the specific example used in this chapter did not indicate the Math in the Middle Institute had a significant impact on participating teachers' effects on student learning, the instruments may not have been designed to detect such changes. Additionally, teacher change from professional development may take time to show an impact on student learning. It is also possible other factors impact a teacher's instruction and his or her ability or perceived ability to change in ways that align with the professional development program (Kromminga, in progress). When utilizing this methodology, determining whether the goals of the program align with what the instruments assess and acknowledging any existing limitations is essential.

Chapter 4 applied curve-of-factors methodology in a value-added context to extend the analysis of student achievement data to situations in which multiple tests with potentially different scales are given each year in a particular subject. Instead of estimating a teacher's effect on changes in a student's scores over time, the curve-of-factors model allowed the estimation of a teacher's effect on changes in some common, latent trait measured by the multiple instruments across years. In the simulation study, the behavior of the curve-of-factors model when all tests were given tended toward that of the alternative Z-score models when a test was missing in both years one and four. With regards to estimating a teacher's true percentile, assuming a curve-of-factors model structure from which the data were originally simulated lost its statistical advantages, such as smaller RMSE, when tests were missing in some of the years. In general, the

curve-of-factors model did not have drastically better performance relative to the alternative Z-score approach, even though the curve-of-factors model was the structure from which the data were simulated.

Together, these two approaches address concerns surrounding the estimation of value-added teacher effects when analyzing less-than-ideal student achievement data. However, considerations should be made when defining what *teacher effects* really describe, and teacher effect estimates should be linked to other valid measures of teacher effectiveness. Although standardized tests can provide useful information about a student's content knowledge in a particular subject, they can only measure a few of the many skills teachers help shape and influence in their students. Teacher evaluation should reflect these many areas of instruction and learning and not be based solely on students' performance on achievement tests. The uncertainty associated with value-added teacher effect estimates suggests such estimates should not be used in isolation of other measures for high-stakes evaluation purposes and decisions. Instead, value-added teacher effects could be one of many different aspects used for teacher development purposes, fostering high quality teaching and motivating informed improvements in education.

References

Ackerman, T., Heafner, T., & Bartz, D. (2006, April). *Teacher effect model for impacting student achievement*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29*(3), 14-22, 43-46.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65.

Borko, H. (2004). Professional development and teaching and learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3-15.

Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K-16*, *8*(1), 3-32.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of meditational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10*(2), 238-262.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1). Retrieved April 7, 2008, from http://epaa.asu.edu/ojs/article/view/392

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81-112.

Doran, H. C. (2003). Adding value to accountability. *Educational Leadership, 61*(3), 55-59.

Drury, D., & Doran, H. C. (2003). The value of value-added analysis. *Policy Research Brief for National School Board Association*, *3*(1), 1-4.

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2000). Risk and protective factors influencing adolescent problem behavior: A multivariate latent growth curve analysis. *Annals of Behavioral Medicine, 22*(2), 103-109.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systematic reform. *Teaching and Teacher Education*, *19*(6), 643-658.

Frome, P., Lasater, B., & Cooney, S. (2005). *Well-qualified teachers and high quality teaching: Are they the same?* Atlanta, GA: Southern Regional Educational Board.

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

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). *ASReml user guide, Release 3.0*. Hernel Hempstead, HP1 1ES, UK: VSN International Ltd.

Green, J. L., Smith, W. M., Heaton, R. M., Jiao, S., & Stroup, W. W. (under review). Estimating the impact of a professional development program on student learning using a value-added model. Submitted to *Journal of Educational and Behavioral Statistics*.

Guskey, T. R. (1994, April). *Professional development in education: In search of the optimal mix*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Hershberg, T., Simon, V.A., & Lea-Kruger, B. (2004). Measuring what matters. *American School Board Journal, 191*(2), 27-31.

Hill, H. C. (2007a). Learning in the teaching workforce. *The Future of Children*, *17*(1), 111-127.

Hill, H. C. (2007b). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, *29*(2), 95-114.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11-30.

Kromminga, B. (in progress). *Down the road of change: A mixed methods study of teacher practices and beliefs following intensive professional development.* Unpublished doctoral dissertation, University of Nebraska-Lincoln, Lincoln, NE.

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

Lewis, W. L., Heaton, R., McGowan, T., & Jacobsen, B. (2004). *Math in the Middle Institute Partnership.* NSF grant proposal project description.

Lissitz, R. W. (Ed.). (2005). *Value added models in education: Theory and applications.* Maple Grove, MN: JAM Press.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd Ed.). Cary, NC: SAS Institute Inc.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 181-211). Mahwah, NJ: Lawrence Erlbaum Associates.

Lockwood, J. R., Louis, T. A., & McCaffrey, D. F. (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics, 27*(3), 255-270.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. P. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44*(1), 47-67.

Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics*, *32*(2), 125-150.

Loucks-Horsley, S., Stiles, K., & Hewson, P. (1996). *Principles of effective professional development for mathematics and science education: A synthesis of standards. NISE Brief, 1*(1), 1-6.

Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics, 31*(1), 35-62.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67-101.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

No Child Left Behind. (2001). Washington, D.C.: U.S. Department of Education.

Pearson. (2008). *Metropolitan Achievement Tests®, eighth edition (MAT 8)*. Retrieved April 28, 2008, from http://harcourtassessment.com/haiweb/cultures/en-us/productdetail.htm?pid=E164C

Presley, J. B., White, B. R., & Gong, Y. (2005). *Examining the distribution and impact of teacher quality in Illinois*. Edwardsville, IL: Illinois Education Research Council.

Rasbash, J., & Browne, W. J. (2008). Non-hierarchical multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 301-334). New York: Springer.

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal for Educational and Behavioral Statistics, 29*(1), 121-129.

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

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science, 6*(1), 15-51.

Roesch, S. C., Norman, G. J., Adams, M. A., Kerr, J., Sallis, J. F., Ryan, S., Calfas, K. J., & Patrick, K. (2009). Latent growth curve modeling of adolescent physical activity: Testing parallel process and mediation models. *Journal of Health Psychology, 14*(2), 313-325.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*(8), 1525-1567.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. J*ournal of Personnel Evaluation in Education*, *14*(4), 329-339.

Sanders, W. L. (2004). *A summary of conclusions drawn from longitudinal analyses of student achievement data over the past 22 years (1982-2004)*. Paper presented at the Governors Education Symposium, Ashville, NC.

Sanders, W. L. (2006, October). *Comparisons among various educational assessment value-added models*. Paper presented at the Battelle for Kids National Value-Added Conference, Columbus, Ohio.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (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 educational measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

SAS Institute Inc. (2008). *SAS/STAT user's guide, Version 9.2*. Cary, NC: SAS Institute Inc.

Shaha, S. H., Lewis, V. K., O'Donnell, T. J., & Brown, D. H. (2004). Evaluating professional development: An approach to verifying program impact on teachers and students. *The Journal of Research in Professional Learning*. Retrieved April 2, 2008 from http://www.nsdc.org/library/publications/research/shaha.pdf

Smith, W. M. (2004). *Nebraska assessments evaluation.* Unpublished project paper, The Math in the Middle Institute Partnership.

Stroup, W. (2007). *UNL research team analysis efforts to date of achievement test data*. Unpublished project paper, The Math in the Middle Institute Partnership.

Stroup, W., & Fang, X. (2006). *Analysis of student math achievement data: Preliminary results*. Unpublished project paper, The Math in the Middle Institute Partnership.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89-122.

Wilson, S. M., & Berne, J. (1999). Teacher knowledge and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, *24*(1), 173-209.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57-67.

Wright, S. P., & Sanders, W. L. (2008, April). *Decomposition of estimates in a layered value-added assessment model*. Paper presented at the National Conference on Value-Added Modeling, Madison, WI.

Wright, S. P., Sanders, W. L., & Rivers, J. C. (2006). Measurement of academic growth of individual students toward variable and meaningful academic standards. In R. Lissitz (Ed.), *Longitudinal and value added models of student performance* (pp. 385-406). Maple Grove, MN: JAM Press.

Appendix A

Z-score Method SAS Analysis Code for Math in the Middle Example[3]

```
libname lps 'C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\LPS Analysis\Data\LPS\analyze LPS using z score';
libname lpss 'C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\LPS Analysis\Data\LPS\reformat dataset';

%macro standardize;
%do i=5 %to 8;
%do j=4 %to 8;
data select;
set lpss.fulldata;
if grade=&i. and year=&j.;
proc sort data=select; by stuid ;
data select_undup;
set select;
by stuid ;
if first.stuid;
proc sql;
create table std_sco_g&i.0&j. as select
case
when std(crt) ne . then (crt-mean(crt))/std(crt)
else (mat-mean(mat))/std(mat)
end
 as std_score, grade-year as track, *
from select_undup;
quit;
%end;
%end;
%mend;
%standardize;

data lps.stddata_full;
set std_sco_g504 std_sco_g505 std_sco_g506 std_sco_g507
std_sco_g508 std_sco_g604 std_sco_g605 std_sco_g606
std_sco_g607 std_sco_g608 std_sco_g704 std_sco_g705
std_sco_g706 std_sco_g707 std_sco_g708 std_sco_g806
std_sco_g807 std_sco_g808;
if tidn30 = "" and M2 = "" then M2 = 0;
run;
```

---

[3] This code was written in collaboration with Shuo Jiao.

```
%let student=stuid;
%let ranvar=tidn30;
%let data_o=lps.stddata_full;
%let zmax=1;
/*This macro creates the layered Z matrix, I created this
macro for the general case when the teachers are not 1, 2,
3, ...., but instead are names.*/
%macro
randvar_full(dataname,w11,w21,w22,w31,w32,w33,w41,w42,w43,w
44);
proc sort data=&data_o; by &ranvar;
data Z(keep=&ranvar);
set &data_o;
by &ranvar;
if first.&ranvar;
run;
data zz;
set z;
rank=_N_;
run;
proc sql;
select max(rank) into: zmax from zz;

data new_1;
merge &data_o zz;
by &ranvar;

proc sort data=new_1; by  track &student grade;

data new_2 (drop=Zmat%eval(&zmax+1) Zmat_a%eval(&zmax+1)
Zmat_b%eval(&zmax+1));
set new_1;
by track &student grade;

/*Create lag variable*/
array reset(3) lagrank1-lagrank3;
array mim(3) lagmim1-lagmim3;
lagrank1=lag(rank);
lagrank2=lag2(rank);
lagrank3=lag3(rank);
lagmim1=lag(m2);
lagmim2=lag2(m2);
lagmim3=lag3(m2);
if first.&student then count=1;
do i=count to 3;
   reset(i)=%eval(&zmax+1);
```

```
    mim(i)=2;
end;
count+1;

%do ii=1 %to %eval(&zmax+1);
Zmat&ii=0;
Zmat_b&ii=0;
Zmat_a&ii=0;
%end;

array Zmat(%eval(&zmax+1)) Zmat1-Zmat%eval(&zmax+1);
if grade=5  then Zmat[rank]+&w11;
else if grade=6 then do; Zmat[rank]+&w21;
Zmat[lagrank1]+&w22; end;
else if grade=7 then do; Zmat[rank]+&w31;
Zmat[lagrank1]+&w32; Zmat[lagrank2]+&w33; end;
else if grade=8 then do; Zmat[rank]+&w41;
Zmat[lagrank1]+&w42; Zmat[lagrank2]+&w43;
Zmat[lagrank3]+&w44; end;

array Zmat_a(%eval(&zmax+1)) Zmat_a1-Zmat_a%eval(&zmax+1);
array Zmat_b(%eval(&zmax+1)) Zmat_b1-Zmat_b%eval(&zmax+1);
if grade=5 then do;
if m2=1 then Zmat_a[rank]+&w11;
if m2=0 then Zmat_b[rank]+&w11;
end;
else if grade=6 then do;
if m2=1 then Zmat_a[rank]+&w21;
if m2=0 then Zmat_b[rank]+&w21;
if lagmim1=1 then Zmat_a[lagrank1]+&w22;
if lagmim1=0 then Zmat_b[lagrank1]+&w22;
end;
else if grade=7 then do;
if m2=1 then Zmat_a[rank]+&w31;
if m2=0 then Zmat_b[rank]+&w31;
if lagmim1=1 then Zmat_a[lagrank1]+&w32;
if lagmim1=0 then Zmat_b[lagrank1]+&w32;
if lagmim2=1 then Zmat_a[lagrank2]+&w33;
if lagmim2=0 then Zmat_b[lagrank2]+&w33;
end;
else if grade=8 then do;
if m2=1 then Zmat_a[rank]+&w41;
if m2=0 then Zmat_b[rank]+&w41;
if lagmim1=1 then Zmat_a[lagrank1]+&w42;
if lagmim1=0 then Zmat_b[lagrank1]+&w42;
if lagmim2=1 then Zmat_a[lagrank2]+&w43;
if lagmim2=0 then Zmat_b[lagrank2]+&w43;
```

```
if lagmim3=1 then Zmat_a[lagrank3]+&w44;
if lagmim3=0 then Zmat_b[lagrank3]+&w44;
end;

sum_z=sum(of Zmat2-Zmat%eval(&zmax)); *Change to two so the
sum of missing teacher not included in standardization of
weights;
run;

data new_3;
set new_2;
%do ii=2 %to &zmax;*Change to two so that the sum of
missing teacher not included in standardization of weights;
if sum_z ne 0 then Zmat&ii=(Zmat&ii/sum_z)**0.5;
else Zmat&ii=0; *Eliminate division by zero;
if sum_z ne 0 then Zmat_b&ii=(Zmat_b&ii/sum_z)**0.5;
else Zmata&ii=0;
if sum_z ne 0 then Zmat_a&ii=(Zmat_a&ii/sum_z)**0.5;
else Zmatb&ii=0;
%end;
run;

data &dataname(drop=count i);
set new_3;
sum_za= sum(of Zmat_a2-Zmat_a%eval(&zmax));
run;
%mend;

%randvar_full(lps.new_std_data1,1,1,1,1,1,1,1,1,1,1);

%macro analysis;
proc sort data=LPS.new_std_data1;
by track stuid grade;
proc glimmix data=LPS.new_std_data1 ;
class track grade stuid;
model std_score=sum_za/solution;
random Zmat_a2-Zmat_a%trim(&zmax)/type=toep(1) solution;
random Zmat_b2-Zmat_b%trim(&zmax)/type=toep(1) solution;
random Zmat2-Zmat%trim(&zmax)/type=toep(1) solution;
random _residual_ /type=un subject=stuid*track;
nloptions om=congra; *Needed to change b/c Quanew
optimization could not be completed with missing teacher
links;
run;
%mend;
%analysis;
```

```
/* blup and covariate */
proc sort data=lps.new_std_data1 out=out;
by rank m2;
data lps.teacher_m2 (keep=tidn30 rank m2);
set out;
by rank;
if last.rank and rank ne 1;
run;


/*extract after effect*/
data aaa (keep=estimate_a StdErrPreda);
set lps.teacher_effect1_UN;
if substr(effect,1,6)="Zmat_a";
estimate_a=estimate ;
StdErrPreda = StdErrPred;
run;


/*extract before effect*/
data bbb (keep=estimate_b StdErrPredb);
set lps.teacher_effect1_UN;
if substr(effect,1,6)="Zmat_b";
estimate_b=estimate;
StdErrPredb = StdErrPred;
run;


/*extract common effect*/
data ccc (keep=estimate_c StdErrPredc);
set lps.teacher_effect1_UN;
if substr(effect,1,6) ne "Zmat_a" and substr(effect,1,6) ne
"Zmat_b" ;
estimate_c=estimate ;
StdErrPredc = StdErrPred;
run;


/*combine all effects*/
data lps.effect_crtz;
merge aaa bbb ccc lps.teacher_m2;
run;


data compare;
set lps.effect_crtz;
before=estimate_b+estimate_c;
after=estimate_a+estimate_c+0.01905; /*substitute this
number with the estimated fixed effect for sum_za*/
diff=after-before;
run;
```

Appendix B

Complete Tests SAS Simulation Code and ASReml Analysis Code

```
*******************************************;
*** Create a Macro to Simulate Data Sets ***;
*******************************************;

Libname Results 'C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Results';

*Delete Old True Teacher Effects File before re-creating
data sets;
Data Results.True_Teacher_Effects;
 Delete;
Run;

%macro
cof_full(nstudent,seed,dataname2,dataname3,dataname4);
%let seed1 = &seed + 3220;
%let seed2 = &seed + 9086;
%let seed3 = &seed + 3;

*************************************;
*** Make Random Teacher Assignments ***;
*************************************;

/* you do this by ranking the random numbers from step 1
   then for year 1 you call ranks 1 to (n/20) teacher 1,
        ranks (n/20)+1 to (2n/20) teacher 2, etc
        for year 2 you call ranks 1 to (n/20) teacher 21,
        ranks (n/20)+1 to (2n/20) teacher 22, etc
        and similarly for years 3 and 4 */

/* step 1: generate random numbers for each student each
year */
data step_1;
 seed=&seed1;
 do student=1 to &nstudent;
  rntyr1=ranuni(seed);
  rntyr2=ranuni(seed);
  rntyr3=ranuni(seed);
```

```
  rntyr4=ranuni(seed);
  output;
 end;

/* step 2: create the ranks using proc rank */
proc rank data=step_1 out=step_2;
 var rntyr1-rntyr4;
 ranks tyr1 tyr2 tyr3 tyr4;
run;

/* step 3: use the ranks to make teacher assignments */
data step_3;
 set step_2;
 real_obs=1;
  teacher_yr1=1;
  if tyr1>((1/20)*&nstudent) then teacher_yr1=2;
  if tyr1>((2/20)*&nstudent) then teacher_yr1=3;
  if tyr1>((3/20)*&nstudent) then teacher_yr1=4;
  if tyr1>((4/20)*&nstudent) then teacher_yr1=5;
  if tyr1>((5/20)*&nstudent) then teacher_yr1=6;
  if tyr1>((6/20)*&nstudent) then teacher_yr1=7;
  if tyr1>((7/20)*&nstudent) then teacher_yr1=8;
  if tyr1>((8/20)*&nstudent) then teacher_yr1=9;
  if tyr1>((9/20)*&nstudent) then teacher_yr1=10;
  if tyr1>((10/20)*&nstudent) then teacher_yr1=11;
  if tyr1>((11/20)*&nstudent) then teacher_yr1=12;
  if tyr1>((12/20)*&nstudent) then teacher_yr1=13;
  if tyr1>((13/20)*&nstudent) then teacher_yr1=14;
  if tyr1>((14/20)*&nstudent) then teacher_yr1=15;
  if tyr1>((15/20)*&nstudent) then teacher_yr1=16;
  if tyr1>((16/20)*&nstudent) then teacher_yr1=17;
  if tyr1>((17/20)*&nstudent) then teacher_yr1=18;
  if tyr1>((18/20)*&nstudent) then teacher_yr1=19;
  if tyr1>((19/20)*&nstudent) then teacher_yr1=20;
  teacher_yr2=21;
  if tyr2>((1/20)*&nstudent) then teacher_yr2=22;
  if tyr2>((2/20)*&nstudent) then teacher_yr2=23;
  if tyr2>((3/20)*&nstudent) then teacher_yr2=24;
  if tyr2>((4/20)*&nstudent) then teacher_yr2=25;
  if tyr2>((5/20)*&nstudent) then teacher_yr2=26;
  if tyr2>((6/20)*&nstudent) then teacher_yr2=27;
  if tyr2>((7/20)*&nstudent) then teacher_yr2=28;
  if tyr2>((8/20)*&nstudent) then teacher_yr2=29;
  if tyr2>((9/20)*&nstudent) then teacher_yr2=30;
  if tyr2>((10/20)*&nstudent) then teacher_yr2=31;
  if tyr2>((11/20)*&nstudent) then teacher_yr2=32;
  if tyr2>((12/20)*&nstudent) then teacher_yr2=33;
```

```
if tyr2>((13/20)*&nstudent) then teacher_yr2=34;
if tyr2>((14/20)*&nstudent) then teacher_yr2=35;
if tyr2>((15/20)*&nstudent) then teacher_yr2=36;
if tyr2>((16/20)*&nstudent) then teacher_yr2=37;
if tyr2>((17/20)*&nstudent) then teacher_yr2=38;
if tyr2>((18/20)*&nstudent) then teacher_yr2=39;
if tyr2>((19/20)*&nstudent) then teacher_yr2=40;
teacher_yr3=41;
if tyr3>((1/20)*&nstudent) then teacher_yr3=42;
if tyr3>((2/20)*&nstudent) then teacher_yr3=43;
if tyr3>((3/20)*&nstudent) then teacher_yr3=44;
if tyr3>((4/20)*&nstudent) then teacher_yr3=45;
if tyr3>((5/20)*&nstudent) then teacher_yr3=46;
if tyr3>((6/20)*&nstudent) then teacher_yr3=47;
if tyr3>((7/20)*&nstudent) then teacher_yr3=48;
if tyr3>((8/20)*&nstudent) then teacher_yr3=49;
if tyr3>((9/20)*&nstudent) then teacher_yr3=50;
if tyr3>((10/20)*&nstudent) then teacher_yr3=51;
if tyr3>((11/20)*&nstudent) then teacher_yr3=52;
if tyr3>((12/20)*&nstudent) then teacher_yr3=53;
if tyr3>((13/20)*&nstudent) then teacher_yr3=54;
if tyr3>((14/20)*&nstudent) then teacher_yr3=55;
if tyr3>((15/20)*&nstudent) then teacher_yr3=56;
if tyr3>((16/20)*&nstudent) then teacher_yr3=57;
if tyr3>((17/20)*&nstudent) then teacher_yr3=58;
if tyr3>((18/20)*&nstudent) then teacher_yr3=59;
if tyr3>((19/20)*&nstudent) then teacher_yr3=60;
teacher_yr4=61;
if tyr4>((1/20)*&nstudent) then teacher_yr4=62;
if tyr4>((2/20)*&nstudent) then teacher_yr4=63;
if tyr4>((3/20)*&nstudent) then teacher_yr4=64;
if tyr4>((4/20)*&nstudent) then teacher_yr4=65;
if tyr4>((5/20)*&nstudent) then teacher_yr4=66;
if tyr4>((6/20)*&nstudent) then teacher_yr4=67;
if tyr4>((7/20)*&nstudent) then teacher_yr4=68;
if tyr4>((8/20)*&nstudent) then teacher_yr4=69;
if tyr4>((9/20)*&nstudent) then teacher_yr4=70;
if tyr4>((10/20)*&nstudent) then teacher_yr4=71;
if tyr4>((11/20)*&nstudent) then teacher_yr4=72;
if tyr4>((12/20)*&nstudent) then teacher_yr4=73;
if tyr4>((13/20)*&nstudent) then teacher_yr4=74;
if tyr4>((14/20)*&nstudent) then teacher_yr4=75;
if tyr4>((15/20)*&nstudent) then teacher_yr4=76;
if tyr4>((16/20)*&nstudent) then teacher_yr4=77;
if tyr4>((17/20)*&nstudent) then teacher_yr4=78;
if tyr4>((18/20)*&nstudent) then teacher_yr4=79;
if tyr4>((19/20)*&nstudent) then teacher_yr4=80;
```

```
Run;

/* Transpose Random Teacher Assignments from Step 3 */
Data Tch_assignments;
Set Step_3;
Array S[4] teacher_yr1 teacher_yr2 teacher_yr3 teacher_yr4;
Do Time = 1 to 4;
     Teacher = S[Time];
Output;
End;
Keep Student Teacher Time;
Run;


******************************;
*** Generate Student Scores ***;
******************************;


/* step 1: generate student scores and output data set */
Proc IML;
Call Randseed(&seed2);
 N=1;
 G={157.5 116.7 99.8 83.3, 116.7 135 98.6 82.7, 99.8 98.6
112.5 80.5, 83.3 82.7 80.5 90}; *AR(1) structure with 1,
0.8, 0.75, 0.7 correlations;
 R=I(4)@{45 0, 0 45}; *20% of total variance attributed to
measurement error;
 lps_sim=j(&nstudent,9,.);
 Do Student=1 to &nstudent;
 D = (Randnormal(1, J(4,1,0), G))`;
 D2 = D@J(2,1,1);
 D = D2`;
 E = Randnormal(1, J(8,1,0), R);
 Mean={62 231 84 234 54 236 70 239}; *MAT/3 to adjust for
variance, (CRT_yr1)*2 to adjust for variance;
 Scores = Mean+D+E;
  lps_sim[student,1]=student;
  lps_sim[student,2]=scores[1,1];
  lps_sim[student,3]=scores[1,2];
  lps_sim[student,4]=scores[1,3];
  lps_sim[student,5]=scores[1,4];
  lps_sim[student,6]=scores[1,5];
  lps_sim[student,7]=scores[1,6];
  lps_sim[student,8]=scores[1,7];
  lps_sim[student,9]=scores[1,8];
End;
Create student_scores from lps_sim;
Append from lps_sim;
```

```
Run;

Data student_scores;
 Set student_scores;
 Student=col1;
 yr1a_score=col2;
 yr1b_score=col3;
 yr2a_score=col4;
 yr2b_score=col5;
 yr3a_score=col6;
 yr3b_score=col7;
 yr4a_score=col8;
 yr4b_score=col9;
 Drop col1-col9;
Run;

/* step 2: Transpose student scores */
Data Sim_Scores;
Set Student_Scores;
Array S[8] yr1a_score yr1b_score yr2a_score yr2b_score
yr3a_score yr3b_score yr4a_score yr4b_score;
Do I = 1 to 8;
     Score = S[I];
Output;
End;
Keep Student I Score;
Run;

/* step 3: create time and test variables */
Proc Sort Data = Sim_Scores;
By I;
Run;
Data Sim_Scores;
Set Sim_Scores;
Time = Ceil(I/2);
If Mod(I,2) = 1 then Test = 'CRT'; Else Test = 'MAT';
Keep Student Time Test Score;
Run;

***************************************************;
*** Merge Student Scores and Teacher Assignments ***;
***************************************************;

Proc Sort Data = Sim_Scores;
By Student Time;
Proc Sort Data = Tch_Assignments;
By Student Time;
```

```
Run;
Data Sim_Scores_n_Tch;
Merge Sim_Scores Tch_Assignments;
By Student Time;
Run;

*****************************;
*** Create Layered Z-matrix ***;
*****************************;

/* step 1: separate MAT & CRT scores to create z-matrix for
each type of test*/
Data Sim_Scores_MAT;
Set Sim_Scores_n_Tch;
If Test = 'MAT';
Run;
Data Sim_Scores_CRT;
Set Sim_Scores_n_Tch;
If Test = 'CRT';
Run;

/* step 2: macro creates the layered z-matrix for the
general case when the teachers are not 1, 2, 3, ...., but
instead are names.*/
%let Student=Student;
%let Ranvar=Teacher;
%let zmax=1;

%macro
Randvar_Full(dataname1,w11,w21,w22,w31,w32,w33,w41,w42,w43,
w44);
Proc Sort data=&data_o; By &ranvar;
Data Z(keep=&ranvar);
Set &data_o;
By &ranvar;
If first.&ranvar;
Run;
Data zz;
Set z;
Rank=_N_;
Run;
Proc sql;
Select max(rank) into: zmax from zz;

Data new_1;
Merge &data_o zz;
By &ranvar;
```

```
Proc Sort Data=new_1; By  &student Time;

Data New_2 (Drop=Zmat%eval(&zmax+1));
Set new_1;
By &student Time;

/*Create lag variable*/
Array reset(3) lagrank1-lagrank3;
lagrank1=lag(rank);
lagrank2=lag2(rank);
lagrank3=lag3(rank);
If first.&student then count=1;
Do i=count to 3;
   Reset(i)=%eval(&zmax+1);
End;
Count+1;

%Do ii=1 %to %Eval(&zmax+1);
Zmat&ii=0;
%End;

Array Zmat(%eval(&zmax+1)) Zmat1-Zmat%eval(&zmax+1);
If time = 1  then Zmat[rank]+&w11;
Else if Time = 2 then do; Zmat[rank]+&w21;
Zmat[lagrank1]+&w22; End;
Else if Time = 3 then do; Zmat[rank]+&w31;
Zmat[lagrank1]+&w32; Zmat[lagrank2]+&w33; End;
Else if Time = 4 then do; Zmat[rank]+&w41;
Zmat[lagrank1]+&w42; Zmat[lagrank2]+&w43;
Zmat[lagrank3]+&w44; End;

Run;

Data &dataname1(drop=count i);
Set New_2;
Run;
%mend;

/* step 3: run macro for both tests */
%Let Data_o=Sim_Scores_CRT;
%Randvar_full(Sim_Scores_CRT1,1,1,1,1,1,1,1,1,1,1);
%Let Data_o=Sim_Scores_MAT;
%Randvar_full(Sim_Scores_MAT1,1,1,1,1,1,1,1,1,1,1);

/* step 4: merge data for both tests */
DATA Sim_Scores;
```

```
  SET Sim_Scores_CRT1 Sim_Scores_MAT1;
Run;
Proc Sort Data = Sim_Scores;
 BY Student Time Test;
Run;


****************************;
*** Create Final Data Set ***;
****************************;

/* step 1: Generate random teacher effects and find ranks
and percentiles by year */

Data Gen_tch_value;
 Seed=&seed3;
 Do Teacher=1 to 80;
  t_eff=rannor(seed)*(sqrt(22.5));
   Output;
 End;
Run;

Data Gen_tch_value;
 Set Gen_tch_value;
 Year = 1;
 If Teacher > 20 then Year = 2;
 If Teacher > 40 then Year = 3;
 If Teacher > 60 then Year = 4;
Run;

Proc Sort Data = Gen_tch_value;
 By Year Teacher;
Run;

Proc Rank Data = Gen_tch_value Out=Rank;
 By Year;
 Ranks True_Rank;
 Var T_Eff;
Run;

Data Gen_tch_value_Temp;
Set Rank;
 Exp = &nsim;
 True_Effect = t_eff;
 True_Percentile=100*(True_Rank/21);
 Drop t_eff;
Run;
```

```
*Send true teacher effects to one permanent file, updated
for each simulation;
%If %eval(&nsim) = 1 %then %do;
Data Results.True_Teacher_Effects;
Set Gen_tch_value_Temp;
Run;
%End;
%Else %do;
Data Results.True_Teacher_Effects;
Set Results.True_Teacher_Effects Gen_tch_value_Temp;
Run;
%End;

/* step 2: merge student data, teacher assignments, and
teacher values into single data set */
Proc IML;
Use Sim_Scores;
Read all var{student score time test teacher rank lagrank1
lagrank2 lagrank3 Zmat1 Zmat2 Zmat3 Zmat4 Zmat5 Zmat6 Zmat7
Zmat8 Zmat9 Zmat10 Zmat11 Zmat12 Zmat13 Zmat14 Zmat15
Zmat16 Zmat17 Zmat18 Zmat19 Zmat20 Zmat21 Zmat22 Zmat23
Zmat24 Zmat25 Zmat26 Zmat27 Zmat28 Zmat29 Zmat30 Zmat31
Zmat32 Zmat33 Zmat34 Zmat35 Zmat36 Zmat37 Zmat38 Zmat39
Zmat40 Zmat41 Zmat42 Zmat43 Zmat44 Zmat45 Zmat46 Zmat47
Zmat48 Zmat49 Zmat50 Zmat51 Zmat52 Zmat53 Zmat54 Zmat55
Zmat56 Zmat57 Zmat58 Zmat59 Zmat60 Zmat61 Zmat62 Zmat63
Zmat64 Zmat65 Zmat66 Zmat67 Zmat68 Zmat69 Zmat70 Zmat71
Zmat72 Zmat73 Zmat74 Zmat75 Zmat76 Zmat77 Zmat78 Zmat79
Zmat80};
Use gen_tch_value;
Read all var{t_eff};
Z =
Zmat1||Zmat2||Zmat3||Zmat4||Zmat5||Zmat6||Zmat7||Zmat8||
Zmat9||Zmat10||Zmat11||Zmat12||Zmat13||Zmat14||Zmat15||
Zmat16||Zmat17||Zmat18||Zmat19||Zmat20||Zmat21||Zmat22||
Zmat23||Zmat24||Zmat25||Zmat26||Zmat27||Zmat28||Zmat29||
Zmat30||Zmat31||Zmat32||Zmat33||Zmat34||Zmat35||Zmat36||
Zmat37||Zmat38||Zmat39||Zmat40||Zmat41||Zmat42||Zmat43||
Zmat44||Zmat45||Zmat46||Zmat47||Zmat48||Zmat49||Zmat50||
Zmat51||Zmat52||Zmat53||Zmat54||Zmat55||Zmat56||Zmat57||
Zmat58||Zmat59||Zmat60||Zmat61||Zmat62||Zmat63||Zmat64||
Zmat65||Zmat66||Zmat67||Zmat68||Zmat69||Zmat70||Zmat71||
Zmat72||Zmat73||Zmat74||Zmat75||Zmat76||Zmat77||Zmat78||
Zmat79||Zmat80;
Score_Total = Score + Z*t_eff;
Test2 = J((&nstudent)*8,1,0);
Do i = 1 to (&nstudent)*8;
```

```
If test[i,1] = 'MAT' then test2[i,1] = 2; Else test2[i,1] =
1;
End;
Scores =
Z||student||score||score_total||time||teacher||test2;
Create tch_student_scores from Scores;
Append from Scores;
Print &nsim;
Run;
Quit;

/* step 3: rename and create variables for PP (Dataname 2)
Dataset*/
Data &dataname2;
 Set tch_student_scores;
 Array Z[80] Z1-Z80;
 Array Col[80] Col1-Col80;
Do i = 1 to 80;
 Z[i] = Col[i];
End;
 Exp = &nsim;
 Student=col81;
 No_Tch_Score=col82;
 Tch_Score_Final=col83;
 If Tch_Score_Final<0 then Tch_Score_Final=0;
 Time=col84;
 Teacher=col85;
 If Col86 = 1 then Test = 'CRT'; Else Test = 'MAT';
 Drop Col1-Col86;
Run;

/* Step 4: Create Zcrt (Dataname 3) and Zmat (Dataname 4)
Datasets */
Data Standard;
 Set &dataname2;
 sum_z=sum(of Z1-Z%eval(&zmax));
Run;

Data New3;
 Set Standard;
 %Do ii=1 %to &zmax;
     if sum_z ne 0 then Z&ii=(Z&ii/sum_z)**0.5; else
Z&ii=0;
 %End;
Run;

Proc Sort Data = New3;
```

```
By Time Test;
Run;

Proc Means Data = New3 Mean Std noprint;
By Time Test;
Var Tch_Score_Final;
Output Out = Stats Mean = Mean Std = Std;
Run;

Data Stats;
Set Stats;
Drop _Type_ _Freq_;
Run;

Data New;
Merge New3 Stats;
By Time Test;
Run;

Data New2;
Set New;
Zscore = (Tch_Score_Final - Mean)/Std;
Drop No_Tch_Score Tch_Score_Final Mean Std Sum_Z;
Run;

Data &Dataname3;
Set New2;
If Test = 'CRT';
Run;

Data &Dataname4;
Set New2;
If Test = 'MAT';
Run;

Proc Sort Data = &Dataname2;
By Student Time Test;
Proc Sort Data = &Dataname3;
By Student Time;
Proc Sort Data = &Dataname4;
By Student Time;
Run;

Quit; *This makes sure a Proc is not still running;
%mend;
```

```
**************************************************;
*** Create Macro to Analyze Simulated Data Sets ***;
**************************************************;


*Delete Old Predicted Teacher Effects File before re-
analyzing data sets;
Data Results.Pred_Teacher_Effects;
 Delete;
Run;

*Delete Old Variance Components File before re-analyzing
data sets;
Data Results.Pred_Variance_pp;
 Delete;
Run;
Data Results.Pred_Variance_zc;
 Delete;
Run;
Data Results.Pred_Variance_zm;
 Delete;
Run;


%macro Analyze(data,model);
*Export simulated SAS data set as a text file;
PROC EXPORT DATA= &data
             OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\AnalyzeThis.dat"
             DBMS=DLM REPLACE;
     DELIMITER='20'x;
     PUTNAMES=YES;
RUN;


%If &model = pp %then %do;
*Create ASReml .as file for Curve of Factors;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.as";
Data _null_;
File tmp;
put 'Curve of Factors for 80 Teacher, 2000 Students';
put ' Z1';
put ' Z2';
put ' Z3';
put ' Z4';
```

```
put ' Z5';
put ' Z6';
put ' Z7';
put ' Z8';
put ' Z9';
put ' Z10';
put ' Z11';
put ' Z12';
put ' Z13';
put ' Z14';
put ' Z15';
put ' Z16';
put ' Z17';
put ' Z18';
put ' Z19';
put ' Z20';
put ' Z21';
put ' Z22';
put ' Z23';
put ' Z24';
put ' Z25';
put ' Z26';
put ' Z27';
put ' Z28';
put ' Z29';
put ' Z30';
put ' Z31';
put ' Z32';
put ' Z33';
put ' Z34';
put ' Z35';
put ' Z36';
put ' Z37';
put ' Z38';
put ' Z39';
put ' Z40';
put ' Z41';
put ' Z42';
put ' Z43';
put ' Z44';
put ' Z45';
put ' Z46';
put ' Z47';
put ' Z48';
put ' Z49';
put ' Z50';
put ' Z51';
```

```
put ' Z52';
put ' Z53';
put ' Z54';
put ' Z55';
put ' Z56';
put ' Z57';
put ' Z58';
put ' Z59';
put ' Z60';
put ' Z61';
put ' Z62';
put ' Z63';
put ' Z64';
put ' Z65';
put ' Z66';
put ' Z67';
put ' Z68';
put ' Z69';
put ' Z70';
put ' Z71';
put ' Z72';
put ' Z73';
put ' Z74';
put ' Z75';
put ' Z76';
put ' Z77';
put ' Z78';
put ' Z79';
put ' Z80';
put ' i';
put ' exp   !I';
put ' student   !I 2000';
put ' notchscore';
put ' tchscorefinal';
put ' time !I';
put ' teacher !I';
put ' Test !A';
put 'AnalyzeThis.dat !SKIP=1 !MAXIT=100 !BRIEF';
put 'tchscorefinal ~ time.Test !r ![ Z1 Z2 Z3 Z4 Z5 Z6 Z7
Z8 Z9 Z10 Z11 Z12 Z13 Z14 Z15 Z16 Z17 Z18 Z19 Z20 ,';
put ' Z21 Z22 Z23 Z24 Z25 Z26 Z27 Z28 Z29 Z30 Z31 Z32 Z33
Z34 Z35 Z36 Z37 Z38 Z39 Z40 Z41 Z42 Z43 Z44 Z45 Z46 ,';
put ' Z47 Z48 Z49 Z50 Z51 Z52 Z53 Z54 Z55 Z56 Z57 Z58 Z59
Z60 Z61 Z62 Z63 Z64 Z65 Z66 Z67 Z68 Z69 Z70 Z71 Z72 ,';
put ' Z73 Z74 Z75 Z76 Z77 Z78 Z79 Z80 !] time.student';
```

```
put '0 0 2     # Measurement error 0 = Default (I_sigma-
squared) 0 = No terms in Direct Product 2 = Two Random
Effects on G-side';

put 'Z1 1';
put '80 0 I 1 !GP';

put 'time.student 2';
put 'time 0  US !GP';
put '1';
put '0 1';
put '0 0 1';
put '0 0 0 1';
put 'student 0 I';
Run;
%End;

%Else %do;
*Create ASReml .as file for Z-score Method;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.as";
Data _null_;
File tmp;
put 'Z-score Method for 80 Teachers, 2000 Students';
put ' Z1';
put ' Z2';
put ' Z3';
put ' Z4';
put ' Z5';
put ' Z6';
put ' Z7';
put ' Z8';
put ' Z9';
put ' Z10';
put ' Z11';
put ' Z12';
put ' Z13';
put ' Z14';
put ' Z15';
put ' Z16';
put ' Z17';
put ' Z18';
put ' Z19';
put ' Z20';
put ' Z21';
```

```
put ' Z22';
put ' Z23';
put ' Z24';
put ' Z25';
put ' Z26';
put ' Z27';
put ' Z28';
put ' Z29';
put ' Z30';
put ' Z31';
put ' Z32';
put ' Z33';
put ' Z34';
put ' Z35';
put ' Z36';
put ' Z37';
put ' Z38';
put ' Z39';
put ' Z40';
put ' Z41';
put ' Z42';
put ' Z43';
put ' Z44';
put ' Z45';
put ' Z46';
put ' Z47';
put ' Z48';
put ' Z49';
put ' Z50';
put ' Z51';
put ' Z52';
put ' Z53';
put ' Z54';
put ' Z55';
put ' Z56';
put ' Z57';
put ' Z58';
put ' Z59';
put ' Z60';
put ' Z61';
put ' Z62';
put ' Z63';
put ' Z64';
put ' Z65';
put ' Z66';
put ' Z67';
put ' Z68';
```

```
put ' Z69';
put ' Z70';
put ' Z71';
put ' Z72';
put ' Z73';
put ' Z74';
put ' Z75';
put ' Z76';
put ' Z77';
put ' Z78';
put ' Z79';
put ' Z80';
put ' i';
put ' exp  !I';
put ' student  !I 2000';
put ' time !I';
put ' teacher !I';
put ' Test !A';
put ' Zscore';
put 'AnalyzeThis.dat !SKIP=1 !MAXIT=100 !BRIEF';
put 'Zscore ~ mu !r ![ Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10 Z11
Z12 Z13 Z14 Z15 Z16 Z17 Z18 Z19 Z20 ,';
put ' Z21 Z22 Z23 Z24 Z25 Z26 Z27 Z28 Z29 Z30 Z31 Z32 Z33
Z34 Z35 Z36 Z37 Z38 Z39 Z40 Z41 Z42 Z43 Z44 Z45 Z46 ,';
put ' Z47 Z48 Z49 Z50 Z51 Z52 Z53 Z54 Z55 Z56 Z57 Z58 Z59
Z60 Z61 Z62 Z63 Z64 Z65 Z66 Z67 Z68 Z69 Z70 Z71 Z72 ,';
put ' Z73 Z74 Z75 Z76 Z77 Z78 Z79 Z80 !]';
put '1 2 1';
put '2000 0 I';
put '4 0 US !GP !S2==1';
put '1';
put '0 1';
put '0 0 1';
put '0 0 0 1';

put 'Z1 1';
put '80 0 I 1 !GP';
Run;
%End;

*Run .as file in ASReml;
%Let asfile=tch_student_scores.as;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NS6 &asfile";
```

```
Run;

*Select and send predicted teacher effects, ranks, and
percentiles to one permanent file, updated for each
simulation;
Filename tmp2 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.sln";
Data Teacher_Temp;
%If &model = pp %then %do;
Infile tmp2 firstobs = 9 obs = 88  lrecl=71;;
Input Teacher 4-5 Pred_Effect 46-59 SE_Pred_Error 62-71;
%End;
%Else %do; *Check for Z-score method output;
Infile tmp2 firstobs = 2 obs = 81  lrecl=71;;
Input Teacher 4-5 Pred_Effect 46-59 SE_Pred_Error 62-71;
%End;
Exp = &nsim;
Model = "&model";
 Year = 1;
 If Teacher > 20 then Year = 2;
 If Teacher > 40 then Year = 3;
 If Teacher > 60 then Year = 4;
Run;

Proc Sort Data = Teacher_Temp;
 By Year Teacher;
Run;

Proc Rank Data = Teacher_Temp Out=Rank;
 By Year;
 Ranks Pred_Rank;
 Var Pred_Effect;
Run;

Data Teacher_Temp;
Set Rank;
 Pred_Percentile=100*(Pred_Rank/21);
Run;

%If %eval(&nsim) = 1 and &model = pp %then %do;
Data Results.Pred_Teacher_Effects;
Set Teacher_Temp;
Run;
%End;
%Else %do;
```

```
Data Results.Pred_Teacher_Effects;
Set Results.Pred_Teacher_Effects Teacher_Temp;
Run;
%End;


*Select and send estimated variance components to one
permanent file, updated for each simulation;

%If &model = pp %then %do;
*Create ASReml .pin file for PP Variance Components;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.pin";
Data _null_;
File tmp;
put 'F error_var 1 * 1';
put 'F teacher_var 2 * 1';
put 'F D1_Var 3 * 1';
put 'F D2D1_Cov 4 * 1';
put 'F D2_Var 5 * 1';
put 'F D3D1_Cov 6 * 1';
put 'F D3D2_Cov 7 * 1';
put 'F D3_Var 8 * 1';
put 'F D4D1_Cov 9 * 1';
put 'F D4D2_Cov 10 * 1';
put 'F D4D3_Cov 11 * 1';
put 'F D4_Var 12 * 1';
Run;


%Let pinfile=tch_student_scores.pin;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NP &pinfile";
Run;


Filename tmp3 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.pvc";
Data Variance_Temp;
Infile tmp3 firstobs = 16 obs = 27  lrecl=55;
Input Error_Var 31-37 Error_SE 45-55 / Teacher_Var 31-37
Teacher_SE 45-55 / D1_Var 31-37 D1_SE 45-55 / D2D1_Cov 31-
37 D2D1_SE 45-55 / D2_Var 31-37 D2_SE 45-55 / D3D1_Cov 31-
```

```
37 D3D1_SE 45-55 / D3D2_Cov 31-37 D3D2_SE 45-55 / D3_Var
31-37 D3_SE 45-55 / D4D1_Cov 31-37 D4D1_SE 45-55 / D4D2_Cov
31-37 D4D2_SE 45-55 / D4D3_Cov 31-37 D4D3_SE 45-55 / D4_Var
31-37 D4_SE 45-55 ;
%End;


%Else %do;
*Create ASReml .pin file for ZCRT and ZMAT Variance
Components;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.pin";
Data _null_;
File tmp;
put 'F E1_var 1 * 1';
put 'F E2E1_cov 2 * 1';
put 'F E2_Var 3 * 1';
put 'F E3E1_Cov 4 * 1';
put 'F E3E2_Cov 5 * 1';
put 'F E3_Var 6 * 1';
put 'F E4E1_Cov 7 * 1';
put 'F E4E2_Cov 8 * 1';
put 'F E4E3_Cov 9 * 1';
put 'F E4_Var 10 * 1';
put 'F Teacher_Var 11 * 1';
Run;


%Let pinfile=tch_student_scores.pin;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NP &pinfile";
Run;


Filename tmp3 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores.pvc";
Data Variance_Temp;
Infile tmp3 firstobs = 15 obs = 25  lrecl=55;;
Input E1_Var 31-37 E1_SE 45-55 / E2E1_Cov 31-37 E2E1_SE 45-
55 / E2_Var 31-37 E2_SE 45-55 / E3E1_Cov 31-37 E3E1_SE 45-
55 / E3E2_Cov 31-37 E3E2_SE 45-55 / E3_Var 31-37 E3_SE 45-
55 / E4E1_Cov 31-37 E4E1_SE 45-55 / E4E2_Cov 31-37 E4E2_SE
```

```
45-55 / E4E3_Cov 31-37 E4E3_SE 45-55 / E4_Var 31-37 E4_SE
45-55 / Teacher_Var 31-37 Teacher_SE 45-55  ;
%End;

Exp = &nsim;
Model = "&model";
Run;

%If %eval(&nsim) = 1 %then %do;
Data Results.Pred_Variance_&model;
Set Variance_Temp;
Run;
%End;
%Else %do;
Data Results.Pred_Variance_&model;
Set Results.Pred_Variance_&model Variance_Temp;
Run;
%End;

%mend;


***********************************************************;
* Create Macro to Simulate & Analyze Multiple Data Sets *;
***********************************************************;

%Macro sim(sims,oseed);
%Do nsim = 1 %to &sims;
%Let seed = &oseed + &nsim;

*Run COF Macro to create PP, Zcrt and Zmat Datasets;
%cof_full(2000,&seed,Tch_Student_Scores_pp,Tch_Student_Scor
es_zcrt,Tch_Student_Scores_zmat);

*Run Analyze Macro to analyze each dataset;
%analyze(Tch_Student_Scores_pp,pp);
%analyze(Tch_Student_Scores_zcrt,zc);
%analyze(Tch_Student_Scores_zmat,zm);

%End;
%mend;
%sim(1000,20060902);

*********************************************************;
*** Create Datasets for Plots and Summary Statistics ***;
*********************************************************;
```

```
*Merge True and Predicted Teacher Effects from the 1000
Simulations;
Proc Sort Data = Results.True_Teacher_Effects;
By Exp Teacher;
Proc Sort Data = Results.Pred_Teacher_Effects;
By Exp Teacher;
Run;


Data Results.Teacher_Effects;
Merge Results.Pred_Teacher_Effects
Results.True_Teacher_Effects;
By Exp Teacher;
Run;


*Create Dataset (Results.Stats_Plot) for RMSE and SEPE
Plots;
Data Results.Teacher_Effects;
Set Results.Teacher_Effects;
SPE = (Pred_Percentile - True_Percentile)**2;
PE = (Pred_Percentile - True_Percentile);
Run;


Proc Sort Data = Results.Teacher_Effects;
By Year True_Percentile Model;
Run;


Proc Means Data = Results.Teacher_Effects Mean noprint;
By Year True_Percentile Model;
Var SPE;
Output Out = Stats1 Mean = MSE;
Run;


Proc Means Data = Results.Teacher_Effects VAR noprint;
By Year True_Percentile Model;
Var PE;
Output Out = Stats2 Var = VPE;
Run;


Data Results.Stats_Plot;
Merge Stats1 Stats2;
By Year True_Percentile Model;
Run;


Data Results.Stats_Plot;
Set Results.Stats_Plot;
Drop _Freq_ _Type_;
RMSE = sqrt(MSE);
```

```
SEPE = sqrt(VPE);
Run;


Proc Sort Data = Results.Stats_Plot;
By Year Model;
Run;


PROC EXPORT DATA= RESULTS.Stats_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Graphs\Stats_Plot.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;


*Create Dataset for Bias Plot;
Proc Sort Data = Results.True_Teacher_Effects;
By Exp Teacher;
Proc Sort Data = Results.Pred_Teacher_Effects;
By Exp Teacher;
Run;


Data Results.Teacher_Effects;
Merge Results.Pred_Teacher_Effects
Results.True_Teacher_Effects;
By Exp Teacher;
Run;


Proc Sort Data=Results.Teacher_Effects;
By Year True_Percentile Model;
Run;


Proc Means Data = Results.Teacher_Effects Mean noprint;
By Year True_Percentile Model;
Var Pred_Percentile;
Output Out = Stats3 Mean = Mean_Pred;
Run;


Data Results.Bias_Plot;
Set Stats3;
Bias = Mean_Pred - True_Percentile;
Drop _Freq_ _Type_;
Run;


PROC EXPORT DATA= RESULTS.Bias_Plot
```

```
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Graphs\Bias_Plot.csv"
            DBMS=CSV REPLACE;
      PUTNAMES=YES;
RUN;


*Create Dataset (Results.Percentile_Plot) for 25th, 50th
and 75th Percentile Sampling Distribution Plots;
Data Results.Percentile;
Set Results.Teacher_Effects;
If Ceil(True_Rank) = 5 or Ceil(True_Rank) = 10 or
Ceil(True_Rank) = 16;
Ceil_True_Rank = Ceil(True_Rank); *CEIL function adjusts
for mean ranks from ties;
Ceil_Pred_Rank = Ceil(Pred_Rank); *CEIL function adjusts
for mean ranks from ties;
Ceil_True_Percentile = (Ceil_True_Rank*100)/21;
Ceil_Pred_Percentile = (Ceil_Pred_Rank*100)/21;
Dummy = 1;
Run;


Proc Sort Data = Results.Percentile;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Run;


Proc Means Data = Results.Percentile Sum noprint;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Var Dummy;
Output out = Stats Sum = sum;
Run;


Data Zero;
Do Ceil_True_Rank = 1 to 20;
Ceil_True_Percentile = 100*(Ceil_True_Rank/21);
 Do Ceil_Pred_Rank = 1 to 20;
 Ceil_Pred_Percentile = 100*(Ceil_Pred_Rank/21);
  Do Year = 1 to 4;
     Do Model2 = 1 to 3;
     Output;
     End;
  End;
 End;
End;
Run;


Data Zero;
```

```
Set Zero;
If Ceil_True_Rank = 5 or Ceil_True_Rank = 10 or
Ceil_True_Rank = 16;
If Model2 = 1 then Model = 'pp';
If Model2 = 2 then Model = 'zc';
If Model2 = 3 then Model = 'zm';
Run;


Data Test;
Merge Stats Zero;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Run;


Data Results.Percentile_Plot;
Set Test;
If Sum = . then Sum = 0;
Relative_Freq = Sum/1000;
Drop _Freq_ _Type_ Model2;
Run;


PROC EXPORT DATA= RESULTS.Percentile_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Graphs\Percentile_Plot.csv"
            DBMS=CSV REPLACE;
    PUTNAMES=YES;
RUN;


*Create Dataset (Results.Prob75_Plot) for 75th Percentile
Classification Probability Plots;
Data Percentile75;
Set Results.Teacher_Effects;
If Pred_Percentile >=75 then Dummy75 = 1;
Else Dummy75 = 0;
Run;


Proc Sort Data = Percentile75;
By True_Percentile Year Model;
Run;


Proc Sort Data =  Results.Teacher_Effects;
By True_Percentile Year Model;
Run;


Proc Means Data = Percentile75 Sum noprint;
By True_Percentile Year Model;
```

```
Var Dummy75;
Output out = Stats75 Sum = sum;
Run;


Data Results.Prob75_Plot;
Set Stats75;
Relative_Freq = Sum/1000;
Drop _Freq_ _Type_;
Run;


PROC EXPORT DATA= RESULTS.PROB75_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Graphs\Prob75_Plot.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;


*Obtain Mean and SD of Teacher Var and Avg Teacher Var SE
for each Model;
Proc Means Data = Results.Pred_Variance_pp Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
Proc Means Data = Results.Pred_Variance_zc Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
Proc Means Data = Results.Pred_Variance_zm Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
```

Appendix C

Missing Tests SAS Simulation Code and ASReml Analysis Code

```
*******************************************;
*** Create a Macro to Simulate Data Sets ***;
*******************************************;

Libname MResults 'C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\MResults';

*Delete Old True Teacher Effects File before re-creating
data sets;
Data MResults.True_Teacher_Effects;
 Delete;
Run;

%macro
cof_missing(nstudent,seed,dataname2,dataname3,dataname4,dat
aname5);
%let seed1 = &seed + 3220;
%let seed2 = &seed + 9086;
%let seed3 = &seed + 3;

*****************************************;
*** Make Random Teacher Assignments ***;
*****************************************;

/* you do this by ranking the random numbers from step 1
   then for year 1 you call ranks 1 to (n/20) teacher 1,
        ranks (n/20)+1 to (2n/20) teacher 2, etc
        for year 2 you call ranks 1 to (n/20) teacher 21,
        ranks (n/20)+1 to (2n/20) teacher 22, etc
        and similarly for years 3 and 4 */

/* step 1: generate random numbers for each student each
year */
data step_1;
 seed=&seed1;
 do student=1 to &nstudent;
  rntyr1=ranuni(seed);
  rntyr2=ranuni(seed);
```

```
  rntyr3=ranuni(seed);
  rntyr4=ranuni(seed);
  output;
 end;

/* step 2: create the ranks using proc rank */
proc rank data=step_1 out=step_2;
 var rntyr1-rntyr4;
 ranks tyr1 tyr2 tyr3 tyr4;
run;

/* step 3: use the ranks to make teacher assignments */
data step_3;
 set step_2;
 real_obs=1;
  teacher_yr1=1;
  if tyr1>((1/20)*&nstudent) then teacher_yr1=2;
  if tyr1>((2/20)*&nstudent) then teacher_yr1=3;
  if tyr1>((3/20)*&nstudent) then teacher_yr1=4;
  if tyr1>((4/20)*&nstudent) then teacher_yr1=5;
  if tyr1>((5/20)*&nstudent) then teacher_yr1=6;
  if tyr1>((6/20)*&nstudent) then teacher_yr1=7;
  if tyr1>((7/20)*&nstudent) then teacher_yr1=8;
  if tyr1>((8/20)*&nstudent) then teacher_yr1=9;
  if tyr1>((9/20)*&nstudent) then teacher_yr1=10;
  if tyr1>((10/20)*&nstudent) then teacher_yr1=11;
  if tyr1>((11/20)*&nstudent) then teacher_yr1=12;
  if tyr1>((12/20)*&nstudent) then teacher_yr1=13;
  if tyr1>((13/20)*&nstudent) then teacher_yr1=14;
  if tyr1>((14/20)*&nstudent) then teacher_yr1=15;
  if tyr1>((15/20)*&nstudent) then teacher_yr1=16;
  if tyr1>((16/20)*&nstudent) then teacher_yr1=17;
  if tyr1>((17/20)*&nstudent) then teacher_yr1=18;
  if tyr1>((18/20)*&nstudent) then teacher_yr1=19;
  if tyr1>((19/20)*&nstudent) then teacher_yr1=20;
  teacher_yr2=21;
  if tyr2>((1/20)*&nstudent) then teacher_yr2=22;
  if tyr2>((2/20)*&nstudent) then teacher_yr2=23;
  if tyr2>((3/20)*&nstudent) then teacher_yr2=24;
  if tyr2>((4/20)*&nstudent) then teacher_yr2=25;
  if tyr2>((5/20)*&nstudent) then teacher_yr2=26;
  if tyr2>((6/20)*&nstudent) then teacher_yr2=27;
  if tyr2>((7/20)*&nstudent) then teacher_yr2=28;
  if tyr2>((8/20)*&nstudent) then teacher_yr2=29;
  if tyr2>((9/20)*&nstudent) then teacher_yr2=30;
  if tyr2>((10/20)*&nstudent) then teacher_yr2=31;
  if tyr2>((11/20)*&nstudent) then teacher_yr2=32;
```

```
if tyr2>((12/20)*&nstudent) then teacher_yr2=33;
if tyr2>((13/20)*&nstudent) then teacher_yr2=34;
if tyr2>((14/20)*&nstudent) then teacher_yr2=35;
if tyr2>((15/20)*&nstudent) then teacher_yr2=36;
if tyr2>((16/20)*&nstudent) then teacher_yr2=37;
if tyr2>((17/20)*&nstudent) then teacher_yr2=38;
if tyr2>((18/20)*&nstudent) then teacher_yr2=39;
if tyr2>((19/20)*&nstudent) then teacher_yr2=40;
teacher_yr3=41;
if tyr3>((1/20)*&nstudent) then teacher_yr3=42;
if tyr3>((2/20)*&nstudent) then teacher_yr3=43;
if tyr3>((3/20)*&nstudent) then teacher_yr3=44;
if tyr3>((4/20)*&nstudent) then teacher_yr3=45;
if tyr3>((5/20)*&nstudent) then teacher_yr3=46;
if tyr3>((6/20)*&nstudent) then teacher_yr3=47;
if tyr3>((7/20)*&nstudent) then teacher_yr3=48;
if tyr3>((8/20)*&nstudent) then teacher_yr3=49;
if tyr3>((9/20)*&nstudent) then teacher_yr3=50;
if tyr3>((10/20)*&nstudent) then teacher_yr3=51;
if tyr3>((11/20)*&nstudent) then teacher_yr3=52;
if tyr3>((12/20)*&nstudent) then teacher_yr3=53;
if tyr3>((13/20)*&nstudent) then teacher_yr3=54;
if tyr3>((14/20)*&nstudent) then teacher_yr3=55;
if tyr3>((15/20)*&nstudent) then teacher_yr3=56;
if tyr3>((16/20)*&nstudent) then teacher_yr3=57;
if tyr3>((17/20)*&nstudent) then teacher_yr3=58;
if tyr3>((18/20)*&nstudent) then teacher_yr3=59;
if tyr3>((19/20)*&nstudent) then teacher_yr3=60;
teacher_yr4=61;
if tyr4>((1/20)*&nstudent) then teacher_yr4=62;
if tyr4>((2/20)*&nstudent) then teacher_yr4=63;
if tyr4>((3/20)*&nstudent) then teacher_yr4=64;
if tyr4>((4/20)*&nstudent) then teacher_yr4=65;
if tyr4>((5/20)*&nstudent) then teacher_yr4=66;
if tyr4>((6/20)*&nstudent) then teacher_yr4=67;
if tyr4>((7/20)*&nstudent) then teacher_yr4=68;
if tyr4>((8/20)*&nstudent) then teacher_yr4=69;
if tyr4>((9/20)*&nstudent) then teacher_yr4=70;
if tyr4>((10/20)*&nstudent) then teacher_yr4=71;
if tyr4>((11/20)*&nstudent) then teacher_yr4=72;
if tyr4>((12/20)*&nstudent) then teacher_yr4=73;
if tyr4>((13/20)*&nstudent) then teacher_yr4=74;
if tyr4>((14/20)*&nstudent) then teacher_yr4=75;
if tyr4>((15/20)*&nstudent) then teacher_yr4=76;
if tyr4>((16/20)*&nstudent) then teacher_yr4=77;
if tyr4>((17/20)*&nstudent) then teacher_yr4=78;
if tyr4>((18/20)*&nstudent) then teacher_yr4=79;
```

```
   if tyr4>((19/20)*&nstudent) then teacher_yr4=80;
Run;


/* Transpose Random Teacher Assignments from Step 3 */
Data Tch_assignments;
Set Step_3;
Array S[4] teacher_yr1 teacher_yr2 teacher_yr3 teacher_yr4;
Do Time = 1 to 4;
     Teacher = S[Time];
Output;
End;
Keep Student Teacher Time;
Run;


******************************;
*** Generate Student Scores ***;
******************************;


/* step 1: generate student scores and output data set */
Proc IML;
Call Randseed(&seed2);
 N=1;
 G={157.5 116.7 99.8 83.3, 116.7 135 98.6 82.7, 99.8 98.6
112.5 80.5, 83.3 82.7 80.5 90}; *AR(1) structure with 1,
0.8, 0.75, 0.7 correlations;
 R=I(4)@{45 0, 0 45}; *20% of total variance attributed to
measurement error;
 lps_sim=j(&nstudent,9,.);
 Do Student=1 to &nstudent;
 D = (Randnormal(1, J(4,1,0), G))`;
 D2 = D@J(2,1,1);
 D = D2`;
 E = Randnormal(1, J(8,1,0), R);
 Mean={62 231 84 234 54 236 70 239}; *MAT/3 to adjust for
variance, (CRT_yr1)*2 to adjust for variance;
 Scores = Mean+D+E;
  lps_sim[student,1]=student;
  lps_sim[student,2]=scores[1,1];
  lps_sim[student,3]=scores[1,2];
  lps_sim[student,4]=scores[1,3];
  lps_sim[student,5]=scores[1,4];
  lps_sim[student,6]=scores[1,5];
  lps_sim[student,7]=scores[1,6];
  lps_sim[student,8]=scores[1,7];
  lps_sim[student,9]=scores[1,8];
End;
Create student_scores from lps_sim;
```

```
Append from lps_sim;
Run;

Data student_scores;
 Set student_scores;
 Student=col1;
 yr1a_score=col2;
 yr1b_score=col3;
 yr2a_score=col4;
 yr2b_score=col5;
 yr3a_score=col6;
 yr3b_score=col7;
 yr4a_score=col8;
 yr4b_score=col9;
 Drop col1-col9;
Run;

/* step 2: Transpose student scores */
Data Sim_Scores;
Set Student_Scores;
Array S[8] yr1a_score yr1b_score yr2a_score yr2b_score
yr3a_score yr3b_score yr4a_score yr4b_score;
Do I = 1 to 8;
     Score = S[I];
Output;
End;
Keep Student I Score;
Run;

/* step 3: create time and test variables */
Proc Sort Data = Sim_Scores;
By I;
Run;
Data Sim_Scores;
Set Sim_Scores;
Time = Ceil(I/2);
If Mod(I,2) = 1 then Test = 'CRT'; Else Test = 'MAT';
Keep Student Time Test Score;
Run;

*************************************************;
*** Merge Student Scores and Teacher Assignments ***;
*************************************************;

Proc Sort Data = Sim_Scores;
By Student Time;
Proc Sort Data = Tch_Assignments;
```

```
By Student Time;
Run;
Data Sim_Scores_n_Tch;
Merge Sim_Scores Tch_Assignments;
By Student Time;
Run;

*****************************;
*** Create Layered Z-matrix ***;
*****************************;

/* step 1: separate MAT & CRT scores to create z-matrix for
each type of test*/
Data Sim_Scores_MAT;
Set Sim_Scores_n_Tch;
If Test = 'MAT';
Run;
Data Sim_Scores_CRT;
Set Sim_Scores_n_Tch;
If Test = 'CRT';
Run;

/* step 2: macro creates the layered z-matrix for the
general case when the teachers are not 1, 2, 3, ...., but
instead are names.*/
%let Student=Student;
%let Ranvar=Teacher;
%let zmax=1;

%macro
Randvar_Full(dataname1,w11,w21,w22,w31,w32,w33,w41,w42,w43,
w44);
Proc Sort data=&data_o; By &ranvar;
Data Z(keep=&ranvar);
Set &data_o;
By &ranvar;
If first.&ranvar;
Run;
Data zz;
Set z;
Rank=_N_;
Run;
Proc sql;
Select max(rank) into: zmax from zz;

Data new_1;
Merge &data_o zz;
```

```
By &ranvar;

Proc Sort Data=new_1; By  &student Time;

Data New_2 (Drop=Zmat%eval(&zmax+1));
Set new_1;
By &student Time;

/*Create lag variable*/
Array reset(3) lagrank1-lagrank3;
lagrank1=lag(rank);
lagrank2=lag2(rank);
lagrank3=lag3(rank);
If first.&student then count=1;
Do i=count to 3;
    Reset(i)=%eval(&zmax+1);
End;
Count+1;

%Do ii=1 %to %Eval(&zmax+1);
Zmat&ii=0;
%End;

Array Zmat(%eval(&zmax+1)) Zmat1-Zmat%eval(&zmax+1);
If time = 1  then Zmat[rank]+&w11;
Else if Time = 2 then do; Zmat[rank]+&w21;
Zmat[lagrank1]+&w22; End;
Else if Time = 3 then do; Zmat[rank]+&w31;
Zmat[lagrank1]+&w32; Zmat[lagrank2]+&w33; End;
Else if Time = 4 then do; Zmat[rank]+&w41;
Zmat[lagrank1]+&w42; Zmat[lagrank2]+&w43;
Zmat[lagrank3]+&w44; End;

Run;

Data &dataname1(drop=count i);
Set New_2;
Run;

%mend;

/* step 3: run macro for both tests */
%Let Data_o=Sim_Scores_CRT;
%Randvar_full(Sim_Scores_CRT1,1,1,1,1,1,1,1,1,1,1);
%Let Data_o=Sim_Scores_MAT;
%Randvar_full(Sim_Scores_MAT1,1,1,1,1,1,1,1,1,1,1);
```

```
/* step 4: merge data for both tests */
DATA Sim_Scores;
 SET Sim_Scores_CRT1 Sim_Scores_MAT1;
Run;
Proc Sort Data = Sim_Scores;
 BY Student Time Test;
Run;

****************************;
*** Create Final Data Set ***;
****************************;

/* step 1: Generate random teacher effects and find ranks
and percentiles by year */

Libname MResults 'C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\MResults';

Data Gen_tch_value;
 Seed=&seed3;
 Do Teacher=1 to 80;
  t_eff=rannor(seed)*(sqrt(22.5));
  Output;
 End;
Run;

Data Gen_tch_value;
 Set Gen_tch_value;
 Year = 1;
 If Teacher > 20 then Year = 2;
 If Teacher > 40 then Year = 3;
 If Teacher > 60 then Year = 4;
Run;

Proc Sort Data = Gen_tch_value;
 By Year Teacher;
Run;

Proc Rank Data = Gen_tch_value Out=Rank;
 By Year;
 Ranks True_Rank;
 Var T_Eff;
Run;

Data Gen_tch_value_Temp;
Set Rank;
```

```
 Exp = &nsim;
 True_Effect = t_eff;
 True_Percentile=100*(True_Rank/21);
 Drop t_eff;
Run;


*Send true teacher effects to one permanent file, updated
for each simulation;
%If %eval(&nsim) = 1 %then %do;
Data MResults.True_Teacher_Effects;
Set Gen_tch_value_Temp;
Run;
%End;
%Else %do;
Data MResults.True_Teacher_Effects;
Set MResults.True_Teacher_Effects Gen_tch_value_Temp;
Run;
%End;


/* step 2: merge student data, teacher assignments, and
teacher values into single data set */
Proc IML;
Use Sim_Scores;
Read all var{student score time test teacher rank lagrank1
lagrank2 lagrank3 Zmat1 Zmat2 Zmat3 Zmat4 Zmat5 Zmat6 Zmat7
Zmat8 Zmat9 Zmat10 Zmat11 Zmat12 Zmat13 Zmat14 Zmat15
Zmat16 Zmat17 Zmat18 Zmat19 Zmat20 Zmat21 Zmat22 Zmat23
Zmat24 Zmat25 Zmat26 Zmat27 Zmat28 Zmat29 Zmat30 Zmat31
Zmat32 Zmat33 Zmat34 Zmat35 Zmat36 Zmat37 Zmat38 Zmat39
Zmat40 Zmat41 Zmat42 Zmat43 Zmat44 Zmat45 Zmat46 Zmat47
Zmat48 Zmat49 Zmat50 Zmat51 Zmat52 Zmat53 Zmat54 Zmat55
Zmat56 Zmat57 Zmat58 Zmat59 Zmat60 Zmat61 Zmat62 Zmat63
Zmat64 Zmat65 Zmat66 Zmat67 Zmat68 Zmat69 Zmat70 Zmat71
Zmat72 Zmat73 Zmat74 Zmat75 Zmat76 Zmat77 Zmat78 Zmat79
Zmat80};
Use gen_tch_value;
Read all var{t_eff};
Z =
Zmat1||Zmat2||Zmat3||Zmat4||Zmat5||Zmat6||Zmat7||Zmat8||
Zmat9||Zmat10||Zmat11||Zmat12||Zmat13||Zmat14||Zmat15||
Zmat16||Zmat17||Zmat18||Zmat19||Zmat20||Zmat21||Zmat22||
Zmat23||Zmat24||Zmat25||Zmat26||Zmat27||Zmat28||Zmat29||
Zmat30||Zmat31||Zmat32||Zmat33||Zmat34||Zmat35||Zmat36||
Zmat37||Zmat38||Zmat39||Zmat40||Zmat41||Zmat42||Zmat43||
Zmat44||Zmat45||Zmat46||Zmat47||Zmat48||Zmat49||Zmat50||
Zmat51||Zmat52||Zmat53||Zmat54||Zmat55||Zmat56||Zmat57||
Zmat58||Zmat59||Zmat60||Zmat61||Zmat62||Zmat63||Zmat64||
```

```
Zmat65||Zmat66||Zmat67||Zmat68||Zmat69||Zmat70||Zmat71||
Zmat72||Zmat73||Zmat74||Zmat75||Zmat76||Zmat77||Zmat78||
Zmat79||Zmat80;
Score_Total = Score + Z*t_eff;
Test2 = J((&nstudent)*8,1,0);
Do i = 1 to (&nstudent)*8;
If test[i,1] = 'MAT' then test2[i,1] = 2; Else test2[i,1] =
1;
End;
Scores =
Z||student||score||score_total||time||teacher||test2;
Create tch_student_scores from Scores;
Append from Scores;
Print &nsim;
Run;
Quit;

/* step 3: rename and create variables for PP_Missing
(Dataname 2) Dataset*/
Data pp_missing;
 Set tch_student_scores;
 Array Z[80] Z1-Z80;
 Array Col[80] Col1-Col80;
Do i = 1 to 80;
 Z[i] = Col[i];
End;
 Exp = &nsim;
 Student=col81;
 No_Tch_Score=col82;
 Tch_Score_Final=col83;
 If Tch_Score_Final<0 then Tch_Score_Final=0;
 Time=col84;
 Teacher=col85;
 If Col86 = 1 then Test = 'CRT'; Else Test = 'MAT';
 Drop Col1-Col86;
Run;

Data &dataname2;
 Set pp_missing;
 If Time = 1 and Test = 'CRT' then delete;
 If Time = 4 and Test = 'MAT' then delete;
Run;

/* Step 4: Create Zcrt_Missing (Dataname3), Zmat_Missing
(Dataname4) and Zmatcrt_missing (Dataname 5) Datasets */
Data Standard;
 Set &dataname2;
```

```
 sum_z=sum(of Z1-Z%eval(&zmax));
Run;

Data New3;
 Set Standard;
 %Do ii=1 %to &zmax;
     if sum_z ne 0 then Z&ii=(Z&ii/sum_z)**0.5; else
Z&ii=0;
 %End;
Run;

Proc Sort Data = New3;
By Time Test;
Run;

Proc Means Data = New3 Mean Std noprint;
By Time Test;
Var Tch_Score_Final;
Output Out = Stats Mean = Mean Std = Std;
Run;

Data Stats;
Set Stats;
Drop _Type_ _Freq_;
Run;

Data New;
Merge New3 Stats;
By Time Test;
Run;

Data New2;
Set New;
Zscore = (Tch_Score_Final - Mean)/Std;
Drop No_Tch_Score Tch_Score_Final Mean Std Sum_Z;
Run;

Data &Dataname3;
Set New2;
If Time = 2 and Test = 'MAT' then delete;
If Time = 3 and Test = 'MAT' then delete;
Run;

Data &Dataname4;
Set New2;
If Time = 2 and Test = 'CRT' then delete;
If Time = 3 and Test = 'CRT' then delete;
```

```
Run;

Data &Dataname5;
Set New2;
If Time = 2 and Test = 'CRT' then delete;
If Time = 3 and Test = 'MAT' then delete;
Run;

Proc Sort Data = &Dataname2;
By Student Time Test;
Proc Sort Data = &Dataname3;
By Student Time;
Proc Sort Data = &Dataname4;
By Student Time;
Proc Sort Data = &Dataname5;
By Student Time;
Run;

Quit; *This makes sure a Proc is not still running;
%mend;


*************************************************;
*** Create Macro to Analyze Simulated Data Sets ***;
*************************************************;

*Delete Old Predicted Teacher Effects File before re-
analyzing data sets;
Data MResults.Pred_Teacher_Effects;
 Delete;
Run;

*Delete Old Variance Components File before re-analyzing
data sets;
Data MResults.Pred_Variance_pp;
 Delete;
Run;
Data MResults.Pred_Variance_zc;
 Delete;
Run;
Data MResults.Pred_Variance_zm;
 Delete;
Run;
Data MResults.Pred_Variance_mc;
 Delete;
Run;

%macro Analyze(data,model);
```

```
*Export simulated SAS data set as a text file;
PROC EXPORT DATA= &data
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\AnalyzeThis.dat"
            DBMS=DLM REPLACE;
     DELIMITER='20'x;
     PUTNAMES=YES;
RUN;

%If &model = pp %then %do;
*Create ASReml .as file for Curve of Factors;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.as";
Data _null_;
File tmp;
put 'Curve of Factors for 80 Teacher, 2000 Students';
put ' Z1';
put ' Z2';
put ' Z3';
put ' Z4';
put ' Z5';
put ' Z6';
put ' Z7';
put ' Z8';
put ' Z9';
put ' Z10';
put ' Z11';
put ' Z12';
put ' Z13';
put ' Z14';
put ' Z15';
put ' Z16';
put ' Z17';
put ' Z18';
put ' Z19';
put ' Z20';
put ' Z21';
put ' Z22';
put ' Z23';
put ' Z24';
put ' Z25';
put ' Z26';
put ' Z27';
```

```
put ' Z28';
put ' Z29';
put ' Z30';
put ' Z31';
put ' Z32';
put ' Z33';
put ' Z34';
put ' Z35';
put ' Z36';
put ' Z37';
put ' Z38';
put ' Z39';
put ' Z40';
put ' Z41';
put ' Z42';
put ' Z43';
put ' Z44';
put ' Z45';
put ' Z46';
put ' Z47';
put ' Z48';
put ' Z49';
put ' Z50';
put ' Z51';
put ' Z52';
put ' Z53';
put ' Z54';
put ' Z55';
put ' Z56';
put ' Z57';
put ' Z58';
put ' Z59';
put ' Z60';
put ' Z61';
put ' Z62';
put ' Z63';
put ' Z64';
put ' Z65';
put ' Z66';
put ' Z67';
put ' Z68';
put ' Z69';
put ' Z70';
put ' Z71';
put ' Z72';
put ' Z73';
put ' Z74';
```

```
put ' Z75';
put ' Z76';
put ' Z77';
put ' Z78';
put ' Z79';
put ' Z80';
put ' i';
put ' exp  !I';
put ' student  !I 2000';
put ' notchscore';
put ' tchscorefinal';
put ' time !I';
put ' teacher !I';
put ' Test !A';
put 'AnalyzeThis.dat !SKIP=1 !MAXIT=100 !BRIEF';
put 'tchscorefinal ~ time.Test !r ![ Z1 Z2 Z3 Z4 Z5 Z6 Z7
Z8 Z9 Z10 Z11 Z12 Z13 Z14 Z15 Z16 Z17 Z18 Z19 Z20 ,';
put ' Z21 Z22 Z23 Z24 Z25 Z26 Z27 Z28 Z29 Z30 Z31 Z32 Z33
Z34 Z35 Z36 Z37 Z38 Z39 Z40 Z41 Z42 Z43 Z44 Z45 Z46 ,';
put ' Z47 Z48 Z49 Z50 Z51 Z52 Z53 Z54 Z55 Z56 Z57 Z58 Z59
Z60 Z61 Z62 Z63 Z64 Z65 Z66 Z67 Z68 Z69 Z70 Z71 Z72 ,';
put ' Z73 Z74 Z75 Z76 Z77 Z78 Z79 Z80 !] time.student';
put '0 0 2';

put 'Z1 1';
put '80 0 I 1 !GP';

put 'time.student 2';
put 'time 0  US !GP';
put '1';
put '0 1';
put '0 0 1';
put '0 0 0 1';
put 'student 0 I';
Run;
%End;

%Else %do;
*Create ASReml .as file for Z-score Method;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.as";
Data _null_;
File tmp;
put 'Z-score Method for 80 Teachers, 2000 Students';
put ' Z1';
```

```
put ' Z2';
put ' Z3';
put ' Z4';
put ' Z5';
put ' Z6';
put ' Z7';
put ' Z8';
put ' Z9';
put ' Z10';
put ' Z11';
put ' Z12';
put ' Z13';
put ' Z14';
put ' Z15';
put ' Z16';
put ' Z17';
put ' Z18';
put ' Z19';
put ' Z20';
put ' Z21';
put ' Z22';
put ' Z23';
put ' Z24';
put ' Z25';
put ' Z26';
put ' Z27';
put ' Z28';
put ' Z29';
put ' Z30';
put ' Z31';
put ' Z32';
put ' Z33';
put ' Z34';
put ' Z35';
put ' Z36';
put ' Z37';
put ' Z38';
put ' Z39';
put ' Z40';
put ' Z41';
put ' Z42';
put ' Z43';
put ' Z44';
put ' Z45';
put ' Z46';
put ' Z47';
put ' Z48';
```

```
put ' Z49';
put ' Z50';
put ' Z51';
put ' Z52';
put ' Z53';
put ' Z54';
put ' Z55';
put ' Z56';
put ' Z57';
put ' Z58';
put ' Z59';
put ' Z60';
put ' Z61';
put ' Z62';
put ' Z63';
put ' Z64';
put ' Z65';
put ' Z66';
put ' Z67';
put ' Z68';
put ' Z69';
put ' Z70';
put ' Z71';
put ' Z72';
put ' Z73';
put ' Z74';
put ' Z75';
put ' Z76';
put ' Z77';
put ' Z78';
put ' Z79';
put ' Z80';
put ' i';
put ' exp  !I';
put ' student  !I 2000';
put ' time !I';
put ' teacher !I';
put ' Test !A';
put ' Zscore';
put 'AnalyzeThis.dat !SKIP=1 !MAXIT=100 !BRIEF';
put 'Zscore ~ mu !r ![ Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10 Z11
Z12 Z13 Z14 Z15 Z16 Z17 Z18 Z19 Z20 ,';
put ' Z21 Z22 Z23 Z24 Z25 Z26 Z27 Z28 Z29 Z30 Z31 Z32 Z33
Z34 Z35 Z36 Z37 Z38 Z39 Z40 Z41 Z42 Z43 Z44 Z45 Z46 ,';
put ' Z47 Z48 Z49 Z50 Z51 Z52 Z53 Z54 Z55 Z56 Z57 Z58 Z59
Z60 Z61 Z62 Z63 Z64 Z65 Z66 Z67 Z68 Z69 Z70 Z71 Z72 ,';
put ' Z73 Z74 Z75 Z76 Z77 Z78 Z79 Z80 !]';
```

```
put '1 2 1';
put '2000 0 I';
put '4 0 US !GP !S2==1';
put '1';
put '0 1';
put '0 0 1';
put '0 0 0 1';

put 'Z1 1';
put '80 0 I 1 !GP';
Run;
%End;


*Run .as file in ASReml;
%Let asfile=tch_student_scores_missing.as;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NS6 &asfile";
Run;


*Select and send predicted teacher effects, ranks, and
percentiles to one permanent file, updated for each
simulation;
Filename tmp2 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.sln";
Data Teacher_Temp;
%If &model = pp %then %do;
Infile tmp2 firstobs = 9 obs = 88  lrecl=71;;
Input Teacher 4-5 Pred_Effect 46-59 SE_Pred_Error 62-71;
%End;
%Else %do;
Infile tmp2 firstobs = 2 obs = 81  lrecl=71;;
Input Teacher 4-5 Pred_Effect 46-59 SE_Pred_Error 62-71;
%End;
Exp = &nsim;
Model = "&model";
 Year = 1;
 If Teacher > 20 then Year = 2;
 If Teacher > 40 then Year = 3;
 If Teacher > 60 then Year = 4;
Run;


Proc Sort Data = Teacher_Temp;
```

```
 By Year Teacher;
Run;

Proc Rank Data = Teacher_Temp Out=Rank;
 By Year;
 Ranks Pred_Rank;
 Var Pred_Effect;
Run;

Data Teacher_Temp;
Set Rank;
 Pred_Percentile=100*(Pred_Rank/21);
Run;

%If %eval(&nsim) = 1 and &model = pp %then %do;
Data MResults.Pred_Teacher_Effects;
Set Teacher_Temp;
Run;
%End;
%Else %do;
Data MResults.Pred_Teacher_Effects;
Set MResults.Pred_Teacher_Effects Teacher_Temp;
Run;
%End;

*Select and send estimated variance components to one
permanent file, updated for each simulation;

%If &model = pp %then %do;
*Create ASReml .pin file for PP Variance Components;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.pin";
Data _null_;
File tmp;
put 'F error_var 1 * 1';
put 'F teacher_var 2 * 1';
put 'F D1_Var 3 * 1';
put 'F D2D1_Cov 4 * 1';
put 'F D2_Var 5 * 1';
put 'F D3D1_Cov 6 * 1';
put 'F D3D2_Cov 7 * 1';
put 'F D3_Var 8 * 1';
put 'F D4D1_Cov 9 * 1';
put 'F D4D2_Cov 10 * 1';
put 'F D4D3_Cov 11 * 1';
```

```
put 'F D4_Var 12 * 1';
Run;


%Let pinfile=tch_student_scores_missing.pin;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NP &pinfile";
Run;


Filename tmp3 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.pvc";
Data Variance_Temp;
Infile tmp3 firstobs = 16 obs = 27  lrecl=55;
Input Error_Var 31-37 Error_SE 45-55 / Teacher_Var 31-37
Teacher_SE 45-55 / D1_Var 31-37 D1_SE 45-55 / D2D1_Cov 31-
37 D2D1_SE 45-55 / D2_Var 31-37 D2_SE 45-55 / D3D1_Cov 31-
37 D3D1_SE 45-55 / D3D2_Cov 31-37 D3D2_SE 45-55 / D3_Var
31-37 D3_SE 45-55 / D4D1_Cov 31-37 D4D1_SE 45-55 / D4D2_Cov
31-37 D4D2_SE 45-55 / D4D3_Cov 31-37 D4D3_SE 45-55 / D4_Var
31-37 D4_SE 45-55 ;
%End;


%Else %do;
*Create ASReml .pin file for ZCrt, ZMat and MatCrt Variance
Components;
Filename tmp "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.pin";
Data _null_;
File tmp;
put 'F E1_var 1 * 1';
put 'F E2E1_cov 2 * 1';
put 'F E2_Var 3 * 1';
put 'F E3E1_Cov 4 * 1';
put 'F E3E2_Cov 5 * 1';
put 'F E3_Var 6 * 1';
put 'F E4E1_Cov 7 * 1';
put 'F E4E2_Cov 8 * 1';
put 'F E4E3_Cov 9 * 1';
put 'F E4_Var 10 * 1';
put 'F Teacher_Var 11 * 1';
Run;
```

```
%Let pinfile=tch_student_scores_missing.pin;
Data _null_;
Call System('cd C:\Users\Jenny\Desktop\Flash Drive - June
2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\Analyze');
X "'C:\Program Files\ASREML3\BIN\asreml.exe' -NP &pinfile";
Run;


Filename tmp3 "C:\Users\Jenny\Desktop\Flash Drive - June 2,
2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\Analyze\tch_student_scores_missing.pvc";
Data Variance_Temp;
Infile tmp3 firstobs = 15 obs = 25  lrecl=55;;
Input E1_Var 31-37 E1_SE 45-55 / E2E1_Cov 31-37 E2E1_SE 45-
55 / E2_Var 31-37 E2_SE 45-55 / E3E1_Cov 31-37 E3E1_SE 45-
55 / E3E2_Cov 31-37 E3E2_SE 45-55 / E3_Var 31-37 E3_SE 45-
55 / E4E1_Cov 31-37 E4E1_SE 45-55 / E4E2_Cov 31-37 E4E2_SE
45-55 / E4E3_Cov 31-37 E4E3_SE 45-55 / E4_Var 31-37 E4_SE
45-55 / Teacher_Var 31-37 Teacher_SE 45-55  ;
%End;


Exp = &nsim;
Model = "&model";
Run;


%If %eval(&nsim) = 1 %then %do;
Data MResults.Pred_Variance_&model;
Set Variance_Temp;
Run;
%End;


%Else %do;
Data MResults.Pred_Variance_&model;
Set MResults.Pred_Variance_&model Variance_Temp;
Run;
%End;

%mend;

***********************************************************;
* Create Macro to Simulate & Analyze Multiple Data Sets *;
***********************************************************;

%Macro sim(sims,oseed);
%Do nsim = 1 %to &sims;
```

```
%Let seed = &oseed + &nsim;

*Run COF Macro to create PP, Zcrt, Zmat and Zmatcrt
Datasets;
%cof_missing(2000,&seed,Tch_Student_Scores_pp,Tch_Student_S
cores_zcrt,Tch_Student_Scores_zmat,Tch_Student_Scores_zmatc
rt);

*Run Analyze Macro to analyze each dataset;
%analyze(Tch_Student_Scores_pp,pp);
%analyze(Tch_Student_Scores_zcrt,zc);
%analyze(Tch_Student_Scores_zmat,zm);
%analyze(Tch_Student_Scores_zmatcrt,mc);

%End;
%mend;
%sim(1000,20060902);


*****************************************************;
*** Create Datasets for Plots and Summary Statistics ***;
*****************************************************;

*Merge True and Predicted Teacher Effects from the 1000
Simulations;
Proc Sort Data = MResults.True_Teacher_Effects;
By Exp Teacher;
Proc Sort Data = MResults.Pred_Teacher_Effects;
By Exp Teacher;
Run;


Data MResults.Teacher_Effects;
Merge MResults.Pred_Teacher_Effects
MResults.True_Teacher_Effects;
By Exp Teacher;
Run;

*Create Dataset (MResults.Stats_Plot) for RMSE and SEPE
Plots;
Data MResults.Teacher_Effects;
Set MResults.Teacher_Effects;
SPE = (Pred_Percentile - True_Percentile)**2;
PE = (Pred_Percentile - True_Percentile);
Run;


Proc Sort Data = MResults.Teacher_Effects;
By Year True_Percentile Model;
Run;
```

```
Proc Means Data = MResults.Teacher_Effects Mean noprint;
By Year True_Percentile Model;
Var SPE;
Output Out = Stats1 Mean = MSE;
Run;

Proc Means Data = MResults.Teacher_Effects VAR noprint;
By Year True_Percentile Model;
Var PE;
Output Out = Stats2 Var = VPE;
Run;

Data MResults.Stats_Plot;
Merge Stats1 Stats2;
By Year True_Percentile Model;
Run;

Data MResults.Stats_Plot;
Set MResults.Stats_Plot;
Drop _Freq_ _Type_;
RMSE = sqrt(MSE);
SEPE = sqrt(VPE);
Run;

Proc Sort Data = MResults.Stats_Plot;
By Year Model;
Run;

PROC EXPORT DATA= MResults.Stats_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\MGraphs\MStats_Plot.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

*Create Dataset for Bias Plot;
Proc Sort Data = MResults.True_Teacher_Effects;
By Exp Teacher;
Proc Sort Data = MResults.Pred_Teacher_Effects;
By Exp Teacher;
Run;

Data MResults.Teacher_Effects;
```

```
Merge MResults.Pred_Teacher_Effects
MResults.True_Teacher_Effects;
By Exp Teacher;
Run;


Proc Sort Data=MResults.Teacher_Effects;
By Year True_Percentile Model;
Run;


Proc Means Data = MResults.Teacher_Effects Mean noprint;
By Year True_Percentile Model;
Var Pred_Percentile;
Output Out = Stats3 Mean = Mean_Pred;
Run;


Data MResults.Bias_Plot;
Set Stats3;
Bias = Mean_Pred - True_Percentile;
Drop _Freq_ _Type_;
Run;


PROC EXPORT DATA= MResults.Bias_Plot
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors Methodology\MGraphs\MBias_Plot.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;


*Create Dataset (MResults.Percentile_Plot) for 25th, 50th
and 75th Percentile Sampling Distribution Plots;
Data MResults.Percentile;
Set MResults.Teacher_Effects;
If Ceil(True_Rank) = 5 or Ceil(True_Rank) = 10 or
Ceil(True_Rank) = 16;
Ceil_True_Rank = Ceil(True_Rank); *CEIL function adjusts
for mean ranks from ties;
Ceil_Pred_Rank = Ceil(Pred_Rank); *CEIL function adjusts
for mean ranks from ties;
Ceil_True_Percentile = (Ceil_True_Rank*100)/21;
Ceil_Pred_Percentile = (Ceil_Pred_Rank*100)/21;
Dummy = 1;
Run;


Proc Sort Data = MResults.Percentile;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Run;
```

```
Proc Means Data = MResults.Percentile Sum noprint;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Var Dummy;
Output out = Stats Sum = sum;
Run;

Data Zero;
Do Ceil_True_Rank = 1 to 20;
Ceil_True_Percentile = 100*(Ceil_True_Rank/21);
 Do Ceil_Pred_Rank = 1 to 20;
 Ceil_Pred_Percentile = 100*(Ceil_Pred_Rank/21);
  Do Year = 1 to 4;
     Do Model2 = 1 to 4;
     Output;
     End;
   End;
 End;
End;
Run;

Data Zero;
Set Zero;
If Ceil_True_Rank = 5 or Ceil_True_Rank = 10 or
Ceil_True_Rank = 16;
If Model2 = 1 then Model = 'mc';
If Model2 = 2 then Model = 'pp';
If Model2 = 3 then Model = 'zc';
If Model2 = 4 then Model = 'zm';
Run;

Data Test;
Merge Stats Zero;
By Ceil_True_Percentile Ceil_Pred_Percentile Year Model;
Run;

Data MResults.Percentile_Plot;
Set Test;
If Sum = . then Sum = 0;
Relative_Freq = Sum/1000;
Drop _Freq_ _Type_ Model2;
Run;

PROC EXPORT DATA= MResults.Percentile_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
```

```
Review\Curve of Factors
Methodology\MGraphs\MPercentile_Plot.csv"
            DBMS=CSV REPLACE;
      PUTNAMES=YES;
RUN;


*Create Dataset (MResults.Prob75_Plot) for 75th Percentile
Classification Probability Plots;
Data Percentile75;
Set MResults.Teacher_Effects;
If Pred_Percentile >=75 then Dummy75 = 1;
Else Dummy75 = 0;
Run;


Proc Sort Data = Percentile75;
By True_Percentile Year Model;
Run;


Proc Sort Data =  MResults.Teacher_Effects;
By True_Percentile Year Model;
Run;


Proc Means Data = Percentile75 Sum noprint;
By True_Percentile Year Model;
Var Dummy75;
Output out = Stats75 Sum = sum;
Run;


Data MResults.Prob75_Plot;
Set Stats75;
Relative_Freq = Sum/1000;
Drop _Freq_ _Type_;
Run;


PROC EXPORT DATA= MResults.PROB75_PLOT
            OUTFILE= "C:\Users\Jenny\Desktop\Flash Drive -
June 2, 2008\PhD\Dissertation\Paper 1 - Intro and Model Lit
Review\Curve of Factors
Methodology\MGraphs\MProb75_Plot.csv"
            DBMS=CSV REPLACE;
      PUTNAMES=YES;
RUN;


*Obtain Mean and SD of Teacher Var and Avg Teacher Var SE
for each Model;
Proc Means Data = MResults.Pred_Variance_pp Mean Std;
Var Teacher_Var Teacher_SE;
```

```
Output Out = VarStats;
Run;
Proc Means Data = MResults.Pred_Variance_zc Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
Proc Means Data = MResults.Pred_Variance_zm Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
Proc Means Data = MResults.Pred_Variance_mc Mean Std;
Var Teacher_Var Teacher_SE;
Output Out = VarStats;
Run;
```