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

This research is supported by the Math in the Middle Institute Partnership. The Math in the Middle Institute Partnership is funded by a Math-Science Partnership grant from the National Science Foundation (NSF MSP Grant #EHR-0142502). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

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, …, 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 g^{th} 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_{eg}^2 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,

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

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,

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

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