COMPARATIVE ANALYSIS OF COMMON STATISTICAL MODELS USED FOR VALUE-ADDED ASSESSMENT OF SCHOOL PERFORMANCE

by

SEDAT SEN

(Under the Direction of Seock-Ho Kim)

ABSTRACT

The purpose of this study is to compare four popular value-added models used in measuring school effectiveness based on their distinguishing characteristics. The simple fixed effects model (SFEM), two hierarchical models (UHLMM and AHLMM), and the layered mixed effects model (LMEM) are the models that have been analyzed based on value-added measures obtained from a common data set with two years standard assessment data. Value-added measures obtained from these models were used to investigate the impact of the differences of each model. Correlational analyses were also conducted to see whether there were remarkably differences among these models. SFEM and UHLMM models produced very similar rank orders of school effects while SFEM and AHLMM have a moderate correlation. Thus there is no much difference between SFEM and two HLM models in terms of the rank orders of schools whereas there was no agreement between LMEM and the other models.

INDEX WORDS: Hierarchical linear models, Layered mixed effects model, School effectiveness, Simple fixed effects model, Value-added assessment.

COMPARATIVE ANALYSIS OF COMMON STATISTICAL MODELS USED FOR VALUE-ADDED ASSESSMENT OF SCHOOL PERFORMANCE

by

SEDAT SEN

B.A., Hacettepe University, Turkey 2006

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2010

COMPARATIVE ANALYSIS OF COMMON STATISTICAL MODELS USED FOR VALUE-ADDED ASSESSMENT OF SCHOOL PERFORMANCE

by

SEDAT SEN

Major Professor: Seock-Ho Kim
Committee: Allan S. Cohen
Jonathan Templin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2010

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION AND THEORETICAL FRAMEWORK

Introduction

Over the past few decades, there has been growing interest in the effectiveness and accountability of schools, especially since the adoption of the No Child Left Behind act of 2001 which requires states to measure student academic achievement and to report on progress using Adequate Yearly Progress (AYP) measures (Beardsley, 2008). This system is based on an approach which gives rewards to schools that make contributions to students' learning and sanctions those that do not make any improvement on student test scores.

Early applications of this state-wide assessment focused on the current status of students. The current-status approach compares different cohorts of students at a single point in time (Doran & Izumi, 2004). It simply uses the percentage of students who passed the state test at the end of the school year.

Educators recognize that a one-time test score is not always a useful way to estimate school effects on student performance. Differences among schools may be due to student and school variables that influence test scores. Current-status methods don't take socioeconomic factors into account when assessing the schools' effectiveness. Although these methods are located at the heart of the state accountability system, there are at least two reasons why they're invalid and inappropriate to use for the purpose of school comparisons.

First, students come to school with different backgrounds. In other words, there is no random assignment of students to schools (Doran & Izumi, 2004). This results in making unfair comparisons between disadvantaged and advantaged schools in terms of socioeconomic status.

Second, current-status methods are cumulative. They reflect the impact of learning obtained from all previous schools on students' performance scores (Doran & Izumi, 2004). We cannot hold only the latest school accountable for his good or poor test score if a student has changed his school during his education. As Ballou et al. (2004) note, holding schools accountable based on mean achievement levels makes no sense when students enter those schools with large mean differences in achievement.

It is widely accepted that status-based accountability systems are likely to be flawed, resulting in inaccurate judgments of school quality. As the shortcomings of this method become apparent, an alternative way of assessing school effectiveness has gained ground in the accountability system. This new method focuses on the improvement students in the school made during the year. Instead of considering how cohort groups have increased in knowledge, measuring student progress over time from one measurement to the next is more reasonable in terms of "learning," which is meant to be "change." Growth models are designed to generate estimates from that kind of data (Doran & Izumi, 2004).

Researchers have come up with the idea of value-added analysis (VAA) which enables them to use individual student achievement scores over time in order to identify effective schools. As defined by Tekwe et al. (2004) “Value-added is a term used to label methods of assessment of school/teacher from one year to the next and then use that measure as the basis for a performance assessment system” (p. 31). Pioneers of VAA claim that VAA generates fairer and more accurate estimates than those generated by state tests that measure only the achievement of a single year.

The primary purpose of VAA is to determine the impact of teachers or schools on the progress of their students (Raudenbush, 2004). Gain scores are computed by taking the differences between students’ scores on state tests from one grade level to the next (Sanders et al., 2002).

The VAA system evaluates schools based simply on how they increased the level of their students’ knowledge. The two basic ideas underlying value-added measurement can be presented as it is calculated for each individual nested within the schools and based on changes in student performance from one year to the next (Ladd & Walsh, 2002). Since each state has an annual assessment system, VAA is readily applicable to the existing state accountability systems. Another advantage of VAA is that, unlike the current-status method, it can control the effect of confounding variables such as student and school socioeconomic status that may influence the test scores. This is simply an attempt to minimize the influence of experiences, privilege, and ethnicity on student performance.

In general, value-added models (VAMs) are a class of statistical model procedures that use students’ standardized test scores over time to identify the degree to which a student’s progress is a function of their own characteristics or that of the characteristics of their school (Doran & Izumi, 2004).

VAMs have recently received a great deal of interest from both policy makers and researchers due to a belief that they can adequately determine how individuals are growing over time while statistically attributing the gain scores to their schools (Sanders et al., 1997). It is a promising research area in education and has a significant role in shaping the school accountability system. VAMs may take different forms including simple gain score models and more complex models with random effects (Lockwood, Doran, & McCaffrey, 2003).

Several approaches have been suggested by researchers to obtain value-added measures of schools. Current-status methods all rely on regression models and they assume that school effects are fixed (Tekwe et al., 2004). They are also confounded with nonschool factors (Sanders, 2000), whereas VAMs require the use of more complex statistical models such as mixed models and hierarchical models with school effects assumed to be random. Hanushek (1972) was the first researcher who attempted to put value-added modeling methods into the accountability system. However, Sanders, who developed the Tennessee Value Added Assessment System (TVAAS), was the first researcher whose work was implemented in a statewide testing system (Stewart, 2006).

According to the RAND report (McCaffrey et al., 2003) early VAM applications (e.g., Murnane 1975; Hanushek 1972) primarily used fixed effects, while more recent applications (including the TVAAS layered model) have used random effects exclusively. Another important model which relies on hierarchical linear model analysis was developed by Raudenbush and Bryk (1986), and Aitkin and Longford (1986). Although there are several VAMs which are based on different assumptions (Braun, 2005; Tekwe et al., 2004), the most popular model has been TVAAS (Olson, 2004). For any of these models to be useful in VAA analysis the test scores must be vertically scaled (Ballou et al., 2004; Doran & Cohen, 2005). In brief, longitudinal data, annual assessment, and vertically equated tests are said to be basic elements of VAMs. Standardized assessment scores are mostly used in VAM studies. Though no VAM has yet been obvious to be clearly superior over another, VAMs are known to be fairer and more accurate than conventional methods.

Since VAMS have the capability of controlling the effects of socioeconomic status and prior performance, they can be preferred by educators rather than current status applications. In this study four popular VAMs will be examined in the context of school effectiveness. A brief description of a simple fixed effects model (SFEM), two hierarchical linear models (HLM), and a layered mixed effects model (LMEM) will be presented in the next sections.

Simple Fixed Effects Model (SFEM)

Fixed effects models (FEM) stands out among other models for the way that they take school effects to be fixed rather than random. This is the simplest VAM, requiring less computation than the others. Estimates of this model are intuitively understood by policymakers and educators with little statistics experience (Wiley, 2006). As Tekwe et al. (2004) stated, the simple fixed effects model (SFEM), an extension of FEM, does not incorporate compositional or student-level covariates and does not apportion variance when students attend multiple schools, thus it does not produce any shrunken estimates. As SFEM uses only two years of data in a single subject, its application is very straightforward.

Model parameterization:

$$d_{ijs} = \beta_{0s} + \sum_{k=1}^{44} \beta_{1ks} S_{kij2} + \varepsilon_{ijs} , \qquad (1)$$

where

$d_{ijs} = \gamma_{ijs2} - \gamma_{ijs1}$,

$d_{ijs} =$ is simple change score obtained from difference between two examinations of a student i in school j on the same subject area s,

$\gamma_{ijst} =$ is the test score on the subject area s $(s = 1, 2)$ at time t $(t = 1, 2)$ for the student j $(j = 1, \cdots, n_j)$ in school i $(i = 1, \cdots, n_i)$,

S_{kij2} = is effect coding at time (t = 2) for school k $(k = 1, \cdots, 44)$ and coding numbers $m(m = 1, \cdots, 43)$,

S_{kij2} = 1 for $k = m$ and$k \neq 44$; 0 for $k \neq m$ and $k \neq 44$; -1 for $k = 44$,

and ε_{ijs} is the random error for student j in school i for subject area s,

it is assumed that $\varepsilon_{ijs} \sim N\left(0, \sigma_{\varepsilon s}^2\right)$.

B_{1ks} coefficient in Equation 1 is value-added in the subject area s and in school k.

For this specification number of schools was taken as 44.

Hierarchical Linear Models

As the name implies, hierarchical linear models (HLM) require using nested data that is ordered hierarchically. In educational data structures students are nested within classes and classes are nested within schools. Due to the nature of the data used in education, HLM has been widely adopted by educators in the assessment of schools. HLM is a special type of the general mixed models family and can be used to obtain value-added measures. These models demand more computation than SFEM. Unlike SFEM, these models produce shrunken effects.

According to Raudenbush and Bryk (1988-89) the HLM analysis consists of four parts as follows:

i. Apportioning variation between and within units
ii. Homogeneity of regression assumption is assessed
iii. Testing for compositional effects
iv. Assessing the effect of the method

Traditional regression methods assume that individuals are independent of each other, but students within the same school might have similar results compared with those from different schools. Due to this violation of assumption HLM may be preferred instead of linear models.

In this study, two different types of HLM will be examined. Both models, unadjusted HLM (UHLMM) with random intercept and adjusted HLM (AHLMM), which consist of two equations called student- and school-level models, are applicable to studying the effects of schools. Simply, the two-level HLM provides an analytical framework for examining the effects of school efforts on student outcome distribution.

Unadjusted Hierarchical Linear Model (UHLMM)

UHLMM uses unadjusted change score with random intercept. This model consists of two level HLM described by the following equations;

Student-level model

$d_{ijs} = \beta_{0is} + \varepsilon_{ijs}$,

where d_{ijs} is the change score defined as in Equation 1, β_{0is} is a random intercept associated with the school *i*, and ε_{ijs} is a random error.

School-level model

$\beta_{0is} = \gamma_{0s} + \xi_{is}$,

where γ_{0s} the mean of the random intercepts, β_{0is}, and ξ_{is} is the random effect of the school *i* on the random intercept for the subject area *s*. It was assumed that β_{0is} and ξ_{is} were independent.

Single equation form

$$d_{ijs} = \beta_{0s} + \xi_{is} + \varepsilon_{ijs}. \qquad (2)$$

.

<u>Adjusted Hierarchical Linear Model (AHLMM)</u>

AHLMM is adjusted for student- and school-level covariates.

Student-level model

$$d_{ijs} = \beta_{1s}\gamma_{ijs1} + \beta_{2s}Min_{ij} + \beta_{3s}Pov_{ij} + \varepsilon_{ijs},$$

where $d_{ijs} = \gamma_{ijs2} - \gamma_{ijs}$, β_{0is} is a random intercept associated with the school *i* and subject area *s*, $Min_{ij} =$ an indicator of minority status (Yes, No) for student *j* in school i, $Pov_{ij} =$ an indicator of poverty (Yes, No) for student *j* in school *i*, β_{1s}, β_{2s}, and β_{3s}, are the fixed effects of intake score, minority status, and poverty on learning gain in subject area *s*, and ε_{ijs} is a random error.

School-level model

$$\beta_{0is} = \gamma_{0s} + \gamma_{1s}Z_{1i} + \gamma_{2s}Z_{2i} + \xi_{is},$$

where Z_{1i} is the mean input score for the i^{th} school, Z_{2i} is the percentage of poverty students in the i^{th} school, ξ_{is} is is the random error associated with the value of the random intercept for the s^{th} subject area test and the i^{th} school in the student level model, and the γ's are fixed coefficient parameters. The assumption concerning the within and between school error terms in this model is that the ε_{ijs} and ξ_{is} are independent.

Single equation form

$$d_{ijs} = \gamma_{0s} + \gamma_{1s}Z_{1i} + \gamma_{2s}Z_{2i} + \beta_{1s}\gamma_{ijs1} + \beta_{2s}Min_{ij} + \beta_{3s}Pov_{ij} + \xi_{is} + \varepsilon_{ijs}, \tag{3}$$

Layered Mixed Effects Model (LMEM)

LMEM (which is also known as TVAAS) is the best-known VAM. It was created by Sanders and his associates to measure educational outcomes in Tennessee. Previous models could only analyze non-missing parts of the data while omitting missing student information on some variables. LMEM, however, enables a longitudinal analysis no matter how sparse or complete the data record for each student is (Sanders & Horn, 1994; Sanders et al., 1997). Sanders and his colleagues called it layered model because the model for later years adds layers to the model for earlier years (Sanders et al., 1997). TVAAS does not attempt to account for confounding factors such as SES, demographics, or other factors that influence achievement. Some advantages of LMEM can be summed up as follows:

- It is capable of including students with missing data
- There is no requirement for information on student or school characteristics
- It has the capacity of analyzing simultaneous assessments of multiple subjects
- It includes information from previous years and uses multiple years' data

In spite of these advantages, it is difficult to comprehend due to the complexity of the model. It is also a special case of mixed models and does produce shrunken effects.

Model parameterization;

The simplest form of the model used in the LMEM analysis for this application is

$$\gamma_{ijst} = \mu_{st} + \sum_{s=1}^{2} \sum_{k=1}^{44} P_{ijk1}\mu_{ks1} + \varepsilon_{ijst}, \tag{4}$$

where γ_{ijst} is defined as in Model 1, and

μ_{st} = the population mean parameter for the test score on subject area *s* at time *t*,

P_{ijkl} = the proportion of academic year time spent by student *j*, who was in the school *i* at time 2 test, in school *k* during the year prior to the test at time *l* (*l* = 1,…, *t)*,

u_{ksl} = the random effect of the school *k* on subject *s* test scores at time 1,

ε_{ijst} = random within school error for student *j* in school *i* for subject area *s* at time *t*.

Number of schools is 44.

Best Linear Unbiased Predictor (BLUP)

The main difference between conventional linear model methods and mixed model methodology is that the latter estimate specific random effects (Little et al., 1996). As stated in the previous sections, the SFEM model assumes school effects as fixed while the other three models assume that school effects are random. When one makes a random effects assumption, these effects are estimated by BLUPs (Raudenbush & Bryk, 2002). Using mixed models, value-added measures are calculated as estimates of best linear unbiased predictors (BLUPs) of the random school effects in each VAM (Tekwe et al., 2004). In our study, mixed models (UHLMM, AHLMM, and LMEM) use students as their own control and include a shrinkage estimate to control school/non-school factors.

BLUP, also known as the "shrinkage estimate," provides some advantages such as the following:

- It protects against spurious estimates due to too little data (Sanders et al., 2006).
- It minimizes the likelihood that a school will be misclassified as extremely good or extremely bad by chance (Franco, 2006).
- Another theoretically proven advantage of BLUP is that it produces the maximum correlation between the estimate and the true effect (Searle et al., 1992).

Statement of the Problem

Although any VAA approach is superior to conventional methods with regard to school accountability, there are a number of criticisms that are commonly leveled against these models. Due to the complexity of the models and their extensive use of computational methods, VAM approaches have not yet been adopted by each state accountability system. Difficulties with user friendliness, insufficient validity evidence, and methodological issues with missing data make it unappealing to educators. Besides, although VAMs attempt to deal with randomization by using adjustments, they do not solve this problem completely (Wiley, 2006).

While advocates of VAA claim that VAMs provide strong tools for school effectiveness, there are still challenges that threaten the validity of the teacher and school effects they are designed to generate (Wiley, 2006). As Kupermintz (2003) noted in his article, TVAAS data should be made available to researchers in order to for them to perform a proper validity study because restriction of access to the data prevents doing external review and validity studies for

TVAAS. Wiley (2006) states that none of the VAA approaches have provided perfectly valid estimates of school contributions to student performance. Sanders & Wright (2008) claim that all of the criticisms made against the validity of VAMs may be made against the validity of any kind of standardized test. As long as the validity of VAM estimates remains questionable, they should not be used as the sole basis for weighty decisions (Wiley, 2006).

Other criticism of VAMs are that "they do not represent perfect strategies for measuring school effects, they more reasonably align with the notion of student learning, do not encourage schools to target instruction for middle-performing students, and set expectations for growth rates for individual students towards an expected learning outcome" (Doran & Izumi, 2004) (p. 28).

Given these disadvantages, there is a common desire for easier and accurate models among practitioners. In order for VAMs to be useful, they should be understood by practitioners and be easy to use. Most of the states still rely on conventional methods due to the complexity of VAA approaches. It is obvious that there is a lack of information on VAMs and disbelief concerning VAA approaches. We believe that this is an important issue to be solved.

Purpose of the Study

There have been numerous studies that show the strengths of the VAMs over the conventional methods. However, reluctance among state accountability systems to use VAMs due to their complexity prevents this method of assessing school effectiveness from gaining popularity. Most VAA approaches remain highly technical, and only a few studies have attempted to provide conclusive evidence that simpler models are as efficient as more complex models (Doran & Fleischman, 2005).

Several models introduced in VAA calculate the value-added measures based on different assumptions. SFEM, UHLMM, and LMEM do not account for school/non-school variables, while AHLMM attempts to control these factors by statistical adjustments. It has been claimed that controlling school/non-school factors has both advantages and disadvantages for the calculation of value-added measures of schools.

Using the Florida Comprehensive Assessment Test (FCAT) scaled scores; this study will investigate the impacts of school/non-school factors on school-level value-added scores and attempt to shed light on the issue of complexity in VAMs. In the context of these two issues four popular VAMs will be examined to determine the most desirable model(s) for practitioners and policy makers. Based on the results obtained from the four models, empirical comparisons of each model will be made.

Questions and Hypothesis

Our primary question is "do we really need complex statistical models for value-added assessment of school effectiveness, or can we assess schools with less complex models as efficiently as with complex models?" We must simply attempt to determine whether there is a difference between simple and complex VAMs in terms of school effectiveness. We will also try to determine the effect of including school and student covariates in VAMs. In the light of these questions all models mentioned in the introductory part will be analyzed and their results will be correlated to show the relationships between each pair of models.

Significance of the Study

There has been an increase in the number of states that have altered their systems basis from conventional methods to VAA methods. However a great number of states have not adopted new approaches in their systems due to the complexity of the new models and data structure(s) needed for their implementation. Amrein-Beardsley (2008) stated "Educators want to use relatively simple, understandable statistical models to analyze educational phenomena, but social complexity demands that statistical models be sophisticated enough to capture reality with integrity (Andrejko, 2004; Callendar, 2004)" (p. 67). The primary criticisms of VAMs noted in Sanders' (2000) article are that the process is too complicated and that there is (still) too much reliance on a single test. Sanders attempted to answer these in his article, but there is still a need for evidence with regard to the advantages of VAMs for school accountability systems.

In this study we will first attempt to determine whether the simple VAMs preferred by practitioners are useful in the assessment of school effects on student performance. If new studies can provide some evidence related to the efficiency of simple models, we believe that there will be an increase in the number of states that accept at least simple models of VAA. Since simple VAMs are easy to implement and understand for non-statisticians, research in this area needs to be increased in the context of school accountability. This study aims to contribute to the solution of the aforementioned educational problem.

Theoretically, like all other models, VAMs are also based on some assumptions. They also differentiate in their mathematical structures. Some of them claim that school/non-school variables affect the value-added estimates, while others claim the reverse of that. However there is no broad consensus on effects of such variables in the calculation of school-level value-added measures. Last but not least, the purpose of this study is to investigate the impact of such variables on school-level value-added scores. In order to help policy makers to decide between the models, the advantages and disadvantages of adding covariates to the models need to be examined in depth.

Finally, it is widely accepted that policy decisions should not be solely based on the results obtained from theoretical models such as VAMs. It is hoped that our findings will, also be helpful for those who make decisions based on non-empirical considerations.

CHAPTER 2

REVIEW OF LITERATURE

To date, several alternative models, which may range from simple gain scores to complex mixed models, have been suggested by researchers with regard to assessment of school effectiveness. However there have been limited numbers of studies which make comparisons among them. If we want to find solutions to problems in accountability systems by adopting new VAA approaches, we should find out which model is most efficient and easiest to implement.

Fortunately, a few important studies have been conducted to determine the most desirable model for computing school effects. *Journal of Educational and Behavioral Statistics* published one volume solely concerning the VAA and popular VAMs. They concluded that there are numerous acceptable models rather than only one superior model.

Tekwe et al. (2004), Ballou et al. (2004), and McCaffrey et al. (2003) are essential studies which show the differences among VAMs. Compared to conventional methods, VAMs are known to be less biased and to produce more precise estimates. In order for VAMs to be accepted by policymakers, several attempts to show the advantages of VAMs have been made with either simulation or real data. Although there is a lack of studies showing which VAM is better than the others, LMEM model has been very popular for accountability systems. Ballou et al. (2004) conducted a simulation study to evaluate the TVAAS (LMEM) model. They noted that the TVAAS uses a highly parsimonious model that omits controls for contextual factors such as SES and demographics that influence achievement.

Unlike the LMEM model, HLM models include school and student variables and attempt to control such factors by statistical adjustment (Bryk & Raudenbush 1992). Sanders et al. (2004) noted that inclusion of these factors in HLM affects the school estimates resulting in biased measures of schools towards zero. Sanders' LMEM model does not account for these variables. His model attempts to eliminate controls for such factors by the use of multiple measures on each student (Ballou, Sanders, & Wright, 2004). In response to criticism that the LMEM model does not include these factors, Sanders conducted a study with the inclusion of these factors to the model and showed that there is no significant difference between the results obtained from both models (Ballou, Sanders, & Wright, 2004).

In fact, Sanders's LMEM model is popular in comparison with other models due to such advantages as being unaffected by students' backgrounds, being more useful with missing data situations, and being the only model that allows for the simultaneous assessment of results from multiple content (Sanders, & Horn, 1998).

However, McCaffrey et al. (2003) performed a simulation study that compared the results of the general model (which is similar to AHLMM) with those of a layered model which is similar to the LMEM. Based on the results, they concluded that AHLMM better fitted with the data than the layered model.

Tekwe et al. conducted an empirical comparison of VAMs by performing simulation data to determine the differences among four popular VAMs (Tekwe et al., 2004). The simple fixed effects model, unadjusted and adjusted hierarchical linear models, and layered mixed effects model were compared using the same data set in their study. They attempted to show that even simpler VAMs can produce results consistent with those obtained from more complex models. Tekwe et al. (2004) claimed that “there is little or no benefit to using the more complex (models)” (p. 31).

Based on the results they recommended the use of the SFEM model due to its simplicity and similar estimates to those of other models in a low-stake accountability system that gives incentives to effective schools (Tekwe et al., 2004). They also found that the AHLM model should be preferred when there is a need for controlling the effects of student and school variables that influence the estimates (Tekwe et al., 2004). They noted that decisions regarding these two models should be based on non-empirical considerations (Tekwe et al., 2004).

CHAPTER 3

PROCEDURE

Instrumentation

In this study, separate analyses were done for each elementary school grade cohort (6th-8th grades) (in Florida) in 2002. Students' scores on the math and reading tests of the Florida Comprehensive Assessment Test (FCAT) that was administered in 2002 and 2003 will be analyzed to obtain estimates of 40 schools located in a large-sized school district. The FCAT is a criterion-referenced test that aims to assess student achievement in the high-order cognitive skills represented in the Sunshine State Standards in reading, mathematics, writing, and science. The FCAT includes three types of questions: multiple choice items, graded response items, and performance tasks. Since VAA requires using vertically scaled scores, FCAT scaled scores reported in the exam data base will be utilized in our analyses.

Sample

Separated analyses were performed for each of three secondary school grade cohorts (i.e., 6th-8th grades) in 2003 in a large sized Florida school district with 40 secondary schools to be graded. Consecutive year reading and math scores on the FCAT from 2002 and 2003 were analyzed. Only standard curriculum students were used in the analyses; exceptional student education (ESE) students and students in the limited English proficiency (LEP) program for two or fewer years were excluded. Examinees that report their ages outside the acceptable age range for a certain grade were excluded from the data.

A total of 60,718 students will be available for analyses after the exclusions (19,611 for the analysis of 6th grade test scores in 2003, 20,433 for the 7th grade analysis, and 20,674 for the 8th analysis). Poverty status information was also provided based upon whether a student takes the free lunch. The other nonschool variable was defined as minority which is based upon black or nonblack ethnicity. Descriptive statistics based on grade and subject combination is presented in table 1. More detailed information related to each school is also presented in Table 2, 3, and 4.

Computer Program and Syntaxes

A data set from 40 schools were analyzed using the statistical analysis software (SAS) program to obtain value-added measures of each school based on model structures provided in the previous sections. For SFEM model analysis the general linear modeling (GLM) procedure of SAS were used, while PROC MIXED statements were used for others due to their special cases of mixed models. Each model was analyzed separately and then the value-added estimates from each model were correlated and compared in terms of school effectiveness. Syntaxes for each analysis are presented in the Appendix.

Table 1

Sample Size, Mean FCAT and Standard Deviation by Subject, Grade and Year, and Percent Minority and Percent Poverty in 2003 by Grade

		Reading			Math			Demographics in 2003	
Grade		2002 score	2003 score	Change score	2002 score	2003 score	Change score	Pov.	Min.
	N	19,611	19,611	19,611	19,611	19,611	19,611	19,611	19,611
6	*M*	1421.32	1527.89	106.57	1566.02	1581.17	15.15	73.7%	28.6%
	SD	368.52	371.85	235.62	294.80	297.80	189.48		
	N	20,433	20,433	20,433	20,433	20,433	20,433	20,433	20,433
7	*M*	1493.98	1623.32	129.33	1554.14	1692.70	138.56	72.2%	28.4%
	SD	385.43	348.92	244.52	293.74	255.18	191.43		
	N	20,674	20,674	20,674	20,674	20,674	20,674	20,674	20,674
8	*M*	1606.93	1782.10	175.16	1675.76	1804.40	128.64	70.3%	28.6%
	SD	345.79	276.42	223.87	274.60	216.95	169.142		

Table 2

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 6 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 6)
1	1706.75	1688.12	13.5	36.5	591
2	1618.40	1597.30	0.2	79.6	476
3	1472.04	1441.21	75.8	80.3	401
4	1517.67	1439.91	36.5	84.9	463
5	1459.90	1369.11	2.5	97	400
6	1590.02	1515.32	27.9	76.7	416
7	1590.97	1549.07	14.9	89.7	349
8	1640.69	1556.01	4.2	70.7	542
9	1490.10	1445.75	71.7	78.6	473
10	1717.36	1676.87	2.9	42.7	375
11	1615.65	1575.67	15.2	61.1	565
12	1582.79	1426.63	3.4	92.9	411
13	1696.44	1654.45	2.8	49	429
14	1673.28	1649.36	4.9	46.3	735
15	1584.88	1500.11	7.2	83.6	305
16	1678.81	1656.90	25.9	49.8	652
17	1569.32	1450.56	24.3	92.9	367
18	1608.35	1555.17	65.3	84.7	562
19	1545.82	1441.85	0.5	93.9	376
20	1509.20	1479.94	45.5	86.7	391
21	1374.63	1258.36	27.9	97.6	340
22	1410.64	1302.53	65.8	92.3	310

Table 2 (continued)

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 6 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 6)
23	1458.91	1407.57	76.7	90.8	348
24	1567.77	1475.65	1.8	92.4	380
25	1502.09	1446.40	57.3	89.3	356
26	1630.69	1622.68	1.6	63.4	506
27	1318.51	1253.47	90.8	96.3	295
28	1648.86	1617.55	5.7	62.6	545
29	1555.20	1482.95	13.6	80.6	589
30	1566.14	1534.95	87.5	79.1	559
31	1630.42	1595.76	73.8	78.2	294
32	1517.82	1427.94	79.4	94.7	452
33	1532.90	1476.56	2.3	86.9	567
34	1779.04	1771.80	11.3	21.5	488
35	1491.55	1458.60	95.7	83.3	282
36	1613.07	1553.91	16.7	72.2	460
37	1522.08	1454.19	30.3	79.4	446
38	1557.37	1509.78	42.2	73.6	481
39	1679.69	1649.27	0.8	69.1	398
40	1506.81	1428.83	1.9	92.7	423
41	1729.57	1725.82	25.1	35.7	541
42	1707.31	1682.37	20.2	51	420
43	1532.65	1492.76	0.8	86	472
44	1381.35	1312.19	80.5	94.7	380

Table 3

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 7 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 7)
1	1808.23	1769.15	14.3	39.1	537
2	1703.48	1684.51	0.2	77.6	527
3	1581.30	1518.76	86.3	89.2	371
4	1541.16	1463.01	37.6	94.1	340
5	1610.87	1515.03	6.7	96.8	464
6	1717.76	1678.97	26.7	71.7	389
7	1724.40	1672.10	18.1	88.1	354
8	1712.08	1664.48	3.5	69.8	605
9	1508.78	1413.85	92.8	91.5	377
10	1839.25	1761.33	2.1	37.2	331
11	1761.40	1702.19	14	62.1	641
12	1699.37	1545.42	2.8	88.5	471
13	1784.82	1728.16	2.1	37.3	469
14	1760.30	1729.24	4.2	45.0	758
15	1704.62	1616.13	7.8	85.8	387
16	1767.13	1725.39	24.3	45.2	672
17	1626.41	1559.43	22.4	85.8	366
18	1711.73	1641.07	62.0	87.1	606
19	1691.20	1548.09	0.7	91.0	401
20	1639.27	1547.24	46.7	84.2	495
21	1532.50	1420.41	38.8	96.1	356
22	1497.85	1434.73	71.4	92.7	370

Table 3 (continued)

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 7 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 7)
23	1638.01	1525.96	77.7	85.7	363
24	1668.44	1570.17	0.5	85.5	620
25	1592.72	1537.57	53.9	85.2	310
26	1745.66	1704.04	1.8	67.1	554
27	1539.05	1467.01	93.0	91.9	271
28	1731.50	1652.01	4.5	66.2	533
29	1650.70	1557.31	12.4	75.3	615
30	1672.40	1628.96	85.3	70.6	599
31	1659.14	1622.47	80.0	74.1	290
32	1629.92	1527.62	81.7	91.7	509
33	1708.24	1583.13	1.6	84.4	569
34	1845.83	1852.14	13.1	22.0	563
35	1611.59	1577.76	95.3	72.4	275
36	1753.95	1679.21	16.8	66.4	428
37	1668.57	1546.17	24.0	76.9	516
38	1667.99	1598.68	44.4	69.5	459
39	1782.66	1735.36	0.4	69.9	492
40	1608.37	1559.92	1.2	91.6	404
41	1825.14	1794.89	25.5	31.8	569
42	1801.09	1758.49	29.9	57.2	374
43	1717.66	1632.43	0.8	82.0	506
44	1531.19	1411.09	85.6	90.2	327

Table 4

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 8 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 8)
1	1859.59	1885.05	13.0	36.3	617
2	1827.83	1839.51	0.2	74.6	564
3	1686.86	1683.18	87.6	82.9	380
4	1718.55	1673.70	45.9	95.8	355
5	1715.30	1663.83	6.8	96.8	440
6	1852.38	1834.64	27.5	71.4	371
7	1830.78	1831.81	12.0	91.5	425
8	1839.67	1826.99	3.4	72.7	645
9	1646.20	1616.66	94.4	87.5	375
10	1944.51	1890.66	0.9	40.2	316
11	1871.51	1839.04	14.3	61.0	631
12	1844.79	1739.19	1.5	87.3	534
13	1860.30	1840.71	3.0	44.7	573
14	1878.75	1890.12	3.3	44.7	695
15	1775.26	1767.98	8.5	80.8	343
16	1872.58	1865.22	25.3	44.5	676
17	1739.94	1714.62	21.6	87.4	348
18	1787.67	1776.67	67.7	85.3	613
19	1785.30	1757.47	0.4	90.0	452
20	1746.37	1691.85	45.2	73.1	458
21	1663.01	1604.76	35.7	97.0	395
22	1681.30	1643.15	70.7	86.2	341

Table 4 (continued)

Mean Scores, Percent Minority and Poverty, and Number of Students of Grade 8 within Schools in 2003

School	Math	Reading	%Minority	%Poverty	*N* (GRADE 8)
23	1762.29	1700.20	75.4	76.1	456
24	1787.31	1747.57	2.3	86.3	598
25	1765.67	1697.11	47.1	83.4	295
26	1848.92	1857.95	1.0	61.0	515
27	1605.62	1566.16	92.2	91.4	269
28	1856.63	1832.92	6.9	63.6	492
29	1771.11	1749.18	11.8	73.8	627
30	1803.83	1803.61	83.6	67.7	609
31	1764.02	1764.75	77.9	72.5	244
32	1733.91	1708.02	81.6	91.1	549
33	1828.46	1792.22	1.2	82.7	595
34	1961.56	1985.95	13.0	17.6	638
35	1722.31	1707.14	90.7	62.5	323
36	1810.74	1764.23	20.3	66.8	364
37	1746.87	1718.41	25.0	74.4	508
38	1801.94	1768.22	44.6	69.6	448
39	1895.16	1864.28	0.5	65.9	437
40	1742.35	1725.07	1.3	88.0	393
41	1917.74	1920.71	27.2	33.3	580
42	1894.53	1882.82	27.4	48.4	343
43	1815.11	1800.31	0.4	84.2	480
44	1668.92	1637.18	83.2	89.3	364

CHAPTER 4

RESULTS

Value-added measures of schools generated from each models were correlated. The results of correlation values among the models are given in Table 5. Since the models each have distinguishing characteristics in their natures we attempted to interpret the results based on these characteristics. Differences in characteristics of the models are presented in Table 6. We also ranked each school based on their value-added estimates from different models. Ranking tables for each grade are also presented in appendices.

Table 5

Table of correlations between value-added measures of the models

	6th grade		7th grade		8th grade	
	Math	Reading	Math	Reading	Math	Reading
SFEM vs. UHLMM Model 1 & Model 2	.99	.99	.99	.99	.99	.99
SFEM vs. AHLMM Model 1 & Model 3	.75	.85	.80	.55	.73	.74
SFEM vs. LMEM Model 1 & Model 4	.45	.25	-.09	-.09	-.34	-.48
UHLMM vs. AHLMM Model 2 & Model 3	.75	.85	.80	.54	.73	.74
UHLMM vs. LMEM Model 2 & Model 4	.49	.27	-.12	-.09	-.36	-.51
AHLMM vs. LMEM Model 3 & Model 4	.01	.01	.00	.00	.00	.00

Note.

SFEM = Simple fixed effects model

UHLMM = Unadjusted hierarchical linear model

AHLMM = Adjusted hierarchical linear model

LMEM = Layered mixed effects model

Table 6

Summary of Models' Distinguished Characteristics

Model identifier	Dependent variable	School effect	Student-level variable	School-level variable	Apportion between schools	Multivariate method
Model 1 (SFEM)	Change score	Fixed	No	No	No	No
Model 2 (UHLMM)	Change score	Random	No	No	No	No
Model 3 (AHLMM)	Change score	Random	Yes	Yes	No	No
Model 4 (LMEM)	Pre/Post-test Scores	Random	No	No	Yes	Yes

Each of the value-added models used in this study is based on different assumptions and has different characteristics. As showed in Table 6 while they differ in some characteristics, they also have some similarities. Value-added measures obtained from all of the models were examined based on the differences and similarities among models. Since they have different assumptions, we cannot make conclusions based on solely correlation values. We just can make conclusions about the rank order of school effects generated from each model.

First of all we attempted to determine the impact of taking school effects as random in the model structure on identifying effective schools. Model 1 is the only one that has taken the school effects as fixed, so we can compare Model 1 to Model 2 which has the same characteristics with Model 1 except for the random school effect. The most important finding that is evident in Table 5 is the very high correlation between SFEM and UHLMM value-added estimates (r = .99) in all cohorts. This suggests that we can have the same school rank orders using two models. We can conclude that there is no difference between taking school effects as random or fixed in terms of rank order of school effects. These models produce similar results and the rank orders are almost same.

The second concern in measuring school effectiveness is to include school and non-school covariates in the models. Among the models, only AHLMM has the capacity of taking student- and school-level variables into account. Apart from this characteristic we can say that AHLMM and UHLMM are identical, so we can make inferences based on the comparison of these two models. As Table 5 showed there are moderate correlations ranging from .54 to .85 between AHLMM and UHLMM through the cohorts. This result indicates that the effect of including school- and non-school variables in the Model 3 had a remarkable impact on VAA of schools.

Another comparison can be made between SFEM and AHLMM in order to see the effect of employing shrinkage or including school- and non-school variables in the Model 3 on value-added assessment of school performance. Correlation analyses between these two models showed moderate values ranging from .55 to .85. These results indicate that there is a notable difference between SFEM and AHLMM. When there is a preference to adjust confounding variables AHLMM can be chosen, yet we cannot recommend the use of one instead of other based on the results. The only thing we can conclude was that there was a difference between the rank orders of schools generated from these two models.

LMEM was the only model that used the multivariate method. The multivariate effect can be assessed when a comparison is made between LMEM and all other models. The correlation of value-added measures from LMEM with those from other models ranged from .0 to .51. We concluded that there was a great discrepancy between LMEM and all other models. Using the multivariate method in value-added models had a great effect on school performance. The difference between LMEM and all other models was notable, and indicate that when the multivariate method is used in the analysis, the estimates change. Therefore the rank orders of schools also change. It is surprising that AHLMM and LMEM were found to be uncorrelated.

The results of the study showed very strong correlations between results generated by SFEM and UHLMM, but much more modest correlation between the results of AHLMM and all other models. However the correlations between LMEM and all other models were found to be much lower in comparison to other models' correlations. We concluded on the basis of these results that there was no much difference between SFEM and hierarchical models in terms of the rank order of school estimates.

After the model was chosen value-added measures of students can be converted to grades. These grades can be used to see the performance of the teachers within each school. The criteria used in Tekwe et al. (2004) study have been used to create this table. In order to get these grades standardized value-added measures from output were divided by their standard errors and assigned gpa values based on the following criteria:

If $z > 2$, then assign a grade of A and 4 growth points;

If $1 < z \leq 2$, then assign a grade of B and 3 growth points;

If $-1 < z \leq 1$, then assign a grade of C and 2 growth points;

If $-2 < z \leq -1$, then assign a grade of D and 1 growth points;

If $z \leq -2$, then assign a grade of F and 0 growth points.

Grading results presented in Table 7.

Since grades from Model 1 and Model 2 are found to be almost same, we did not show the Model 2 values in the table. The grades assigned to each school based on Model4 were remarkably different from those assigned based on Model 1 and 3. However Model 1 and Model 3 produced more similar results than model 4 did.

These results showed that large schools with higher value-added estimates can have lower gpa values than smaller schools with lower value-added estimates. It is also possible that large schools with lower value-added estimates can have higher gpa values.

Table 7

Notes.

M1 = Model 1

M3 = Model 3

M4 = Model 4

M = Math GPA, averaged over grades.

R = Reading GPA, averaged over grades.

T = Total GPA, averaged over grades and subjects.

6 = Sixth grade GPA, averaged over subjects.

7 = Seventh grade GPA, averaged over subjects.

8 = Eighth grade GPA, averaged over subjects.

Table 7

Growth Point Averages for Each School Based on Value-Added Measures from Each of Three Models

	M1						M3						M4					
School	M	R	G6	G7	G8	T	M	R	G6	G7	G8	T	M	R	G6	G7	G8	T
1	0.00	0.66	0.50	0.50	0.00	0.33	0.33	1.33	0.50	1.50	0.50	0.83	4.00	4.00	4.00	4.00	4.00	4.00
2	3.00	3.66	2.00	4.00	4.00	3.33	2.33	2.33	1.50	2.50	3.00	2.33	3.66	4.00	4.00	3.50	4.00	3.83
3	1.00	2.33	0.50	1.00	3.50	1.66	2.33	2.66	2.00	2.00	3.50	2.5	0.00	0.00	0.00	0.00	0.00	0.00
4	2.00	1.00	2.00	0.50	2.00	1.50	2.00	1.66	2.50	1.00	2.00	1.83	0.00	0.00	0.00	0.00	0.00	0.00
5	3.33	3.00	2.00	4.00	3.50	3.16	2.66	1.66	2.00	2.50	2.00	2.16	0.00	0.00	0.00	0.00	0.00	0.00
6	2.66	1.66	2.50	2.50	1.50	2.16	2.66	2.00	3.00	2.00	2.00	2.33	3.00	3.33	2.00	3.50	4.00	3.16
7	1.00	2.33	1.50	1.00	2.50	1.66	1.66	2.00	2.00	1.50	2.00	1.83	3.33	3.66	2.5	4.00	4.00	3.50
8	3.00	2.66	4.00	2.00	2.50	2.83	1.66	2.00	2.00	1.50	2.00	1.83	3.66	3.66	3.5	3.50	4.00	3.66
9	2.00	1.00	0.00	1.00	3.50	1.50	2.33	2.00	2.00	2.00	2.50	2.16	0.00	0.00	0.00	0.00	0.00	0.00
10	3.66	3.00	4.00	3.00	3.00	3.33	3.00	2.33	2.50	2.50	3.00	2.66	4.00	4.00	4.00	4.00	4.00	4.00
11	2.33	1.66	3.50	2.50	0.00	2.00	2.33	1.33	2.00	2.50	1.00	1.83	4.00	4.00	4.00	4.00	4.00	4.00
12	3.00	2.33	3.00	2.50	2.50	2.66	1.66	1.66	2.00	1.50	1.50	1.66	3.00	0.33	1.00	1.50	2.50	1.66
13	2.66	2.00	4.00	2.00	1.00	2.33	2.33	2.33	3.00	2.00	2.00	2.33	4.00	4.00	4.00	4.00	4.00	4.00
14	1.66	2.00	4.00	1.50	0.00	1.83	2.00	2.00	2.00	2.00	2.00	2.00	4.00	4.00	4.00	4.00	4.00	4.00
15	1.00	2.33	2.00	1.00	2.00	1.66	1.00	2.00	2.00	1.00	1.50	1.50	2.00	2.00	2.00	2.50	1.50	2.00
16	0.00	1.00	1.00	0.00	0.50	0.50	1.33	2.00	1.00	2.00	2.00	1.66	4.00	4.00	4.00	4.00	4.00	4.00
17	0.33	1.00	1.50	0.00	0.50	0.66	0.66	1.33	2.00	0.50	0.50	1.00	0.66	0.00	1.00	0.00	0.00	0.33
18	0.00	1.33	0.00	0.00	2.00	0.66	2.00	2.66	2.00	2.00	3.00	2.33	2.66	2.66	3.00	3.00	2.00	2.66
19	3.33	3.00	3.50	4.00	2.00	3.16	2.66	1.66	3.00	2.50	1.00	2.16	1.66	0.66	0.50	1.00	2.00	1.16
20	1.66	1.66	1.00	2.00	2.00	1.66	2.00	2.00	2.00	2.00	2.00	2.00	0.00	0.33	0.50	0.00	0.00	0.16
21	1.00	1.00	0.00	1.50	1.50	1.00	0.66	1.00	1.00	1.50	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00

Table 7 (continued)

	M1						M3						M4					
School	M	R	G6	G7	G8	T	M	R	G6	G7	G8	T	M	R	G6	G7	G8	T
22	2.00	2.33	2.00	0.50	4.00	2.16	2.00	2.66	2.50	1.50	3.00	2.33	0.00	0.00	0.00	0.00	0.00	0.00
23	3.00	2.00	1.50	3.00	3.00	2.50	3.66	2.33	3.00	3.00	3.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00
24	1.33	2.66	1.50	3.00	1.50	2.00	0.66	2.00	2.00	1.00	1.00	1.33	2.00	1.00	1.50	1.50	1.50	1.50
25	3.33	2.66	4.00	2.50	2.50	3.00	3.33	2.33	4.00	2.00	2.50	2.83	0.33	0.00	0.00	0.00	0.50	0.16
26	1.00	1.66	2.50	1.50	0.00	1.33	0.66	1.33	1.50	1.50	0.00	1.00	4.00	4.00	4.00	4.00	4.00	4.00
27	2.00	2.33	1.50	3.50	1.00	2.16	2.33	2.66	2.50	3.50	1.50	2.50	0.00	0.00	0.00	0.00	0.00	0.00
28	1.66	2.66	3.50	1.00	2.00	2.16	1.66	2.00	2.00	1.50	2.00	1.83	4.00	3.66	4.00	3.50	4.00	3.83
29	3.66	2.66	3.00	3.00	3.50	3.16	2.00	2.00	2.00	2.00	2.00	2.00	1.33	0.66	1.50	0.50	1.00	1.00
30	0.66	2.33	0.50	1.00	3.00	1.50	2.66	3.00	2.00	2.50	4.00	2.83	2.00	2.33	2.00	2.00	2.50	2.16
31	1.66	2.66	2.00	1.50	3.00	2.16	2.66	2.66	2.50	2.00	3.50	2.66	2.00	2.66	4.00	1.50	1.50	2.33
32	2.33	3.00	1.50	2.50	4.00	2.66	3.33	3.33	3.00	3.00	4.00	3.33	0.00	0.00	0.00	0.00	0.00	0.00
33	2.00	2.66	0.00	4.00	3.00	2.33	1.33	1.66	0.00	2.50	2.00	1.50	2.66	1.66	1.00	2.00	3.50	2.16
34	1.33	1.33	4.00	0.00	0.00	1.33	2.00	2.00	2.00	2.00	2.00	2.00	4.00	4.00	4.00	4.00	4.00	4.00
35	0.00	0.66	0.00	0.00	1.00	0.33	1.33	2.00	1.00	1.50	2.50	1.66	0.00	0.66	0.50	0.50	0.00	0.33
36	2.33	1.66	1.50	2.50	2.00	2.00	2.00	1.33	1.00	2.00	2.00	1.66	3.66	3.00	3.50	4.00	2.50	3.33
37	3.33	1.33	3.00	2.50	1.50	2.33	2.66	1.33	2.50	2.50	1.00	2.00	0.66	0.00	0.00	1.00	0.00	0.33
38	2.66	0.66	1.00	2.00	2.00	1.66	2.66	1.66	1.50	2.50	2.50	2.16	2.00	2.00	2.00	2.00	2.00	2.00
39	2.66	3.33	4.00	2.50	2.50	3.00	2.00	2.33	2.00	2.00	2.50	2.16	4.00	4.00	4.00	4.00	4.00	4.00
40	2.66	2.66	1.50	3.50	3.00	2.66	1.33	1.33	1.50	1.50	1.00	1.33	0.00	0.00	0.00	0.00	0.00	0.00
41	1.33	0.66	3.00	0.00	0.00	1.00	2.66	1.66	2.50	1.50	2.50	2.16	4.00	4.00	4.00	4.00	4.00	4.00
42	2.00	1.66	2.50	2.50	0.50	1.83	2.00	2.00	1.50	2.50	2.00	2.00	4.00	4.00	4.00	4.00	4.00	4.00
43	2.00	0.66	0.00	3.00	1.00	1.33	1.00	1.00	0.50	1.50	1.00	1.00	2.66	2.33	1.50	3.00	3.00	2.50
44	-	-	-	-	-	-	2.33	2.33	2.00	2.50	2.50	2.33	0.00	0.00	0.00	0.00	0.00	0.00

Individual school estimates from the models are shown in tables from Appendices E to J together with their ranks. The ranks of the school estimates from Model 1 are very similar to the ranks of the school estimates from Model 2. It is noted that estimates from both models are almost the same. This result also indicated that there was a little difference in taking school effects as random and fixed. Among the models Model 4 produced the greatest estimates for each individual school and it did not show any sign of relation to the other models. Model 3 was found to have moderate agreement with Model 1 and Model 3. Overall, our analysis demonstrated that VAM based rankings of schools are highly unstable across different grades. When results are compared grade by grade they are very consistent with the results of correlational analyses. The same conclusions can be made based on solely looking at school ranks.

Estimates shown in tables 8 through 13 are the value-added estimates generated from each models. For Model 1, the estimates can be interpreted as the difference between the school specific sample average change and the average of these average changes. Model 2 estimates are shrunken versions of the estimates of school effects in Model 1. They can be calculated as estimates of best linear unbiased predictors (BLUPs) of the random school effects in each school and each grade. Value-added estimates of Model 3 and Model 4 are also calculated as estimates of best linear unbiased predictors.

CHAPTER 5

SUMMARY AND DISCUSSION

Summary

The purpose of the present study is to show whether there was a remarkable differences among the models used in value-added assessment of schools. As its name implies the simplest model we used was the simple fixed effects model that assumes school effects as random. Two hierarchical linear models and a layered mixed effects model, special cases of mixed effects models, were other models used in this study. Each model has distinguishing characteristics and different assumptions. Value-added estimates of individual schools obtained from these models were analyzed to see the effects of the different characteristics on the estimates and school identification.

The primary question was to investigate whether simpler models such as SFEM are as effective as the more complex models such as AHLMM and LMEM in terms of school rankings. Previous research has found that there is a little difference between the results of simple and complex models (Tekwe et al. 2004). Correlation between SFEM and AHLMM ranged from .55 to .85 while SFEM and LMEM correlation values ranged between .09 and .48. The result of this study was partially consistent with their conclusions. While the simple model produced similar rank orders of school effects with AHLMM it did not show any agreement with LMEM. We also concluded that simple models are as effective as complex models and that they can be substituted in place of more complex models except for LMEM.

Another concern in value-added studies is to determine the impact of the inclusion of school and student background variables into models on value-added estimates. Among the models only AHLMM has statistical adjustment for these confounding variables that can affect the school estimates. Tekwe et al. stated that both inclusion and exclusion of these variable during the analysis result in biased estimates of schools. In our study AHLMM model was compared to other models in order to see the effect of these covariates. There were no remarkable differences between the results of AHLMM and the results of UHLMM and SFEM while LMEM was not in agreement with AHLMM. Correlation values between AHLMM and SFEM ranged from .55 to .85, AHLMM and UHLMM ranged from .54 to .85, and AHLMM and LMEM was found to be almost zero. Only the LMEM and AHLMM relationship was not consistent with the previous studies. Overall we can conclude that inclusion of these covariates had a great effect on value-added estimates.

The present study also reported gpa grading and rankings of each school based on value-added estimates obtained from each model. These results also showed consistency with the correlational analysis. Overall VAM based rankings of schools were found to be unstable over the grades.

In conclusion we showed that VAMs have an important role in shaping accountability system. We believe that showing the advantages and disadvantages of each model used in VAA would be helpful for those who have to make decisions on identification of successful schools.

Discussion

The goal of this study was to examine the questions raised by the application of value-added modeling of student performance as a means of evaluating schools. Four popular value-added models were used to obtain value-added measures of schools by analyzing two years standard assessment data. However the purpose was to show the differences between VAMs rather than to find the best VAM. Tekwe et al. (2004) have also conducted a similar study to compare the four VAMs used in my study. Since they used a narrow data set their conclusions remained limited. We also attempted to interpret our results on the basis of their findings. Overall the results from previous studies were mostly consistent with the results from the current study. In this study each model pair was compared to determine the effect of the different characteristics of these models such as random effects, adjustment of confounding variables, and multivariate method on the value-added measures and rankings of schools. Several conclusions can be drawn from the results of the current study concerning the application of these four models for value-added assessment of secondary school performance. The key conclusions of the study regarding these issues may be summarized as follows.

The primary question in this study was whether the simpler fixed effect model produces similar "value-added" effects than a more complex random effects model. Based on the statistical analysis, similar results were obtained from SFEM and HLM models. As Tekwe et al. concluded that there is little difference between the use of simpler and more complex models in value-added assessment of schools, our results were similar to their argument. Since there is a desire for using simpler statistical models among the public, these results may support the use of SFEM in accountability system with the prove of further research.

Some researchers concluded that the possibility of bias from the exclusion of covariates is also problematic for all VAM evaluations (McCaffrey et al., 2003). In our study we used AHLMM to show the effect on VAM estimates of including covariates in the model. In our analyses inclusion of these covariates in the Model 3 was found to have a remarkable effect on value-added estimates of schools. It is noted that this result was also consistent with the comments of Tekwe et al. (2004). This study supported the theory that the omission of covariates that contribute to outcomes can bias parameter estimates when students are stratified by those covariates (McCaffrey et al., 2003). From the results we can recommend that when students come from different backgrounds the use of AHLMM should be preferred over other VAMs.

Another important finding from the study was the conclusion pertaining to the use of multivariate method in school assessment. Based on the results the effect of using the multivariate LMEM compared to other models was notable. There was a disagreement in value-added results from the multivariate LMEM and other models. Measures from LMEM and those from others were not highly correlated. The finding of the current study is not consistent with the conclusions of Tekwe et al. (2004) in terms of LMEM estimates. We also used two years data for this analysis as Tekwe et al. did. We came up with different results. There is a little difference found between estimates from LMEM and those from SFEM in Tekwe et al. (2004) and Weiss (2006) studies. It should be noted that none of the students in our data changed their school during the years of analysis. Since LMEM measures the effect of multiple schools on student's performance it is not surprising that estimates may be affected by the presence of stable students. LMEM is a strong model that takes multiple years, subjects, and schools into account. We could not see the true effect of LMEM in our study due to two years data and stable students. We highly recommend that further research be done using different data to see the impact of the multivariate method on school effectiveness.

In this study the minority status of a student was defined as Black or non-Black race. There may be some other way to capture minority. Descriptive statistics showed that we can easily see that majority of the population comes from Hispanic ethnicity. We could have also defined the minority as Hispanic or non-Hispanic. We believe that there will be slight differences in estimations when the minority status is based on Hispanic race. Impact of the race on estimates is one of the possible researches.

According to other researchers the vertical scaling process is very challenging itself and introduces more error in longitudinal studies (Doran & Cohen, 2005). It should also be noted that test scores must be vertically scaled in order for VAMs to be useful in identifying effective schools (Ballou et al., 2004). We cannot just compare 6th graders test scores with 7th graders test scores. It is like measuring a child's weight in pounds one year and in kilograms the next year. We need to connect all tests to each other and put them into a common scale. The scaling issue seemed to be a threat to value-added studies (McCaffrey et al., 2003). McCaffrey et al. also concluded that the variance in vertically linked scale scores does not increase by grade and that this causes incompatibility between data and model.

While states have been gradually adopting VAMs in their systems the importance of VAMs in policy decisions is still questionable. Many researchers have claimed that VA methodology is not ready to support high-stakes decisions. Evidence pertaining to the reliability and validity of these models remains controversial. Our conclusion concerning the use of VAMs in accountability systems is that decisions should not be made based solely on empirical studies.

From literature we can see that VAMs were not found to have sufficient precision to be useful for high-stakes decisions (Kupermintz, 2003). Tekwe et al. (2004) also stated that the choice of model should be based on non-empirical issues. According to Tekwe et al. the choice of model should be SFEM in low-stakes decisions which do not sanction schools or teachers. In higher stakes decisions AHLMM would be the choice of model that takes student and school backgrounds into account. We also concluded that until VAMs are proven to be valid in accountability systems they should be used to identify the most and the least effective schools rather than giving sanctions and rewards based on the value-added measures generated from these models.

While this study examined a variety of questions raised by various applications of VAM estimates for school effectiveness, several important issues were not examined. Other important issues that are not examined in this study that can be listed are the effect of missing data, number of years of data, and multiple schools. All the models analyzed in this study used two years' data and they estimated parameters by assuming there was no missing data.

The following suggestions for future research should be taken into account:

Even though we produced value-added measures from these models, we cannot be sure about the precision of these values. We can easily see that small schools have less precise estimates due to the large value of standard error. Further research should be conducted to determine whether these values reflect school effectiveness correctly. Weiss (2006) also suggested that future research in school effectiveness should focus on the precision of estimates for making policy decisions.

Missing data is also problematic in value-added studies as in other research. We have only one model (LMEM) that utilizes the data no matter how sparse it is. Other VAMs do not have the capacity to deal with missing data. Students get sick, miss exams, and change schools. Their absence from exams affects the value-added estimates of schools.

Since value-added assessment is based on each individual student's performance the tracking of students should be another concern in value-added studies. Many state data systems are not capable of tracking students who change schools. States should develop a good system that can track students even when they change their schools. McCaffrey et al. (2003) stated that there is little knowledge about the impact of missing data on school and teacher effectiveness. There is a need for further research on these issues in value-added studies.

REFERENCES

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher, 37*(2), 65–75.

Andrejko, L. (2004). Value-added assessment: A view from a practitioner. *Journal of Educational and Behavioral Statistics*, *29*, 7–9.

Aitkin, M., & Longford, N. (1986). Statistical modeling in school effectiveness studies. *Journal of the Royal Statistical Society, A 149,* 1-43.

Ballou, D., Sanders, W. L., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37–66.

Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, *37*(2), 65–75.

Braun, H. I. (2004, December). *Value-added modeling: What does due diligence require?* Princeton, NJ: Educational Testing Service.

Bryk, A. S., & S.W. Raudenbush. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Callendar, J. (2004). Value-added student assessment. *Journal of Educational and Behavioral Statistics, 29,* 5.

Doran, H., & L.T. Izumi. (2004). *Putting education to the test: A value-added model for California. San Francisco: Pacific Research Institute.*

Doran, H. C., & Cohen, J. (2005). The confounding effect of linking bias on gains estimated from value-added models. In R. Lissitz (Ed.), *Value-added models* in *education: Theory and application.* Maple Grove, MN: JAM Press.

Doran, H. C., & S. Fleischman. 2005. Challenges of value-added assessment. *Educational Leadership*, 63(3): 85–87.

Franco, S. (2006). *The relationships among building level school/non school factors and value-added scores in Ohio.* Unpublished doctoral dissertation, University of Cincinnati, Ohio.

Hanushek, E. A. (1972), *Education and race, Lexington Books: Lexington, MA.*

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value-Added Assessment System. *Educational Evaluation and Policy Analysis, 25*, 287–298.

Ladd, H. F., & Walsh, R. P. (2002), Implementing value-added measures of school effectiveness: Getting the incentives right, *Economics of Education Review*, *21*, 1-17.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *Preface to SAS system for mixed models,* Cary, NC: SAS Institute Inc.

Lockwood, J., Doran, H. C., & McCaffrey, D. F. (2003). Using R for estimating longitudinal student achievement models. *The Newsletter of the R Project*, *3*(3), 17-23.

McCaffrey, D., Lockwood, J.R., Koretz, D, & Hamilton , L. (2003). *Evaluating value-added models for teacher accountability*. Washington, DC: RAND.

Murnane, R. J. (1975). *The impact of school resources on the learning of children.* Cambridge, MA: Ballinger Publishing Co.

Olson, L. (2004, November 16). "Value added" models gain in popularity. *Education Week.* Retrieved January 16, 2006, fromhttp://www.edweek.org/ew/articles /2004/11/17/12value.h24.html

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1-17.

Raudenbush, S., & Bryk, A. (1988-89). Methodological advances in studying effects of schools and classrooms on student learning. In E. Z. Roth (Ed.), *Review of research in education* (pp. 423-475). Washington, DC: American Educational Research Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

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

Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Sanders, W. L., Saxton, A . M., & Horn, S. P. (1997). The Tennessee Value-Added Educational Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162), Thousand Oaks, CA: Corwin Press.

Sanders, W. L., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*, 247–256.

Sanders W. L. (2000). Annual CREATE Jason Millman Memorial Lecture: Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*, 329–339.

Sanders, W. L., Saxton, A., Schneider, J., Dearden, B., Wright, S. P., & Horn, S. (2002). *Effects of building change on indicators of student achievement growth: Tennessee Value-Added Assessment System.* Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L. (2004). *How can value-added assessment lead to greater accountability?* [Biographical sketch with article]. Report presented at First Annual Policy Conference of the New York State Educational Conference Board (Investment and Accountability for Student Success), Albany, NY. Retrieved February 10, 2007, from http://www.nysecb.org/2004conference/04sanders.html

Sanders, W. L. (2006) October, *Comparisons among various educational assessment value-added models.* Paper presented at The Power of Two – National Value-Added Conference, Columbus, Ohio.

Sanders, W. L., & Wright, S. P. (2008, April 14). *A response to Amrein-Beardsley (2008): "Methodological concerns about the Education Value-Added Assessment System."* Available: www.sas.com/govedu/edu/services/Sanders_Wright_response_to_Amrein-Beardsley_4_14_2008.pdf

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York: John Wiley & Sons.

Stewart, B. E. (2006). *Value-added modeling: The challenge of measuring educational outcomes. New York Carnegie Corporation:*

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., et al. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*, 11–36.

Wiley, E. W. (2006). *A practitioner´s guide to value added assessment.* Tempe, AZ: Educational Policy Studies Laboratory, Arizona State University.

Willms, D. J., & S. W. Raudenbush. (1989). A Longitudinal Hierarchical Linear Models for Estimating School Effects and Their Stability. *Journal of Educational Measurement, 26:*209-232.

APPENDICES

A. Sas codes for SFEM model

```
PROC GLM DATA=;

MODEL CHANGE = S1 – S43/SOLUTION;

RUN;
```

B. Sas codes for UHLM model

```
PROC MIXED DATA=;

CLASS STUDENT;

MODEL CHANGE =;

RANDOM INTERCEPT / TYPE = UN SUB = SCHOOL SOLUTION;

REPEATED / TYPE = UN SUB = STUDENT;

RUN;
```

C. Sas codes for AHLM model

```
PROC MIXED DATA=;

CLASS STUDENT MIN POV;

MODEL CHANGE = Z1 Z2 Y1 MIN POV;

RANDOM INTERCEPT/TYPE= UN SUB = SCHOOL SOLUTION;

REPEATED/TYPE = UN SUB = STUDENT;

RUN;
```

D. Sas codes for LMEM model

```
PROC MIXED DATA=;

CLASS STUDENT;

MODEL Y = X11 X21 X12 X22/NOINT SOLUTION;

RANDOM ZM1_1-ZM1_22/TYPE=TOEP(1) SOLUTION;

RANDOM ZM2_1-ZM2_22/TYPE=TOEP(1) SOLUTION;

RANDOM ZR1_1-ZR1_22/TYPE=TOEP(1) SOLUTION;

RANDOM ZR2_1-ZR2_22/TYPE=TOEP(1) SOLUTION;

RUN;
```

E. Grade 6 Math Estimates

Table 8

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 6 Math Results

	Model 1		Model 2		Model 3		Model 4	
Rank	Estimate	School	Estimate	School	Estimate	School	Estimate	School
1	54.126	34	47.539	25	43.963	25	205.400	34
2	50.883	10	45.621	13	22.978	19	156.970	41
3	46.729	41	41.297	19	19.502	6	144.010	10
4	43.380	39	39.370	34	18.071	22	134.640	1
5	32.629	42	29.861	41	17.040	13	134.470	42
6	32.055	14	28.624	6	14.723	32	123.850	13
7	31.476	16	27.787	10	14.085	37	107.250	39
8	25.660	11	23.992	14	11.225	23	107.220	16
9	24.598	13	22.195	11	10.468	31	101.890	14
10	24.186	1	21.982	8	10.446	41	77.441	28
11	24.036	28	21.971	42	10.196	4	69.383	8
12	22.312	6	19.878	12	9.751	27	59.463	26
13	21.740	26	19.528	37	9.287	12	58.542	31
14	19.308	12	17.133	39	9.253	34	47.321	2
15	10.485	8	9.613	29	7.079	11	44.734	11
16	10.254	7	9.203	36	7.000	42	42.055	36
17	9.985	33	8.964	4	6.872	30	37.532	18
18	8.621	2	7.843	28	5.885	8	20.219	7
19	7.804	43	6.678	22	4.198	38	19.363	6
20	6.766	36	5.929	24	3.468	10	14.216	15
21	3.580	30	3.196	26	2.105	29	12.268	12
22	1.377	38	1.181	38	1.489	39	-0.950	17

Table 8 (continued)

	Model 1		Model 2		Model 3		Model 4	
Rank	Estimate	School	Estimate	School	Estimate	School	Estimate	School
23	0.695	18	0.552	5	1.349	36	-2.464	24
24	-4.603	24	-4.034	15	1.149	14	-4.088	30
25	-6.922	19	-6.003	31	1.092	9	-12.702	38
26	-9.650	15	-8.645	17	-1.146	24	-14.890	29
27	-9.718	29	-8.779	32	-3.735	17	-23.957	19
28	-10.151	25	-9.158	23	-4.549	5	-36.867	33
29	-10.366	31	-9.277	2	-4.838	3	-37.009	43
30	-13.212	23	-11.659	7	-5.679	28	-47.359	37
31	-13.489	37	-12.163	40	-6.250	20	-51.556	32
32	-18.218	20	-16.900	1	-6.839	18	-51.726	4
33	-19.810	40	-17.065	27	-7.952	15	-59.865	20
34	-20.228	17	-18.407	9	-9.202	44	-62.298	40
35	-21.194	32	-19.345	20	-9.377	7	-66.704	25
36	-21.681	35	-19.563	30	-10.226	26	-76.567	35
37	-24.274	4	-22.656	16	-12.559	35	-78.855	9
38	-32.380	5	-28.937	3	-14.295	40	-96.330	3
39	-33.237	3	-29.809	21	-15.309	2	-108.140	5
40	-34.008	22	-30.654	18	-17.635	21	-108.880	23
41	-50.386	44	-41.325	44	-23.673	16	-155.640	22
42	-53.935	21	-45.737	43	-26.979	1	-185.050	44
43	-58.680	9	-49.734	33	-39.858	43	-191.170	21
44	-	27	-50.094	35	-42.581	33	-245.150	27

F. Grade 6 Reading Estimates

Table 9

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 6 Reading Results

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
1	60.918	25	44.015	25	-24.438	36	253.15	34
2	48.315	10	35.401	10	-23.498	33	208.32	41
3	36.844	13	27.928	13	-16.870	43	171.45	1
4	27.022	34	21.091	34	-14.180	1	164.81	42
5	24.526	39	19.427	14	-13.321	21	159.03	10
6	23.067	14	18.243	28	-11.845	42	140.86	16
7	22.844	28	18.239	39	-10.927	35	137.51	13
8	22.459	8	17.915	8	-10.611	38	133.62	14
9	17.803	26	13.995	26	-9.648	24	132.23	39
10	15.573	19	12.195	11	-8.487	6	106.69	26
11	15.465	2	11.993	2	-6.880	9	101.77	28
12	15.173	11	11.402	19	-6.373	16	81.64	2
13	12.764	29	10.336	29	-5.099	4	79.26	31
14	11.373	27	7.840	20	-4.015	41	60.57	11
15	10.610	20	7.753	27	-4.011	17	41.18	8
16	10.071	12	7.536	12	-2.407	18	40.38	18
17	8.149	5	6.051	5	-1.944	40	39.01	36
18	6.386	32	4.879	32	-1.771	15	34.05	7
19	5.408	40	4.064	40	-0.895	5	20.47	30
20	3.358	23	2.388	23	-0.093	37	1.13	6
21	3.189	15	2.180	15	0.115	7	-4.30	38
22	2.815	31	1.899	31	0.587	12	-13.66	15

Table 9 (continued)

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
23	1.086	7	0.756	7	2.018	3	-21.016	43
24	0.765	37	0.560	37	2.194	22	-30.766	29
25	-1.526	41	-1.245	41	2.909	19	-33.484	20
26	-6.618	22	-4.619	22	3.048	44	-37.043	33
27	-7.420	16	-6.164	16	3.343	26	-37.656	24
28	-9.302	17	-6.806	17	3.457	29	-53.914	35
29	-12.984	4	-9.036		3.732	2	-58.836	37
30	-13.007	3	-9.660	24	3.736	11	-62.140	17
31	-13.095	24	-9.730	3	4.714	30	-66.159	25
32	-14.541	1	-10.055	4	5.163	14	-67.186	9
33	-16.027	30	-11.841	1	6.266	8	-70.691	19
34	-17.307	6	-12.910	30	6.417	34	-71.413	3
35	-19.570	42	-13.058	6	6.687	28	-72.888	4
36	-22.921	18	-14.798	42	7.251	39	-83.628	40
37	-24.321	43	-18.471	18	8.767	20	-84.608	32
38	-25.337	38	-18.894	43	9.931	31	-85.729	12
39	-26.937	21	-19.247	21	10.142	23	-104.010	23
40	-29.138	9	-19.764	38	11.042	13	-141.980	5
41	-35.153	33	-22.641	9	13.489	27	-197.470	44
42	-44.516	36	-28.362	33	14.161	32	-205.890	22
43	-54.027	35	-34.359	36	15.269	10	-249.450	21
44	-	-	-36.434	35	32.868	25	-253.270	27

G. Grade 7 Math Estimates

Table 10

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 7 Math Results

	Model 1		Model2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
1	57.870	23	49.492	23	33.291	23	162.360	34
2	53.031	5	46.823	5	20.699	37	154.510	10
3	47.795	37	42.700	37	20.364	27	141.950	41
4	40.406	19	35.029	19	14.043	10	125.150	1
5	36.508	33	32.936	33	13.832	44	117.440	42
6	36.151	27	29.464	27	12.164	5	101.850	13
7	30.674	2	27.457	2	9.736	32	99.798	39
8	24.923	29	22.643	29	9.464	33	84.818	16
9	18.898	10	18.669	44	8.441	19	79.106	11
10	18.225	43	16.234	43	7.951	38	78.147	14
11	12.827	40	15.925	10	7.334	2	71.348	36
12	11.716	20	11.116	40	7.116	42	63.428	26
13	11.066	11	10.404	20	6.847	11	49.425	28
14	9.111	25	10.078	11	6.232	29	42.135	7
15	9.068	38	7.978	38	5.996	30	35.743	43
16	7.255	39	7.586	25	5.763	20	35.694	6
17	6.635	36	6.429	39	3.821	25	30.305	8
18	5.726	32	5.782	36	3.697	36	29.958	18
19	1.864	12	5.089	32	2.904	13	26.491	33
20	1.816	6	1.627	12	1.383	6	22.782	15
21	0.091	42	1.547	6	1.150	31	21.766	2
22	-2.547	13	0.057	42	0.852	43	17.688	12

Table 10 (continued)

	Model 1		Model2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
23	-3.870	8	-2.275	13	0.743	9	9.613	19
24	-6.633	24	-3.537	8	0.344	39	-8.907	30
25	-8.816	9	-6.059	24	-0.773	41	-12.671	37
26	-9.798	14	-7.605	9	-1.741	18	-12.828	24
27	-12.150	26	-9.088	14	-2.536	16	-13.224	38
28	-15.031	4	-10.963	26	-2.886	14	-21.744	31
29	-15.451	18	-12.757	4	-6.716	3	-30.363	29
30	-16.327	30	-14.055	18	-6.718	34	-41.547	20
31	-16.867	21	-14.411	21	-7.294	8	-42.572	23
32	-16.981	28	-14.834	30	-7.469	40	-50.790	32
33	-19.894	16	-15.253	28	-8.690	12	-53.956	17
34	-20.113	31	-16.626	31	-9.894	26	-68.064	35
35	-24.199	15	-18.255	16	-10.866	35	-69.483	5
36	-24.806	7	-20.913	15	-11.895	28	-71.777	40
37	-27.363	41	-21.167	7	-12.107	1	-86.708	25
38	-28.194	3	-24.219	3	-14.069	4	-98.231	3
39	-29.706	22	-24.726	41	-14.497	22	-137.380	4
40	-33.060	17	-25.507	22	-14.731	24	-138.720	27
41	-38.604	34	-28.340	17	-15.526	21	-146.000	21
42	-39.861	1	-34.838	34	-15.698	7	-147.020	44
43	-53.587	35	-35.802	1	-18.492	15	-169.440	9
44	-	-	-43.844	35	-21.575	17	-180.100	22

H. Grade 7 Reading Estimates

Table 11

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 7 Reading Results

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
1	48.914	5	35.798	5	7.404	32	237.380	34
2	35.783	33	27.574	33	7.362	30	181.120	41
3	32.365	40	22.794	40	6.361	27	155.640	1
4	29.638	19	20.832	19	6.111	42	146.380	10
5	25.913	2	19.622	2	4.798	10	144.080	42
6	23.872	24	18.772	24	4.345	3	122.240	39
7	23.095	6	16.150	12	4.282	6	117.020	14
8	21.937	12	16.096	6	4.072	5	115.070	13
9	21.214	39	15.800	39	3.317	25	113.030	16
10	21.087	42	14.523	42	3.183	31	91.709	26
11	18.842	27	11.599	27	2.844	33	90.099	11
12	16.881	10	11.181	10	2.102	44	72.450	2
13	14.798	32	11.134	32	1.983	36	66.967	36
14	14.434	25	9.352	25	1.876	23	66.594	6
15	12.230	36	8.793	36	1.696	34	59.747	7
16	11.797	21	8.128	15	1.272	39	52.893	8
17	11.643	15	8.021	21	1.127	2	40.536	28
18	6.524	7	4.450	7	1.115	13	29.844	18
19	3.875	8	3.080	8	0.931	1	21.275	43
20	1.583	3	1.135	3	0.460	40	17.925	30
21	-1.223	43	-0.857	43	0.298	21	11.347	31
22	-2.885	11	-2.217	11	0.194	35	5.246	15

Table 11 (continued)

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
23	-4.239	26	-3.181	26	-0.013	19	-11.825	38
24	-4.913	29	-3.785	29	-0.444	22	-27.172	33
25	-5.868	30	-4.436	13	-0.517	24	-31.892	35
26	-6.130	13	-4.506	30	-0.656	14	-39.973	24
27	-8.157	31	-5.083	31	-1.023	12	-49.667	40
28	-8.449	14	-6.100	23	-1.114	11	-50.021	17
29	-8.564	28	-6.158	44	-1.426	38	-52.630	29
30	-9.049	23	-6.426	28	-2.102	15	-61.211	19
31	-12.412	1	-6.832	14	-2.275	7	-62.315	20
32	-12.988	20	-9.195	22	-2.371	8	-63.417	37
33	-13.521	22	-9.356	1	-2.404	9	-64.037	12
34	-18.339	38	-9.596	20	-2.487	20	-71.004	25
35	-18.648	37	-13.306	38	-2.811	16	-81.610	32
36	-19.994	18	-13.950	37	-3.258	18	-82.632	23
37	-21.213	34	-15.535	18	-4.081	26	-89.699	3
38	-25.766	17	-16.213	34	-4.141	29	-93.799	5
39	-27.813	16	-17.510	17	-4.318	28	-138.870	27
40	-29.875	4	-19.828	4	-5.623	37	-143.730	4
41	-37.379	9	-22.111	16	-5.817	43	-171.610	22
42	-37.844	35	-23.288	35	-5.945	41	-185.390	21
43	-51.707	41	-25.664	9	-7.016	4	-192.060	9
44	-	-	-39.701	41	-7.291	17	-194.030	44

I. Grade 8 Math Estimates

Table 12

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 8 Math Results

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
1	53.349	5	45.623	5	23.913	23	165.800	34
2	36.285	25	30.461	23	19.343	10	147.420	10
3	35.378	23	28.975	25	15.672	30	122.290	41
4	27.223	2	24.135	2	15.548	25	99.598	39
5	26.707	10	21.658	10	15.064	32	98.591	42
6	23.448	9	20.330	29	10.501	9	83.880	14
7	22.629	29	19.632	9	9.625	31	77.750	16
8	21.139	40	17.851	40	9.545	5	76.634	11
9	19.244	32	17.050	32	8.943	38	65.474	13
10	18.413	22	15.191	22	8.518	3	64.824	1
11	13.297	3	11.220	3	8.455	2	61.724	28
12	10.692	12	9.507	12	6.574	41	57.292	6
13	10.421	38	9.053	38	5.924	22	54.149	26
14	8.449	8	7.710	8	4.938	6	50.096	12
15	6.316	31	7.149	44	3.405	29	45.129	8
16	3.241	4	4.944	31	3.319	44	36.135	7
17	1.820	28	2.810	4	3.198	39	33.996	33
18	1.641	30	1.724	28	2.041	12	33.358	2
19	1.251	20	1.611	30	1.594	8	20.737	43
20	1.146	39	1.220	20	1.524	18	16.358	36
21	0.618	6	1.122	39	1.386	28	9.637	30
22	-0.244	19	0.655	6	-0.848	20	7.733	38

Table 12 (continued)

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
23	-2.572	36	-0.062	19	-1.972	36	-6.354	18
24	-2.980	21	-1.985	36	-2.074	13	-6.713	24
25	-4.100	37	-2.354	21	-2.314	16	-8.668	19
26	-4.842	43	-3.416	37	-3.194	34	-18.489	15
27	-4.875	7	-3.988	7	-3.266	4	-22.750	29
28	-5.745	13	-4.030	43	-3.385	11	-27.821	25
29	-6.468	33	-4.920	13	-3.438	40	-29.307	31
30	-9.813	15	-5.582	33	-3.938	42	-31.361	23
31	-13.777	18	-7.889	15	-4.122	7	-46.631	37
32	-14.160	24	-12.102	18	-4.435	35	-47.068	20
33	-14.372	41	-12.408	24	-4.895	14	-50.914	40
34	-15.427	11	-12.552	41	-7.882	33	-53.171	17
35	-15.474	14	-13.613	11	-8.420	43	-59.493	32
36	-18.354	27	-13.791	14	-8.524	37	-70.385	35
37	-19.509	17	-14.170	27	-9.398	19	-74.200	4
38	-20.792	16	-15.865	17	-11.435	27	-77.661	5
39	-27.466	35	-18.533	16	-11.685	15	-105.450	3
40	-30.793	42	-22.086	35	-14.047	21	-110.710	22
41	-37.675	34	-25.057	42	-15.835	24	-123.010	44
42	-38.433	26	-33.331	26	-16.464	17	-128.990	21
43	-43.317	1	-33.504	34	-17.581	1	-145.400	9
44	-	-	-38.404	1	-19.881	26	-184.070	27

J. Grade 8 Reading Estimates

Table 13

Estimates of the School Effects Obtained from Models 1, 2, 3, and 4 Based on Grade 8 Reading Results

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
1	65.271	22	55.060	22	30.828	31	213.600	34
2	55.096	32	49.532	32	30.666	32	149.000	41
3	53.641	31	42.580	31	28.159	30	119.120	14
4	42.242	33	39.095	44	23.278	22	118.000	10
5	31.507	2	38.358	33	19.037	18	113.930	1
6	30.905	3	28.528	2	17.309	3	110.580	42
7	21.270	9	26.651	3	17.157	44	94.473	16
8	19.985	18	18.392	9	14.916	2	92.933	39
9	19.846	30	18.353	18	14.157	33	86.963	26
10	19.651	7	18.216	30	13.495	35	70.105	13
11	16.306	29	17.315	7	9.190	39	68.903	2
12	13.512	15	15.060	29	7.755	10	68.546	11
13	12.609	39	11.622	15	7.516	16	63.633	6
14	12.048	24	11.250	39	7.082	9	62.279	28
15	11.755	35	11.152	24	6.479	42	56.663	8
16	10.649	5	10.047	35	3.202	7	43.366	7
17	5.351	27	9.554	5	3.190	28	33.549	30
18	2.886	28	4.565	27	0.293	29	30.192	43
19	0.098	4	2.832	28	-0.216	6	22.307	33
20	-0.130	12	0.354	4	-0.455	15	6.969	18
21	-0.199	8	0.169	12	-0.779	8	-1.367	38
22	-1.267	10	0.110	8	-1.470	23	-1.593	15

Table 13 (continued)

	Model 1		Model 2		Model 3		Model 4	
School	Estimate	Rank	Estimate	Rank	Estimate	Rank	Estimate	Rank
23	-3.905	20	-0.784	10	-1.654	41	-4.705	31
24	-6.663	36	-3.137	20	-2.709	38	-5.265	36
25	-7.420	19	-5.378	36	-2.980	14	-11.922	19
26	-7.697	40	-6.204	19	-5.573	36	-20.163	29
27	-10.004	23	-6.326	40	-5.735	27	-21.740	24
28	-10.629	16	-8.471	23	-5.794	24	-29.956	12
29	-13.596	38	-9.400	16	-6.159	13	-43.634	40
30	-13.953	17	-11.481	17	-6.940	20	-50.390	37
31	-14.403	6	-11.588	38	-7.605	34	-53.728	17
32	-15.412	37	-11.977	6	-8.682	4	-60.683	32
33	-15.430	42	-12.695	42	-11.891	25	-60.930	35
34	-15.944	21	-13.374	37	-12.783	12	-68.184	23
35	-17.490	25	-13.409	21	-14.226	1	-70.552	25
36	-20.386	13	-14.061	25	-14.246	19	-76.398	20
37	-23.153	43	-18.015	13	-14.866	37	-84.613	3
38	-25.715	14	-20.100	43	-15.325	5	-93.752	4
39	-44.643	11	-23.215	14	-15.518	17	-103.860	5
40	-45.980	41	-40.170	11	-16.403	43	-123.500	22
41	-52.446	26	-41.044	41	-17.013	11	-129.530	44
42	-54.671	1	-46.266	26	-17.054	40	-149.700	9
43	-69.274	34	-49.154	1	-21.451	26	-161.510	21
44	-	-	-62.558	34	-26.186	21	-197.460	27