A COMPARISON OF VALUE-ADDED ACCOUNTABILITY SYSTEMS WITH ACCOUNTABILITY SYSTEMS THAT USE THE SUCCESS OF ECONOMICALLY DISADVANTAGED STUDENTS AS A KEY ACCOUNTABILITY INDICATOR

A Dissertation

by

KEVIN LYNN BARLOW

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Bruce Thompson
Committee Members,	Mary Margaret Capraro
	Oi-Man Kwok
	Victor L. Willson
Head of Department,	Victor L. Willson

December 2014

Major Subject: Educational Psychology

ABSTRACT

An effective and acceptable accountability system for education continues to puzzle educators and researchers. The focus of the present study was on two high profile accountability system models: status-based models and value-added models. Status-based models are those models that only considered the status (pass/fail) of students on achievement tests. Value-added models are those that make an attempt to show the achievement gains that students are making and therefore attribute those gains to the students' educational experiences (e.g., teachers and schools). In particular, the present study investigated how the two models might rank campuses differently when accounting for the low SES indicators of both students and schools.

The present study investigated three independent, but connected studies of these two models. First, a comparison of value-added models is presented. Second, a Monte Carlo study is presented comparing the rankings of a status-based accountability system model and a value-added accountability system model. Third, the results from a field study where data from a Texas school district was used to compare the rankings of a status-based accountability system model and a value-added accountability system model are presented. In the latter two studies, evidence is presented showing that the two models ranked campuses differently within each district. This was especially evident when a district had a wide distribution of the campus percentages of low SES students throughout the district. The effect low SES students had on accountability is noticeable in the ranking of the campuses. Campuses with a high percentage of low SES students performed lower under the status-based accountability system model than under the value-added accountability system model.

Identifying the correct accountability system model to use in evaluating our schools is extremely important in this age of accountability. Teachers, principals, and superintendents are evaluated on the performance of their students, schools, and districts, and in some cases pay incentives are also attached to these evaluations. The present study shows the need for more research in accountability systems in order to ensure that these evaluations are fair.

DEDICATION

This document is dedicated to my parents, Alva Jean Barlow and Ramona Madalene Barlow. Although my father passed away several years ago, my mother passed away in the midst of me writing this document. It would have meant so much to me if she would have been able to see the finished product. I thank both of them for all of their support.

ACKNOWLEDGEMENTS

I am so grateful to have had the opportunity to study at Texas A&M University. The faculty in the Educational Psychology Department are of the utmost quality – they want to see students succeed. I want to especially thank my committee members: Dr. Bruce Thompson, Dr. Mary Margaret Capraro, Dr. Oi-Man Kwok, and Dr. Victor L. Willson. Specifically, I want to thank my committee chair, Dr. Bruce Thompson. It's amazing to me how everything he asks you to do has its place in the program. He has truly been an inspiration to me.

Most importantly, I want to express my appreciation to my wife, Glenna. She has been so supportive through this entire process. I was away from home for long hours at class and spent many more hours writing the dissertation, but she was always there to help. Thanks for all you do, Glenna.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

LIST OF TABLES

Page

Page

Page

CHAPTER I

INTRODUCTION

Accountability in education is continually in the news and is a major topic of discussion among legislators, policy makers, local school board members, school administrators, and the public in general (Nuttall, Goldstein, Prosser, & Rasbash, 1989; TEA, 2014; USDE, 2002). The present study compared the rankings of campuses using value-added accountability systems with the rankings of campuses using accountability systems that consider the success of economically disadvantaged students (i.e., students of low socio-economic status or low SES) as a key accountability indicator. The current (2004 to 2011) accountability system in Texas is one such accountability system that uses the success of economically disadvantaged students as a key accountability indicator. If a campus or district does not achieve the accountability rating it desires, the reason is typically because of this indicator (e.g., the economically disadvantaged students in science did not perform well). The current accountability system in Texas does not do an adequate job of accounting for the effects of economically disadvantaged students on the performance of schools. Schools with a high percentage of economically disadvantaged students have a higher probability of receiving a low accountability rating than those schools that have a low percentage of economically disadvantaged students. According to Ladd and Walsh (2002), schools with a large percentage "of disadvantaged students and that do not have sufficient compensatory resources to offset the educational challenges that such students pose… may be deemed ineffective despite using their insufficient resources more productively and efficiently than other schools" (p. 16).

With the rise of accountability demands in education, school districts in general, and Texas school districts specifically, struggle to determine ways to measure student improvement and the effects that teachers have on that student improvement. Providing an accountability system that takes into account the individual differences of students and the effects that teachers have on the achievement of those individual students is vital. The challenge before educational researchers is determining which system best achieves this. Currently, only a few school districts in Texas have attempted to construct a value-added system to account for the effects that teachers have on student achievement in parallel to (or in lieu of) the Texas accountability system. Evidence of this is shown in Table 1 focusing on the 41 districts in Texas with at least 30,000 students enrolled in the 2012-2013 school year. These districts were selected to be surveyed because they are the districts most likely to have the available resources to construct such a system.

Accountability in education has been discussed for several years in the United States (TEA, 2014; USDE, 2002). Texas entered the era of accountability when "the Texas Legislature in 1993 enacted statutes that mandated the creation of the Texas public school accountability system to rate school districts and evaluate campuses" (TEA, 2014, p. 3). This accountability system was possible, in part, because of the assessment and data infrastructure already in place in Texas which began with the Texas Assessment of Basic Skills (TABS) exam in 1979. The accountability system of 1993 was modified in 2003 and again in 2013.

Table 1

Survey of Districts (at least 30,000 Students Enrolled in 2013-2014) on Value-Added Models

District	Student Enrollment	Value-added Model	Associated With Teacher Pay	Associated With Teacher Evaluations
Aldine ISD	65,415	Yes	Yes	Yes
Alief ISD	45,748	Yes	No	No
Amarillo ISD	33,130	Yes	No	No
Arlington ISD	64,913	NA	NA	NA
Austin ISD	86,233	Yes	Yes	Yes
Brownsville ISD	49,129	Yes	No	No
Clear Creek ISD	39,480	No	No	No
Conroe ISD	53,632	Yes	No	No
Corpus Christi ISD	39,142	NA	NA	NA
Cypress-Fairbanks ISD	109,733	NA	NA	NA
Dallas ISD	158,680	Yes	Yes	Yes
Edinburg CISD	33,582	No	No	No
El Paso ISD	62,884	NA	NA	NA
Fort Bend ISD	69,123	No	No	No
Fort Worth ISD	83,255	NA	NA	NA
Frisco ISD	42,530	NA	NA	NA
Garland ISD	57,914	NA	NA	NA
Houston ISD	202,586	Yes	Yes	Yes
Humble ISD	36,867	No	No	No
Irving ISD	34,961	Yes	No	No
Katy ISD	64,408	NA	NA	NA
Keller ISD	33,254	NA	NA	NA
Killeen ISD	41,687	No	No	No
Klein ISD	46,778	Yes	No	No
Leander ISD	34,265	No	No	No
Lewisville ISD	52,399	No	No	No
Mansfield ISD	32,831	No	No	No
Mesquite ISD	39,028	No	No	No
North East ISD	67,701	Yes	No	No
Northside ISD	99,426	Yes	No	No
Pasadena ISD	53,449	Yes	No	No
Pharr-San Juan-Alamo	32,025	NA	NA	NA
Plano ISD	54,921	Yes	No	No
Richardson ISD	37,954	NA	NA	NA
Round Rock ISD	45,588	Yes	Yes	Yes
San Antonio ISD	54,236	Yes	No	No
Socorro ISD	44,054	NA	NA	NA
Spring Branch ISD	34,778	Yes	No	No
Spring ISD	36,028	No	No	No
United ISD	42,753	NA	NA	NA
Ysleta ISD	43,512	No	No	No

Eisele-Dyrli (2010) presented a historical perspective of the link between the assessment systems and accountability systems in Texas, summarized in the abbreviated timeline below:

> 1979: State senate bill requires students to take the TABS [Texas Assessment of Basic Skills] exam, which takes all student groups into account and highlights achievement gaps.1990: TAAS [Texas Assessment of Academic Skills], a more rigorous state exam, is introduced and remains in place for 12 years.
>
> 2001: State legislature toughens accountability with TAKS [Texas Assessment of Knowledge and Skills], a more rigorous state exam.
>
> 2002: President and former Texas governor George W. Bush signs NCLB [No Child Left Behind] into law. (pp. 35-36)

Intertwined with accountability in education is the performance of economically disadvantaged students on standardized tests. Research shows that economically disadvantaged students perform lower on standardized tests (Caldas, 1993; Darandari, 2004; Kennedy & Mandeville, 2000; Willms, 1992). Teachman (1987) found that “Family background has been prominent in models of educational attainment. In most research, family background has been measured by socioeconomic indicators (e.g., parents' education and family income), to the exclusion of other family characteristics that also affect educational attainment” (p. 548). This relationship has led state (Texas is a good example) and national governments to develop accountability systems with the performance of economically disadvantaged students as one of the key indicators.

However, there may be an alternative explanation for this relationship. Marks (2006) explained that student achievement may not be correlated to the economic status of the students, but rather to the schools where economically disadvantaged students find themselves enrolled. Parents with higher socio-economic status have the means to place their children in better schools than do poorer parents with similar children.

The concept of economically disadvantaged students performing lower on standardized tests warrants an accountability system that accounts for this predisposition to low performance of these students. In other words, a teacher with only economically disadvantaged students will most likely have lower tests scores than a teacher with no economically disadvantaged students, regardless of the quality of instruction, thus setting up a system of inequality with teacher accountability. Research has shown a need for a system that holds teachers accountable for the individual students they have in class; however, evaluating teachers using student performance has been quite challenging for administrators, researchers, and policymakers (Kupermintz, 2003; Millman & Schalock 1997; Shinkfield & Stuffelbeam, 1995).

Regardless of what accountability system is developed and used, there needs to be a way to validate that the scores the system is yielding have some meaning with respect to what is expected—in this case a measure of student achievement. Crocker and Algina (1986) explained that criterion-related validation is “a study of the relationship between test scores and a practical performance criterion” (p. 238) and that finding a suitable criterion for such studies is sometimes difficult. Hill, Kapitula, and Umland (2011) added that studies into the reliability of value-added scores have shown low reliability while similar studies of the validity of value-added scores have shown mixed results. The researchers explained that this could be the result of spurious correlations due to other student characteristics not included in the model.

Validity of scores produced by any instrument in its purest form would indicate that some accepted instrument had produced scores that measured a desired construct and that the validity study showed comparable scores from some new instrument to the scores from the accepted instrument. The present study compared ranks of campuses derived from value-added accountability systems with ranks of campuses derived from accountability systems that use the success of economically disadvantaged students as a key accountability indicator. Convergence would have indicated the two sets of scores were generating similar results; however, divergence would not have necessarily indicated which method was producing more valid campus ranks with respect to student achievement. Thus, the present study compares two types of accountability systems (value-added accountability systems and accountability systems that use the success of economically disadvantaged students as a key accountability indicator) rather than a validity study of value-added accountability systems.

Using teacher-observation data would be one criterion for a validity study of value-added accountability systems; however, finding a common instrument across campuses in which the scores have been examined for inter-rater reliability would have been challenging. The Texas Teacher Appraisal System (TTAS) was implemented in Texas schools during the 1986-87 school year (Setliff, 1989); however, scores from this instrument are not examined for inter-rater reliability in practice and scores show very little variability. In addition, the instrument has been revised over the years and school districts may use a local instrument instead of the State adopted instrument for teacher appraisals. Bertrand and Leclerc (1985) found “it hard to believe that there were so many unreliable variables” in their study of teacher observation data, even though other studies had found results similar to theirs (p. 197). For the purposes of the present study, I chose to use the success of economically disadvantaged students as an indicator to rank schools rather than teacher-observation data. Time and resources permitting, an in-depth study using such teacher observation data would be warranted. According to Peterson, Micceri, and Smith (1985), "The main ingredients required for validation studies are valid and reliable measures of both teacher performance and student outcomes" (p. 76).

Economically Disadvantaged Student Performance

The economically disadvantaged status of students has a direct effect on student academic performance. The poorer a student's family is, the lower the student will tend to perform on achievement tests, and vice versa. Aikens and Barbarin (2008) explained that "the disadvantages that low-SES children face across contexts are great" (p. 250). This is especially true in reading and mathematics. Roberts and Bryant (2011) found the performance of students on mathematics achievement tests to be highly correlated with the SES of students, and Aikens and Barbarin (2008) found similar results for student performance on reading achievement tests.

The effect of SES on student academic performance is sometimes confused with the effects of the English Language Learner (ELL) status on student academic performance. However, English Language Learners have similar correlations between SES and student academic performance. The SES status of students is the stronger predictor of the two outcomes. Krashen and Brown (2005) found in their study of ELL students that "the higher performing ELLs were of high SES" (p. 192). The disadvantages of low SES students are evident during the early years of a child's educational experience. According to Stipek and Ryan (1997), "disadvantaged children are every bit as eager to learn as their more economically advantaged peers. They do, however, have much further to go in terms of their intellectual skills" (p. 722). Willms (2000) added that geographical areas "with high average literacy skills tend to be those with smaller inequalities between advantaged and disadvantaged backgrounds" (p. 246). Caldas (1993) conducted a study in Louisiana that examined several input factors and their relationship to student achievement and found socio-economic status to be a very strong predictor of student achievement at the school level. In 1997, Caldas and Bankston developed multiple models in an attempt to isolate the effect of SES on achievement at the student level and found a correlation of -0.045, or accounted for a variance of 0.203%. Wang (2006) attributed 17.0% of the variance of student achievement scores to the campus level SES effect. Thus the effect of SES on student scores is observed at both the student level and the campus level. This phenomenon was the focus of the present study. Traditional status-based accountability systems typically do not account for the correlation between SES and student achievement scores, but rather use a low SES group score with the expectation that districts/campuses will be held accountable for the group's performance as a whole, even though research shows that these students will perform lower than their high SES counterparts. Districts/campuses with a higher percentage of low SES students will have lower scores than districts/campuses with a lower percentage of low SES students.

To show the effect of economically disadvantaged students on accountability ratings in Texas empirically, the following analyses were conducted using data from the Texas Education Agency (2011) website. Statewide campus accountability data for 8,075 campuses from 2011 were analyzed under the 25 "traditional" indicators used in Texas Accountability Systems (i.e., the 5 subjects × 5 student groups listed below). Ratings were adjusted using the absolute standards.

Reading/ELA (All Students, African American, Hispanic, White, Econ. Disadv.)

Writing (All Students, African American, Hispanic, White, Econ. Disadv.)

Social Studies (All Students, African American, Hispanic, White, Econ. Disadv.)

Mathematics (All Students, African American, Hispanic, White, Econ. Disadv.)

Science (All Students, African American, Hispanic, White, Econ. Disadv.)

First, there were 1,501 campuses for which the campus missed exactly 1 of these 25 indicators and therefore were not moved up to the next campus rating level. For example, the campus met 24 indicators at the Recognized level and 1 indicator at the Acceptable level and therefore received an overall rating of Acceptable. Out of these 1,501 campuses, the indicator missed for 45% of the campuses was 1 of the 5 economically disadvantaged indicators. Thus 20% of the indicators (5 of the 25 indicators) accounted for 45% of the campuses missing a campus rating level by exactly one indicator.

Second, these 1,501 campuses included campuses that did not meet the minimum size requirements for all 5 of the economically disadvantaged groups. Excluding these campuses left 235 campuses for which the campus missed exactly 1 of these 25 indicators and therefore were not moved up to the next campus rating level. Out of these 235 campuses, the indicator missed for 51% of the campuses was 1 of the 5 economically disadvantaged indicators. Thus 20% of the indicators (5 of the 25 indicators) accounted for 51% of the campuses missing a campus rating level by exactly one indicator.

Third, there were 6,261 campuses for which the campus missed at least 1 of these 25 indicators (regardless of size requirements). Out of these 6,261 campuses, the 5 economically disadvantaged indicators accounted for 36% of the missed indicators.

From these empirical data one can see that the economically disadvantaged indicators disproportionally affect campus ratings with respect to all 25 indicators. The current accountability system, with its reliance on binary indicators, appears to inadequately characterize the performance of schools. A value-added accountability system might better account for this disparity between the two groups (high socio-economic students and low socio-economic students) and statistically level the playing field between them.

Research Questions

My primary research questions in the present study were as follows:

1. How do prominent value-added accountability systems compare?
2. Using simulated data, how do the campus rankings generated using value-added accountability systems compare to the campus rankings generated by accountability systems that use the success of economically disadvantaged students as a key accountability indicator (SES accountability systems)?
3. Using field data, how do the campus rankings generated using value-added accountability systems compare to the campus rankings generated by accountability systems that use the success of economically disadvantaged students as a key accountability indicator (SES accountability systems)?

Organization of the Document

This dissertation is divided into five distinct but related chapters. Excluding the first and last chapters, the three remaining chapters are ready for submission for publication in peer-reviewed journals. Below is a description of each chapter:

- Chapter I presents a brief introduction to value-added models and the way they relate to the current accountability environment as well as a foundation for the three embedded individual, but related, studies.
- Chapter II presents a literature review and a comparison of prominent value-added accountability systems. This second chapter represents the first journal article.
- Chapter III presents findings from a Monte Carlo simulation comparing the rankings of campuses generated using value-added accountability systems compared to the rankings of campuses generated by accountability systems that use the success of economically disadvantaged students as a key accountability indicator (SES accountability systems). This third chapter represents the second journal article.
- Chapter IV presents findings from a field study comparing the campus rankings generated using value-added accountability systems to the rankings of campuses generated by accountability systems that use the success of economically disadvantaged students as a key accountability indicator (SES accountability systems). This fourth chapter represents the third journal article.
- Chapter V discusses the three individual but related studies.

CHAPTER II

LITERATURE REVIEW: COMPARISON OF PROMINENT VALUE-ADDED ACCOUNTABLITY SYSTEMS

Although most researchers tend to define value-added models (VAM) as models that account for the effect teachers (or collective units of teachers such as schools or campuses) have on student achievement, they disagree on what variables are needed in the model to measure this effect. Thomas (1998) defined VAMs in simple terms as models that "show whether some schools are performing markedly better or worse than other schools having taken into account intake factors" (p. 92). What researchers disagree on, however, is how to account for those intake factors, which generally include student background variables (e.g., student socio-economic status) and student prior achievement scores. Goldstein and Spiegelhalter (1996) wrote that this "is a very important debate over the best choice of indicator measures and their validity as measures of effectiveness" (p. 386). Moreover, researchers tend to look for the most parsimonious model that explains an acceptable level of variance. This is especially true with value-added models because the majority of users will be teachers, principals, parents, and policy makers who are often most comfortable with simpler models. Tekwe et al. (2004) stated that, "it is generally accepted among experts that value-added systems are desirable. The theoretically preferred methods, however, are quite complex and produce value-added measures that are not readily understandable" (p. 25).

In order to maximize the benefit of value-added models with respect to school improvement, these stakeholders must buy into the value-added process (Olson, 1998). Lenkeit (2013) added that "the judgement of teachers' and schools' work seems to be increasingly principled by political rather than pedagogical convictions" (p. 41). These political dimensions add to the complexity of adopting value-added models.

Other definitions of value-added models include those models that compare "outcomes after adjusting for varying intake achievement and reflect the relative boost a school gives to a pupil's previous level of attainment in comparison to similar pupils in other schools" (Thomas, 1998, p. 94); models that "adequately adjust effect measures for intake differences between schools (e.g., in previous achievement and SES [socio-economic status] of pupils)" (Teddlie, Reynolds, & Sammons, 2000, p. 64); models that estimate achievement by controlling for background variables only (Stevens, 2005); "any method of assessment that adjusts for a valid measure of incoming knowledge or ability" (Tekwe et al., 2004, p. 13); models that evaluate whether the school has enhanced the performance of its students between their entry and exit (Goldstein & Spiegelhalter, 1996); "methods of analyzing gains, growth in scores, or the amount of knowledge added from year to year as students progress through school" (Amrein-Beardsley, 2008, p. 65); and "the effect on students' outcomes associated with attendance at a particular school, net of the effects associated with student family background and wider social and economic factors that lie outside the control of teachers or school administrators" (Willms, 2000, p. 240).

Status-Based Accountability Systems

Status-based accountability systems are those systems that rate districts/campuses primarily on the passing status of a group (or groups) of students (e.g., the percentage of all students who passed the Mathematics portion of the State of Texas Assessments of Academic Readiness). Romer (1997) stated that "content standards are a compilation of specific statements of what students should know or be able to do" (p. 8). Status-based accountability systems typically measure those standards at the end of the school year with no accounting of what students knew at the beginning of the school year. According to Linn (2008), “the typical status approach sets annual achievement targets and compares the status of current student achievement to those targets” (p. 700). Rowan, Correnti, and Miller (2002) showed that status-based accountability systems measure the accumulation of student achievement over years whereas teachers typically have students for only one year; they contend that “most analysts don’t want to analyze teacher effects on achievement status, preferring instead to examine teacher effects on students’ academic growth” (p. 1536). Green (2010) agreed, stating that “estimating the change in a teacher’s effect on student achievement” is preferred over a status measure (p. 39).

According to Tekwe et al. (2004), status-based accountability systems have been around for over 35 years. The earlier methods used only students’ scores from one year to estimate the effects a school had on student performance (Coleman, Campbell, & Kilgore, 1982). Tekwe et al. (2004) added that “the distinguishing characteristic of status-based methods is the absence of adjustment for students’ incoming knowledge level” (p. 12). Willms (2000) found that “state and district monitoring systems usually report school mean test scores, or the percentage of students achieving some criteria, without regard to the characteristics of students entering the school” (p. 240). Wyse, Zeng, and Martineau (2011) added that status models present many challenges when student characteristics are not considered.

Researchers adept in accountability systems tend to agree that these systems must account for intake variables and possibly context variables. Ladd and Walsh (2002) stated “we emphasize that, as a measure of school effectiveness, gains in student performance are far superior to the alternative of relying on the average level of student achievement" (p. 16). Value-added systems estimate growth while accounting for student and/or school variables; status-based accountability systems are driven primarily by the percentage of students who pass a particular achievement test. Typically, low SES students perform lower than high SES students on achievement tests (Aikens & Barbarin, 2008; Roberts & Bryant, 2011); thus, in many cases the variable that has the most impact on status-based accountability systems is the SES student variable.

Purpose of Value-Added Accountability Systems

For many years, schools have been compared to each other using the proportion of students who attain some level on a standardized test (i.e., status scores); however, this is not a fair comparison because it does not take into account any student characteristics, such as socio-economic status or prior student achievement. A much fairer process to determine the effect that schools have on student achievement is to focus on gains in student achievement scores (value-added) rather than the absolute score itself, which is the underlying purpose of using value-added accountability systems (Olson, 1998; Stevens, 2005; Tekwe et al., 2004; Thomas, Nuttall, & Goldstein, 1992). These systems provide a fairer comparison between teachers and between schools. Thomas (1998) endorsed the argument of Desmond Nuttall that schools should be held accountable for only what they can control or influence, not for student differences. Most teachers and principals recognize that the academic level of students as they enter a school affects their academic level as they leave the school. Students who come in at a high academic level will leave at a high academic level and vice versa; thus schools with students already achieving at higher levels will do better on achievement tests when using only the absolute score of the achievement test. However, there may be schools that have a greater effect on student achievement gains even though these improvements don't stand out with respect to a status score. Tekwe et al. (2004) stated that "it is widely accepted among educators and researchers that value-added assessment of school performance is better than an assessment based on status-scores alone" (p. 29). Thomas (1998) identified the following points as the rationale (or purpose) for using a value-added accountability system.

- It presents test results of teachers and schools in a more fair and meaningful way.
- It provides detailed and aggregate data that can be used by stakeholders.
- It provides a system to evaluate school improvement projects by using value-added trend data.
- It provides academic measures that can be compared to non-academic measures such as parent perception surveys.
- It provides monitoring data for individual students and student groups.

Moreover, Thomas (1998) also identified the needs of academic research, local education agencies (LEA), and individual schools as the driving forces behind the value-added movement.

Although value-added models attempt to measure achievement gains students make from one year to the next in consideration of numerous variables that affect student achievement, it is still unclear as to whether these models are more beneficial for parents in deciding which schools to send their children to or for school leaders in deciding which educational practices to implement. Raudenbush (2004) explained the difference between these two viewpoints and pointed out that value-added models may be more beneficial to parents than to school leaders.

Value-Added Models vs. Traditional League Tables

League tables have been used extensively to show school rankings, typically using some form of absolute achievement score (either adjusted or unadjusted). One major disadvantage of using league tables in this way is that they do not typically account for student differences or intake variables. According to Thomas (1998), “schools with high achieving intakes will tend to do well for that reason alone. Neither the initially high achieving nor the initially low achieving school is assisted by the publication of raw league tables” (p. 92). Even though some accountability systems have tried to use adjusted league tables, Goldstein and Spiegelhalter (1996) argued that adjusted league tables have many of the same inadequacies as unadjusted ones. Using league tables based on valued-added scores may be more meaningful than league tables based on raw achievement scores.

Thomas, Nuttall, and Goldstein (1992) illustrated the differences between using league tables based on value-added scores and league tables based on raw achievement scores. By using value-added scores, they re-ranked 402 schools whose rankings using raw achievement scores were published in *The Guardian* as released by the Department for Education (DfE) of the United Kingdom (UK). Each school had previously been ranked using raw achievement scores with an A, B, or C (A being the highest rank). Each school was ranked again using value-added scores with an A, B, or C (A being the highest rank). These rankings did not match one-to-one. Some schools with a rank of A using raw achievement schools were given a rank of C using the value-added scores, and vice versa. This illustration showed “that crude DfE based league tables will be misleading concerning the real effectiveness of schools or colleges, and that, therefore, they are best ignored” (p. 96). Thomas (1998) summarized the results as shown in Table 2.

Table 2

The Guardian *Analysis, 1992*

Value-Added Score Ranks	Raw Achievement Score Ranks	Number of Post-16 institutions (n = 402)
A	C	1
A	B	48
C	A	6
C	B	39

Notes. A = scores in top 25%; B = scores in middle 50%; C = scores in bottom 25%. Adapted from Thomas (1998, p. 93).

Although this table does not show the movement of all schools from one level to another, it does show that schools can move from the top 25% to the bottom 25%, and vice versa, when using a value-added methodology over raw achievement scores. Thomas and Mortimore (1996) argued that reporting only test results is not adequate in showing "the effectiveness of individual schools, and that a fairer and more accurate method is to present examination results in a way that takes account of the characteristics of students attending each school" (p. 5).

Educational Effectiveness Research (EER) and Accountability

Educational Effectiveness Research (EER) and accountability in education seem to be on parallel courses. While EER "represents an integration of the fields of school effectiveness (school organisation and educational policy) and research aimed at the classroom level (teacher behavior, instruction methods, and curriculum analyses)" (Lenkeit, 2013, p. 40.), accountability in education seems to imply punitive consequence for not attaining a predefined performance level or "considerable social importance" (Goldstein & Spiegelhalter, 1996, p. 386). Moreover, most educational effectiveness researchers agree that a fair comparison of the performance of schools (or teachers) must control for the differences in student intakes (Ballou, Sanders, & Wright, 2004). These "fair comparisons" imply the capacity to measure "value-added" in that they are designed to account for as many factors as possible outside the educator's realm of control and thus measure the value that teachers, schools, and/or districts are adding to a student's education. According to McPherson (1993), test results alone do not show the value-added by a school. Student intakes and school context must be considered.

However, researchers disagree on ways to control for these differences, such as the extent to which background variables should be used in value-added accountability systems. For example, the Tennessee Value-Added Assessment System (TVAAS) attempts to account for the influence of a student's background variables on the school effects through the student's prior achievement measured over multiple points in time (Ballou, Sanders, & Wright, 2004); nevertheless, even though there is a correlation between the differences in such background variables (e.g., socio-economic background) and prior achievement, "the magnitude of the typical correlation cannot justify using the [prior] achievement variable as the sole proxy for background variables" (Kupermintz, 2003, pp. 294-295).

With any accountability system, there will be a risk of incorrectly placing a student, teacher, or school in a performance-level group, such as running the analysis and then incorrectly placing a teacher in a low-performing group of teachers. Whatever accountability system is used, this risk needs to be considered and accounted for appropriately. If an accountability system is being used to place students into remedial groups (low-stakes accountability), then it is perhaps safer to err on the side of placing too many students in the remedial groups. If an accountability system is being used to provide merit pay to teachers (high stakes accountability), then accuracy is much more important. A value-added model might not be necessary in low stakes accountability; however, it may be a necessity in high stakes accountability where teachers' salaries may be affected (Tekwe al, 2004). Researchers in educational effectiveness research seem to be more concerned with models that measure effectiveness while those involved in accountability seem to be more concerned with attaching some status to a teacher, school, or district.

Level of Analysis

Another element to be considered when working with value-added models is the level of data to analyze. Most models seem to start with student data and then analyze the data at various levels (e.g., teacher level, school level, and district level). When the unit of analysis begins with student data, schools are "able to use this information to focus their attention on individual pupils, groups of pupils or subject departments" (Thomas, 1998, p. 102). However, the process of finding the appropriate level of analysis for accountability can be difficult. Lenkeit (2013) asserted that we may only be able to hold the school accountable and ought use no smaller unit of analysis. Some evidence shows that including multiple academic subjects in the model would be beneficial. Sammons, Nuttall, and Cuttance (1993) stated that "variation from year to year and for different subjects/outcomes mean that no single measure of school effectiveness should be given undue emphasis" (p. 402). Hill and Rowe (1996) agreed that no one single measure should be emphasized; however, the more measures used, the more complex the statistical procedure becomes.

There is a multitude of possible combinations when constructing a value-added model. One limitation to be considered is sample size. Sampling error will be larger in a model that uses more small groups of data versus one that uses fewer large groups of data. For example, a model that has sampled 1,000 data points will produce larger standards errors when the data are analyzed at the classroom level as opposed to the campus level. Thomas

(1998) emphasized that “one limitation of looking at individual subjects is that sometimes there are only a few pupils taking a particular examination (particularly at 'A'-levels) and therefore the statistical uncertainty of the value-added score may be relatively large” (p. 100). This is less of a problem as sample sizes increase. Most status-based accountability systems use subject scores (e.g., mathematics and reading) separately. Value-added accountability systems may also be well-served by this as long as they account for intake variables and possibly context variables. Tekwe et al. (2004) stressed that “regardless of the methods chosen for value-added assessment, it is preferable to hold each school accountable for each subject by grade combination separately” (p. 32).

Student Intakes

Researchers agree that student intakes are highly correlated with student outcomes. Goldstein and Spiegelhalter (1996) asserted that “There is now a large body of evidence which demonstrates that the single most important predictor of subsequent achievement in school is obtained by using measures of intake” (p. 395). However, as previously mentioned, researchers do not agree which student intakes should be included in value-added models. Aitkin and Longford (1986) cited three minimum “requirements of an adequate analysis of school differences” of which one is “pupil-level data on outcome, intake and relevant background variables, together with relevant school- and LEA-level variables” (p. 25). Most, if not all, student intakes fall into two major categories: student background variables and prior attainment. Other intakes are subsumed by one of these two broad categories. For example, growth can be an intake, but it is derived from multiple prior attainment measures. The mean school SES can be an intake, but it is an aggregate of the student SES background variable. Although most researchers agree that value-added models should be used to “gauge the impact of schooling on student learning *net* of the effect of student background variables" (Lenkeit, 2013, p. 39), there is no agreement as to which of these two broad categories are most impactful on student outcomes. In addition, these two broad categories are highly correlated, and with this correlation one intake can be used to account for the other intake (Lenkeit, 2013). Moreover, with respect to prior attainment, there is also no agreement among researchers about the number of measurement points that might be needed in the value-added model. For the purpose of the present study, the following were defined as student intakes:

- Student Background – non-academic student variables (e.g., student socio-economic status, number of books at home, gender, first language, and highest parent education) that may influence student outcomes.
- Prior Attainment – achievement levels, as measured in previous school years, that students bring in to the current school year.

Extensions of these two categories are as follows:

- School Level Variables – variables that characterize the school as a whole (e.g., mean SES).
- Achievement Growth – the measured difference between two or more prior attainment levels.

Background Variables

Some researchers argue that background variables (especially when prior attainment data are not available) should be used as student intake variables in value-added models to determine the level of value-added, both the background variables (e.g., socioeconomic status) of the individual student as well as the collective background variables of the school (e.g., mean socioeconomic status of the students enrolled in the school). Choi and Seltzer (2003) found that "it is necessary to control or adjust for various student intake factors—student SES, home resources, and other family background factors—and overall school intake characteristics such as school mean SES, available school resources, facilities, and so on" (pp. 35-36). Both the social class of the student and the social class of the other students in the school need to be considered (Blakey & Heath, 1992; Teddlie, Stringfield, & Reynolds, 2000). Fortier, Vallerand, and Guay (1995) recognized that many variables, such as cognitive engagement, parental variables, and interest in subject matter, might affect student achievement. Lenkeit (2013) argued "that student background characteristics function as good substitutes for prior attainment, when only cross-sectional data are available to predict academic achievement levels" (p. 44). To compare student performance among schools, one must include both individual student variables as well as the "overall composition and context of the school" (Willms, 2000, p. 244). Dreeben and Barr (1988) contended that "the difficulty of classes, as indicated by large size, low mean aptitude, and a large number of low-aptitude students, will constrain both the arrangement of classes into groups and their instruction" (p. 133). Thomas and Mortimore (1996) compared multiple value-added models and found that the best models controlled for pupils' prior attainments, gender, age, ethnicity, mobility, and socioeconomic status, and "by comparing the significant explanatory variables of different value-added models (the basic, refined, differential, prior attainment only and no prior attainment models)" (p. 26) showed that background and context variables could be used in place of prior-attainment data, but were not as useful. In summary, the best value-added models include both individual and school level background data, such as gender, age, ethnicity, and socioeconomic status.

Prior Attainment

Researchers advocating use of prior attainment scores in value-added models argue that prior attainment scores have the most impact on determining value-added because these scores are already closely correlated to background variables (e.g., socio-economic status). Lenkeit (2013) argued that “because of this strong and systematic association of student’s socioeconomic background variables and their academic achievement, it is reasonable to assume that the former function as adequate controls for the prediction of academic achievement” (pp. 42-43). Sanders and Horn (1994) added that “by focusing on measures of academic gain, each student serves as his or her own ‘control’ or, in other words, each child can be thought of as a ‘blocking factor’” (p. 305). Thomas and Mortimore (1996) compared multiple value-added models and found that “when prior attainment data are available no school context factors are significant and the fit of the model is substantially improved” (p. 26). Thomas (1998) showed that “the level of attainment an individual pupil has when s/he begins at a school is the key component in valid value-added analyses” (p. 98). Lenkeit (2013) also added that “including prior achievement scores in the prediction of achievement status considerably contributes to explaining differences between students and schools in comparison to a model with family background characteristics only” (p. 53). Researchers acknowledge that background variables act as good student intake variables for value-added models; however, using prior attainment scores can add to or replace background variables as student intake variables. Mortimore, Sammons, and Thomas (1994) suggested that prior attainment information is necessary in value-added models. Martineau (2006) added that such models would also be more effective if they included a “measurement of a given grade-level’s content in both the grade below and the appropriate grade level” (p. 57).

Statistical Models

From the previous discussion, several variables that could be used in value-added models were identified. Along with determining which variables to consider in value-added models, the researcher must also decide which statistical model to use. According to Lenkeit (2013), the approaches researchers take "to measure effectiveness can be differentiated by their conviction of the nature of student intake, on the one hand, and the respective modelling techniques, on the other hand" (p. 41). Two prominent statistical techniques are discussed below: multiple regression analysis and hierarchical linear models analysis.

Multiple Regression Analysis

Multiple regression analysis uses predictor variables to estimate a single outcome variable. Thompson (2006) defined multiple regression analysis as "a statistical technique that can be used to investigate relationships between a single outcome variable and two or more predictor variables" (p. 216). According to Tekwe et al. (2004), "status-based methods typically rely on regression models, which include school effects that are assumed fixed" (p. 12). When used in value-added accountability systems, the predictor variables are one or more of the intake (or context) variables mentioned previously and the outcome variable is the predicted score of the student. The predicted score is then subtracted from the student's actual score. A positive difference implies the district/campus added value to student's achievement while a negative score implies that the district/campus did not. The general form of a multiple regression analysis is:

$$\hat{Y} = \beta_0 + \beta_1 \mathrm{x}_0 + \beta_2 \mathrm{x}_1 + \varepsilon_i \qquad (1)$$

In accountability models, the *x* variables are the background and prior attainment variables, the β weights are fixed, ε_i is the individual score error, and $\hat{Y}$ is the individual predicted score. One major challenge with multiple regression analysis is that it does not account for nested data. Raudenbush and Bryk (1988) asserted that "three threats to valid inference have plagued analyses of nested data: misestimated precision, aggregation bias, and heterogeneity of regression" (pp. 428-429). In addition, Ott and Longnecker (2001) stated the following assumptions need to be met when considering multiple regression analysis:

1. The relation is, in fact, linear, so that the errors all have expected value zero: $E(\varepsilon_i) = 0$ for all i.
2. The errors all have the same variance: $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2$ for all i.
3. The errors are independent of each other.
4. The errors are all normally distributed; ε_i is normally distributed for all i. (p. 534)

When these assumptions are violated, the predictive model can be inaccurate. Violating these assumptions is not as serious when using hierarchical linear modeling. Status-based models using multiple regression analysis have often been used in the past to determine school accountability and/or value-added; however, researchers have been faced with many challenges when using these models, thus researchers are continually looking for better models.

Hierarchical Linear Models Analysis

Hierarchical Linear Models (HLM), or Multi-Level Models (MLM), analysis is often used by researchers with nested data. Nested data is defined as data that can be grouped in a hierarchical manor (e.g., students within classrooms, classrooms within campuses, and campuses within districts) (Ciarleglio & Makuch, 2007; De Leeuw & Kreft, 1995; Draper, 1995). HLM analysis is most often used when the researcher would like to account for variation of the outcome variable at different levels of the nested model. According to Willms (2000), HLM analysis "can be used to partition the variation in an outcome measure into the components associated with each level of the schooling hierarchy" (p. 240). Thus HLM analysis is becoming the statistical model of choice in value-added accountability systems (Aitkin & Longford, 1986; Goldstein, 1997; Raudenbush & Bryk, 1986; Sammons & Luyten, 2009; Weerasinghe & Orsak, 1998). The multi-level structure of current educational systems necessitates the use of statistical techniques that go beyond the traditional multiple regression analysis. As stated by Goldstein and Spiegelhalter (1996), "the data structures that we are concerned with are hierarchical in nature" (p. 390). Tekwe et al. (2004) added that "Hierarchical Linear Models (HLM) have been used extensively for value-added analysis, adjusting for important student and school-level covariates such as socioeconomic status" (p. 11). Aitkin and Longford (1986) went on to cite three minimum "requirements of an adequate analysis of school differences" of which one was "explicit modelling of the multi-level structure through variance components at each sampling level-child, school and LEA" (p. 25).

In one study of value-added models, Thomas and Mortimore (1996) found it crucial to consider and evaluate "a variety of different models for measuring school effectiveness using sophisticated statistical techniques (multilevel analysis)" (p. 6). There are a few Texas school districts that use HLM in their local accountability systems regardless of what model the state accountability system is using. Dallas Independent School District is one such district (Babu & Mendro, 2003; Olson, 1998; Webster & Mendro, 1997; Weerasinghe & Orsak, 1998). See Table 1.

Below is the intercept-only model according to Hox (2010), also known as the empty model or the simplest form of hierarchical linear modeling. This model has no explanatory variables.

$$Y_{ij} = \beta_{0j} + e_{ij} \qquad (2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

By substitution:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \text{ (p. 15)}$$

If this model was used for predicting the State of Texas Assessments of Academic Readiness (STAAR) Mathematics scores, Y_{ij} would be the dependent score of i^{th} student on j^{th} campus, β_{0j} would be the mean score of the j^{th} campus, γ_{00} would be the grand mean of the scores, u_{0j} would be the random effect of campus j, and e_{ij} would be the error score of i^{th} student on j^{th} campus.

Modeling nested data using HLM techniques reduces the seriousness of violating the assumption of independence required of multiple regression analysis. Hox (2010) stated that erroneous statistically significant results can occur when "the assumption of independence of the observations" is violated (p. 4). In addition, Sanders and Horn (1994) explained other problems associated with the traditional regression analysis that HLM analysis can alleviate, such as "missing student records, various modes of teaching (self-contained classroom versus departmentalized instruction versus team teaching), teachers changing assignments over years, transient students, [and] regression to the mean" (p. 299).

In summary, value-added models should include student background variables, school background variables, and/or prior attainment variables input into a hierarchal linear model with multiple measurement points. Some researchers argue that value-added models require a minimum of two measurement points (Tekwe et al., 2004) and others argue that three or more measurement points are needed (Willet, 1988). Value-added accountability models have many advantages over the traditional accountability models. Advantages of the Tennessee Value-Added Accountability System (TVAAS), according to Sanders, Saxton, and Horn (1997), include the following:

- Individual students act as their own blocking variable.
- The model works even when data is missing.
- Longitudinal data of the model allows for efficiency of estimating the model parameters.
- Longitudinal data across subjects provide benefits to the model.
- Scores can be used to show how teachers influence gains.
- Shrinkage estimates of the efforts of teachers are considered.

The model used in the present study was a Hierarchical Linear Model (HLM) that included prior attainment scores, an SES intake variable, and the campus SES context variable. Prior attainment scores and SES variables capture the majority of the variance in value-added measures (Kennedy & Mandeville, 2000; Willms, 1992).

Summary of Models

Below is a comparison of prominent value-added models in use and/or being studied by Sanders, Saxton, and Horn (1997), Webster and Mendro (1997), and Thomas and Mortimore (1996). Table 3 compares many of the features discussed earlier: statistical

Table 3

Accountability Models

Accountability Model	Statistical Model	Independent Variables	Dependent Variable	Number of Measurement Points	Level of Analysis
TVAAS (EVAAS) System[a]	HLM	Student Test Score, Subject, Grade, Year, System	Gain Score (or Residual)	Up to 5 Years	System
TVAAS (EVAAS) School[a]	HLM	Student Test Score, Subject, Grade, Year, School, System	Gain Score (or Residual)	Up to 5 Years	School
TVAAS (EVAAS) Teacher[a]	HLM: Layered	Student Test Score, Subject, Grade, Year, Teacher School, System	Gain Score (or Residual)	Up to 5 Years	Teacher
Dallas ISD[b]	Stage 1: Regression	Ethnicity/Language Proficiency, Gender, Free-Lunch Status, Census Income, Census Poverty, Census College Attendance	Achievement Residual, Attendance Residual	1 Year	
	Stage 2: HLM	Achievement Residual from Stage 1, Attendance Residual from Stage 1	School Level Residual on Achievement and Attendance,	2 Years	School
		School-Level Variables Including: Mobility, Crowdedness, Percentage Minority, Percentage Black, Percentage Hispanic, Percentage on Free-Lunch program, Average Census Variables	Student-Level Residual on Achievement and Attendance		Teacher

Table 3 Continued

Accountability Model	Statistical Model	Independent Variables	Dependent Variable	Number of Measurement Points	Level of Analysis
Basic Model[c]	HLM	Gender, Ethnicity, Age, CAT Scores: Verbal, Quantitative, Non-verbal, Free School Meals, # of Years in UK Secondary Schools, Multiple School Indicator	GSCE: Total Score Residual, Math Score Residual, English Score Residual	1 Year	School
Refined Model[c]	HLM	Gender, Ethnicity, Age, CAT Scores: Verbal, Quantitative, Non-verbal, Free School Meals, # of Years in UK Secondary Schools, Multiple School Indicator Percentage of Households with Head RGV, Percentage of Persons 18+ with HE Qualifications	GSCE: Total Score Residual, Math Score Residual, English Score Residual	1 Year	School

Table 3 Continued

Accountability Model	Statistical Model	Independent Variables	Dependent Variable	Number of Measurement Points	Level of Analysis
Differential Model[c]	HLM	Gender, Ethnicity, Age, CAT Scores by School Bands (1-3): Verbal, Quantitative, Non-verbal, Free School Meals, # of Years in UK Secondary Schools, Multiple School Indicator Percentage of Households with Head RGV, Percentage of Persons 18+ with HE Qualifications	GSCE: Total Score Residual, Math Score Residual, English Score Residual	1 Year	School
Prior Attainment Only Model[c]	HLM	CAT Scores: Verbal, Quantitative, Non-verbal,	GSCE: Total Score Residual, Math Score Residual, English Score Residual	1 Year	School

Table 3 Continued

Accountability Model	Statistical Model	Independent Variables	Dependent Variable	Number of Measurement Points	Level of Analysis
No Prior Attainment Model[c]	HLM	Gender, Ethnicity, Age, Free School Meals, # of Years in UK Secondary Schools, Multiple School Indicator, Special Educational Needs Indicator	GSCE: Total Score Residual, Math Score Residual, English Score Residual	1 Year	School
		Percentage of Chinese, Percentage of Households with Head RGI, Percentage of Households with Head RGII, Percentage of Households owner-Occupied, Percentage of Households Lacking Toilet/Bathroom and no CH with Dependent Children, Percentage on Free School Meals			

[a] Sanders, Saxton, and Horn in Millman (1997); [b] Webster and Mendro in Millman (1997); [c] Thomas and Mortimore (1996). *Note*. GCSE: General Certificate of Secondary Education; CAT: Cognitive Abilities Test

model, independent variables (e.g., prior attainment variables and intake variables), dependent variable, number of measurement points, and the level of analysis. When comparing these value-added models, the researcher should consider that a model using student intake variables extensively will not always need to include school level variables. In addition, value-added models that do not include prior attainment, but for which the residuals for some schools change signs (positive to negative or vice versa) when using value-added models that do include prior attainment, are being misclassified by the value-added model not including prior attainment (Thomas & Mortimore, 1996). Thomas and Mortimore added that "the value-added results that take account of prior attainment (for example, the basic and revised models) provide the best fitting models, are the most comprehensive and are therefore most likely to provide the best measures" (p. 25).

Recommendations for Future Studies

As previously noted, there are multiple approaches to modeling and measuring value-added education; however, one piece that is consistently missing with many, if not all, models is a validity study validating the models against some outside measure. Criterion-related validity seems most appropriate. Crocker and Algina (1986) defined criterion-related validity as comparing "performance on some real behavioral variable of practical importance" (p. 218). Studies showing comparisons of the results of value-added models to outside variables, such as principal evaluations or student surveys, would prove most useful in supporting the value-added movement. Amrein-Beardsley (2008) added that "to generate criterion-related evidence of validity, it is also necessary to assess whether teachers who post large gains from year to year are the teachers deemed most effective through other, independent measures of teacher quality" (p. 67).

Multiple studies show comparisons between models (Tekwe et al., 2004; Thomas & Mortimore, 1996), but very few studies compare model results to an outside measure. Amrein-Beardsley (2008) explored "in depth the shortage of external reviews and validity studies" (p. 65) of value-added models. Kupermintz (2003) asserted that "in light of the potential threats to the validity of TVAAS teacher evaluation information, a serious research program is urgently needed" (p. 296). For example, how do the results of ranking teachers using a value-added model compare to the results of ranking teachers using principal evaluations? For the most part these validity studies are nonexistent. On the other hand, one challenge with validity studies is determining which model measures what we think it is measuring. If a comparison between a value-added model and principal evaluations do not match, which is the more valid measure? We are trying to measure how much teachers add to the value of a student's education. Principal evaluations may be influenced by outside factors (e.g., how well the teacher is liked). Are principal evaluations of teachers or value-added models more effective measuring value-added education? These are the questions that remain unanswered by researchers with respect to value-added models. Regardless of which approach is taken, models need to be validated for measuring teacher effectiveness. Thomas and Mortimore (1996) added that "new research is vital to investigate and describe the relationship between negative value-added results and measures of school processes, in particular, the quality of teaching and learning" (p. 28).

Lastly, none of the models investigated uses a causal framework (Pearl, 1995; Rosenbaum & Rubin, 1983; Rubin, 1978; Rubin, 2004) that explicitly accounts for selection bias. When using non-experimental data, the outcomes of multiple groups (e.g., multiple classes of students) may differ systemically (Heckman, Ichimura, Smith, & Todd, 1998). The intent of value-added models is to show that teacher performance causes student achievement; however, without using a causal framework this relationship is implied only. An in-depth study of value-added models using causal frameworks is recommended.

CHAPTER III

MONTE CARLO SIMULATION: A VALUE-ADDED MODEL

The present Monte Carlo study focused on the relationship between various levels of "value-added" by campuses (how much value campuses added to a student's academic achievement) and various distributions of economically disadvantaged (low socio-economic status or low SES) students on each campus within a district. Five values of each variable were presented resulting in a 5 by 5 design.

Value-added was simulated by varying the Correlation of Campus Random Effects (CCRE) between grade levels. Each CCRE used a different Pearson *r* correlation between the grade level random variables designated for campus value-added effectiveness. Each district varied by the distribution of low SES students on each of their campuses. Scores (grade 2 to grade 8) were generated for 300 students on each of the 50 campuses of each of the 5 districts of the 5 by 5 design. These scores were modeled using the Mathematics portion of the Texas Assessment of Knowledge and Skills (TAKS). For the purpose of the present study, students had no missing data, continued to be enrolled on the same campus (grade 2 through grade 8) each year, and continued to be classified as either economically disadvantaged or non-economically disadvantaged throughout the study once identified as such in grade 2. These simulated scores were used to rank campuses twice within each district within the 5 by 5 design. First, campuses were ranked within the district using a value-added accountability system model with 1 as the highest rank and 50 as the lowest rank. Second, campuses were ranked within the district using an SES accountability system model with 1 as the highest rank and 50 as the lowest rank. After each campus was assigned a rank based on each of these two systems, a Pearson *r* correlation was calculated between the two campus rankings for the district. Note that a Spearman rho equals a Pearson *r* if the two variables have identical distribution shapes. The Pearson *r* correlations varied from district to district; therefore; the number of times each specific Pearson *r* correlation was calculated was summed after conducting 100 cycles (or runs) of each variation of the 5 by 5 design. These sums were then tabulated in a series of 5 tables, one for each CCRE. This chapter presents the methodology used for this Monte Carlo study and is divided into five sections: model specifications, simulation conditions, simulation parameters, simulation procedures, and analysis.

Model Specifications

The present study specified three models: a model to simulate student end-of-year scores from grade 2 to grade 8, a value-added accountability system model (hierarchical linear model to calculate campus value-added residuals) to analyze these scores and rank campuses, and a traditional SES accountability system model to rank these same campuses a second time. A Pearson *r* correlation was then calculated between the campus rankings of these last two models for each district of the 5 by 5 design.

Model to Simulate Student End-of-Year Scores

Using a variation of Wang's (2006) methodology, student end-of-year scores were simulated for grades 2, 3, 4, 5, 6, 7, and 8. This model assumed that a student's score change (increase or decrease) from one grade to the next was due to the natural growth of the student (the sum of the grand mean of individual student natural growth, β_{10}, the product of the fixed effect of the student SES indicator, β_{11}, and the SES indicator of the student, SES_{ti}, and the random individual student natural growth of the student, r_{1i}) and the growth added by the specific campus where the student was enrolled (the sum of the product of the fixed effect of the campus SES, γ_{01}, and the percent of low SES students on the specific campus where the student was enrolled, SES_j, and the value-added effect of the campus, u_{tj}). These student score changes were added cumulatively for each grade level to the baseline score at grade 3 to determine the student score for a specific grade level.

The baseline score of each student at grade 3 (student growth curve intercept) was attributed to the sum of the product of the fixed effect of Grade 2 scores on the baseline intercept at Grade 3, β_{01}, and the student's grade 2 score, $Grade\_2_{ij}$, the student's random effect on the baseline intercept at Grade 3, r_{0i}, and the campus's random effect on the baseline intercept at Grade 3, u_{0j}. Bosker and Witziers (1995) found an intraclass correlation of 0.210 (σ^2 = 0.210, σ = 0.458) between grade level campus mean scores within a district for student scores having a standard normal distribution. Darandari (2004) found a correlation of -0.930 between grade level campus mean scores and the campus percentage of low SES students. The present study investigated various distributions of low SES students within a district. The mean low SES for the district was held constant at 0; however, the variance between campuses was allowed to vary from District 1 (σ^2 = 0.100, σ = 0.316) to District 5 (σ^2 = .900, σ = .949) by incrementing the variance by 0.200 for each district. The values for SES_j were then transformed into campus percentages of low SES by re-centering the low SES campus mean to 50 and adding the product of the various low SES standard deviations and 12.5, thus resulting in 5 different low SES distributions with each having a mean of 50. This model yields both very narrow distributions of low SES students within a district and very wide distributions of low SES students within a district. Each of these 5 different distributions was referenced as a school district, as detailed in Table 4. When simulating the grade 2 student scores, the distribution of grade 2 student scores was standard normal (μ = 0,

Table 4

Sample of District Low SES Distributions (Modified from RUN_ID=50 and CCRE_ID=1)

	District				
	1	2	3	4	5
	μ=0	μ=0	μ=0	μ=0	μ=0
	μ_{rc}=50 after re-centering				
	σ^2=0.10	σ^2=0.30	σ^2=0.50	σ^2=0.70	σ^2=0.90
Campus	σ_{rc}=12.5×σ after re-centering				
1	39.44	36.09	32.17	32.41	24.70
2	43.41	38.29	35.29	32.44	35.91
3	43.86	39.31	35.32	32.53	37.38
4	45.38	42.26	37.46	33.63	37.42
5	45.41	42.46	38.55	34.75	38.20
6	45.70	42.46	39.51	37.44	38.34
7	45.71	42.59	40.37	37.97	38.46
8	46.20	42.95	41.02	38.33	38.60
9	46.27	43.35	41.55	39.23	39.01
10	46.79	43.83	42.91	41.26	40.12
…					
41	54.10	55.32	57.88	61.81	55.10
42	54.23	56.07	57.97	61.82	56.04
43	54.79	57.00	59.03	61.95	56.40
44	54.86	57.67	60.02	62.83	60.19
45	55.20	58.09	61.15	63.42	60.23
46	55.43	59.15	61.49	63.45	60.82
47	56.29	59.49	62.18	63.47	63.65
48	56.73	60.25	63.52	64.96	72.47
49	56.84	65.03	65.58	65.68	75.83
50	57.58	67.52	66.31	77.12	81.79

Note: rc: Re-Centered

σ = 1, also meaning β_{00} = 0), grade campus mean score intraclass correlations were set to 0.210 (Bosker & Witziers, 1995), the correlation between the campus percentage of low SES students and the grade 2 campus mean scores was set to -0.930 (Darandari, 2004), and the campus low SES distributions varied.

Adapted from Wang's (2006) model, the algebraic form of the model to simulate student end-of-year scores from grade 3 to grade 8 for the present study is stated below:

$$Y_{tij} = \beta_{00} + \beta_{01} * Grade\_2_{ij} + r_{0i} + u_{0j} \tag{3}$$

$$+ (\beta_{10} + \beta_{11} * SES_{ti} + r_{1i}) * time$$

$$+ \sum_{1}^{t}(\gamma_{01} * SES_{j} + u_{tj})$$

$$+ e_{tij}$$

where:

Y_{tij} = standardized TAKS Math score at time t, for student i, on campus j

B_{00} = grand mean of Grade 2 scores

B_{01} = fixed effect of Grade 2 scores on the baseline intercept at Grade 3

$Grade\_2_{ij}$ = Grade 2 score for student i on campus j

r_{0i} = student random effect on the baseline intercept at Grade 3 of student i

u_{0j} = campus random effect on the baseline intercept at Grade 3 of campus j

β_{10} = grand mean of individual student natural growth

β_{11} = fixed effect of student SES indicator

SES_{ti} = SES indicator at time t for student i

r_{1i} = random individual student natural growth for student i

time = Grade 3 (t = 0) through Grade 8 (t = 5)

γ_{01} = fixed effect of campus SES

SES_j = percent of low SES students on campus j

u_{tj} = value-added effect at time t of campus j

e_{tij} = the error score at time t, for student i, on campus j

The value-added effects of the campuses were simulated by correlating the random variable, u_{tj}, with itself from one year to the next—the higher this correlation, the less random the variable, u_{tj}. Campuses that had higher value-added effects one year had identical value-added effects the following year and vice versa. Campuses that consistently had higher value-added effects (where u_{tj} was positive) than other campuses from one year to the next year were considered to be adding value to the growth of their students. Campuses with inconsistent values of u_{tj} from one year to the next or with values consistently negative were not considered to be adding value. Therefore the focus of value-added studies becomes this key variable, u_{tj}. When all other variables that could affect a student's score are accounted for in the model, u_{tj} should have a positive correlation with the student's final score. To simulate this effect, a set of five random variables (one for each grade level interval) with a pairwise correlation representing campus value-added growth consistency from one year to the next was generated. For the purpose of the present study the correlation was allowed to vary so that various correlations could be analyzed. Five different correlations, r_{tt}, were used (where r_{tt}^2 = 0.100, 0.300, 0.500, 0.700, and 0.900), each designated as a different CCRE (1, 2, 3, 4, and 5), as detailed in Table 5. Under CCRE 1, the five variables were multivariate normal with pairwise bivariate correlations such that r_{tt}^2 equaled 0.100. Likewise, under CCRE 5, these five variables were multivariate normal with bivariate pairwise correlations such that r_{tt}^2 equaled 0.900. Thus the campus value-added effects for CCRE 5 were much more consistent than for CCRE 1. A campus performing high the first year continued to perform high the following years and vice versa. This would be consistent with what would be expected in public schools if some campuses added value and some campuses did not.

Table 5

Correlation of Campus Random Effects

CCRE	1	2	3	4	5
$r_{tt}^{2\,a}$	0.10	0.30	0.50	0.70	0.90

[a] Effect size of the pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5

Value-Added Accountability System Model

After the student scores were simulated, they were formatted so that each student had six records in the Stud_Scores_Array table—one record corresponding to each grade level score (i.e., grades 3 through 8). See Table 6 for an example. Grade levels were converted to time: Grade 3 (t=0) and Grade 8 (t=5). These transformed data were then used in an hierarchical linear model to calculate how much a campus added value to the student's growth as measured by student residuals. These data were then used to determine Beta values in the following HLM (Hox, 2010).

$$Y_{ti} = \pi_{0i} + \pi_{1i} * time_{ti} + \pi_{2i} * SES_{ti} + e_{ti} \tag{4}$$

$$\pi_{0i} = \beta_{00} + \beta_{01} * SES_j + u_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * SES_j + u_{1i}$$

$$\pi_{2i} = \beta_{20} + \beta_{21} * SES_j + u_{2i}$$

which yields:

$$Y_{ti} = \beta_{00} + \beta_{01} * SES_j + u_{0i} + (\beta_{10} * time_{ti} + \beta_{11} * SES_j * time_{ti} + u_{1i} * time_{ti}) \quad (5)$$
$$+ (\beta_{20} * SES_{ti} + \beta_{21} * SES_j * SES_{ti} + u_{2i} * SES_{ti}) + e_{ti}$$

which yields:

$$Y_{ti} = \beta_{00} + \beta_{01} * SES_j + \beta_{10} * time_{ti} + \beta_{11} * time_{ti} * SES_j + \beta_{20} * SES_{ti} + \beta_{21} * SES_{ti} * SES_j + u_{0i} + u_{1i} * time_{ti} + u_{2i} * SES_{ti} + e_{ti} \quad (6)$$

which yields:

$$Y_{ti} = \beta_{00} + \beta_{10} * time_{ti} + \beta_{20} * SES_{ti} + \beta_{01} * SES_j + \beta_{11} * time_{ti} * SES_j + \beta_{21} * SES_{ti} * SES_j + u_{1i} * time_{ti} + u_{2i} * SES_{ti} + u_{0i} + e_{ti} \quad (7)$$

Table 6

Stud_Scores_Array Table: Example of the HLM Data Format

Run_ID	CCRE_ID	Dist_ID	Camp_ID	Stud_ID	Camp_Econ_Dis	SES_{ti}	Time	Score
1	1	1	1	1	45	0	0	2.5
1	1	1	1	1	45	0	1	2.6
1	1	1	1	1	45	0	2	2.3
1	1	1	1	1	45	0	3	3.6
1	1	1	1	1	45	0	4	3.0
1	1	1	1	1	45	0	5	4.6
…								
100	5	5	50	300	100	1	0	-1.5
100	5	5	50	300	100	1	1	-1.6
100	5	5	50	300	100	1	2	-0.5
100	5	5	50	300	100	1	3	-1.7
100	5	5	50	300	100	1	4	-2.0
100	5	5	50	300	100	1	5	-0.5

Substituting these Beta values into Equation 8 below, along with the original school and student predictors, SES_j, SES_{ti}, and $time_{ti}$, a grade 8 predicted score, Y_{5i}, was generated for each student.

$$Y_{ti} = \beta_{00} + \beta_{01} * SES_j + \beta_{10} * time_{ti} + \beta_{11} * time_{ti} * SES_j + \beta_{20} * SES_{ti} + \beta_{21} * SES_{ti} * SES_j \quad (8)$$

where $t = 5$

A residual score was then generated for each student by subtracting the predicted grade 8 score from the simulated grade 8 score. A positive residual score showed that the campus added value to the student's growth while a negative residual score showed that the campus detracted from the student's growth. The residuals were then averaged for the campus to yield a mean residual score which was used to rank the 50 campuses. The higher the mean residual score, the more value the campus was adding to student growth and the higher the campus ranking.

Traditional SES Accountability System Model

A third model was developed using the student scores generated by the first model. In this model, the percentage of low SES students with a passing score on each campus was used to generate campus rankings—the higher the percentage of low SES students passing, the higher the campus ranking. This is analogous to many of the current accountability system models that are based on a passing percentage of students (which many times comes down to how the lower SES students performed). In Texas, the area that has historically kept a campus from moving up to the next accountability level was the percentage of lower SES students passing either math or science or both on their campus, thus the reason for the focus on how well lower SES students perform on the Mathematics portion of the TAKS in this study.

Simulation Conditions

This Monte Carlo study investigated the relationship between various levels of "value-added" by campuses and various distributions of low SES students on each campus within a district and the way campus rankings generated under these various conditions compared to rankings generated under the traditional SES accountability system models. Five values of each variable (value-added and SES) were investigated resulting in a 5 by 5 design.

Value-added was simulated by varying the Correlation of Campus Random Effects (CCRE) between grade levels. These value-added effects were measured with a pairwise bivariate correlation, $r_{tt,}$ where r_{tt}^2 ranged from 0.100 to 0.900 with increments of 0.200 and were referenced as CCRE 1 through CCRE 5. The low SES campus mean percent in each district was 50 (recentered from 0). In addition, the distribution of low SES students within each district were transformed and varied from a standard deviation of 3.950 (0.316, or $\sigma, \times$ 12.500) to a standard deviation of 11.863 (0.949, or $\sigma, \times$ 12.500) and was referenced as District 1 through District 5. See again Table 4. Each district had 50 campuses and each campus had 300 students. Each combination was run 100 times.

This 5 by 5 design allowed for trends in the data to be analyzed. The extremes of these trends are:

Correlation of Campus Random Effects (r_{tt}^2=0. 100), SES_j (μ = 50, σ = 3.950)

Correlation of Campus Random Effects (r_{tt}^2=0. 100), SES_j (μ = 50, σ = 11.863)

Correlation of Campus Random Effects (r_{tt}^2=0. 900), SES_j (μ = 50, σ = 3.950)

Correlation of Campus Random Effects (r_{tt}^2=0. 900), SES_j (μ = 50, σ = 11.863)

Simulation Parameters

To generate student scores representative of real-life data to use in this Monte Carlo study, parameters were selected to be used in the model to simulate student end-of-year scores as explained in Table 7. These parameters can be divided into three major categories: campus level parameters, student level parameters, and correlations between various parameters. The total campus effect was divided into two components: the campus SES effect, $\gamma_{01} * SES_j$; and the campus random (or value-added) effect, u_{tj}. The total student effect,

Table 7

Parameter Values of Fixed Effects, Variance Components, Correlations, and Other Variables

Parameter	Description	Value
	Campus Level Fixed Effect	
β_{00}	Grand mean of intercept	$\mu = 0.000$
γ_{01}	Fixed effect of campus SES	-0.414
	Campus Level Random Effect	
u_{0j}	Variance of campus random effect on the baseline intercept at grade 3 for campus *j*	0.100
u_{tj}	Variance of value-added effect for campus *j* at time *t* where *t*=1, *t*=2, *t*= 3, *t*=4, *t*=5	0.317
	Campus Level Predictor	
$SES_j\_Percent$	Percent of low SES students on campus *j*	$\mu = 50$, σ = (3.950 thru 11.863)
	Student Level Fixed Effect	
β_{01}	Fixed effect of Grade 2 scores on the baseline intercept at Grade 3	0.762
β_{10}	Grand mean of individual student natural growth	0.630
β_{11}	Fixed effect of student SES indicator	-0.045
	Student Level Random Effect	
r_{0i}	Variance of student random effect on the baseline intercept at Grade 3	0.220
r_{1i}	Variance of random individual student natural growth	0.325
	Student Level Predictors	
$Grade\_2_{ij}$	Grade 2 scores	$\mu = 0$, $\sigma^2 = 1.0$
SES_{ti}	SES indicator	0 or 1
	Student Level Outcomes	
Y_{0ij}	Standardized TAKS Math score at time *0*, for student *i*, on campus *j* (Grade 3 Score)	$\mu = 0$, $\sigma = 1.0$
Y_{tij}	Standardized TAKS Math score at time *t*, for student *i*, on campus *j* where *t*=1, *t*=2, *t*= 3, *t*=4, *t*=5	Varies
	Correlations	
$r_{ses*grade\ 2}$	Correlation between campus SES and school mean grade 2 test score	-0.930
r_{tt}	Pairwise bivariate Pearson *r* correlation of u_{tj} where *t*=0, *t*=1, *t*=2, *t*= 3, *t*=4, *t*=5	0.316 thru 0.949
$r_{grade\ 2}$	Intraclass correlation of grade 2 test scores	0.210
	Other	
$time_{ti}$	Grade 3 (t = 0) through Grade 8 (t = 5)	0 thru 5
e_{tij}	The error score at time *t*, for student *i*, on campus *j*	$\mu = 0$, $\sigma^2 = 0.100$
r_{1i}	Student SES Effect ($\beta_{11} * SES_{ti}$) + Student Random Effect	0.325
Total Campus Effect	Campus SES Effect ($\gamma_{01} * SES_j$) + Campus Random (or Value-Added) Effect (u_{tj})	Varies

Adapted from Wang (2006).

r_{1i}, was also divided into two components: the student SES effect, $\beta_{11} * SES_{ti}$, and the student random effect.

The campus level parameters included fixed effects, random effects, and predictors. In generating student scores, one has to have a beginning point. In the present study, the beginning point was the grand mean of grade 2 scores (β_{00}), which was set to zero with an intraclass correlation ($r_{grade\ 2}$) of 0.210 (Bosker & Witziers, 1995). The percent of low SES students on each campus was derived from this initial distribution of grade 2 scores by using a correlation ($r_{ses*grade2}$) of -0.930 (Darandari, 2004) between the grade 2 campus mean scores and the low SES campus percentages. This correlation is negative because campus SES is usually measured as the percent of students with a low socio-economic status and is therefore negatively correlated with student achievement. In the present Monte Carlo study, the mean campus low SES was set to a constant of 0 (then re-centered to 50) for each district, but the campus low SES variance varied as follows: 0.100, 0.300, 0.500, 0.700, and 0.900 (and was then transformed using a multiple of 12.5). For example, a low SES campus variance of 0.500 would be transformed to 58.838% (50 + 0.707*12.5). These variances were selected to demonstrate a range of low SES distributions that are similar to what might be seen in authentic campus data. Campuses can range from 0 percent low SES to 100 percent low SES in the real world. Typically, real campuses within real districts will have varying distributions of low SES students. This distribution model maximizes the variation of campus SES, thus reflecting real campus data. The campus SES fixed effect, γ_{01}, was set to -0.414 (Wang, 2006).

The last of the campus-level parameters needed were the campus random effects on student growth. These were estimated at each grade level from grade 3 through grade 8, thus there were six parameters. The variance of the grade 3 campus random effects, u_{0j}, was set to 0.100, reflecting the findings of Bosker and Witziers (1995), who indicated that the school selection effect explained 10% of the total variance of test scores. Wang (2006) stated that the variance of the total campus effect was 0.487 (σ=0.698) where 65%, or 0.317 (σ=0.563), was attributed to campus random (or value-added) effects and 35%, or 0.170 (σ=0.412), was attributed to the campus SES effect; therefore, the variance of each of the remaining campus random (or value-added) effects (u_{1j}, u_{2j}, u_{3j}, u_{5j}, u_{5j}) was set to 0.317. The pairwise correlation, r_{tt}, between each of these campus effects at each grade level varied such that r_{tt}^2 equaled 0.100, 0.300, 0.500, 0.700, and 0.900, as depicted earlier in Table 5. The focus of the present study was the trend from one extreme of campus value-added effects to the other, rather than one specific pairwise correlation between these campus value-added effects at each grade level.

The student level parameters included fixed effects, random effects, and predictors. The grade 2 scores, $Grade\_2_{ij}$, were set to a mean of 0 and a standard deviation of 1 as well as the grade 3 scores, Y_{0ij}. The calculation of the grade 3 scores followed the parameter settings described by Wang (2006):

- β_{01}: previous year scores explained 58.0% (σ^2=0.580, σ=0.762) of the total variance of the baseline year scores
- u_{0j}: school selection effect explained 10.0% (σ^2=0.100, σ=0.316) of the total variance of baseline year scores
- r_{0i}: student random effect explained 22.0% (σ^2=0.220, σ=0.469) of the total variance of baseline year scores

- e_{tij}: measurement error explained 10.0% (σ^2=0.100, σ=0.316) of the total variance of baseline year scores

According to Ponisciak and Bryk's (2005) research in the Chicago public schools, the mean student natural growth rate of all students, β_{10}, is assumed to be linear. Each student's individual natural growth should be calculated by summing the mean student growth rate, β_{10}, (which they found to be 0.630 logits) and the individual student's random effect on natural growth, r_{1i}, (which they found to have a variance of 0.325).

The variance of the total student effect, r_{1i}, was divided into two components: the student SES effect, $\beta_{11} * SES_{ti}$; and the student random effect. The SES indicator for the student was set to either "1" for "Economically Disadvantaged" (or low SES) or "0" for "Not Economically Disadvantaged" (or high SES), the traditional method of coding this indicator. Caldas and Bankston (1997) investigated the impact that both individual and family background characteristics had on academic achievement and found a correlation of -.0045 between student SES and academic achievement. Therefore, the fixed effect of the student SES indicator, β_{11}, was set to -0.045.

Using these parameters, the subsequent end-of-year Mathematics TAKS scores were allowed to change (increase or decrease) from year to year based primarily on student natural growth, campus effects, and student effects. The focus of the present Monte Carlo study was the campus effects.

Simulation Procedure

The simulation procedure can be divided into five parts: generating student test scores from grade 3 to grade 8; estimating campus value-added coefficients (Beta values); ranking campuses by mean student residuals (value-added accountability model); ranking campuses by low SES passing percentages (tradition SES accountability model); and calculating a Pearson *r* correlation between the two rankings. See Appendix A for the complete SAS code developed by the author for the simulation procedure. Pieces of this code are explained in the following sections.

Generating Student Test Scores from Grade 3 to Grade 8

Generating student test scores involved three steps: creating and populating campus level variables, creating and populating student level variables, and updating student level variables from previously created and populated variables.

Creating and populating campus level variables.

Five CCRE's and five districts were used in this 5 by 5 design. Each district consisted of 50 campuses, thus 1,250 (or 5*5*50) campus level records were created for these variables in each cycle (or run) of the simulation where each record was unique to each CCRE, district, and campus combination. Each record began with the variables *Run_ID*, *CCRE_ID*, *Dist_ID*, and *Camp_ID* to identify these 1,250 unique campus level records. In the first record these variables were set such that: *Run_ID*=1, *CCRE_ID*=1, *Dist_ID*=1, and *Camp_ID*=1. In the last record these variables were set such that: *Run_ID*=100, *CCRE_ID*=5, *Dist_ID*=5, and *Camp_ID*=50. The *Run_ID* was used to identify the number of times the simulation was replicated. The Camp_Data table was created with the campus level variables below:

Run_ID (9)
Dist_ID
CCRE_ID
Camp_ID
Grade_2_Mean_j
SES_j
γ_{01}

Campus_SES_Effect
Campus_Effect0j - Campus_Effect5j
u_{0j} - u_{5j}

The first step in the simulation was to populate the *$Grade\_2\_Mean_j$* variable using an intraclass correlation (or variance of the classes or campus means) of 0.210 (Bosker & Witziers, 1995). The *$Grade\_2\_Mean_j$* variable was populated using the SAS code below:

```
Grade_2_Mean_j = 0.458 * rannor(-5);                                    (10)
```

The present study focused on 5 different districts with 5 different distributions of low SES students on each campus, *SESj*. The mean *SES_j* for each district was set to 0 with a correlation between *SES_j* and *$Grade\_2\_Mean_j$* of -0.930 (Darandari, 2004) along with the varying standard deviations for each of the 5 districts as previously described. The values for *SESj* were then transformed into campus percentages, *$SES_j\_Percent$*, of low SES by re-centering the low SES campus mean to 50 and adding the product of the various low SES standard deviations and 12.5. The *SES_j* and the *$SES_j\_Percent$* variables were populated using the SAS code below:

```
SES_j = -(((.93**2)*(.1+(Dist_ID-1)*.2)/.21)**.5)                       (11)
*Grade_2_Mean_j-(((.1+(Dist_ID-1)*.2)-((.93**2)
*(.1+(Dist_ID-1)*.2)))**.5*rannor (-5));
```

where the Dist_ID varied from 1 to 5

```
SES_j_Percent = 50 + 12.5 * SES_j;
```

Increasing the variance of the distribution of *SESj* by 0.200 for each district maintained the normal distribution, but allowed for comparisons between narrower *SESj* distributions and wider *SESj* distributions, which would occur in the real world.

The fixed effect of campus SES_j, γ_{01}, was set to -0.414 (Wang, 2006). The *Campus_SES_Effect* was populated using the SAS code below:

$$\text{Campus\_SES\_Effect} = \gamma_{01} * \text{SES}_j; \tag{12}$$

In addition to the present study's focus on 5 different districts with 5 different distributions of SES_j, it also focused on 5 different CCRE's with 5 different campus effects (measured by correlations of student scores between grade levels—the higher the correlation the more consistent the campus effect). For CCRE 1, all campus effect pairwise bivariate correlations, r_{tt}, were set such that r_{tt}^2 equaled 0.100. For each progression to the next CCRE, r_{tt}^2 was increased by 0.200, thus for CCRE 5 the campus effect pairwise bivariate correlations, r_{tt}, were set such that r_{tt}^2 equaled 0.900. The remaining campus variables to be populated were the $Campus\_Effect_{tj}$ variables and the value-added campus effect variables: u_{0j}, u_{1j}, u_{2j}, u_{3j}, u_{4j}, and u_{5j}. The variance of the campus random intercept, u_{0j}, was set to 0.100 (σ=0.316) (Bosker & Witziers, 1995) and the variances of the remaining campus random effects were set to 0.317 (σ=0.563) (Wang, 2006). The variables, u_{0j}, u_{1j}, u_{2j}, u_{3j}, u_{4j}, and u_{5j} were populated using the SAS code below:

```
%Macro Gen_Camp_Effect (Start=0.1, Stop=0.9, Step=.2);                    (13)
%do i=&Sim_Start %to &Sim_Finish;
    %do j=&District_Start %to &District_Finish;
    proc sql;
    create table Temp_2 (
    Col1 numeric,
    Col2 numeric,
    Col3 numeric,
    Col4 numeric,
    Col5 numeric,
    Col6 numeric);
    quit;
```

```
	%let r2=&Start;
		%do %until(%sysevalf(&r2 gt &Stop));
		%let jj = %sysevalf((&r2**.5)*(0.317**.5)*(0.317**.5));
		%let jjj = %sysevalf((&r2**.5)*(0.317**.5)*(0.100**.5));
			proc iml;
mu={ 0, 0, 0, 0, 0, 0};
sigma= {0.100	&jjj	&jjj	&jjj	&jjj	&jjj,
	&jjj	0.317	&jj	&jj	&jj	&jj,
	&jjj	&jj	0.317	&jj	&jj	&jj,
	&jjj	&jj	&jj	0.317	&jj	&jj,
	&jjj	&jj	&jj	&jj	0.317	&jj,
	&jjj	&jj	&jj	&jj	&jj	0.317};
				call vnormal(et, mu, sigma, &Campus_Finish);
				create Temp_1 from et;
				append from et;
				proc append base=Temp_2 data=Temp_1;
				run;
			proc sql;
				insert into Camp_Effect (
					u0j,
					u1j,
					u2j,
					u3j,
					u4j,
					u5j)
				select col1, col2, col3, col4, col5, col6
				from Temp_2;
			quit;
			proc sql; delete * from Temp_2; quit;
			%let r2 =%sysevalf(&r2+&Step);
		%end;
	%end;
%end;
%Mend Gen_Camp_Effect;
```

The *Campus_Effect*$_{tj}$ variables were populated using the SAS code below:

$$\begin{aligned}
&Campus_Effect_{0j} = Campus_SES_Effect + u_{0j};\\
&Campus_Effect_{1j} = Campus_SES_Effect + u_{1j};\\
&Campus_Effect_{2j} = Campus_SES_Effect + u_{2j};\\
&Campus_Effect_{3j} = Campus_SES_Effect + u_{3j};\\
&Campus_Effect_{4j} = Campus_SES_Effect + u_{4j};\\
&Campus_Effect_{5j} = Campus_SES_Effect + u_{5j};
\end{aligned} \tag{14}$$

Creating and populating student level variables.

The Stud_Scores table was created next which contained all the variables needed for the simulation, including the campus level variables. The focus at this point was populating the student level variables of this table.

Five CCRE's and five districts were used in this 5 by 5 design. Each district consisted of 50 campuses with each campus having 300 students, thus 375,000 (or 5*5*50*300) student level records were created for these variables in each cycle (or run) of the simulation where each record was unique to each CCRE, district, campus, and student combination. Each record began with the variables *Run_ID*, *CCRE_ID*, *Dist_ID*, *Camp_ID*, and *Stud_ID* to identify these 375,000 unique student records. In the first record, these variables were set such that: *Run_ID*=1, *CCRE_ID*=1, *Dist_ID*=1, *Camp_ID*=1, and *Stud_ID*=1. In the last record, these variables were set such that: *Run_ID*=100, *CCRE_ID*=5, *Dist_ID*=5, *Camp_ID*=50, and *Stud_ID*=300. The complete set of Stud_Scores table variables are listed below:

Run_ID (15)
CCRE_ID
Dist_ID
Camp_ID
Stud_ID
Grade_2_Mean_j
SES_j
$\text{SES}_j\text{\_Percent}$
γ_{01}
Campus_SES_Effect
u_{0j} - u_{5j}
Grade_2_{ij}
SES_{ti}
B_{01}
β_{10}
β_{11}

r_{0i}
r_{1i}
e_{0ij} - e_{5ij}
$Stud\_SES\_Effect_{ti}$
$Student\_Random\_Effect_{ti}$
$Student\_Effect_{ti}$
$Campus\_Effect_{0j}$ - $Campus\_Effect_{5j}$
$Grade\_3_{ij}$ - $Grade\_8_{ij}$
Grade_8_Mean;
Grade_8_STD;
Grade_8_zscore;
Grade_8_Pass
Intercept_Beta (β_{00})
Camp_Econ_Dis_Beta (β_{01})
Time_Beta (β_{10})
Camp_Econ_Dis_Time_Beta (β_{11})
Stud_Econ_Dis_Beta (β_{20})
Camp_Econ_Dis_Stud_Econ_Dis_Beta (β_{21})
Predicted_Score
Residual

Campus level variables were populated with the campus level data from the Camp_Data table using various SQL update statements and joining on the *CCRE_ID*, *Dist_ID*, and *Camp_ID* variables. The data for each campus were replicated 300 times to populate each of the 300 student records for each campus with the same campus data.

The first step in the simulation was to consider the Grade_2 score for each student. As stated earlier, the mean of Grade 2 scores, $Grade\_2\_Mean_j$, for each campus had an intraclass correlation of 0.210, thus the remaining variance of scores (σ^2=0.790, σ=0.889) was attributed to students. The Grade 2 score, $Grade\_2_{ij}$, for each student was populated using the SAS code below:

$Grade\_2_{ij}$ = Grade_2_Meanj + ((1-.21)**.5)*rannor(-5); (16)

Each campus had already been identified with a specified percentage of low SES students as previously discussed; however, each student also needed to be identified as low SES (coded with a 1) or high SES (coded with a 0). These students were identified by multiplying the already specified campus percentages, *SESj_Percent*, of low SES students by 300 which yields the number of students needed to be identified on each campus and then setting the SES_{ti} variable of all students on each campus with a *Stud_ID* less than this number to 1. The *Stud_ID*'s on each campus began with 1 and ended with 300. All campus data was random so identifying the students using this method still provided the needed randomness. The SES_{ti} variable was initially set to 0 and then updated using the SAS code below:

$$\mathrm{SES}_{ti} = 1 \text{ where Stud\_ID} <= (\mathrm{SES}_{j}\text{\_Percent}/100) * 300; \tag{17}$$

The variables β_{01}, β_{10}, β_{11}, and r_{1i} were set to 0.762 (Wang, 2006), 0.630 (Ponisciak & Bryk, 2005), -0.045 (Caldas & Bankston, 1997), and 0.325 (Ponisciak & Bryk, 2005), respectively, as previously discussed. The student random effect on the student's baseline intercept at Grade 3 (r_{0i}) was set to have a variance of 0.220 (Wang, 2006) upon creation using the SAS code below:

$$r_{0i} = 0.469 * \text{rannor}(-5); \tag{18}$$

The error scores, e_{tij}, were set to have a variance of 0.100 (σ=0.316) (Wang, 2006) using the SAS code below:

$$\begin{aligned} e_{0ij} &= 0.316 * \text{rannor}(-5); \\ e_{1ij} &= 0.316 * \text{rannor}(-5); \\ e_{2ij} &= 0.316 * \text{rannor}(-5); \\ e_{3ij} &= 0.316 * \text{rannor}(-5); \\ e_{4ij} &= 0.316 * \text{rannor}(-5); \\ e_{5ij} &= 0.316 * \text{rannor}(-5); \end{aligned} \tag{19}$$

The Student_Random_Effect$_{ti}$ variable was populated using the SAS code below:

```
%Macro Gen_Stud_Effect;                                                    (20)

proc sql;
        create table Stud_Effect (
        Student_Random_Effectti numeric); quit;
%do i=&Sim_Start %to &Sim_Finish;
        %do j=&District_Start %to &District_Finish;
                %do CCRE_ID=&CCRE_Start %to &CCRE_Finish;
                        %do Camp_ID=&Campus_Start %to &Campus_Finish;

                                proc sql;
                                create table Temp_2 (
                                Col1 numeric,
                                Col2 numeric,
                                Col3 numeric);
                                quit;

                                proc iml;
                                mu={ 0, 0, 0};
                                sigma= {0.001 0.001          0.011,
                                        0.001          0.324          0.001,
                                        0.011          0.001          0.250};
                                        call vnormal(et, mu, sigma, &Student_Finish);
                                                create Temp_1 from et;
                                                append from et;

                                                proc append base=Temp_2 data=Temp_1;
                                                run;
                                        proc sql;
                                                insert into Stud_Effect (
                                                        Student_Random_Effectti)
                                                select col2
                                                from Temp_2;
                                        quit;
                                        proc sql;
                                        delete * from Temp_2;
                                        quit;
                        %end;
                %end;
        %end;
%end;
%Mend Gen_Stud_Effect;
```

Updating student level variables.

At this point, all of the campus level and student level parameters had been set. The remaining student level variables were derived based on those parameters. In order to standardize the *Grade_8*$_{ij}$ scores, the mean and standard deviation of the *Grade_8*$_{ij}$ scores, along with their corresponding z-scores, were calculated for each CCRE using the SAS code below:

```
proc corr data=Work.zscores_temp noprint outp=Pearson;                    (21)
var Grade_8ij;
run;

Grade_8_Mean = (select Grade_8ij from Pearson where _TYPE_ = 'MEAN');
Grade_8_STD = (select Grade_8ij from Pearson where _TYPE_ = 'STD');
Grade_8_zscore = (Grade_8ij-Grade_8_Mean)/Grade_8_STD;
```

A z-score of -0.709, corresponding to that of the Mathematics TAKS Grade 8 passing score, was used as the passing score. The remaining variables were updated using the SAS code below:

```
Stud_SES_Effectti = B11*SESti;                                            (22)
Student_Effectti = Stud_SES_Effectti + Student_Random_Effectti;
Grade_3ij = B01 * Grade_2ij + u0j + r0i + e0ij;
Grade_4ij = 0.63 + Grade_3ij - e0ij + Student_Effectti + Campus_Effect1j + e1ij;
Grade_5ij = 0.63 + Grade_4ij - e1ij + Student_Effectti + Campus_Effect2j + e2ij;
Grade_6ij = 0.63 + Grade_5ij - e2ij + Student_Effectti + Campus_Effect3j + e3ij;
Grade_7ij = 0.63 + Grade_6ij - e3ij + Student_Effectti + Campus_Effect4j + e4ij;
Grade_8ij = 0.63 + Grade_7ij - e4ij + Student_Effectti + Campus_Effect5j + e5ij;
Grade_8_Pass = 1 where Grade_8_zscore >= -0.709;
Grade_8_Pass = 0 where Grade_8_zscore < -0.709
```

Estimating Campus Valued-Added Coefficients (Beta Values)

To conduct a Hierarchical Linear Model (HLM) analysis on these data in the Stud_Scores table, these data had to be converted so that each student had six records—one for each grade level score. The Stud_Scores_Array table was created and populated with SES_j_Percent (or Camp_Econ_Dis), SES_{ti}, Time, and Score data, so that grade levels corresponded to time, denoted by 0 through 5. See Table 6 earlier for an example. Using the SAS procedure for conducting an HLM analysis, PROC MIXED, value-added coefficients (Beta values) were calculated using the SAS code below:

```
PROC MIXED DATA=Work.Stud_Scores_Array_1                    (23)
COVTEST NOCLPRINT  method = ML;
CLASS Stud_ID;
MODEL Score = Time SESti Camp_Econ_Dis Camp_Econ_Dis*Time
Camp_Econ_Dis*SESti/ SOLUTION;
RANDOM intercept Time SESti;
REPEATED /subject = Stud_ID type = cs rcorr;

ods output solutionf=Work.SF(keep=effect estimate rename=(estimate=overall));
run;
```

Ranking Campuses by Mean Student Residuals (Value-Added Accountability Model)

After the value-added coefficients (Beta values) were generated using the PROC MIXED procedure above, they were used to update the following fields in the Stud_Scores table using the SAS code below.

```
Intercept_Beta  (β00) = intercept;                          (24)
Camp_Econ_Dis_Beta (β01) = Camp_Econ_Dis ;
Time_Beta (β10) = Time;
Camp_Econ_Dis_Time_Beta  (β11) = Camp_Econ_Dis*Time;
Stud_Econ_Dis_Beta  (β20) = SESti;
Camp_Econ_Dis_Stud_Econ_Dis_Beta  (β21)= Camp_Econ_Dis*SESti;
```

The Predicted_Score and Residual fields were then updated using the SAS code below:

```
Predicted_Score=Intercept_Beta + Camp_Econ_Dis_Beta*SESj    (25)
+ Time_Beta*5 + Camp_Econ_Dis_Time_Beta*SESj*5
+ Camp_Econ_Dis_Stud_Econ_Dis_Beta*SESj*SESti
```

+ Stud_Econ_Dis_Beta* SES_{ti};

Residual = $Grade\_8_{ij}$ - Predicted_Score; (26)

The residuals were then averaged for each campus and the campuses were ranked within each district under the 5 by 5 design such that the campus with the highest mean residual was ranked 1 and the campus with the lowest mean residual was ranked 50.

Ranking Campuses by Low SES Passing Percentages (Traditional SES Accountability Model)

The students in the Stud_Scores table already had both SES_{ti} and the *Grade_8_Pass* indictors populated. These two fields were then used to generate the percent of low SES students passing on each campus. The campuses were then ranked a second time within each district under the 5 by 5 design such that the campus with the highest percent of low SES students with a passing score was ranked 1 and the campus with the lowest percent of low SES students with a passing score was ranked 50. These data for the two campus rankings were collected in the Camp_Rank table with the fields below:

Run_ID (27)
CCRE_ID
Dist_ID
Camp_ID
Camp_VA_Average
Camp_VA_Rank
Camp_Econ_Average
Camp_Econ_Rank

Calculating a Pearson r Correlation Between the Two Rankings

Each campus now had two rankings within each district under the 5 by 5 design—one using a value-added accountability system model and the other using a traditional SES accountability system model. These two rankings were then used to calculate a Pearson *r* correlation for each district with the SAS code below:

```
proc corr data=Work.Camp_Rank_Temp noprint outp=Pearson;                (28)
var Camp_VA_Rank Camp_Econ_Rank; run;
```

These data were collected in the Correlations table with the fields below:

```
Run_ID                                                                  (29)
Sim_ID
Dist_ID
Correlation
```

Analysis

The present study focused on a comparison of campus rankings using two different accountability system models under the 5 by 5 design. The first model ranked the 50 campuses of each district using a campus value-added accountability system model. The second model ranked the 50 campuses of each district using the traditional SES accountability system model. A Pearson *r* correlation was then calculated between the campus rankings of these two models. Note that a Spearman rho equals a Pearson *r* if the two variables have identical distribution shapes. It is this Pearson *r* correlation that is of interest in this analysis. Each of the 5 CCRE's was investigated using 5 districts with different distributions of low SES students resulting in 25 different combinations (e.g., CCRE 1, District 1; CCRE 1, District 2;…CCRE 5, District 4; CCRE 5, District 5). Each combination, or design, was run 100 times so when reviewing results, the aggregate can be thought of as a count (whole number) or a percentage of the runs. The correlations were summed by grouped intervals of 0.10 (e.g., .00 to < .10 and .10 to < .20). The aggregates of these simulations are tabulated in Table 8 through Table 12 below.

Table 8

CCRE 1: Pairwise Correlation (r_{tt}) Between Campus Value-Added Variables (u_{tj}): r_{tt}^2 = .100.

	District				
r^a	1	2	3	4	5
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	1	1	3
.50 to <.60	0	1	7	14	28
.60 to <.70	2	10	32	37	48
.70 to <.80	10	40	38	42	20
.80 to <.90	51	44	22	6	1
.90 to 1.00	37	5	0	0	0
Mean: r^2	0.75	0.63	0.53	0.48	0.41
Standard Deviation: r^2	0.12	0.12	0.12	0.11	0.10

[a] Correlation between model 2 and model 3 campus rankings

Table 9

CCRE 2: Pairwise Correlation (r_{tt}) Between Campus Value-Added Variables (u_{tj}): r_{tt}^2 = .300.

	District				
r^a	1	2	3	4	5
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	0	0	0
.50 to <.60	0	0	0	7	10
.60 to <.70	0	3	12	21	44
.70 to <.80	6	32	42	55	40
.80 to <.90	27	49	38	15	6
.90 to 1.00	67	16	8	2	0
Mean: r^2	0.82	0.69	0.63	0.54	0.48
Standard Deviation: r^2	0.09	0.11	0.12	0.11	0.10

[a] Correlation between model 2 and model 3 campus rankings

Table 10

CCRE 3: Pairwise Correlation (r_{tt}) Between Campus Value-Added Variables (u_{tj}): r_{tt}^2 = .500.

	District				
r^a	1	2	3	4	5
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	0	0	0
.50 to <.60	0	0	1	0	8
.60 to <.70	0	2	8	19	27
.70 to <.80	5	27	31	51	42
.80 to <.90	26	52	53	27	22
.90 to 1.00	69	19	7	3	1
Mean: r^2	0.83	0.71	0.65	0.58	0.53
Standard Deviation: r^2	0.10	0.11	0.11	0.11	0.11

[a] Correlation between model 2 and model 3 campus rankings

Table 11

CCRE 4: Pairwise Correlation (r_{tt}) Between Campus Value-Added Variables (u_{tj}): r_{tt}^2 = .700.

	District				
r^a	1	2	3	4	5
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	1	0	0
.50 to <.60	0	0	0	1	2
.60 to <.70	0	1	9	11	20
.70 to <.80	1	15	28	42	47
.80 to <.90	25	54	50	43	27
.90 to 1.00	74	30	12	3	4
Mean: r^2	0.85	0.74	0.66	0.61	0.57
Standard Deviation: r^2	0.08	0.11	0.13	0.11	0.12

[a] Correlation between model 2 and model 3 campus rankings

Table 12

CCRE 5: Pairwise Correlation (r_{tt}) Between Campus Value-Added Variables (u_{tj}): r_{tt}^2 = .900.

	District				
r[a]	1	2	3	4	5
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	0	0	0
.50 to <.60	0	0	0	2	0
.60 to <.70	0	0	4	9	16
.70 to <.80	1	9	27	38	39
.80 to <.90	28	55	52	45	39
.90 to 1.00	71	36	17	6	6
Mean: r^2	0.85	0.76	0.69	0.63	0.61
Standard Deviation: r^2	0.09	0.09	0.11	0.12	0.12

[a] Correlation between model 2 and model 3 campus rankings

These 25 combinations, or designs, were then analyzed. First, one CCRE was analyzed at a time at the different distribution levels of low SES. Second, one distribution level of low SES was analyzed at a time at the different levels of CCRE. Third, the extremes (i.e., low CCRE, narrow distribution of low SES to high CCRE, wide distribution of low SES) were analyzed. The Pearson *r* correlation between the two rankings of the two models is the variable of interest in these analyses. The simulation was run 100 times so there were 100 Pearson *r* correlations to analyze under each of the 25 combinations.

Results

Accuracy of Estimated Parameters

The present study used a Monte Carlo simulation to generate the data for analysis. These data were first analyzed for accuracy in representing the desired correlations and population distributions. The total data set consists of 37,500,000 records so only data in run 50 (out of runs 1 through 100) were analyzed for the accuracy of estimated parameters. See

Tables 13 through 15. In most cases, the sample estimate was close to the true value of the parameter; however, there were some exceptions. For example, the skewness and kurtosis of *SESj_Percent* where *Run ID* equals 50 and *Dist_ID* equals 5 are 0.306 and 0.799, respectively. The true value of each of these is zero. In addition the five pairwise bivariate Pearson *r* correlations of u_{tj} where *Run_ID* equals 50 were higher than expected; however, sampling other runs of the simulation shows these estimates to be more within range.

Varying Only the Value-Added Variable

The value-added parameter was the correlation of campus random effects (CCRE), or r_{tt} (pairwise bivariate Pearson *r* correlation of u_{tj} where t=0 through 5 and j=the campus). The variable, r_{tt}, was allowed to vary between five values such that r_{tt}^2 =.100, r_{tt}^2=.300, r_{tt}^2=.500, r_{tt}^2=.700, and r_{tt}^2=.900. In this first analysis, trends were examined where only r_{tt} varied within each district. District 1 data in Tables 8 through 12 were analyzed by aggregating these data; see Table 16. As r_{tt} increased, the mean of the effect sizes, r^2, of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model also increased for District 1. This is also true when examining Districts 2 through 5 independently. The difference in these distributions of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model were also tested with a Pearson Chi Square test which showed the difference in these distributions to be statistically significant at α=.01. See Table 17. However, some of the cells for this test have very small expected values which can be problematic when conducting the Pearson Chi Square test. Cochran (1952) stated that "since χ^2 has been established as the limiting distribution of X^2 in large samples, it is customary to recommend, in applications of the test, that the smallest expected number in any class should

Table 13

Descriptive Statistics of Variables Using Run_ID =50 as a Sample

Parameter	Description	n	Mean	SD	Skewness	Kurtosis
u_{0j}	Variance of campus random effect on the baseline intercept at grade 3	375,000	0.03[a] (0.00)[b]	0.35 (0.32)	0.23 (0.00)	0.04 (0.00)
u_{1j}	Variance of value-added effect at time 1	375,000	0.03 (0.00)	0.63 (0.56)	0.02 (0.00)	0.13 (0.00)
u_{2j}	Variance of value-added effect at time 2	375,000	0.01 (0.00)	0.61 (0.56)	0.03 (0.00)	-0.27 (0.00)
u_{3j}	Variance of value-added effect at time 3	375,000	0.05 (0.00)	0.63 (0.56)	0.12 (0.00)	-0.19 (0.00)
u_{4j}	Variance of value-added effect at time 4	375,000	0.04 (0.00)	0.60 (0.56)	-0.03 (0.00)	0.05 (0.00)
u_{5j}	Variance of value-added effect at time 5	375,000	0.03 (0.00)	0.64 (0.56)	-0.19 (0.00)	-0.22 (0.00)
r_{0i}	Variance of student random effect on the baseline intercept at Grade 3	375,000	0.00 (0.00)	0.47 (0.47)	0.00 (0.00)	0.02 (0.00)
r_{1i}	Variance of random individual student natural growth	375,000	-0.02 (0.00)	0.57 (0.57)	0.03 (0.00)	-0.03 (0.00)
$Grade\_2_{ij}$	Grade 2 scores	375,000	0.00 (0.00)	1.00 (1.00)	-0.01 (0.00)	-0.00 (0.00)
Y_{0ij}	Standardized TAKS Math score at time *0*, for student *i*, on campus *j* (Grade 3 Score)	375,000	0.03 (0.00)	1.02 (1.00)	0.02 (0.00)	0.00 (0.00)
e_{0ij}	Error score at grade 3	375,000	-0.00 (0.00)	0.32 (0.32)	-0.00 (0.00)	-0.00 (0.00)
e_{1ij}	Error score at grade 4	375,000	-0.00 (0.00)	0.32 (0.32)	0.01 (0.00)	-0.01 (0.00)
e_{2ij}	Error score at grade 5	375,000	0.00 (0.00)	0.32 (0.32)	-0.01 (0.00)	0.00 (0.00)
e_{3ij}	Error score at grade 6	375,000	-0.00 (0.00)	0.32 (0.32)	0.00 (0.00)	0.01 (0.00)
e_{4ij}	Error score at	375,000	0.00	0.32	0.01	-0.01

Table 13 Continued

Parameter	Description	n	Mean	SD	Skewness	Kurtosis
	grade 7		(0.00)	(0.32)	(0.00)	(0.00)
e_{5ij}	Error score at grade 8	375,000	0.00 (0.00)	0.32 (0.32)	-0.00 (0.00)	0.00 (0.00)

[a] Sample estimate; [b] True value

Table 14

Descriptive Statistics of Variables Using Run_ID=50 and Dist_ID=1 through 5 as Samples

Parameter	Description	*Dist_ID*	*n*	Mean	SD	Skewness	Kurtosis
SES_j_Percent	Percent of low SES students on campus *j*	1	75,000	50.39[a] (50.00)[b]	4.06 (3.95)	-0.06 (0.00)	-0.09 (0.00)
SES_j_Percent	Percent of low SES students on campus *j*	2	75,000	49.61 (50.00)	6.87 (6.85)	-0.03 (0.00)	-0.02 (0.00)
SES_j_Percent	Percent of low SES students on campus *j*	3	75,000	50.63 (50.00)	8.45 (8.84)	0.14 (0.00)	-0.48 (0.00)
SES_j_Percent	Percent of low SES students on campus *j*	4	75,000	50.58 (50.00)	10.34 (10.46)	-0.00 (0.00)	0.06 (0.00)
SES_j_Percent	Percent of low SES students on campus *j*	5	75,000	48.65 (50.00)	10.70 (11.86)	0.31 (0.00)	0.80 (0.00)

[a] Sample estimate; [b] True value

be 10 or (with some writers) 5” (p. 328). The Fisher Exact test (FET) provides an alternative to the Chi-Square test when cells have these low expected values (Bolboaca, Jantschi, Sestras, Sestras, & Pamfil, 2011; Fisher, 1935; Roscoe & Byars, 1971).

Therefore, the Fisher Exact test was conducted in addition to the Pearson Chi-Square test to account for these cells with small expected values. As one can see from the SPSS output in Table 17, both tests produced the same final results. The difference in the distributions of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model were statistically significant at $\alpha=.01$. One can see similar results in Tables 18 through 25 for Districts 2 through 5. The distribution of these Pearson *r* correlations are dependent on the value-added parameter represented by the correlation of campus random effects (CCRE), or r_{tt}, for each of the five districts.

Table 15

Correlations of Variables Using Run_ID=50 as a Sample

Parameter	Description	*Run_ID*	*CCRE_ID*	*Dist_ID*	*n*	*r*
$r_{ses*grade\ 2}$	Correlation between campus SES and school mean grade 2 test score	50	All	1	75,000	-0.94[a] (-0.93)[b]
r_{tt}	Pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5	50	1	All	75,000	0.26[c] to 0.55[d] (0.32)
r_{tt}	Pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5	50	2	All	75,000	0.53 to 0.73 (0.55)
r_{tt}	Pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5	50	3	All	75,000	0.70 to 0.83 (0.71)
r_{tt}	Pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5	50	4	All	75,000	0.84 to 0.91 (0.84)
r_{tt}	Pairwise bivariate Pearson correlation of u_{tj} where t=0, t=1, t=2, t= 3, t=4, t=5	50	5	All	75,000	0.95 to 0.97 (0.95)
$r_{grade\ 2}$	Intraclass correlation of grade 2 test scores	50	All	1	75,000	0.21 (0.21)

[a] Sample estimate; [b] True value; [c] Smallest correlation; [d] Largest correlation

Table 16

District 1: Distribution of Pearson r Correlations for each Pairwise Correlation (r_{tt})

r[a]	r_{tt}^2 0.10	0.30	0.50	0.70	0.90
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	0	0	0
.50 to <.60	0	0	0	0	0
.60 to <.70	2	0	0	0	0
.70 to <.80	10	6	5	1	1
.80 to <.90	51	27	26	25	28
.90 to 1.00	37	67	69	74	71
Mean: r^2	0.75	0.82	0.83	0.85	0.85
Standard Deviation: r^2	0.12	0.09	0.10	0.08	0.09

[a] Correlation between model 2 and model 3 campus rankings

Table 17

District 1: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	50.21[a]	12	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	49.76	12	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	45.84			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]10 cells (50%) have expected count less than 5. The minimum expected count is .40. [b]Based on 10000 sampled tables with starting seed 1314643744.

Table 18

District 2: Distribution of Pearson r Correlations for Each Pairwise Correlation (r_{tt})

	r_{tt}^2				
r^a	0.10	0.30	0.50	0.70	0.90
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	0	0	0	0	0
.50 to <.60	1	0	0	0	0
.60 to <.70	10	3	2	1	0
.70 to <.80	40	32	27	15	9
.80 to <.90	44	49	52	54	55
.90 to 1.00	5	16	19	30	36
Mean: r^2	0.63	0.69	0.71	0.74	0.76
Standard Deviation: r^2	0.12	0.11	0.11	0.11	0.09

[a] Correlation between model 2 and model 3 campus rankings

Table 19

District 2: Pearson Chi-Square Test and Fisher's Exact Test

				Monte Carlo Sig. (2-sided)		
			Asymp. Sig.		99% Confidence	
Test	Value	df	(2-sided)	Sig.	Lower	Upper
Pearson Chi-Square	78.78[a]	16	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	82.15	16	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	78.07			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]10 cells (40%) have expected count less than 5. The minimum expected count is .20. [b]Based on 10000 sampled tables with starting seed 2000000.

Table 20

District 3: Distribution of Pearson r Correlations for each Pairwise Correlation (r_{tt})

r[a]	r_{tt}^2 0.10	0.30	0.50	0.70	0.90
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	1	0	0	1	0
.50 to <.60	7	0	1	0	0
.60 to <.70	32	12	8	9	4
.70 to <.80	38	42	31	28	27
.80 to <.90	22	38	53	50	52
.90 to 1.00	0	8	7	12	17
Mean: r^2	0.53	0.63	0.65	0.66	0.69
Standard Deviation: r^2	0.12	0.12	0.11	0.13	0.11

[a] Correlation between model 2 and model 3 campus rankings

Table 21

District 3: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	102.86[a]	20	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	103.59	20	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	92.45			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]10 cells (33.3%) have expected count less than 5. The minimum expected count is .40. [b]Based on 10000 sampled tables with starting seed 92208573.

Table 22

District 4: Distribution of Pearson r Correlations for Each Pairwise Correlation (r_{tt})

	r_{tt}^2				
r^a	0.10	0.30	0.50	0.70	0.90
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	1	0	0	0	0
.50 to <.60	14	7	0	1	2
.60 to <.70	37	21	19	11	9
.70 to <.80	42	55	51	42	38
.80 to <.90	6	15	27	43	45
.90 to 1.00	0	2	3	3	6
Mean: r^2	0.48	0.54	0.58	0.61	0.63
Standard Deviation: r^2	0.11	0.11	0.11	0.11	0.12

[a] Correlation between model 2 and model 3 campus rankings

Table 23

District 4: Pearson Chi-Square Test and Fisher's Exact Test

				Monte Carlo Sig. (2-sided)		
			Asymp. Sig.		99% Confidence	
Test	Value	df	(2-sided)	Sig.	Lower	Upper
Pearson Chi-Square	111.35[a]	20	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	117.14	20	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	109.89			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]15 cells (50%) have expected count less than 5. The minimum expected count is .20. [b]Based on 10000 sampled tables with starting seed 2000000.

Table 24

District 5: Distribution of Pearson r Correlations for Each Pairwise Correlation (r_{tt})

	r_{tt}^2				
r^a	0.10	0.30	0.50	0.70	0.90
<0.0	0	0	0	0	0
.00 to <.10	0	0	0	0	0
.10 to <.20	0	0	0	0	0
.20 to <.30	0	0	0	0	0
.30 to <.40	0	0	0	0	0
.40 to <.50	3	0	0	0	0
.50 to <.60	28	10	8	2	0
.60 to <.70	48	44	27	20	16
.70 to <.80	20	40	42	47	39
.80 to <.90	1	6	22	27	39
.90 to 1.00	0	0	1	4	6
Mean: r^2	0.41	0.48	0.53	0.57	0.61
Standard Deviation: r^2	0.10	0.10	0.11	0.12	0.12

[a] Correlation between model 2 and model 3 campus rankings

Table 25

District 5: Pearson Chi-Square Test and Fisher's Exact Test

				Monte Carlo Sig. (2-sided)		
			Asymp. Sig.		99% Confidence	
Test	Value	df	(2-sided)	Sig.	Lower	Upper
Pearson Chi-Square	164.86[a]	20	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	177.73	20	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	161.91			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]10 cells (33.3%) have expected count less than 5. The minimum expected count is .60. [b]Based on 10000 sampled tables with starting seed 1314643744.

In this analysis, the distribution of the percentage of low SES students was held constant while the campus value-added effect, $r_{tt,}$ varied. A possible explanation of the observed trends is that when the campus value-added effect, $r_{tt,}$ is small, the distribution of the percentage of low SES students in a district has a greater effect on campus level test score aggregates than campus value-added effects have, thus the two models yield different rankings. As the campus value-added effect increases, the campus value-added effect offsets the effect of the distribution of the percentage of low SES students in the district, thus yielding similar rankings between the two models. This was the case regardless of the distribution of the percentage of low SES students within a district; however, this was more evident with the wider distributions of the percentage of low SES students.

Varying Only the Standard Deviation of SES_j_Percent

The percentage of low SES students parameter was SES_j*_Percent*. The distribution of the variable, SES_j*_Percent*, was allowed to vary between five values: σ^2=.100, σ^2=.300, σ^2=.500, σ^2=.700 and σ^2=.900, identified by Districts 1 through 5, respectively. In this second analysis, trends were examined when only σ^2 varied (from district to district) for each campus value-added effect, r_{tt}. For example, r_{tt}^2=.100 data in Table 8 were analyzed. As one can see by visual inspection of Districts 1 through 5, as σ^2 increased, the mean of the effects sizes, r^2, of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model decreased for r_{tt}=.100. This is also true when examining r_{tt}^2 = .300, .500, .700, and .900, independently. The difference in these distributions of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model were also tested with a Pearson Chi Square test which showed the difference in these distributions to be statistically significant at α=.01.

See Table 26. However, again some of the cells for this test have very small expected values which can be problematic when conducting the Pearson Chi Square test.

Therefore, the Fisher Exact test was again conducted in addition to the Pearson Chi-Square test to account for these cells with small expected values. As seen from the SPSS output in Table 26, both tests produced the same final results. The difference in the distributions of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model was statistically significant at α=.01. One can see similar results in Tables 27 through 30 for r_{tt}^2 = .300, .500, .700, and .900. The distribution of these Pearson *r* correlations are dependent on the distribution of the percentage of low SES students within each district for each of the five value-added parameters represented by the correlation of campus random effects (CCRE), or r_{tt}.

Table 26

CCRE 1: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	346.92[a]	20	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	365.81	20	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	341.38			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]5 cells (16.7%) have expected count less than 5. The minimum expected count is 1.00. [b]Based on 10000 sampled tables with starting seed 2000000.

Table 27

CCRE 2: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	352.62[a]	16	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	357.18	16	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	335.88			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]5 cells (20.0%) have expected count less than 5. The minimum expected count is 3.40. [b]Based on 10000 sampled tables with starting seed 1314643744.

Table 28

CCRE 3: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	301.65[a]	16	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	290.57	16	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	273.65			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]5 cells (20.0%) have expected count less than 5. The minimum expected count is 1.80. [b]Based on 10000 sampled tables with starting seed 1502173562.

Table 29

CCRE 4: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	256.82[a]	20	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	268.39	20	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	256.40			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]10 cells (33.3%) have expected count less than 5. The minimum expected count is .20. [b]Based on 10000 sampled tables with starting seed 112562564.

Table 30

CCRE 5: Pearson Chi-Square Test and Fisher's Exact Test

Test	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Lower	Upper
Pearson Chi-Square	212.38[a]	16	<.01	<.01[b]	<.01	<.01
Likelihood Ratio	226.07	16	<.01	<.01[b]	<.01	<.01
Fisher's Exact Test	213.00			<.01[b]	<.01	<.01
N of Valid Cases	500					

[a]5 cells (20.0%) have expected count less than 5. The minimum expected count is .40. [b]Based on 10000 sampled tables with starting seed 92208573.

In this analysis, the campus value-added effect, $r_{tt,}$ was held constant while the distribution of the percentage of low SES students varied. A possible explanation of the observed trends is that when the distribution of the percentage of low SES students in a district is narrow, the campus value-added effect, $r_{tt,}$ has a similar effect on campus level test score aggregates as the distribution of the percentage of low SES students in the district have, thus the two models yield similar rankings. As the distribution of the percentage of low SES students increases, the campus value-added effect has a lesser effect compared to the effect of the distribution of the percentage of low SES students, thus yielding dissimilar rankings between the two models. This was the case regardless of the campus value-added effect; however, this was less evident with the greater values of campus value-added effect.

Varying Both the Value-Added Variable and the Standard Deviation of SES_j_Percent

When both the campus value-added parameter, $r_{tt,}$ and the distribution of the percentage of low SES students in the district, *SES_j_Percent*, were allowed to vary, the trends of most interest were between the extremes of the models: a narrow distribution of the percentage of low SES students in the district, *SES_j_Percent*, with a small campus value-added parameter, $r_{tt,}$ to a wide distribution of the percentage of low SES students in the district, SES_j*_Percent*, with a large campus value-added parameter, $r_{tt,}$ (see Figure 1, Trend 1); and, a large campus value-added parameter, $r_{tt,}$ with a narrow distribution of the percentage of low SES students in the district, SES_j*_Percent,* to a small campus value-added parameter, $r_{tt,}$ with a wide distribution of the percentage of low SES students in the district, SES_j*_Percent* (see Figure 1, Trend 2).

Conceptually, these two extremes are of most interest because as one variable changes, it may interact or counteract the effects of the other variable. By visual inspection of Tables 7 through 11, we can see this is true. Comparing the distribution of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model in Table 8 for District 1 where the mean of r^2 equals 0.75 to that of Table 12 District 5 where the mean of r^2 equals 0.61 (which corresponds to Figure 1,

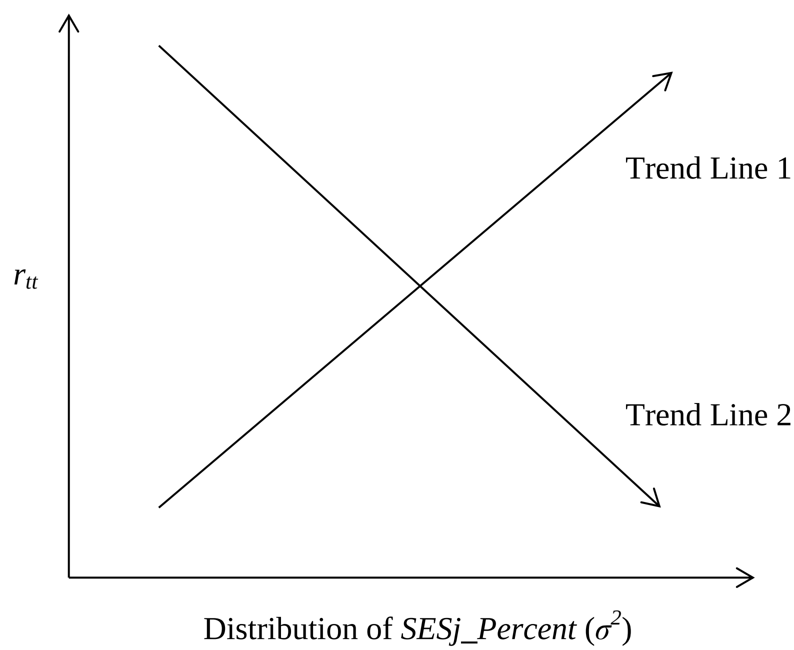

Figure 1. Distribution of *SESj_Percent* (σ^2) vs r_{tt}

Trend 1), one can see the distributions are different, but not too very different. This could represent a counteracting effect of the two variables. However, when comparing the distribution of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model in Table 8 for District 5 where the mean of r^2 equals 0.41 to that of Table 12 for District 1 where the mean of r^2 equals 0.85 (which corresponds to Figure 1, Trend 2), one can see the distributions are very different. This could represent an interaction effect of the two variables. This last comparison represents the most extreme differences in the distribution of the Pearson *r* correlations between the value-added accountability system model and the SES accountability system model of all cases examined in the present study.

Conclusions and Discussions

Implications for Practice

The main conclusion of the present Monte Carlo study is that there is a mismatch between the rankings resulting from value-added accountability system models and the rankings resulting from SES accountability system models. Research has been published on the effects of low SES students on standardized test scores (Aikens & Barbarin, 2008; Roberts & Bryant, 2011); however, many accountability models use a status-based indicator that is affected by the SES level of students. This lends itself to the possibility that such models are measuring SES (at both the campus level and the student level) rather than the achievement of students enrolled at the campus. By contrast, even though value-added models are far from perfect, they do attempt to measure achievement gained by students in one year, accounting for outside factors that may affect such achievement. To this end, this mismatch may indicate that although value-added models are not perfect, status-based models are even less desirable.

Campuses with a small percentage of low SES students have an unfair advantage over campuses with a large percentage of low SES students under the SES accountability system model. The evidence of this is two-fold. First, as mentioned above, the literature shows that low SES students do not perform well on standardized tests. Second, in the present study, the mismatch between the rankings resulting from value-added accountability system models and the rankings resulting from SES accountability system models was the greatest when the distribution of the percentage of low SES students in the district was wide (i.e., the district had campuses with a low percentage of low SES students and at the same time had campuses with a high percentage of low SES students), the campus value-added effect was small, and all other variables were held constant. For example, the average effect size, r^2, of the Pearson *r* correlations between the two rankings for District 1 (the narrowest distribution of the percentage of low SES students) in Table 8 (where the campus value-added effect was the smallest) was 0.75 and the average effect size, r^2, of the Pearson *r* correlations between the two rankings for District 5 (the widest distribution of the percentage of low SES students) in Table 8 was 0.41. The only variable that was allowed to vary in this case was the distribution of the percentage of low SES students. This is the most extreme case in the present study; however, this case demonstrates what could happen in real life with a real collection of campuses where the distribution of the percentage of low SES students can be very wide. Similar trends are seen for the remaining campus value-added effects examined in Tables 9 thru 12: 0.82 vs. 0.48, 0.83 vs. 0.53, 0.85 vs. 0.57, and 0.85 vs.0.61. District 1 consistently has a higher Pearson *r* correlation between the two rankings than does District 5. A discrepancy between the ways the two systems rank campuses in practice could have a detrimental effect on teacher and/or principal evaluations.

SES accountability system models have a tendency to place campuses with smaller percentages of low SES students at higher rankings than campuses with larger percentages of low SES students—these could be false positives. Conversely, under these same conditions, SES accountability system models have a tendency to place campuses with larger percentages of low SES students at lower rankings than campuses with smaller percentages of low SES students— these could be false negatives. As an example, Table 31 shows campuses categorized into rankings of less than or equal to 25 and rankings of more than 25 (using the SES accountability system model) based on the percentage of low SES students on the campus for runs 41 through 50 where CCRE equals 1 and District equals 5. As previously mentioned, cases where CCRE equals 1 and District equals 5 have the lowest Pearson *r* correlations between the value-added accountability system model and the SES accountability system model. Data show the advantage campuses with less than 50% low SES have under the SES accountability system model when all other parameters are held constant. For campuses with less than or equal to 50% low SES, 197 had a rank of less than or equal to 25 (i.e., ranked better than half the campuses). For campuses with more than 50% low SES, 75 had a rank of less than or equal to 25. This discrepancy is evidence of the effect low SES has on campus rankings under a status-based model.

Table 31

Ranks Using the SES Accountability System Model for Runs 41 Through 50, CCRE=1, District=5

Percentage of low SES	Campuses with a rank less than or equal to 25	Campuses with a rank more than 25
Campuses with less than or equal to 50% low SES	197	59
Campuses with more than 50% low SES	75	169

In practice, these false positives and false negatives can be quite troubling. Annually (by August 8th), the Texas Education Agency publishes ratings based off, in part, assessment scores for every campus and district in Texas (Texas Education Code, 2014). These ratings may affect the evaluations of teacher, principals, and superintendents, and are thus considered high stakes ratings. Many educators recognize the false positives and false negatives intuitively. For example, a teacher with a class of no low SES students may not show any real gains in assessment scores, but have all of his/her students pass the assessment. At the same time, a teacher with a class of all low SES students may have all students show gains in assessment scores, but have only a portion of his/her students pass the assessment. Under the status-based systems (e.g., SES accountability system models), the former teacher will receive more positive feedback in informal or formal evaluations than the latter. These inequities cause concerns for teachers, principals, and superintendents. The advantage of a value-added system is that it at least makes an attempt to balance these inequities.

Limitations and Future Studies

As with any study, there are limitations that need to be considered and investigated in future studies. The present Monte Carlo study had three major limitations: ecological validity, appropriate variable determination, and validity using an outside measure.

Limitation 1: Ecological validity.

Although efforts were made to ensure the data generated for usage in the present Monte Carlo study were ecologically valid, there are inherent problems to Monte Carlo studies with respect to ecological validity. Parameters used were supported in previous research (Bosker & Witziers, 1995; Caldas & Bankston, 1997; Darandari, 2004; Ponisciak & Bryk, 2005; Wang, 2006); however, simulation data can rarely substitute perfectly for field data (MacCallum, 2003; Myers, Ahn, & Ying, 2011). For example, assumptions were made in the present Monte Carlo study that the data were normally distributed. Rarely in field data would the data perfectly fit a normal distribution. By extension, rarely would one set of field data perfectly fit the distribution of another set of field data. Even if the distributions in the present Monte Carlo study were set to the parameters of a specific set of field data, they still may not be generalizable to all field data. A future study may include an investigation of various distributions of data rather than just the normal distribution.

Limitation 2: Appropriate variable determination.

Although the present Monte Carlo study focused on prior attainment and the low SES variable at both the campus level and the student level, there are many other variables that could have been used in the model as well. Research has been conducted on value-added models using numerous variables (Choi & Seltzer, 2003; Fortier, Vallerand, & Guay, 1995; Teddlie, Stringfield, & Reynolds, 2000; Thomas & Mortimore, 1996; Willms, 2000); however, prior attainment and low SES are recognized as two of the most impactful (Aikens & Barbarin, 2008; Roberts & Bryant, 2011), thus their selection for the model in the present Monte Carlo study. Family education, neighborhoods, and Limited English Proficient (LEP) status could have been included in the present Monte Carlo study. A future study may include an investigation of other variables in the value-added model as well.

Limitation 3: Validity using an outside measure.

At the center of any accountability system model should be student achievement; however, determining how to measure student achievement is quite challenging. The present Monte Carlo study assumes that student achievement is somehow measured by achievement tests; however, comparing scores from achievement tests (and by extension teacher evaluations derived from those scores) to an outside measure would provide some validity (low or high) to that assumption (Amrein-Beardsley, 2008). For example, including a third accountability system model that uses principal evaluations or an accountability system model that uses the success of students after they leave public schools might prove useful. A future study may include an investigation of various outside measures other than achievement test scores. However, these data are much more difficult to simulate (or obtain), thus they were not included in the present Monte Carlo study.

CHAPTER IV

FIELD STUDY: A VALUE-ADDED MODEL

The present field study focused on the correlation between campus rankings of a district when the campuses were ranked using two different accountability systems: a valued-added accountability system and an accountability system that uses the success of economically disadvantaged students as a key accountability indicator (i.e., an SES accountability system model). The campus data used in the present field study were collected from the 9 middle school campuses in a large school district in Texas. This district will be referenced as the Field Study District.

Grade 3 scores through grade 8 scores were collected on each student of the 9 middle schools as well as the SES indicator of each student in the Field Study District. This allowed for longitudinal data on each student ($time_{0i}$ for grade 3 through $time_{5i}$ for grade 8) to be included in the analysis as well as the SES of the student and the aggregate SES of each campus. The scores of interest were scores from the Mathematics portion of the Texas Assessment of Knowledge and Skills (TAKS). For the purpose of the present study, students did not have missing data, were assumed to be continuously enrolled on the same campus (grade 3 through grade 8) each year using the middle school of enrollment as their campus, and were continuously classified as either economically disadvantaged or non-economically disadvantaged throughout the study. These collected scores were used in the following Hierarchical Linear Model (HLM) (Hox, 2010) to determine Beta values that could be used to generate a predicted grade 8 score and a grade 8 residual (grade 8 score – grade 8 predicted score) for each student.

$$Y_{ti} = \pi_{0i} + \pi_{1i} * time_{ti} + \pi_{2i} * SES_{ti} + e_{ti} \quad (30)$$

$$\pi_{0i} = \beta_{00} + \beta_{01} * SES_j + u_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * SES_j + u_{1i}$$
$$\pi_{2i} = \beta_{20} + \beta_{21} * SES_j + u_{2i}$$

where Y_{ti} = Mathematics TAKS (NCE) score at time t, for student i

$time_{ti}$ = Grade 3 (t = 0) through Grade 8 (t = 5)

SES_{ti} = Student SES indicator

SES_j = Percent of low SES students on campus j

The residuals from this value-added accountability system model were then used to rank the 9 campuses. The campus with the highest campus mean residual was ranked 1 and the campus with the lowest campus mean residual was ranked 9. The campuses were then ranked a second time using an SES accountability system model. The campus with the highest percentage of low SES students with a passing score was ranked 1 and the campus with the lowest percentage of low SES students with a passing score was ranked 9. A Pearson r correlation was then calculated between the campus rankings of these two models. Note that a Spearman rho equals a Pearson r if the two variables have identical distribution shapes. This chapter is divided into two sections: field study procedures and analysis.

Field Study Procedure

The field study procedures can be divided into five parts: collecting student test scores from grade 3 to grade 8; estimating campus value-added coefficients (or β values); ranking campuses by mean student residuals (i.e., value-added accountability system model); ranking campuses by SES passing percentages (i.e., traditional SES accountability system model); and calculating a Pearson r correlation between the two rankings. See Appendix B for the complete SAS code developed by the author for the field study procedure.

Collecting Student Test Scores from Grade 3 to Grade 8

The Field Study District was contacted using the contact procedures on their district website for conducting research. After submitting the appropriate forms and being approved through the appropriate processes, I received a data file with student Mathematics Texas Assessment of Knowledge and Skills (TAKS) scores and imported it into the Field Study District table. See Table 32 for an example of the format of the Field Study District table.

Table 32

Field Study District File: Example of the Format

Camp ID	Stud ID	SES	Grade 3 Raw Score	Grade 4 Raw Score	Grade 5 Raw Score	Grade 6 Raw Score	Grade 7 Raw Score	Grade 8 Raw Score
1	1	1	20	25	24	30	32	40
1	2	0	25	26	25	31	35	40
…								
9	299	0	28	30	34	33	35	41
9	300	1	29	29	33	33	35	38

The Stud_Scores table was created next which contained all the variables needed for the simulation, including the campus level variables. The complete set of Stud_Scores table variables are listed below:

Run_ID (31)
CCRE_ID
Dist_ID
Camp_ID
Stud_ID
SES_j
SES_{ti}
$Grade\_3_{ij}$
$Grade\_4_{ij}$

$Grade\_5_{ij}$
$Grade\_6_{ij}$
$Grade\_7_{ij}$
$Grade\_8_{ij}$
Grade_8_Pass
Intercept_Beta
Camp_Econ_Dis_Beta
Time_Beta
Camp_Econ_Dis_Time_Beta
Stud_Econ_Dis_Beta
Camp_Econ_Dis_Stud_Econ_Dis_Beta
Predicted_Score
Residual

The District Field Study table data were then inserted into the following fields of the Stud_Scores table.

Run_ID (32)
CCRE_ID
Dist_ID
Camp_ID
Stud_ID
SES_i
SES_{ti}
$Grade\_3_{ij}$
$Grade\_4_{ij}$
$Grade\_5_{ij}$
$Grade\_6_{ij}$
$Grade\_7_{ij}$
$Grade\_8_{ij}$

The campus percentage of low SES students, SES_j, was then updated using the SAS code below:

SES_j = (select 100*(sum(b.SES_{ti})/count(b.SES_{ti})) (33)
from Stud_Scores_1 as b where (a.Camp_ID = b.Camp_ID));

The Texas Student Assessment Program: Technical Digests for years 2004 to 2009 (TEA, 2003-2004; TEA, 2004-2005; TEA, 2005-2006; TEA, 2006-2007; TEA, 2007-2008;

TEA, 2008-2009), the years corresponding to the data to be used for the present field study, were examined and two concerns about using Mathematics TAKS scores in the present field study were noted:

- The lack of normality of the Mathematics TAKS scale scores; these data are skewed to the left. See Table 33.
- The weak, if any, linkage between Mathematics TAKS scale scores from one grade level to the next (i.e., no vertical scale).

Table 33

Scale Score Distributions and Statistics

Grade	Year	N	Mean	Standard Deviation	Skewness	Kurtosis
3[a]	2004	271,275	2246.70	178.59	0.30	0.17
4[b]	2005	278,466	2255.65	194.13	0.22	-0.15
5[c]	2006	295,119	2292.90	235.09	0.24	-0.25
6[d]	2007	299,437	2291.46	245.38	0.38	-0.13
7[e]	2008	318,687	2218.88	183.67	0.44	0.11
8[f]	2009	317,831	2240.68	198.52	0.50	0.26

[a] TEA (2003-2004); [b] TEA (2004-2005); [c] TEA (2005-2006); [d] TEA (2006-2007); [e] TEA (2007-2008); [f] TEA (2008-2009)

To address these two concerns, the student Mathematics TAKS raw scores were winsorized and then converted into normal curve equivalent (NCE) scores before entering them into the value-added model.

Winsorizing the scores allowed for some degree of management of the skewness of these data to the left. Wilcox (1995) defined the basic formula for winsorizing a set of data as the following:

$$\bar{X}_w = \frac{1}{n}\left((g+1)X_{(g+1)} + X_{(g+2)} + \cdots + X_{(n-g-1)} + (g+1)X_{(n-g)}\right) \quad (34)$$

> where $X_{(1)} < \bullet \bullet \bullet < X_{(n)}$ are the observations written in ascending order, and $g=[kn]$, where k is some predetermined constant between 0 and .5, and the notation $[kn]$ indicates that kn is rounded down to the nearest integer. (p. 61)

Research shows that trimmed means (winsorized means being special types of trimmed means) are preferred over non-trimmed means when working with skewed data (Barnett, 1978; Gross, 1976; Keselman, Kowalchuk, & Lix, 1998; Keselman, Lix, & Kowalchuk, 1998; Ramsey & Ramsey, 2007; Tukey & McLaughlin 1963; Wilcox, 1994; Wilcox, 1995). Wilcox (1994) listed some of the many reasons why trimmed means are preferred:

> more power when testing hypotheses, the sample trimmed mean has a higher finite sample breakdown point, and the population trimmed mean, μ_1, is closer to the "bulk" of a skewed distribution suggesting that it is a better measure of location than the mean (p. 272).

Keselman, Kowalchuk, and Lix, (1998) added that "the standard error of the trimmed mean is less affected by departures from normality because extreme observations, that is, observations in the tails of a distribution, are censored or removed" (p. 146).

The Mathematics TAKS scores in this left tail are scores of students who did not pass the Mathematics TAKS of which some possible reasons would be guessing, not having enough time, and misconceptions about the content. The probability of these scores representing the students' true scores is very low, thus measurement error would be considered very high in this left tail. Two methods were considered to account for this left tail: trimming the left side of the data and winsorizing the left side of the data. Trimming did not seem to be an appealing option because it would eliminate up to 20.9% of the failed scores depending on the grade level. See Table 34. Thus winsorizing was considered the more appropriate alternative.

Table 34

Trimming Effects

Grade Level	Percent of Scores Lost	Percent of Failed Scores Lost
3	2.7%	15.6%
4	3.1%	16.4%
5	3.5%	18.7%
6	9.9%	20.9%
7	1.0%	4.0%
8	2.1%	10.0%

A special type of winsorizing was used where the "guessing" scores were redistributed in the left tail of the distribution. Scores were considered to be "guessing" scores if they were within two standard errors of measurement of the score associated with guessing. For example, on average a student with no knowledge taking a test with 40 items having four multiple choice answer stems would get 10 out of the 40 questions correct and thus make a raw score of 10. Scores within two standard errors of measurement of 10 were considered guessing scores. TEA (2003) defined the formula for the standard error of measurement (SEM) as:

$$S_E = S_X \sqrt{1 - r_{xx}} \tag{35}$$

where S_E is the standard error of measurement, S_X is the standard deviation of raw scores, and r_{xx} is the reliability of the test scores.

Table 35 shows the standard error of measurement along with the raw scores considered "guessing" scores for each grade level. The Mathematics TAKS raw scores of students with Mathematics TAKS raw scores less than the corresponding guessing raw score plus two SEMs were randomly redistributed among the raw scores in the left tail between the guessing raw score plus two SEMs and the guessing raw score plus four SEMs for the corresponding grade level, thus redistributing 95% of the guessing scores.

Converting the students' Mathematics TAKS raw scores to Normal Curve Equivalent (NCE) scores allowed for some degree of linkage between grade levels. Linn (2000) defined NCEs as "simply normalized standard scores with a mean of 50 and a standard deviation of 21.06, which happens to be the standard deviation that makes NCEs coincide with National

Table 35

"Guessing" Score Redistribution

Grade Level	Reliability	SD	SEM	Raw Score Guessing	"Guessing" Raw Score + 2 SEMs	"Guessing" Raw Score + 4 SEMs
3	.85	5.96	2.29	10	15	19
4	.89	7.10	2.34	11	16	20
5	.90	7.28	2.35	11	16	20
6	.91	8.26	2.52	12	17	22
7	.92	9.81	2.79	12	18	23
8	.91	9.11	2.81	13	19	24

This redistribution of Mathematics TAKS raw scores was completed using the SAS code below:

```
Grade_3ij = 15 + floor((5)*rand("Uniform")) where Grade_3ij<15;          (36)
Grade_4ij = 16 + floor((5)*rand("Uniform")) where Grade_4ij<16;
Grade_5ij = 16 + floor((5)*rand("Uniform")) where Grade_5ij<16;
Grade_6ij = 17 + floor((6)*rand("Uniform")) where Grade_6ij<17;
Grade_7ij = 18 + floor((6)*rand("Uniform")) where Grade_7ij<18;
Grade_8ij = 19 + floor((6)*rand("Uniform")) where Grade_8ij<19;
Grade_8_Pass = 1 where Grade_8ij>=30;
```

Grade_8_Pass = 0 where Grade_8ij<30;

Percentile ranks at three points, namely 1, 50, and 99" (p. 5). Black (2012) showed that NCE scores are a "reliable and valid alternative for use in comparison to other methods of performance evaluation and accountability" (p. 122). Goldschmidt, Choi, and Martinez (2004) found that "NCEs may be more comparable across tests than IRT based scale scores" (p. 23). Furthermore, Goldschmidt, Choi, Martinez, and Novak (2010) went on to say that "when using longitudinal models for school accountability purposes, researchers and policymakers have more flexibility—inferences based on NCEs provide statistically and substantively similar results to inferences based on scale scores" (pp. 351-352). Shaw (2012) added that "the longitudinally invariant parallel IRT and NCE score models demonstrated strong rank-order consistency with the multivariate static model" (p. 182). Goldschmidt, Choi, and Martinez (2004) also stated that "if the objective of the evaluation is to rank schools then the choice of NCE or scale score will not change the ensuing school ranking. NCEs can accurately rank schools" (p. 23).

Using statewide Mathematics TAKS raw score data from the respective Texas Student Assessment Program: Technical Digests for years 2004 to 2009 (TEA, 2003-2004; TEA, 2004-2005; TEA, 2005-2006; TEA, 2006-2007; TEA, 2007-2008; TEA, 2008-2009), a standard deviation was calculated for each set of grade level raw scores. The standard deviation at each grade level was then used to convert the raw scores at each grade level into z-scores. Using Linn's (2000) formula, the z-scores were then converted into NCE scores:

$$\text{NCE Score} = 50 + 21.06 * \text{z-score} \qquad (37)$$

There was now a one-to-one relationship between the raw scores, z-scores, scale scores, and NCE scores for each grade level. See Appendices C through H. The value-added accountability system model of the present study required time-series data for each student (i.e., a score of interval data for each student at periodic time intervals that have some interpretable meaning). This conversion to NCE scores provides time-series data (annual intervals) that are linked, thus providing data with interpretable meaning. Using the look-up tables in Appendices C through H, each student's Mathematics TAKS raw score was then converted to its corresponding NCE score. This conversion of scores was completed using the SAS code below:

```
Update Stud_Scores as a set Grade_3ij = (select b.NCE_Score from NCE as b (38)
where (b.Time=0 and a.Grade_3ij=b.Raw_Score));

Update Stud_Scores as a set Grade_4ij = (select b.NCE_Score from NCE as b
where (b.Time=1 and a.Grade_4ij=b.Raw_Score));

Update Stud_Scores as a set Grade_5ij = (select b.NCE_Score from NCE as b
where (b.Time=2 and a.Grade_5ij=b.Raw_Score));

Update Stud_Scores as a set Grade_6ij = (select b.NCE_Score from NCE as b
where (b.Time=3 and a.Grade_6ij=b.Raw_Score));

Update Stud_Scores as a set Grade_7ij = (select b.NCE_Score from NCE as b
where (b.Time=4 and a.Grade_7ij=b.Raw_Score));

Update Stud_Scores as a set Grade_8ij = (select b.NCE_Score from NCE as b
where (b.Time=5 and a.Grade_8ij=b.Raw_Score));
```

The student Mathematics TAKS scale scores from the Field Study District were then winsorized and converted into NCE scores to address the two concerns mentioned earlier: lack of normality and lack of a vertical scale.

Student scores were then formatted so that each student had six records in the Stud_Scores_Array table—one record corresponding to each grade level score (i.e., grades 3 through 8). See Table 36 for an example of the format of the Stud_Scores_Array table containing NCE scores.

Table 36

Stud_Scores_Array Table (NCE Scores): Example of the Format

Camp_ID	Stud_ID	Camp_Econ_Dis	SES_{ti}	Time	NCE Score
1	1	45	0	0	50
1	1	45	0	1	55
1	1	45	0	2	54
1	1	45	0	3	60
1	1	45	0	4	62
1	1	45	0	5	70
…					
9	300	100	1	0	59
9	300	100	1	1	59
9	300	100	1	2	63
9	300	100	1	3	63
9	300	100	1	4	65
9	300	100	1	5	68

Estimating Campus Valued-Added Coefficients (Beta Values)

To conduct a Hierarchical Linear Model (HLM) analysis, the format of the winsorized raw scores in the Stud_Scores table were transformed so that each student had six records—one for each grade level score. The Stud_Scores_Array table had been created and populated with SES_j, (or Camp_Econ_Dis), SES_{ti}, Time, and Raw Score data, so that grade levels corresponded to time, denoted by 0 through 5. These raw scores had also been transformed into NCE scores.

Using the SAS procedure for conducting an HLM analysis, PROC MIXED, value-added coefficients (Beta values) were calculated using the SAS code below:

```
PROC MIXED DATA=Work.Stud_Scores_Array_1                                    (39)
COVTEST NOCLPRINT  method = ML;
CLASS Stud_ID;
MODEL Score = Time SES_ti Camp_Econ_Dis Camp_Econ_Dis*Time
Camp_Econ_Dis*SES_ti/ SOLUTION;
RANDOM intercept Time SES_ti;
REPEATED /subject = Stud_ID type = cs rcorr;

ods output solutionf=Work.SF(keep=effect estimate rename=(estimate=overall));
run;
```

Ranking Campuses by Mean Student Residuals (Value-Added Accountability System Model)

After the value-added coefficients (Beta values) were generated using the PROC MIXED procedure above, they were used to update the following fields in the Stud_Scores table using the SAS code below.

Intercept_Beta (β_{00}) = intercept; (40)
Camp_Econ_Dis_Beta (β_{01}) = Camp_Econ_Dis ;
Time_Beta (β_{10}) = Time;
Camp_Econ_Dis_Time_Beta (β_{11}) = Camp_Econ_Dis*Time;
Stud_Econ_Dis_Beta (β_{20}) = SES_{ti};
Camp_Econ_Dis_Stud_Econ_Dis_Beta (β_{21})= Camp_Econ_Dis*SES_{ti};

The Predicted_Score and Residual fields were then updated using the SAS code below:

Predicted_Score=Intercept_Beta + Camp_Econ_Dis_Beta*SES_j (41)
+ Time_Beta*5 + Camp_Econ_Dis_Time_Beta*SES_j*5
+ Camp_Econ_Dis_Stud_Econ_Dis_Beta*SES_j*SES_{ti}
+ Stud_Econ_Dis_Beta* SES_{ti};

Residual = $Grade\_8_{ij}$ - Predicted_Score; (42)

The residuals were then averaged for each campus and the campuses were ranked such that the campus with the highest mean residual was ranked 1 and the campus with the lowest mean residual was ranked 9.

Ranking Campuses by SES Passing Percentages (Traditional SES Accountability System Model)

At this time, the students in the Stud_Scores table had both SES_{ti} and the *Grade_8_Pass* indictors populated. These two fields were then used to generate the percent of low SES students passing on each campus. The campuses were then ranked a second time within the district such that the campus with the highest percentage of low SES students with a passing score was ranked 1 and the campus with the lowest percentage of low SES students with a passing score was ranked 9. These data for the two campus rankings were collected in the Camp_Rank table with the fields below:

Run_ID (43)
CCRE_ID
Dist_ID
Camp_ID
Camp_VA_Average
Camp_VA_Rank
Camp_Econ_Average
Camp_Econ_Rank

Calculating a Pearson r Correlation between the Two Rankings

Each campus now had two rankings within the district—one using a value-added accountability system model and the other using a traditional SES accountability system model. These two rankings were then used to calculate a Pearson *r* correlation for the district with the SAS code below:

```
proc corr data=Work.Camp_Rank_Temp noprint outp=Pearson;        (44)
var Camp_VA_Rank Camp_Econ_Rank;
run;
```

These data were collected in the Correlations table with the fields below:

Run_ID (45)
Sim_ID
Dist_ID
Correlation

Analysis

The present study focused on a comparison of campus rankings using two different accountability system models. The first model ranked the 9 campuses of the Field Study District using a campus value-added accountability system model. The second model ranked the 9 campuses of the Field Study District using a traditional SES accountability system model. The two rankings of each of the 9 campuses were then used to calculate a Pearson *r* correlation between the two rankings produced by the two different accountability system models.

Model Fit

After receiving the data from the Field Study District, descriptive statistics were calculated as seen in Table 37 and Table 38. The campus percentages of low SES students in the Field Study District had a mean of 11.818 and a standard deviation of 6.885 which best fit District 2 of Table 4 where the mean had been set to 50 and the standard deviation had been set to 6.847. However, note that the difference in the mean of the campus percentages of low SES students of the Field Study District and District 2 of Table 4 is quite large. As with most field data, the descriptive statistics of the Field Study District do not match our assumptions perfectly.

Table 37 shows that winsorizing the raw scores from the Field Study District and then converting them to NCE scores helped resolve the issues of skewness and lack of a vertical scale slightly. The skewness and kurtosis of the winsorized data were a little closer to zero than the raw data for most grade levels. The mean and standard deviation of the NCE scores are slightly more consistent than the raw data from grade 3 to grade 8.

In addition, the correlation between the campus percentages of low SES students and school mean of grade 3 test scores was -0.827. This corresponds to Darandari's (2004) value of -0.930. The intraclass correlation of grade 3 test scores was 0.075, corresponding to Bosker and Witziers's (1995) value of 0.210. See Table 38.

Results

The present field study focused on the correlation between campus rankings of the Field Study District when the campuses were ranked using two different accountability systems: a valued-added accountability system model and an SES accountability system model. The calculated Pearson *r* correlation between these two rankings of the 9 middle school campuses was 0.360.

The Pearson *r* correlation is lower than desired which could be explained by the distribution of the percentage of low SES students amongst the campuses in the district. The campuses with a lower percentage of low SES students tend to do better on standardized tests than campuses with a higher percentage of low SES students. See Table 38 for empirical data of the correlation between the campus percentages of low SES students and school mean of grade 3 test scores. Thus campuses with a low percentage of low SES students could do

Table 37

Descriptive Statistics of Variables for District Field Data

Parameter	Description	*n*	Mean	SD	Skewness	Kurtosis
SES_j	Percent of low SES students on campus *j*	9	11.82	6.89	2.01	5.31
Grade_3ij	TAKS Math score at time *0*, for student *i*, on campus *j* (Grade 3 Score)	1,547	60.63[a] (34.97)[b]	14.39 (4.10)	-1.21 (-1.22)	1.43 (1.66)
Grade_4ij	TAKS Math score at time *1*, for student *i*, on campus *j* (Grade 4 Score)	1,547	61.63 (37.28)	13.03 (4.49)	-1.47 (-1.55)	2.30 (3.21)
Grade_5ij	TAKS Math score at time *2*, for student *i*, on campus *j* (Grade 5 Score)	1,547	61.30 (39.43)	13.14 (4.55)	-1.59 (-1.67)	2.91 (3.61)
Grade_6ij	TAKS Math score at time *3*, for student *i*, on campus *j* (Grade 6 Score)	1,547	60.43 (39.37)	14.79 (5.83)	-1.14 (-1.18)	0.90 (1.14)
Grade_7ij	TAKS Math score at time *4*, for student *i*, on campus *j* (Grade 7 Score)	1,547	64.91 (40.78)	12.77 (5.98)	-1.12 (-1.10)	0.88 (0.86)
Grade_8ij	TAKS Math score at time *5*, for student *i*, on campus *j* (Grade 8 Score)	1,547	63.33 (42.69)	13.21 (5.74)	-1.10 (-1.18)	0.97 (1.43)

[a]After raw scores had been Winsorized and converted to NCE scores. [b]Orginal raw scores.

Table 38

Correlations of Variables for District Field Data

Parameter	Description	n	r
$r_{ses*grade\ 3}$	Correlation between campus SES and school mean grade 3 test score	9	-0.83 (-0.93)[a]
$r_{grade\ 3}$	Intraclass correlation of grade 3 test scores	9	0.08 (0.21)[b]

[a] Darandari (2004). [b] Bosker and Witziers (1995).

well on the standardized test regardless of how much value the campus added to their educational experience, and vice versa. On the other hand, value-added models attempt to account for how much value the campus added to the educational experience of the students regardless of the campus percentages of low SES students. These two models do not necessarily rank campuses in the same order, thus the low Pearson r correlation between the two rankings. These field data provide evidence that the SES accountability system models are biased towards campuses with low percentages of low SES students.

Conclusions and Discussion

Implications for Practice

The main conclusion of the present field study is that there is a mismatch between the rankings resulting from value-added accountability system models and the rankings resulting from SES accountability system models. Research has been published on the effects of low SES students on standardized test scores (Aikens & Barbarin, 2008; Roberts & Bryant, 2011); however, many accountability models use a status-based indicator that is affected by the SES level of students. This lends itself to the possibility that such models are measuring SES (at both the campus level and the student level) rather than the achievement of students enrolled at the campus. By contrast, even though value-added models are far from perfect, they do attempt to measure achievement gained by students in one year, accounting for outside factors that may affect such achievement. To this end, this mismatch may indicate that although value-added models are not perfect, status-based models are even less desirable.

The calculated Pearson *r* correlation between the rankings of the nine middle school campuses of the present field study using the value-added accountability system model and the SES accountability system model was 0.360. A Pearson *r* correlation of this size is evidence that the systems are ranking the campuses differently. Campuses with a high rank under one system are receiving a lower rank under the other system, and vice versa - these could be false positives and/or false negatives. In the present field study, seven of the nine campuses received different rankings between the two systems. The ranking of the campus with the lowest percentage of low SES students (4.367%) changed the most - six positions. This campus ranked first under the SES accountability system model and ranked seventh under the value-added accountability system model. This change in the ranking of this campus is consistent with the research showing that high SES students have an advantage over low SES students on standardized tests (Caldas, 1993; Darandari, 2004; Kennedy & Mandeville, 2000; Willms, 1992) and represents a false positive.

These false positives and false negatives can have detrimental effects on educators. Educators may not be receiving credit for the work they are doing with respect to student achievement or they may be receiving undo credit. Annually (by August 8th), the Texas Education Agency publishes ratings based off, in part, assessment scores for every campus and district in Texas (Texas Education Code, 2014). These ratings may affect the evaluations of teachers, principals, and superintendents, and are thus considered high stakes ratings. Many educators recognize these false positives and false negatives intuitively. For example, a teacher with a class of no low SES students may not show any real gains in assessment scores, but have all of their students pass the assessment. At the same time, a teacher with a class of all low SES students may have all students show gains in assessment scores, but have only a portion of their students pass the assessment. Under status-based systems (e.g., SES accountability system models), the former teacher will receive more positive feedback in informal or formal evaluations than the latter. These inequities cause concerns for teachers, principals, and superintendents. The advantage of a value-added system is that it at least makes an attempt to balance these inequities. An accurate model for determining campus ratings and teacher, principal, and superintendent evaluations is crucial.

Lastly, an issue tangentially related to the use of status-based accountability systems is that most of these systems use assessment instruments that do not produce normally distributed scores. In other words, there are far too many students who make the maximum score (and far too few students who make the minimum score) and as a result there is very little information gained about these students or the schools these students attend. The focus of the school leaders then becomes getting the few students who failed the assessment the year before to pass the assessment in the current year at the expense of the many students who made the maximum score on the assessment the year before. This could mean that advanced classes are not offered to these students. A value-added accountability system is less likely to be a catalyst for this mindset as all students must make progress under a value-added accountability system. An instrument for which the scores are normally distributed would work well in a value-added accountability system and would also provide information on these high-performing students.

Limitations and Future Studies

As with any study, there are limitations that need to be considered and investigated in future studies. The present field study had three major limitations: lack of normality of the Mathematics TAKS scale scores; weak, if any, linkage between Mathematics TAKS scale scores from one grade level to the next (i.e., no vertical scale); and the small sample of middle schools.

Limitation 1: Lack of normality of the Mathematics TAKS scale scores.

As mentioned earlier, the Mathematics TAKS scale scores do not conform very well to a normal distribution. This field study attempted to adjust the normality by winsorizing the data; however, this was only partially effective. A future study may duplicate the current field study using standardized test scores that are more normally distributed.

Limitation 2: Weak, if any, linkage between Mathematics TAKS scale scores from one grade level to the next (i.e., no vertical scale).

As mentioned earlier, there is a weak to no linkage between Mathematics TAKS scale score form one grade level to the next as no vertical scale is available. This field study attempted to adjust for this by translating the Mathematics TAKS scores into NCE scores. This allowed some linking between Mathematics TAKS scores from grade level to grade level; however, it would have been better if the test scores were linked using Item Response Theory (IRT). A future study may duplicate the current field study using standardized test scores that are linked from one grade level to the next. State of Texas Assessments of Academic Readiness (STAAR) data (Texas Education Agency, 2013) would have been good for this; however, currently only three years of STAAR data are available.

Limitation 3: Small sample of middle schools.

The present field study was limited by the number of middle schools and the number of low SES students on each middle school campus. The number of low SES students on these middle school campuses ranged from 2 to 26 and in three cases all the low SES students on the campus passed the Mathematics TAKS thus limiting the distribution of the low SES students passing percentages. A future study may involve a larger district with more middle schools and more low SES students.

CHAPTER V

SUMMARY AND CONCLUSIONS

Accountability continues to be an important, but ambiguous topic in education. Most people involved with education (e.g., legislators, local school board members, and school administrators) agree that school systems should be held accountable; however, few agree on how to hold school systems accountable. Educational researchers continue to experiment with accountability models, but few agree on the best model to use. However, two types of models seem to dominate the research: models that are status-based (i.e., uses a cutoff test score to determine pass/fail) and models that are value-added based (i.e., uses an achievement gained score). The present three studies focused on models that are value-added based and the way they compare to models that are status-based. The first study presented a literature review and a comparison of prominent value-added accountability systems. The second study presented findings from a Monte Carlo simulation comparing the rankings of campuses generated using a value-added accountability system model compared to the rankings of campuses generated using an accountability system model that uses the success of economically disadvantaged students as a key accountability indicator (SES accountability system model). The third study presented findings from a field study comparing the campus rankings generated using a value-added accountability systems model to the rankings of campuses generated by an accountability system that uses the success of economically disadvantaged students as a key accountability indicator (SES accountability system model).

Study One

Findings from the first study show that value-added accountability systems use various statistical models, independent variables (e.g., prior attainment variables and intake variables), dependent variables, number of measurement points, and levels of analyses. The researcher should consider that a model using student intake variables extensively will not always need to include school level variables. The effects of student intake variables can sometimes be adequately accounted for without including school level variables, depending on the model. Prior attainment and SES variables are prominent variables to consider in value-added models (Kennedy & Mandeville, 2000; Willms, 1992). Value-added models that do not include prior attainment, but for which the residuals for some schools change signs (positive to negative or vice versa) when using value-added models that do include prior attainment, are misclassifying the rankings of campuses by the value-added model not including prior attainment (Thomas & Mortimore, 1996). Thomas (1998) showed that ranking campuses using a status-based model and then ranking them again using a value-added model does not always provide identical rankings. Many campuses will have different rankings under the two models. The second study is an extension of this research by Thomas.

Study Two

The second study contributes to the literature review of study one by offering a simulation that parallels the research by Thomas (1998). In study two, a Monte Carlo study was conducted where student test scores were generated for 50 campuses within five districts. Prior attainment, the percentage of low SES students on the campus, and the SES level of the student were the key variables considered. This simulation was then run under five different value-added parameters and five different distributions of the percentage of low SES students within the district (a different distribution for each of the five districts)—totaling 25 different combinations in all. These 50 campuses were then ranked using an SES accountability system model (status-based model) and then a second time using a value-added accountability system model. Findings showed that the two models ranked campuses differently. When the value-added parameter was held constant for all five districts, one might assume that the rankings would be similar using both models; however, this was not the case. The Pearson *r* correlation between the rankings was less than 0.50 in some cases. This phenomenon was more evident when the distribution of the percentage of low SES students within the district was wider. Thus when the difference of the percentages of low SES students on campuses was the greatest, there were some campuses composed of almost no low SES students and some campuses composed of almost all low SES students. The Pearson *r* correlations between the rankings were the lowest when these distributions of the percentages of low SES students were the widest, thus providing evidence that SES students are affecting the way campuses are ranked. Status-based models make no attempt to adjust for low SES students, thus these models give an advantage to campuses with a lower percentage of low SES students over campuses with a higher percentage of low SES students. Value-added models, although not perfect, at least make an attempt to account for the level of SES of the students. This advantage of the campuses with a lower percentage of SES students was not as great under the value-added accountability system model. These results mirror the results found by Thomas (1998). This simulation was the foundation for the analysis of the field data in study three.

Study Three

The third and final study was identical to the second study with the exception that the analysis was run only one time and that it was run with field data from the 9 middle school campuses in a large district in Texas rather than simulated data. Results from the third study are similar to those of the second study. The distribution of the percentage of low SES students of this Field Study District data best fit the distribution of the percentage of low SES students of District 2 in Table 4 and also yielded similar results to that of District 2 with CCRE 2 in Table 8. However, note that the lowest Pearson *r* correlation range for District 2 with CCRE 1 was 0.50 to less than 0.60, and the Pearson *r* correlation between the two rankings of the Field Study District data was 0.360. This similarity in results provides evidence that the simulation yielded results which might mirror what is happening in the field; however, future studies need to be conducted. Campuses with a lower percentage of low SES students have an advantage using status-based models.

In conclusion, educational researchers must be careful not to place too much emphasis on accountability systems that yield ratings that are correlated to external student variables, such as SES level. Such models may be measuring those external student variables rather than the achievement of students enrolled at the campus. Even though value-added models are far from perfect, they do make an attempt to measure achievement gained by students in one year accounting for outside factors that may affect such achievement. To this end, this mismatch may indicate that although value-added models are not perfect, status-based models are even less desirable.

REFERENCES

Aikens, N., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*, 235-251.

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

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

Babu, S., & Mendro, R. (2003, April). *Teacher accountability: HLM-based teacher effectiveness indices in the investigation of teacher effects on student achievement in a state assessment program.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

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

Barnett, V. (1978). The study of outliers: Purpose and model. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 27*, 242-250.

Bertrand, R., & Leclerc, M. (1985). Reliability of observational data on teaching practices in secondary school mathematics. *Teaching & Teacher Education, 1*, 187-198.

Black, A. (2012). *A comparison of value-added, ordinary least squares regression, and the California STAR accountability indicators.* (Order No. 3513721, University of

Southern California). *ProQuest Dissertations and Theses*, 162. Retrieved from http://search.proquest.com/docview/1027916812?accountid=7108.

Blakey, L. & Heath, A. (1992). Differences between comprehensive schools: Some preliminary findings. In D. Reynolds & P. Cuttance (Eds), *School Effectiveness: Research, Policy and Practice* (pp. 121–133). London: Cassell.

Bolboaca, S., Jantschi, L., Sestras, A., Sestras, R., & Pamfil, D., (2011). Pearson-Fisher chi-square statistic revisited, *Information*, *2*, 528-545.

Bosker, R., & Witziers, B. (1995). *A meta analytical approach regarding school effectiveness: The true size of school effects and the effect size of educational leadership* (Tech. Rep.). The Netherlands: University of Twente, Department of Education, Division of Education Administration.

Caldas, S. (1993). Reexamination of input and process factor effects on public school achievement. *The Journal of Educational Research, 86*, 206-214.

Caldas, S., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research*, *90*, 269-277.

Choi, K., & Seltzer, M. (2003). Addressing questions concerning equity in longitudinal studies of school effectiveness and accountability: Modelling heterogeneity in relationships between initial status and rates of change. Los Angeles, CA: Center for the Study of Evaluation, University of California. Retrieved from http://www.cse.ucla.edu/products/Reports/R614.pdf.

Ciarleglio, M., & Makuch, R. (2007). Hierarchical linear modeling: An overview. *Child Abuse & Neglect*, *31*, 91–98.

Cochran, W. (1952). The χ^2 Test of Goodness of Fit. *Annals of Mathematical Statistics, 23,* 315-345.

Coleman, J., Campbell, T., & Kilgore, S. (1982). *High achievement: Public, catholic, and other private schools compared.* New York, NY: Basic.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Belmont, CA: Wadsworth Group/Thomson Learning.

Darandari, E. (2004). *Robustness of hierarchical linear model parameter estimates under violations of second-level residual homoskedasticity and independence assumptions.* (Order No. 3137420, The Florida State University). *ProQuest Dissertations and Theses,* 231-231. Retrieved from http://search.proquest.com/docview/305182787?accountid=7108

De Leeuw, J., & Kreft, I. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics, 20*, 171–189.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics, 20*, 115-147.

Dreeben, R., & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education, 61*, 129-142.

Eisele-Dyrli, K. (2010). Lone star. *District Administration, 46*(5), 33-36, 38.

Fisher, R. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, 98*, 39-82.

Fortier, M.S., Vallerand, R.J., & Guay, F. (1995). Academic motivation and school performance: Toward a structural model. *Contemporary Educational Psychology, 20,* 257-274.

Goldschmidt, P., Choi, K., & Martinez, F. (2004). *Using hierarchical growth models to monitor school performance over time: Comparing NCE to Scale Score results (CSE Technical Report).* Los Angeles, CA: National Center for Research on Evaluation Standards and Student Testing, University of California.

Goldschmidt, P., Choi, K., Martinez, F., & Novak, J. (2010). Using growth models to monitor school performance: Comparing the effect of the metric and the assessment. *School Effectiveness and School Improvement, 21*(3), 337-357.

Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement, 8,* 369-395.

Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*, 385-443.

Green, J. (2010). Estimating teacher effects using value-added models. (Order No. 3412864, The University of Nebraska - Lincoln). *ProQuest Dissertations and Theses,*160. Retrieved from http://search.proquest.com/docview/748220889?accountid=7108

Gross, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *Journal of the American Statistical Association, 71,* 409-416.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66*, 1017-1098.

Hill, H., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48,* 794-831.

Hill, P., & Rowe, K. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement, 7*, 1-34.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Kennedy, E. & Mandeville, G. (2000). Some methodological issues in school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 189–205). London, UK: Routledge.

Keselman, H., Kowalchuk, R., & Lix, L. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed means. *Psychometrika, 63,* 145-163.

Keselman, H., Lix, L., & Kowalchuk, R. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods, 3,* 123-141.

Krashen, S., & Brown, C. (2005). The ameliorating effects of high socioeconomic status: A secondary analysis. *Bilingual Research Journal*, *29*, 185-196.

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

Lenkeit, J. (2013). Effectiveness measures for cross-sectional studies: A comparison of value added models and contextualised attainment models. *School Effectiveness and School Improvement, 24*, 39-63.

Linn, R. (2000). Assessments and accountability. *Educational Researcher*, *20*(2), 4–16.

Linn, R. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, *40*, 699-711.

MacCallum, R. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38*, 113-139.

Marks, G. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research, 48*, 21-40.

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

McPherson, A. (1993). Measuring added value in schools. *Education Economics, 1*, 43-51.

Millman, J., & Schalock, H. (1997). Beginnings and introduction. In J. Millman (Ed.). *Grading teachers, grading schools: Is student achievement a valid measure?* (pp. 3–10). Thousand Oaks, CA: Corwin Press, Inc.

Mortimore, P., Sammons, P., & Thomas, S. (1994). School effectiveness and value-added measures. *Assessment in Education: Principles, Policy & Practice, 1*(3), 315-332.

Myers, N., Ahn, S., & Ying, J. (2011). Sample size and power estimates for a confirmatory factor analysis model in exercise and sport: A Monte Carlo approach. *Research Quarterly for Exercise and Sport, 82*, 412-423.

Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research, 13*, 769-776.

Olson, L. (1998). A question of value. *Education Week on the Web, 17*(35), Retrieved from http://web.ebscohost.com.lib-ezproxy.tamu.edu:2048/ehost/detail?sid=e8d7109d-33bd-4c95-9a31-

ddcc854d3399%40sessionmgr4&vid=1&hid=25&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=tfh&AN=638865.

Ott, L., & Longnecker, M. (2001). *An introduction to statistical methods and data analysis* (5th ed). Pacific Grove, CA; Thomson Learning.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*, 669-710.

Peterson, D., Micceri, T., & Smith, O. (1985). Measurement of teacher performance: A study in instrument development. *Teaching & Teacher Education, 1*, 63-77.

Ponisciak, S., & Bryk, A. (2005). Value-added analysis of the Chicago public schools: An application of hierarchical models. In R.W. Lissitz (Ed.), *Value-added models in education: Theory and application* (pp. 19-40). Maple Grove, MN: JAM Press.

Ramsey, P., & Ramsey, P. (2007). Optimal trimming and outlier elimination. *Journal of Modern Applied Statistical Methods, 6*, 355-360.

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

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

Raudenbush, S., & Bryk, A. (1988). Chapter 10: Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education, 15*, 423-475.

Roberts, R., & Bryant, D. (2011). Early mathematics achievement trajectories: English-language learner and native English-speaker estimates, using the early childhood longitudinal survey. *Development Psychology, 47*, 916-930.

Romer, R. (1997). Today standards—tomorrow success. *National Association of Secondary School Principals Bulletin, 81*, 7-11.

Roscoe, J., & Byars, J. (1971) An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. *Journal of the American Statistical Association. 66*, 755-759.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

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

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics, 6*, 34-58.

Rubin, D. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics, 31*, 161-170.

Sammons, P., & Luyten, H. (2009). Editorial article for special issue on alternative methods for assessing school effects and schooling effects. *School Effectiveness and School Improvement, 20*, 133-143.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal, 19,* 381-405.

Sanders, W., & Horn, S. (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., Saxton, A., & Horn, S. (1997). The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment. In Millman

(Ed.), *Grading teachers, grading school: Is student achievement a valid evaluation measure?* (pp. 137-162). Thousand Oaks, CA: Corwin.

Setliff, B. (1989). The effects of the Texas teacher appraisal system on the climate of six small school systems. (Texas Tech University - Lubbock). Retrieved from https://repositories.tdl.org/ttu-ir/bitstream/handle/2346/15396/31295005847982.pdf?sequence=1

Shaw, L. (2012). Incorporating latent variable outcomes in value-added assessment: An evaluation of univariate and multivariate measurement model structures. (Order No. 3546638, The University of Nebraska - Lincoln). *ProQuest Dissertations and Theses*, 430. Retrieved from http://search.proquest.com/docview/1266848200?accountid=7108

Shinkfield, A., & Stufflebeam, D. (1995). *Teacher evaluation: guide to effective practice.* Boston, MA: Kluwer Academic Publishers.

Stevens, J. (2005). The study of school effectiveness as a problem of research design. In R. Lissitz (Ed.), *Value added models in education: Theory and applications* (pp. 166–208). Maple Grove, MN: JAM Press.

Stipek, D., & Ryan, R. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology, 33*, 711-723.

Teachman, J. (1987). Family background, educational resources, and educational attainment. *American Sociological Review*, *52*, 548-557.

Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The* *international handbook of school effectiveness research* (pp. 55–133). London, UK: Routledge.

Teddlie, C., Stringfield, S., & Reynolds, D. (2000). Context issues within school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 160–186). London, UK: Routledge.

Tekwe, C., Carter, R., Ma, C., Algina, J., Lucas, M., 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.

Texas Education Agency (TEA). (2003). *Standard error of measurement* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/standards

Texas Education Agency (TEA). (2003-2004). *Texas student assessment program: Technical digest for the academic year 2003-2004* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0304

Texas Education Agency (TEA). (2004). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2004 administration grade 3* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr04/

Texas Education Agency (TEA). (2004-2005). *Texas student assessment program: Technical digest for the academic year 2004-2005* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0405

Texas Education Agency (TEA). (2005). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2005 administration grade 4* (Tech. Rep.).

Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr05/

Texas Education Agency (TEA). (2005-2006). *Texas student assessment program: Technical digest for the academic year 2005-2006* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0506

Texas Education Agency (TEA). (2006). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2006 administration grade 5* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr06/

Texas Education Agency (TEA). (2006-2007). *Texas student assessment program: Technical digest for the academic year 2006-2007* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0607

Texas Education Agency (TEA). (2007). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2007 administration grade 6* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr07/

Texas Education Agency (TEA). (2007-2008). *Texas student assessment program: Technical digest for the academic year 2007-2008* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0708

Texas Education Agency (TEA). (2008). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2008 administration grade 7* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr08/

Texas Education Agency (TEA). (2008-2009). *Texas student assessment program: Technical digest for the academic year 2008-2009* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0809

Texas Education Agency (TEA). (2009). *Texas assessment of knowledge and skills raw score conversion table: Mathematics - spring 2009 administration grade 8* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/student.assessment/taks/convtables/yr09/

Texas Education Agency (TEA). (2011). *2011 Accountability system: District data download* (Tech. Rep.). Austin, TX: Retrieved from http://ritter.tea.state.tx.us/perfreport/account/2011/dist_download.html

Texas Education Agency (TEA). (2013). *State of Texas assessments of academic readiness (STAAR^{TM}): Vertical scale technical report* (Tech. Rep.). Austin, TX: Retrieved from http://www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=25769806053&libID=25769806056

Texas Education Agency (TEA). (2014). *2014 Accountability manual for Texas public school districts and campuses* (Tech. Rep.). Austin, TX: Retrieved from http://ritter.tea.state.tx.us/perfreport/account/2014/manual/index.html

Texas Education Code. (2014). *Chapter 39.054a* (Tech. Rep.). Austin, TX: Retrieved from http://www.statutes.legis.state.tx.us/Docs/ED/htm/ED.39.htm#C

Thomas, S. (1998). Value-added measures of school effectiveness in the United Kingdom. *Prospects, 28,* 91-108.

Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary school effectiveness. *Research Papers in Education (London), 11*(1), 5-33.

Thomas, S., Nuttall, D., & Goldstein, H. (1992, October, 20). The *Guardian* survey (of A-level examination results). *The Guardian (Manchester)*, 96-102.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach.* New York, NY: Guilford.

Tukey, J., & McLaughlin, D. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorizing 1. *Sankhya*, A25, 331-352.

United States Department of Education (USDE). (2002). *Strategic plan 2002-2007* (Tech. Rep.). Washington, DC: Retrieved from http://www2.ed.gov/about/reports/strat/plan2002-07/plan.pdf.

Wang, Z. (2006). Comparison of value-added models for school ranking and classification: A Monte Carlo study. (Order No. 3296572, Texas A&M University). *ProQuest Dissertations and Theses,* 145. Retrieved from http://search.proquest.com/docview/304931319?accountid=7108.

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

Weerasinghe, D. & Orsak, T. (1998, April). Can hierarchical linear modeling be used to rank schools: A simulation study with conditions under which hierarchical linear modeling is applicable. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Wilcox, R. R. (1994). Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed. *Biornetrical Journal, 36,* 259-273.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, 65(1), 51-77.

Willet, J. (1988). Chapter 9: Questions and answers in the measurement of change. *Review of Research in Education, 15*, 345-422.

Willms, J. (1992). *Monitoring School Performance: A Guide for Educators*. London: Falmer Press.

Willms, J. (2000). Monitoring school performance for "standards-based reform." *Evaluation & Research in Education, 14,* 237-253.

Wyse, A., Zeng, J., & Martineau, J. (2011). A graphical transition table for communicating status and growth. *Practical Assessment, Research & Evaluation, 16*(11), 1-8.

APPENDIX A

SAS CODE: MONTE CARLO SIMULATION

```
%Monitor_Start;

data CCRE_Dist_Camp_ID;

%let Sim_Start=1;
%let Sim_Finish=1;

%let CCRE_Start=1;
%let CCRE_Finish=5;

%let District_Start=1;
%let District_Finish=5;

%let Campus_Start=1;
%let Campus_Finish=50;

%let Student_Start=1;
%let Student_Finish=300;

do Run_ID=&Sim_Start to &Sim_Finish;
      do Dist_ID=&District_Start to &District_Finish;
            do CCRE_ID=&CCRE_Start to &CCRE_Finish;
                  do Camp_ID=&Campus_Start to &Campus_Finish;

/*Intraclass correlation of Grade_2_Meanj (Grade_2ij campus means).*/

                        Grade_2_Meanj=.458*rannor(-5);

/*Correlation between SESj and Grade_2_Meanj.*/
/*Sets correlation to .93 while keeping the intraclass correlation of
Grade_2_Meanj set to .21*/

                        SESj = -(((.93**2)*(.1+(Dist_ID-
1)*.2)/.21)**.5)*Grade_2_Meanj-(((.1+(Dist_ID-1)*.2)-
((.93**2)*(.1+(Dist_ID-1)*.2)))**.5*rannor (-5));

/*Campus SES fixed effect (-0.414) on SESj*/

                        Y01=-0.414;

                        Campus_SES_Effect=Y01*SESj;

                        Campus_Effect0j=0;
                        Campus_Effect1j=0;
                        Campus_Effect2j=0;
                        Campus_Effect3j=0;
                        Campus_Effect4j=0;
                        Campus_Effect5j=0;

                        output;
```

```
                  end;
            end;
      end;
end;

run;

/*Generate Campus Effect*/

proc sql;
      create table Camp_Effect (
      u0j numeric,
      u1j numeric,
      u2j numeric,
      u3j numeric,
      u4j numeric,
      u5j numeric);
quit;

%Macro Gen_Camp_Effect (Start=0.1, Stop=0.9, Step=.2);

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&District_Start %to &District_Finish;

      proc sql;
      create table Temp_2 (
      Col1 numeric,
      Col2 numeric,
      Col3 numeric,
      Col4 numeric,
      Col5 numeric,
      Col6 numeric);
      quit;

      %let r2=&Start;

            %do %until(%sysevalf(&r2 gt &Stop));
            %let jj = %sysevalf((&r2**.5)*(0.317**.5)*(0.317**.5));
            %let jjj = %sysevalf((&r2**.5)*(0.317**.5)*(0.100**.5));
                  proc iml;
                        mu={ 0, 0, 0, 0, 0, 0};
sigma= {0.100     &jjj        &jjj        &jjj        &jjj        &jjj,
      &jjj        0.317       &jj         &jj         &jj         &jj,
      &jjj        &jj         0.317       &jj         &jj         &jj,
      &jjj        &jj         &jj         0.317       &jj         &jj,
      &jjj        &jj         &jj         &jj         0.317       &jj,
      &jjj        &jj         &jj         &jj         &jj         0.317};

                              call vnormal(et, mu, sigma, &Campus_Finish);
                              create Temp_1 from et;
                              append from et;

                              proc append base=Temp_2 data=Temp_1;
                              run;

                        proc sql;
                              insert into Camp_Effect (
```

```
                                      u0j,
                                      u1j,
                                      u2j,
                                      u3j,
                                      u4j,
                                      u5j)
                                select col1, col2, col3, col4, col5, col6
                                from Temp_2;
                          quit;

                          proc sql;
                          delete * from Temp_2;
                          quit;

                          %let r2 =%sysevalf(&r2+&Step);
            %end;
      %end;
%end;

%Mend Gen_Camp_Effect;
%Gen_Camp_Effect;

/*%Gen_Camp_Data;*/

data Camp_Data;
   set CCRE_Dist_Camp_ID;
   set Camp_Effect;
run;

proc sql;
/*Total Campus Effect Variance*/

      Update Camp_Data Set Campus_Effect0j = Campus_SES_Effect + u0j;
      Update Camp_Data Set Campus_Effect1j = Campus_SES_Effect + u1j;
      Update Camp_Data Set Campus_Effect2j = Campus_SES_Effect + u2j;
      Update Camp_Data Set Campus_Effect3j = Campus_SES_Effect + u3j;
      Update Camp_Data Set Campus_Effect4j = Campus_SES_Effect + u4j;
      Update Camp_Data Set Campus_Effect5j = Campus_SES_Effect + u5j;

quit;

/*Generate Student Score Parameters*/

data Stud_Scores;

do Run_ID=&Sim_Start to &Sim_Finish;
      do CCRE_ID=&CCRE_Start to &CCRE_Finish;
            do Dist_ID=&District_Start to &District_Finish;
                  do Camp_ID=&Campus_Start to &Campus_Finish;
                        do Stud_ID=&Student_Start to &Student_Finish;
                              Grade_2_Meanj=0;
                              SESj=0;
                              SESj_Percent=0;
                              Y01=0;
                              Campus_SES_Effect=0;
                              u0j=0;
                              u1j=0;
```

```
                              u2j=0;
                              u3j=0;
                              u4j=0;
                              u5j=0;
                              Grade_2ij=0;
                              SESti=0;
                              B01=0.762;
                              B10=0.630;
                              B11=-0.045;
                              r0i=0.469*rannor(-5);
                              r1i=0.325;
                              e0ij=0.316*rannor(-5);
                              e1ij=0.316*rannor(-5);
                              e2ij=0.316*rannor(-5);
                              e3ij=0.316*rannor(-5);
                              e4ij=0.316*rannor(-5);
                              e5ij=0.316*rannor(-5);
                              Stud_SES_Effectti=0;
                              Student_Random_Effectti=0;
                              Student_Effectti=0;
                              Campus_Effect0j=0;
                              Campus_Effect1j=0;
                              Campus_Effect2j=0;
                              Campus_Effect3j=0;
                              Campus_Effect4j=0;
                              Campus_Effect5j=0;
                              Grade_3ij=0;
                              Grade_4ij=0;
                              Grade_5ij=0;
                              Grade_6ij=0;
                              Grade_7ij=0;
                              Grade_8ij=0;
                              Grade_8_Mean=0;
                              Grade_8_STD=0;
                              Grade_8_zscore=0;
                              Grade_8_Pass=0;
                              Intercept_Beta=0;
                              Camp_Econ_Dis_Beta=0;
                              Time_Beta=0;
                              Camp_Econ_Dis_Time_Beta=0;
                              Stud_Econ_Dis_Beta=0;
                              Camp_Econ_Dis_Stud_Econ_Dis_Beta=0;
                              Predicted_Score=0;
                              Residual=0;
                              output;
                        end;
                  end;
            end;
      end;
end;

run;

/*Update Stude_Scores;*/

proc sql;
```

```
      update Stud_Scores as a set SESj = (select b.SESj from Camp_Data as
b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set Y01 = (select b.Y01 from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set Grade_2_Meanj = (select b.Grade_2_Meanj
from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u0j = (select b.u0j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u1j = (select b.u1j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u2j = (select b.u2j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u3j = (select b.u3j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u4j = (select b.u4j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set u5j = (select b.u5j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores as a set Campus_SES_Effect = (select
b.Campus_SES_Effect from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;
proc sql;
      update Stud_Scores set SESj_Percent = 50 + 12.5 * SESj;
quit;
proc sql;
```

```
      update Stud_Scores set SESti = 1 where Stud_ID<=(SESj_Percent/100) * 
&Student_Finish;
quit;
/*Gen_Stud_Effect*/

%Macro Gen_Stud_Effect;

proc sql;
      create table Stud_Effect (
      Student_Random_Effectti numeric);
quit;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&District_Start %to &District_Finish;
            %do CCRE_ID=&CCRE_Start %to &CCRE_Finish;
                  %do Camp_ID=&Campus_Start %to &Campus_Finish;

                        proc sql;
                        create table Temp_2 (
                        Col1 numeric,
                        Col2 numeric,
                        Col3 numeric);
                        quit;

                        proc iml;
                        mu={ 0, 0, 0};
                        sigma= {0.001     0.001       0.011,
                              0.001       0.324       0.001,
                              0.011       0.001       0.250};
                              call vnormal(et, mu, sigma, 
&Student_Finish);
                                    create Temp_1 from et;
                                    append from et;

                                    proc append base=Temp_2 data=Temp_1;
                                    run;

                              proc sql;
                                    insert into Stud_Effect (
                                          Student_Random_Effectti)
                                    select col2
                                    from Temp_2;
                              quit;

                              proc sql;
                              delete * from Temp_2;
                              quit;

                  %end;
            %end;
      %end;
%end;

%Mend Gen_Stud_Effect;
%Gen_Stud_Effect;

data Stud_Scores;
```

```
   set Stud_Scores;
   set Stud_Effect;
run;
proc sql;
      update Stud_Scores set Stud_SES_Effectti = B11*SESti;
      update Stud_Scores set Student_Effectti = Stud_SES_Effectti +
Student_Random_Effectti;
quit;
proc sql;
      update Stud_Scores as a set Campus_Effect0j = (select
b.Campus_Effect0j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores as a set Campus_Effect1j = (select
b.Campus_Effect1j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores as a set Campus_Effect2j = (select
b.Campus_Effect2j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores as a set Campus_Effect3j = (select
b.Campus_Effect3j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores as a set Campus_Effect4j = (select
b.Campus_Effect4j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores as a set Campus_Effect5j = (select
b.Campus_Effect5j from Camp_Data as b
      where (a.Run_ID=b.Run_ID and a.CCRE_ID=b.CCRE_ID and
a.Dist_ID=b.Dist_ID and a.Camp_ID=b.Camp_ID));
quit;

proc sql;
      update Stud_Scores set Grade_2ij = Grade_2_Meanj + ((1-
.21)**.5)*rannor(-5);
      update Stud_Scores set Grade_3ij = B01 * Grade_2ij + u0j + r0i +
e0ij;
      update Stud_Scores set Grade_4ij = 0.63 + Grade_3ij - e0ij +
Student_Effectti + Campus_Effect1j + e1ij;
```

```
    update Stud_Scores set Grade_5ij = 0.63 + Grade_4ij - e1ij +
Student_Effectti + Campus_Effect2j + e2ij;
    update Stud_Scores set Grade_6ij = 0.63 + Grade_5ij - e2ij +
Student_Effectti + Campus_Effect3j + e3ij;
    update Stud_Scores set Grade_7ij = 0.63 + Grade_6ij - e3ij +
Student_Effectti + Campus_Effect4j + e4ij;
    update Stud_Scores set Grade_8ij = 0.63 + Grade_7ij - e4ij +
Student_Effectti + Campus_Effect5j + e5ij;
quit;

%Macro zscores;

proc sql;

create table zscores_temp
      (Run_ID numeric,
      CCRE_ID numeric,
      Dist_ID numeric,
      Camp_ID numeric,
      Grade_8ij numeric);
quit;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&CCRE_Start %to &CCRE_Finish;
            proc sql;
                  insert into Work.zscores_temp
                        (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Grade_8ij)
                  select Run_ID, CCRE_ID, Dist_ID, Camp_ID, Grade_8ij
                  from Work.Stud_Scores where Run_ID=&i and CCRE_ID=&j;
            quit;

            proc corr data=Work.zscores_temp noprint outp=Pearson;
                  var Grade_8ij;
            run;

            proc sql;
                  update Work.Stud_Scores Set Grade_8_Mean = (select
Grade_8ij from Pearson where _TYPE_ = 'MEAN') where Run_ID=&i and
CCRE_ID=&j;
                  update Work.Stud_Scores Set Grade_8_STD = (select
Grade_8ij from Pearson where _TYPE_ = 'STD') where Run_ID=&i and
CCRE_ID=&j;
                  update Work.Stud_Scores Set Grade_8_zscore = (Grade_8ij-
Grade_8_Mean)/Grade_8_STD where Run_ID=&i and CCRE_ID=&j;
            quit;

            proc sql;
                  delete * from work.zscores_temp;
            quit;
      %end;
%end;

%Mend zscores;
%zscores;

proc sql;
```

```
      update Stud_Scores set Grade_8_Pass = 1 where Grade_8_zscore >= -
0.709;
      update Stud_Scores set Grade_8_Pass = 0 where Grade_8_zscore < -
0.709;
quit;

proc append base=Mylib.Stud_Scores data=Work.Stud_Scores;
run;

/*Create Work.Stud_Scores_Array Table.*/

proc sql;
      create table Work.Stud_Scores_Array
            (Run_ID numeric,
            CCRE_ID numeric,
            Dist_ID numeric,
            Camp_ID numeric,
            Stud_ID numeric,
            Camp_Econ_Dis numeric,
            SESti numeric,
            Time numeric,
            Score numeric);
quit;

/*Update Work.Stud_Scores_Array;*/

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 0,
Grade_3ij
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 1,
Grade_4ij
      from Work.Stud_Scores;
quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 2,
Grade_5ij
      from Work.Stud_Scores;
quit;

proc sql;
```

```
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 3,
Grade_6ij
      from Work.Stud_Scores;
quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 4,
Grade_7ij
      from Work.Stud_Scores;
quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select
            Run_ID, CCRE_ID, Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 5,
Grade_8ij
      from Work.Stud_Scores;
quit;

/*Create Work.SF_All Table.*/

proc sql;
      create table Work.SF_All
            (Effect char(19),
            Overall numeric);
quit;

/*Perform HLM Analysis.*/

%Macro HLM;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&CCRE_Start %to &CCRE_Finish;
            %do k=&District_Start %to &District_Finish;
                  data Stud_Scores_Array_1;
                        set Stud_Scores_Array;
                        if Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                        run;

                        PROC MIXED DATA=Work.Stud_Scores_Array_1 COVTEST
NOCLPRINT  method = ML;
                        CLASS Stud_ID;
                        MODEL Score = Time SESti Camp_Econ_Dis
Camp_Econ_Dis*Time Camp_Econ_Dis*SESti/ SOLUTION;
                        RANDOM intercept Time SESti;
                        REPEATED /subject = Stud_ID type = cs rcorr;
```

```
                        ods output solutionf=Work.SF(keep=effect estimate
rename=(estimate=overall));
                        run;

                        proc append base=SF_All data=SF;
                        run;
            %end;
      %end;
%end;

%mend HLM;
%HLM;


/*Create Work.SF_ALL_ID Table.*/

Data SF_ALL_ID;

do Run_ID= &Sim_Start to &Sim_Finish;
      do CCRE_ID=&CCRE_Start to &CCRE_Finish;
            do Dist_ID=&District_Start to &District_Finish;
                  do z=1 to 6;
                  output;
                  end;
            end;
      end;
end;

run;

/*Generate SF_ALL_Data Table.*/

Data SF_ALL_Data;
      Set SF_All_ID;
      Set SF_ALL;
run;


/*Update Stud_Scores with HLM Betas*/

%Macro HLM_Betas;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&CCRE_Start %to &CCRE_Finish;
            %do k=&District_Start %to &District_Finish;

/*Insert Beta Parameters, Predicted_Scores, and Residuals into Stud_Scores
Table*/

                              proc sql;
                                    UPDATE Work.Stud_Scores SET
Intercept_Beta = (select overall from Work.SF_ALL_DATA where
effect="Intercept" and Run_ID=&i and CCRE_ID=&j and Dist_ID=&k) where
Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                              quit;
                              proc sql;
```

```
                                UPDATE Work.Stud_Scores SET
Camp_Econ_Dis_Beta = (select overall from Work.SF_ALL_DATA where
effect="Camp_Econ_Dis" and Run_ID=&i and CCRE_ID=&j and Dist_ID=&k) where
Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET
Camp_Econ_Dis_Time_Beta = (select overall from Work.SF_ALL_DATA where
effect="Time*Camp_Econ_Dis" and Run_ID=&i and CCRE_ID=&j and Dist_ID=&k)
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET Time_Beta
= (select overall from Work.SF_ALL_DATA where effect="Time" and Run_ID=&i
and CCRE_ID=&j and Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET
Stud_Econ_Dis_Beta = (select overall from Work.SF_ALL_DATA where
effect="SESti" and Run_ID=&i and CCRE_ID=&j and Dist_ID=&k) where
Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET
Camp_Econ_Dis_Stud_Econ_Dis_Beta = (select overall from Work.SF_ALL_DATA
where effect="SESti*Camp_Econ_Dis" and Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET
Predicted_Score=Intercept_Beta + Camp_Econ_Dis_Beta*SESj + Time_Beta*5 +
Camp_Econ_Dis_Time_Beta*SESj*5 +
Camp_Econ_Dis_Stud_Econ_Dis_Beta*SESj*SESti + Stud_Econ_Dis_Beta* SESti;
                          quit;
                          proc sql;
                                UPDATE Work.Stud_Scores SET Residual =
Grade_8ij - Predicted_Score;
                          quit;
            %end;
      %end;
%end;

%Mend HLM_Betas;
%HLM_Betas;

/*Create Work.Camp_Rank Table.*/

data Work.Camp_Rank;

do Run_ID=&Sim_Start to &Sim_Finish;
      do CCRE_ID=&CCRE_Start to &CCRE_Finish;
            do Dist_ID=&District_Start to &District_Finish;
                  do Camp_ID=&Campus_Start to &Campus_Finish;
                        Camp_VA_Average=0;
                        Camp_VA_Rank=0;
                        Camp_Econ_Average=0;
```

```
                        Camp_Econ_Rank=0;
                        OUTPUT;
                  end;
            end;
      end;
end;

run;

/*Generate Campus Ranks*/

%Macro Camp_Rank;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&CCRE_Start %to &CCRE_Finish;
            %do k=&District_Start %to &District_Finish;
                  %do m=&Campus_Start %to &Campus_Finish;

/*Insert Campus Averages grouped by VA into Work.Camp_Rank Table*/

                              proc sql;
                                    update Work.Camp_Rank set
Camp_VA_Average=(select distinct avg (Residual) from Work.Stud_Scores
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k and Camp_ID=&m)
                                    where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k  and Camp_ID=&m;
                              quit;

/*Insert campus Averages grouped by Econ_Dis into Camp_Rank Table*/

                              proc sql;
                                    update Work.Camp_Rank set
Camp_Econ_Average=(select distinct avg (Grade_8_Pass) from
Work.Stud_Scores where SESti = 1 and Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k and Camp_ID=&m)
                                    where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k and Camp_ID=&m;
                              quit;

                  %end;

/*Rank campuses by VA and insert rankings into Camp_Rank Table*/

                  data Work.Camp_Rank_Temp;
                        Set Work.Camp_Rank;
                  run;

                  proc sql;
                        update Work.Camp_Rank as a set Camp_VA_Rank =
(select count(distinct b.Camp_VA_Average) from Work.Camp_Rank_Temp as b
where b.Camp_VA_Average > a.Camp_VA_Average and b.Run_ID=&i and
b.CCRE_ID=&j and b.Dist_ID=&k) + 1 where a.Run_ID=&i and a.CCRE_ID=&j and
a.Dist_ID=&k;
                  quit;

                  proc sql; delete * from work.camp_rank_temp; quit;
```

```
/*Rank campuses by Econ_Dis and insert rankings into Camp_Rank Table*/

                  data Work.Camp_Rank_Temp;
                        Set Work.Camp_Rank;
                  run;

                  proc sql;
                        update Work.Camp_Rank as a set Camp_Econ_Rank =
(select count(distinct b.Camp_Econ_Average) from Work.Camp_Rank_Temp as b
where b.Camp_Econ_Average > a.Camp_Econ_Average and b.Run_ID=&i and
b.CCRE_ID=&j and b.Dist_ID=&k) + 1 where a.Run_ID=&i and a.CCRE_ID=&j and
a.Dist_ID=&k;
                  quit;

                  proc sql; delete * from work.camp_rank_temp; quit;

            %end;
      %end;
%end;

%Mend Camp_Rank;
%Camp_Rank;

/*Create Work.Final_Report Table.*/
/*MyLib.Work.Final_Report Table will be used to report data.*/

data Work.Final_Report;
      do i=&CCRE_Start to &CCRE_Finish;
            do p=-100 to 100;
                  CCRE_ID=i;
                  Correlation_Range="-1.00 to <=-1.00";
                  Correlation_Range_Order=0;
                  Correlation=p/100;
                  Dist_ID_01=0;
                  Dist_ID_02=0;
                  Dist_ID_03=0;
                  Dist_ID_04=0;
                  Dist_ID_05=0;
                  output;
            end;
      end;
run;

proc sql;
      update Work.Final_Report set Correlation_Range_Order=1,
Correlation_Range= "-1.00 to <-.90" where Correlation >=-1.00 and
Correlation <-.90;
      update Work.Final_Report set Correlation_Range_Order=2,
Correlation_Range= "-.90 to <-.80" where Correlation >=-.9 and Correlation
<-.80;
      update Work.Final_Report set Correlation_Range_Order=3,
Correlation_Range= "-.80 to <-.70" where Correlation >=-.8 and Correlation
<-.70;
      update Work.Final_Report set Correlation_Range_Order=4,
Correlation_Range= "-.70 to <-.60" where Correlation >=-.7 and Correlation
<-.60;
```

```
	update Work.Final_Report set Correlation_Range_Order=5,
Correlation_Range= "-.60 to <-.50" where Correlation >=-.6 and Correlation
<-.50;
	update Work.Final_Report set Correlation_Range_Order=6,
Correlation_Range= "-.50 to <-.40" where Correlation >=-.5 and Correlation
<-.40;
	update Work.Final_Report set Correlation_Range_Order=7,
Correlation_Range= "-.40 to <-.30" where Correlation >=-.4 and Correlation
<-.30;
	update Work.Final_Report set Correlation_Range_Order=8,
Correlation_Range= "-.30 to <-.20" where Correlation >=-.3 and Correlation
<-.20;
	update Work.Final_Report set Correlation_Range_Order=9,
Correlation_Range= "-.20 to <-.10" where Correlation >=-.2 and Correlation
<-.10;
	update Work.Final_Report set Correlation_Range_Order=10,
Correlation_Range= "-.10 to <.00" where Correlation >=-.1 and Correlation
<.0;
	update Work.Final_Report set Correlation_Range_Order=11,
Correlation_Range= ".00 to <.10" where Correlation >=.0 and Correlation
<.10;
	update Work.Final_Report set Correlation_Range_Order=12,
Correlation_Range= ".10 to <.20" where Correlation >=.1 and Correlation
<.20;
	update Work.Final_Report set Correlation_Range_Order=13,
Correlation_Range= ".20 to <.30" where Correlation >=.2 and Correlation
<.30;
	update Work.Final_Report set Correlation_Range_Order=14,
Correlation_Range= ".30 to <.40" where Correlation >=.3 and Correlation
<.40;
	update Work.Final_Report set Correlation_Range_Order=15,
Correlation_Range= ".40 to <.50" where Correlation >=.4 and Correlation
<.50;
	update Work.Final_Report set Correlation_Range_Order=16,
Correlation_Range= ".50 to <.60" where Correlation >=.5 and Correlation
<.60;
	update Work.Final_Report set Correlation_Range_Order=17,
Correlation_Range= ".60 to <.70" where Correlation >=.6 and Correlation
<.70;
	update Work.Final_Report set Correlation_Range_Order=18,
Correlation_Range= ".70 to <.80" where Correlation >=.7 and Correlation
<.80;
	update Work.Final_Report set Correlation_Range_Order=19,
Correlation_Range= ".80 to <.90" where Correlation >=.8 and Correlation
<.90;
	update Work.Final_Report set Correlation_Range_Order=20,
Correlation_Range= ".90 to <1.00" where Correlation >=.9 and Correlation
<=1.00;
quit;

/*Create Work.Correlations Table.*/

proc sql;
	create table Work.Correlations
		(Run_ID numeric,
		CCRE_ID numeric,
		Dist_ID numeric,
```

```
            Correlation numeric);
quit;

%Macro Final_Report_Run;

/*Insert correlations into Correlations Table*/

proc sql;
      insert into Work.Correlations (Run_ID, CCRE_ID, Dist_ID,
Correlation)
      (select distinct Run_ID, CCRE_ID, Dist_ID, 0 from Work.Camp_Rank);
quit;

%do i=&Sim_Start %to &Sim_Finish;
      %do j=&CCRE_Start %to &CCRE_Finish;
            %do k=&District_Start %to &District_Finish;
                  proc sql;
                        insert into Work.Camp_Rank_Temp
                              (Run_ID, CCRE_ID, Dist_ID, Camp_ID,
Camp_VA_Average, Camp_VA_Rank, Camp_Econ_Average, Camp_Econ_Rank)
                        select Run_ID, CCRE_ID, Dist_ID, Camp_ID,
Camp_VA_Average, Camp_VA_Rank, Camp_Econ_Average, Camp_Econ_Rank
                        from Work.Camp_Rank where Run_ID=&i and CCRE_ID=&j
and Dist_ID=&k;
                  quit;

                  proc corr data=Work.Camp_Rank_Temp noprint outp=Pearson;
                        var Camp_VA_Rank Camp_Econ_Rank;
                  run;

                  proc sql;
                        update Work.Correlations Set Correlation = (select
round (Camp_VA_Rank,.01) from Pearson where _name_ = 'Camp_Econ_Rank')
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                  quit;

                  proc sql;
                        delete * from Work.Camp_Rank_Temp;
                  quit;
            %end;
      %end;
%end;

proc append base=Mylib.Correlations data=Work.Correlations;
run;

/*Update Work.Final_Report Table from Work.Correlations Table.*/

      %do j=&CCRE_Start %to &CCRE_Finish;
            %do y=-100 %to 100;
                  %do k=1 %to 5;
                        proc sql;
                        update Work.Final_Report set
Dist_ID_0&k=(select(count(Run_ID)) from Mylib.Correlations where
Correlation =&y/100 and CCRE_ID =&j and Dist_ID=&k group by CCRE_ID,
Dist_ID) where CCRE_ID=&j and Correlation=&y/100;
                        quit;
```

```
                  %end;
            %end;
      %end;
      %do kk=1 %to 5;
            proc sql;
            update Work.Final_Report set Dist_ID_0&kk = 0 where
Dist_ID_0&kk = .;
            quit;
      %end;
run;

/*Copy Work.Final_Report Table to Mylib.Final_Report Table.*/

Data Mylib.Final_Report;
      Set Work.Final_Report;
run;

/*Run Mylib.Final_Report Report.*/

proc sql;
      select
            CCRE_ID,
            Correlation_Range,
            Sum (Dist_ID_01) AS D1,
            Sum (Dist_ID_02) AS D2,
            Sum (Dist_ID_03) AS D3,
            Sum (Dist_ID_04) AS D4,
            Sum (Dist_ID_05) AS D5

      from Mylib.Final_Report group by CCRE_ID, Correlation_Range_Order,
Correlation_Range order by CCRE_ID, Correlation_Range_Order;
quit;

%Mend Final_Report_Run;

%Final_Report_Run;

%Monitor_Finish;

proc printo log; run;
```

APPENDIX B

SAS CODE: FIELD STUDY

```
/*Import CCISD Data File*/

PROC IMPORT OUT= WORK.CCISD_File
            DATAFILE= "C:\Users\barlowk\Documents\CCISD_File.txt"
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

/*Import NCE Conversion File*/

PROC IMPORT OUT= WORK.NCE
            DATAFILE= "C:\Users\barlowk\Documents\NCE.txt"
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

/*Generate Run, CCRE, District, and Campus ID's*/

%let GRun=1;
%let GCCRE=1;
%let GDistrict=1;
%let GCampus=9;

/*Create Stud_Scores Table*/

proc sql;
      create table Work.Stud_Scores
            (Run_ID numeric,
            CCRE_ID numeric,
            Dist_ID numeric,
            Camp_ID numeric,
            Stud_ID numeric,
            SESj numeric,
            SESti numeric,
            Grade_3ij numeric,
            Grade_4ij numeric,
            Grade_5ij numeric,
            Grade_6ij numeric,
            Grade_7ij numeric,
            Grade_8ij numeric,
            Grade_8_Pass numeric,
            Intercept_Beta numeric,
            Camp_Econ_Dis_Beta numeric,
            Time_Beta numeric,
            Camp_Econ_Dis_Time_Beta numeric,
            Stud_Econ_Dis_Beta numeric,
            Camp_Econ_Dis_Stud_Econ_Dis_Beta numeric,
            Predicted_Score numeric,
```

```
            Residual numeric);
quit;

/*Populate Stud_Scores Table from CCISD_File*/

proc sql;
      insert into Work.Stud_Scores (
            Run_ID,
            CCRE_ID,
            Dist_ID,
            Camp_ID,
            Stud_ID,
            SESj,
            SESti,
            Grade_3ij,
            Grade_4ij,
            Grade_5ij,
            Grade_6ij,
            Grade_7ij,
            Grade_8ij,
            Grade_8_Pass,
            Intercept_Beta,
            Camp_Econ_Dis_Beta,
            Time_Beta,
            Camp_Econ_Dis_Time_Beta,
            Stud_Econ_Dis_Beta,
            Camp_Econ_Dis_Stud_Econ_Dis_Beta,
            Predicted_Score,
            Residual)
      select &GRun,
            &GCCRE,
            &GDistrict,
            Camp_ID,
            Stud_ID,
            0,
            SES,
            Grade_3_Raw_Score,
            Grade_4_Raw_Score,
            Grade_5_Raw_Score,
            Grade_6_Raw_Score,
            Grade_7_Raw_Score,
            Grade_8_Raw_Score,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0,
            0
      from CCISD_File;
quit;

proc sql;
      update Work.Stud_Scores Set Camp_ID=1 where Camp_ID=84910004;
      update Work.Stud_Scores Set Camp_ID=2 where Camp_ID=84910042;
```

```
	update Work.Stud_Scores Set Camp_ID=3 where Camp_ID=84910043;
	update Work.Stud_Scores Set Camp_ID=4 where Camp_ID=84910044;
	update Work.Stud_Scores Set Camp_ID=5 where Camp_ID=84910045;
	update Work.Stud_Scores Set Camp_ID=6 where Camp_ID=84910046;
	update Work.Stud_Scores Set Camp_ID=7 where Camp_ID=84910047;
	update Work.Stud_Scores Set Camp_ID=8 where Camp_ID=84910048;
	update Work.Stud_Scores Set Camp_ID=9 where Camp_ID=84910049;
quit;


/*Create Stud_Scores_1 Table*/

data Work.Stud_Scores_1;
	set Work.Stud_Scores;
run;

/*Calculate SESj from SESti*/

proc sql;
	update Stud_Scores as a Set SESj = (select
100*(sum(b.SESti)/count(b.SESti)) from Stud_Scores_1 as b where (a.Camp_ID
= b.Camp_ID));
quit;

/*Update the "Guessing" Raw Scores with Random Raw Scores between 2SEMs
and 4SEMs greater than "Guessing". */

proc sql;
	update Stud_Scores set Grade_3ij = 15 + floor((5)*rand("Uniform"))
where Grade_3ij<15;
	update Stud_Scores set Grade_4ij = 16 + floor((5)*rand("Uniform"))
where Grade_4ij<16;
	update Stud_Scores set Grade_5ij = 16 + floor((5)*rand("Uniform"))
where Grade_5ij<16;
	update Stud_Scores set Grade_6ij = 17 + floor((6)*rand("Uniform"))
where Grade_6ij<17;
	update Stud_Scores set Grade_7ij = 18 + floor((6)*rand("Uniform"))
where Grade_7ij<18;
	update Stud_Scores set Grade_8ij = 19 + floor((6)*rand("Uniform"))
where Grade_8ij<19;
	update Stud_Scores set Grade_8_Pass = 1 where Grade_8ij>=30;
	update Stud_Scores set Grade_8_Pass = 0 where Grade_8ij<30;
quit;


proc sql;
	update Stud_Scores as a set Grade_3ij = (select b.NCE_Score from NCE
as b
	where (b.Time=0 and a.Grade_3ij=b.Raw_Score));

	update Stud_Scores as a set Grade_4ij = (select b.NCE_Score from NCE
as b
	where (b.Time=1 and a.Grade_4ij=b.Raw_Score));

	update Stud_Scores as a set Grade_5ij = (select b.NCE_Score from NCE
as b
	where (b.Time=2 and a.Grade_5ij=b.Raw_Score));
```

```
      update Stud_Scores as a set Grade_6ij = (select b.NCE_Score from NCE
as b
      where (b.Time=3 and a.Grade_6ij=b.Raw_Score));

      update Stud_Scores as a set Grade_7ij = (select b.NCE_Score from NCE
as b
      where (b.Time=4 and a.Grade_7ij=b.Raw_Score));

      update Stud_Scores as a set Grade_8ij = (select b.NCE_Score from NCE
as b
      where (b.Time=5 and a.Grade_8ij=b.Raw_Score));

quit;


/*Create Stud_Scores_Array Table*/

proc sql;
      create table Work.Stud_Scores_Array
            (Run_ID numeric,
            CCRE_ID numeric,
            Dist_ID numeric,
            Camp_ID numeric,
            Stud_ID numeric,
        Camp_Econ_Dis numeric,
            SESti numeric,
            Time numeric,
            Score numeric);
quit;

/*Populate Stud_Scores_Array from Stud_Scores.*/

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 0,
Grade_3ij
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 1,
Grade_4ij
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 2,
Grade_5ij
```

```
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 3,
Grade_6ij
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 4,
Grade_7ij
      from Work.Stud_Scores;
      quit;

proc sql;
      insert into Work.Stud_Scores_Array
            (Run_ID, CCRE_ID, Dist_ID,Camp_ID, Stud_ID, Camp_Econ_Dis,
SESti, Time, Score)
      select Run_ID, CCRE_ID,Dist_ID, Camp_ID, Stud_ID, SESj, SESti, 5,
Grade_8ij
      from Work.Stud_Scores;
      quit;


/*Create Work.SF_All Table.*/

proc sql;
      create table Work.SF_All
            (Effect char(19),
            Overall numeric);
quit;

/*Perform HLM Analysis.*/

%Macro HLM;

%do i=1 %to &GRun;
      %do j=1 %to &GCCRE;
            %do k=1 %to &GDistrict;
                  data Stud_Scores_Array_1;
                        set Stud_Scores_Array;
                        if Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                        run;

                        PROC MIXED DATA=Work.Stud_Scores_Array_1 COVTEST
NOCLPRINT  method = ML;
                        CLASS Stud_ID;
                        MODEL Score = Time SESti Camp_Econ_Dis
Camp_Econ_Dis*Time Camp_Econ_Dis*SESti/ SOLUTION;
                        RANDOM intercept Time SESti;
```

```
                        REPEATED /subject = Stud_ID type = cs rcorr;
                        ods output solutionf=Work.SF(keep=effect estimate
rename=(estimate=overall));
                        run;

                        proc append base=SF_All data=SF;
                        run;
            %end;
      %end;
%end;

%mend HLM;
%HLM;

/*Create Work.SF_ALL_ID Table.*/

Data SF_ALL_ID;

do Run_ID=1 to &GRun;
      do CCRE_ID=1 to &GCCRE;
            do Dist_ID=1 to &GDistrict;
                  do z=1 to 6;
                  output;
                  end;
            end;
      end;
end;

run;

/*Generate SF_ALL_Data Table.*/

Data SF_ALL_Data;
      Set SF_All_ID;
      Set SF_ALL;
run;

/*Update Stud_Scores with HLM Betas*/

%Macro HLM_Betas /*(Run=1, CCRE=1,  District=1)*/;

%do i=1 %to &GRun;
      %do j=1 %to &GCCRE;
            %do k=1 %to &GDistrict;

/*Insert Beta Parameters, Predicted_Scores, and Residuals into Stud_Scores
Table*/

                  proc sql;
                        UPDATE Work.Stud_Scores SET Intercept_Beta =
(select overall from Work.SF_ALL_DATA where effect="Intercept" and
Run_ID=&i and CCRE_ID=&j and Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j
and Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET Camp_Econ_Dis_Beta =
(select overall from Work.SF_ALL_DATA where effect="Camp_Econ_Dis" and
```

```
Run_ID=&i and CCRE_ID=&j and Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j
and Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET
Camp_Econ_Dis_Time_Beta = (select overall from Work.SF_ALL_DATA where
effect="Time*Camp_Econ_Dis" and Run_ID=&i and CCRE_ID=&j and Dist_ID=&k)
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET Time_Beta = (select
overall from Work.SF_ALL_DATA where effect="Time" and Run_ID=&i and
CCRE_ID=&j and Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET Stud_Econ_Dis_Beta =
(select overall from Work.SF_ALL_DATA where effect="SESti" and Run_ID=&i
and CCRE_ID=&j and Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET
Camp_Econ_Dis_Stud_Econ_Dis_Beta = (select overall from Work.SF_ALL_DATA
where effect="SESti*Camp_Econ_Dis" and Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k) where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET
Predicted_Score=Intercept_Beta + Camp_Econ_Dis_Beta*SESj + Time_Beta*5 +
Camp_Econ_Dis_Time_Beta*SESj*5 +
Camp_Econ_Dis_Stud_Econ_Dis_Beta*SESj*SESti + Stud_Econ_Dis_Beta* SESti;
                  quit;
                  proc sql;
                        UPDATE Work.Stud_Scores SET Residual = Grade_8ij -
Predicted_Score;
                  quit;

            %end;
      %end;
%end;

%Mend HLM_Betas;
%HLM_Betas;

/*Create Work.Camp_Rank Table.*/

data Work.Camp_Rank;

do Run_ID=1 to &GRun;
      do CCRE_ID=1 to &GCCRE;
            do Dist_ID=1 to &GDistrict;
                  do Camp_ID=1 to &GCampus;
                        Camp_VA_Average=0;
                        Camp_VA_Rank=0;
                        Camp_Econ_Average=0;
                        Camp_Econ_Rank=0;
                        OUTPUT;
```

```
                  end;
            end;
      end;
end;

run;


/*Generate Campus Ranks*/

%Macro Camp_Rank /*(Run=1, CCRE=1,  District=1, Campus=9)*/;

%do i=1 %to &GRun;
      %do j=1 %to &GCCRE;
            %do k=1 %to &GDistrict;
                  %do m=1 %to &GCampus;

/*Insert Campus Averages grouped by VA into Work.Camp_Rank Table*/

                              proc sql;
                                    update Work.Camp_Rank set
Camp_VA_Average=(select distinct avg (Residual) from Work.Stud_Scores
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k and Camp_ID=&m)
                                    where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k  and Camp_ID=&m;
                              quit;

/*Insert campus Averages grouped by Econ_Dis into Camp_Rank Table*/

                              proc sql;
                                    update Work.Camp_Rank set
Camp_Econ_Average=(select distinct avg (Grade_8_Pass) from
Work.Stud_Scores where SESti = 1 and Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k and Camp_ID=&m)
                                    where Run_ID=&i and CCRE_ID=&j and
Dist_ID=&k and Camp_ID=&m;
                              quit;

                  %end;

/*Rank campuses by VA and insert rankings into Camp_Rank Table*/

                  data Work.Camp_Rank_Temp;
                        Set Work.Camp_Rank;
                  run;

                  proc sql;
                        update Work.Camp_Rank as a set Camp_VA_Rank =
(select count(distinct b.Camp_VA_Average) from Work.Camp_Rank_Temp as b
where b.Camp_VA_Average > a.Camp_VA_Average and b.Run_ID=&i and
b.CCRE_ID=&j and b.Dist_ID=&k) + 1 where a.Run_ID=&i and a.CCRE_ID=&j and
a.Dist_ID=&k;
                  quit;

                  proc sql; delete * from work.camp_rank_temp; quit;
```

```
/*Rank campuses by Econ_Dis and insert rankings into Camp_Rank Table*/

                  data Work.Camp_Rank_Temp;
                        Set Work.Camp_Rank;
                  run;

                  proc sql;
                        update Work.Camp_Rank as a set Camp_Econ_Rank =
(select count(distinct b.Camp_Econ_Average) from Work.Camp_Rank_Temp as b
where b.Camp_Econ_Average > a.Camp_Econ_Average and b.Run_ID=&i and
b.CCRE_ID=&j and b.Dist_ID=&k) + 1 where a.Run_ID=&i and a.CCRE_ID=&j and
a.Dist_ID=&k;
                  quit;

                  proc sql; delete * from work.camp_rank_temp; quit;

            %end;
      %end;
%end;

%Mend Camp_Rank;
%Camp_Rank;

/*Create Work.Correlations Table.*/

proc sql;
      create table Work.Correlations
            (Run_ID numeric,
            CCRE_ID numeric,
            Dist_ID numeric,
            Correlation numeric);
quit;

%Macro Final_Report_Run;

/*Insert correlations into Correlations Table*/

proc sql;
      insert into Work.Correlations (Run_ID, CCRE_ID, Dist_ID,
Correlation)
      (select distinct Run_ID, CCRE_ID, Dist_ID, 0 from Work.Camp_Rank);
quit;

%do i=1 %to &GRun;
      %do j=1 %to &GCCRE;
            %do k=1 %to &GDistrict;
                  proc sql;
                        insert into Work.Camp_Rank_Temp
                              (Run_ID, CCRE_ID, Dist_ID, Camp_ID,
Camp_VA_Average, Camp_VA_Rank, Camp_Econ_Average, Camp_Econ_Rank)
                        select Run_ID, CCRE_ID, Dist_ID, Camp_ID,
Camp_VA_Average, Camp_VA_Rank, Camp_Econ_Average, Camp_Econ_Rank
                        from Work.Camp_Rank where Run_ID=&i and CCRE_ID=&j
and Dist_ID=&k;
                  quit;

                  proc corr data=Work.Camp_Rank_Temp noprint outp=Pearson;
```

```
                        var Camp_VA_Rank Camp_Econ_Rank;
                  run;

                  proc sql;
                        update Work.Correlations Set Correlation = (select
round (Camp_VA_Rank,.01) from Pearson where _name_ = 'Camp_Econ_Rank')
where Run_ID=&i and CCRE_ID=&j and Dist_ID=&k;
                  quit;

            %end;
      %end;
%end;

%Mend Final_Report_Run;
%Final_Report_Run;
```

APPENDIX C

GRADE 3 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2004

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
3	0	0	1228	-5.368	-63
3	0	1	1378	-5.200	-60
3	0	2	1487	-5.032	-56
3	0	3	1554	-4.864	-52
3	0	4	1603	-4.696	-49
3	0	5	1643	-4.528	-45
3	0	6	1678	-4.360	-42
3	0	7	1708	-4.192	-38
3	0	8	1735	-4.024	-35
3	0	9	1760	-3.856	-31
3	0	10	1783	-3.688	-28
3	0	11	1805	-3.521	-24
3	0	12	1826	-3.353	-21
3	0	13	1845	-3.185	-17
3	0	14	1864	-3.017	-14
3	0	15	1883	-2.849	-10
3	0	16	1901	-2.681	-6
3	0	17	1918	-2.513	-3
3	0	18	1936	-2.345	1
3	0	19	1953	-2.177	4
3	0	20	1986	-2.009	8
3	0	21	1987	-1.841	11
3	0	22	2004	-1.673	15
3	0	23	2021	-1.505	18
3	0	24	2041	-1.337	22
3	0	25	2055	-1.169	25
3	0	26	2073	-1.002	29
3	0	27	2100	-0.834	32
3	0	28	2111	-0.666	36
3	0	29	2131	-0.498	40
3	0	30	2152	-0.330	43
3	0	31	2174	-0.162	47
3	0	32	2198	0.006	50
3	0	33	2224	0.174	54
3	0	34	2253	0.342	57
3	0	35	2286	0.510	61

Appendix C continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
3	0	36	2326	0.678	64
3	0	37	2400	0.846	68
3	0	38	2440	1.014	71
3	0	39	2548	1.182	75
3	0	40	2697	1.349	78

Adapted from Texas Education Agency (2004).

APPENDIX D

GRADE 4 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2005

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
4	1	0	1280	-4.708	-49
4	1	1	1424	-4.567	-46
4	1	2	1529	-4.426	-43
4	1	3	1592	-4.286	-40
4	1	4	1638	-4.145	-37
4	1	5	1676	-4.004	-34
4	1	6	1707	-3.863	-31
4	1	7	1735	-3.722	-28
4	1	8	1760	-3.581	-25
4	1	9	1783	-3.440	-22
4	1	10	1804	-3.300	-19
4	1	11	1823	-3.159	-17
4	1	12	1842	-3.018	-14
4	1	13	1860	-2.877	-11
4	1	14	1877	-2.736	-8
4	1	15	1893	-2.595	-5
4	1	16	1909	-2.454	-2
4	1	17	1925	-2.314	1
4	1	18	1940	-2.173	4
4	1	19	1955	-2.032	7
4	1	20	1970	-1.891	10
4	1	21	1997	-1.750	13
4	1	22	2000	-1.609	16
4	1	23	2015	-1.468	19
4	1	24	2030	-1.328	22
4	1	25	2047	-1.187	25
4	1	26	2061	-1.046	28
4	1	27	2077	-0.905	31
4	1	28	2100	-0.764	34
4	1	29	2110	-0.623	37
4	1	30	2128	-0.482	40
4	1	31	2146	-0.342	43
4	1	32	2166	-0.201	46
4	1	33	2186	-0.060	49
4	1	34	2209	0.081	52
4	1	35	2233	0.222	55

Appendix D continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
4	1	36	2261	0.363	58
4	1	37	2292	0.504	61
4	1	38	2329	0.645	64
4	1	39	2400	0.785	67
4	1	40	2438	0.926	70
4	1	41	2541	1.067	72
4	1	42	2684	1.208	75

Adapted from Texas Education Agency (2005).

APPENDIX E

GRADE 5 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2006

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
5	2	0	1111	-4.886	-53
5	2	1	1284	-4.748	-50
5	2	2	1407	-4.611	-47
5	2	3	1482	-4.474	-44
5	2	4	1537	-4.336	-41
5	2	5	1580	-4.199	-38
5	2	6	1617	-4.061	-36
5	2	7	1650	-3.924	-33
5	2	8	1678	-3.786	-30
5	2	9	1705	-3.649	-27
5	2	10	1729	-3.512	-24
5	2	11	1752	-3.374	-21
5	2	12	1773	-3.237	-18
5	2	13	1793	-3.099	-15
5	2	14	1813	-2.962	-12
5	2	15	1832	-2.825	-9
5	2	16	1850	-2.687	-7
5	2	17	1868	-2.550	-4
5	2	18	1885	-2.412	-1
5	2	19	1902	-2.275	2
5	2	20	1919	-2.138	5
5	2	21	1936	-2.000	8
5	2	22	1953	-1.863	11
5	2	23	1978	-1.725	14
5	2	24	1986	-1.588	17
5	2	25	2003	-1.451	19
5	2	26	2037	-1.313	22
5	2	27	2038	-1.176	25
5	2	28	2056	-1.038	28
5	2	29	2074	-0.901	31
5	2	30	2100	-0.764	34
5	2	31	2112	-0.626	37
5	2	32	2132	-0.489	40
5	2	33	2154	-0.351	43
5	2	34	2176	-0.214	45
5	2	35	2200	-0.076	48

Appendix E Continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
5	2	36	2226	0.061	51
5	2	37	2255	0.198	54
5	2	38	2287	0.336	57
5	2	39	2324	0.473	60
5	2	40	2400	0.611	63
5	2	41	2422	0.748	66
5	2	42	2497	0.885	69
5	2	43	2620	1.023	72
5	2	44	2792	1.160	74

Adapted from Texas Education Agency (2006).

APPENDIX F

GRADE 6 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2007

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
6	3	0	1101	-4.269	-40
6	3	1	1278	-4.148	-37
6	3	2	1406	-4.027	-35
6	3	3	1483	-3.906	-32
6	3	4	1540	-3.785	-30
6	3	5	1585	-3.664	-27
6	3	6	1624	-3.543	-25
6	3	7	1657	-3.422	-22
6	3	8	1688	-3.301	-20
6	3	9	1715	-3.180	-17
6	3	10	1740	-3.059	-14
6	3	11	1764	-2.938	-12
6	3	12	1786	-2.817	-9
6	3	13	1808	-2.696	-7
6	3	14	1828	-2.575	-4
6	3	15	1848	-2.454	-2
6	3	16	1867	-2.333	1
6	3	17	1885	-2.212	3
6	3	18	1904	-2.091	6
6	3	19	1921	-1.970	9
6	3	20	1939	-1.849	11
6	3	21	1956	-1.728	14
6	3	22	1974	-1.607	16
6	3	23	1994	-1.486	19
6	3	24	2008	-1.365	21
6	3	25	2025	-1.244	24
6	3	26	2046	-1.123	26
6	3	27	2060	-1.002	29
6	3	28	2078	-0.881	31
6	3	29	2100	-0.760	34
6	3	30	2115	-0.639	37
6	3	31	2134	-0.518	39
6	3	32	2153	-0.397	42
6	3	33	2174	-0.276	44
6	3	34	2195	-0.155	47
6	3	35	2217	-0.034	49

Appendix F Continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
6	3	36	2241	0.087	52
6	3	37	2266	0.208	54
6	3	38	2294	0.329	57
6	3	39	2323	0.450	59
6	3	40	2357	0.571	62
6	3	41	2400	0.692	65
6	3	42	2440	0.813	67
6	3	43	2497	0.934	70
6	3	44	2574	1.055	72
6	3	45	2701	1.176	75
6	3	46	2878	1.297	77

Adapted from Texas Education Agency (2007).

APPENDIX G

GRADE 7 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2008

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
7	4	0	1362	-3.448	-23
7	4	1	1498	-3.346	-20
7	4	2	1597	-3.244	-18
7	4	3	1657	-3.142	-16
7	4	4	1701	-3.040	-14
7	4	5	1736	-2.938	-12
7	4	6	1765	-2.836	-10
7	4	7	1791	-2.734	-8
7	4	8	1814	-2.632	-5
7	4	9	1835	-2.530	-3
7	4	10	1854	-2.428	-1
7	4	11	1872	-2.326	1
7	4	12	1889	-2.224	3
7	4	13	1905	-2.122	5
7	4	14	1920	-2.020	7
7	4	15	1935	-1.918	10
7	4	16	1949	-1.816	12
7	4	17	1963	-1.714	14
7	4	18	1976	-1.612	16
7	4	19	1989	-1.510	18
7	4	20	2002	-1.408	20
7	4	21	2023	-1.306	22
7	4	22	2027	-1.204	25
7	4	23	2039	-1.102	27
7	4	24	2061	-1.000	29
7	4	25	2064	-0.898	31
7	4	26	2076	-0.796	33
7	4	27	2100	-0.694	35
7	4	28	2101	-0.592	38
7	4	29	2114	-0.490	40
7	4	30	2126	-0.388	42
7	4	31	2139	-0.286	44
7	4	32	2153	-0.184	46
7	4	33	2167	-0.082	48
7	4	34	2181	0.020	50
7	4	35	2196	0.122	53

Appendix G continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
7	4	36	2211	0.224	55
7	4	37	2228	0.326	57
7	4	38	2245	0.428	59
7	4	39	2264	0.530	61
7	4	40	2284	0.632	63
7	4	41	2307	0.734	65
7	4	42	2332	0.836	68
7	4	43	2360	0.938	70
7	4	44	2400	1.040	72
7	4	45	2437	1.142	74
7	4	46	2496	1.244	76
7	4	47	2593	1.346	78
7	4	48	2728	1.448	80

Adapted from Texas Education Agency (2008).

APPENDIX H

GRADE 8 MATHEMATICS TAKS NCE SCORE CONVERSION: SPRING 2009

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
8	5	0	1242	-4.056	-35
8	5	1	1398	-3.946	-33
8	5	2	1509	-3.836	-31
8	5	3	1576	-3.727	-28
8	5	4	1624	-3.617	-26
8	5	5	1663	-3.507	-24
8	5	6	1696	-3.397	-22
8	5	7	1724	-3.288	-19
8	5	8	1749	-3.178	-17
8	5	9	1772	-3.068	-15
8	5	10	1793	-2.959	-12
8	5	11	1813	-2.849	-10
8	5	12	1831	-2.739	-8
8	5	13	1848	-2.629	-5
8	5	14	1865	-2.520	-3
8	5	15	1881	-2.410	-1
8	5	16	1896	-2.300	2
8	5	17	1911	-2.190	4
8	5	18	1925	-2.081	6
8	5	19	1939	-1.971	8
8	5	20	1953	-1.861	11
8	5	21	1967	-1.752	13
8	5	22	1980	-1.642	15
8	5	23	1994	-1.532	18
8	5	24	2007	-1.422	20
8	5	25	2020	-1.313	22
8	5	26	2033	-1.203	25
8	5	27	2047	-1.093	27
8	5	28	2060	-0.984	29
8	5	29	2074	-0.874	32
8	5	30	2100	-0.764	34
8	5	31	2101	-0.654	36
8	5	32	2115	-0.545	39
8	5	33	2130	-0.435	41
8	5	34	2144	-0.325	43
8	5	35	2160	-0.216	45

Appendix H Continued

Grade_Level	Time	Raw_Score	Scale_Score	Z-Score	NCE_Score
8	5	36	2176	-0.106	48
8	5	37	2192	0.004	50
8	5	38	2210	0.114	52
8	5	39	2228	0.223	55
8	5	40	2247	0.333	57
8	5	41	2268	0.443	59
8	5	42	2291	0.553	62
8	5	43	2317	0.662	64
8	5	44	2345	0.772	66
8	5	45	2400	0.882	69
8	5	46	2416	0.991	71
8	5	47	2465	1.101	73
8	5	48	2531	1.211	76
8	5	49	2642	1.321	78
8	5	50	2796	1.430	80

Adapted from Texas Education Agency (2009).