8-1999

The impact of teacher effect on student math competency achievement

June C. Rivers Sanders

To the Graduate Council:

I am submitting herewith a dissertation written by June C. Rivers Sanders entitled "The impact of teacher effect on student math competency achievement." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Educational Administration.

Gerald C. Ubben, Major Professor

We have read this dissertation and recommend its acceptance:

Mary Jane Connelly, John Ray, Arnold Saxton

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by June C. Rivers Sanders entitled "The Impact of Teacher Effect on Student Math Competency Achievement." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Leadership Studies in Education.

Gerald C. Ubben, Major Professor

We have read this dissertation and recommend its acceptance:

Mary Jane Connelly

Arnold M Saxton

Accepted for the Council:

Associate Vice Chancellor and Dean of The Graduate School

THE IMPACT OF TEACHER EFFECT ON STUDENT MATH COMPETENCY ACHIEVEMENT

A Dissertation
Presented for the
Doctor of Education
Degree
The University of Tennessee, Knoxville

June C. Rivers Sanders
August 1999

DEDICATION

This dissertation is dedicated to my parents

Mary and Alfred O'Guin and the late James Cook

who challenged and guided me,

my children, Brian and Mark Rivers and Sarah and Will Sanders,

who remind me daily that quality teaching matters,

and my husband, Bill, whose patience supported me

throughout this process.

ACKNOWLEDGMENTS

Many colleagues, friends, and associates made this work possible, and I am grateful to each of them.

My graduate committee, Dr. Gary Ubben, Dr. Mary Jane Connelly, Dr. John Ray, and Dr. Arnold Saxton, for their guidance, their encouragement, their patience, and their expertise.

John Schneider of the University of Tennessee Value-Added Research and Assessment Center who merged the many data sources to create the data set this study required.

Dr. Soo-Hee Park and my colleagues at Test Processing Services whose professionalism and courage have meant much to me during this trying year.

The two school districts who made their Competency data available.

ABSTRACT

Using 4th grade student achievement math scores and Tennessee Value-Added Assessment System (TVAAS) math teacher effectiveness estimates for grades 5-8 (matched to individual students) as predictor variables, this study's purpose was to ascertain the residual and cumulative effects of these teachers on student 9th grade Competency scores. Students from 2 large districts who were 9th graders in the fall of 1997 with complete information were included. From the base model ANOVA, partial sums of squares for teacher estimates for all grades were significant, indicating the presence of residual effects free of partial confounding by other variables: 5th, F=8.75, p-value = 0.0031; 6th, F=14.82, p-value=0.0001; 7th, F=25.75, p-value = 0.0001; 8th, F=73.43, p-value = 0.0001. The differential effect of 5th and 6th grade teachers on 4th grade prior achievement levels produced mean Competency 9th grade scores for students in the bottom quartile ranging from 59-65 (se=0.63 to 0.81), depending on the level of teacher effectiveness students encountered in 5th grade. Competency means for similar low-achieving students assigned to 4 consecutive very ineffective teachers were 53.6 (se=0.733) versus 71.8 (se=.939) for 4 very effective teachers (70 is required for passing). Although a significant ethnic effect was apparent, it paled in comparison to this student Competency score response to 4 similarly effective teachers. Substituting a success variable in place of the 9th grade Competency score as the dependent variable allowed the determination of passing probabilities for various sequences of teachers with each quartile of 4th grade student prior achievement. As the assumed cut score was raised, student passing probability at 9th grade grew increasingly more dependent on 4th grade scores and teacher effectiveness estimates for grades 5-8: F=82.12 at 60 versus 129.83 at 80 (p-values=0.0001). Although all children benefited from highly effective teachers, the lower 50% benefited most. A change in level of effectiveness of teacher sequence increased the probability of passing to at least 50/50 for at least one quartile of students for each assumed cut score. Conclusions: (1) Tennessee's current high stakes student test is holding students accountable for something beyond their control, and the punitive policies associated with it should be reviewed for appropriateness. (2) the availability of effective teachers for lower achieving students is a critical determinant in their ability to pass. (3) Broad variability in teacher effectiveness suggests a need for improved teacher preparation and additional staff development opportunities for practicing teachers with improved accountability measures for teacher preparation.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Propelled into the national spotlight by the National Governors' Summit and President Clinton's education agenda, many states have sought to improve public education by focusing on standards for students. One strategy of the standards emphasis is an attempt to raise the learning expectation for students, and by so doing, to raise their achievement level. The high stakes consequences students experience when they fail to adequately demonstrate what they should know and be able to do by the completion of designated grade levels have life-changing potential for them—retention, failure to graduate, etc.

These consequences would be more palatable if achievement were solely the product of student effort. However, if the effectiveness of educational delivery has a measurable effect on student achievement, equity of educational opportunity becomes an issue. A primary fallacy in the standards movement might be the inability of public education to provide equitable learning opportunities for all students. If it is possible that some of the explanation for the variation in student learning comes from the effectiveness of the assigned classroom teachers, then some of the explanation of student failure might also come from the student's assignment to an ineffective classroom teacher or a sequence of ineffective teachers.

Background

Amid national attention to raising standards, Tennessee, along with several states, has endorsed a "get tough with kids" stance in educational policy. Superficially, it doesn't sound bad—kids don't try; provide them with a serious consequence to make them work harder, and they will. Students may exist who are not serious in their attempts at school or at testing, these students may benefit from this approach; however, calls and letters from parents and teachers suggest that some students try very hard and remain unsuccessful in their attempts at passing tests required for graduation.

In school year 1997-98, after two attempts at the Tennessee TCAP Competency Test, approximately 30,000 ninth graders failed to earn a passing math score and 25,000 were unsuccessful in language arts (State Testing and Evaluation Center, 1998). These figures represent more than one third of the students entering 9^{th} grade each year. The percentage of students per ethnic group passing the math subtest in 1995-96 is 73.1% Caucasian (N=45,908), 46.0% African-American (N=12,269), and 69% other (N=1,668) (State Testing and Evaluation Center, 1997).

Students graduating in Tennessee must earn a passing score on both subtests to receive a regular diploma. The items for these subtests are constructed to measure the student's mastery of 50 objectives (19 in math and 31 in language arts); the objectives were determined by a committee of educational practitioners to be necessary for success in high school course work. The objectives represent successful mastery of 8th grade material (State Testing and Evaluation Center, 1997). Recently approved State Board of Education policy will examine student performance at the 10th grade level (Minutes of the September 1998 Meeting of the Tennessee State Board of Education). In 2001, incoming freshman must successfully complete tests in Algebra I, language arts, and biology to earn a high school diploma. Does the responsibility for high failure rates on these tests lie solely with the student, or is there some evidence that teacher effectiveness might be a contributory factor?

Problem Statement

Tennessee has recently decided that all students must pass more difficult tests than the TCAP Competency test prior to graduation. The rationale is that setting a high standard will improve student performance. The underlying assumption is that the student is totally responsible for his/her performance on the test, and that failure to pass is the student's fault. This logic omits the possibility that school districts bear some responsibility for students to perform adequately on the TCAP Competency and that a student's sequence of ineffective or effective teachers might have a bearing on student performance. These effects may impact student TCAP Competency test pass/fail rates.

Purpose of the Study

The purpose of this study was to determine the degree to which teacher assignment might affect students' ability to pass the TCAP Competency test. This study estimated the effect of a student's series of effective/ineffective teachers on his/her TCAP Competency math score (a passing score of 70 is a graduation requirement). An intermediate step included the determination of the degree of residual teacher effect on the achievement of two cohorts of students, each in a large metropolitan district, across four years. Further, this study investigated the effect of varying the cut score for the test on the pass/fail probabilities for cohorts of students experiencing different sequences of teacher effectiveness.

Research Questions

1. What is the measurable effect of teacher effectiveness sequences on student Competency math scores?

2. What are the interactions among gender, ethnicity and prior levels of student achievement and sequences of teacher effectiveness?

3. How does varying the cut score required for passing the Competency test affect the mean probability of passing for students within specific teacher sequences?

Significance

Overall, this study provided valuable information for policy makers about educators' liability in student success/failure on high stakes student tests. It added to the body of knowledge concerning the way teacher effectiveness and sequences of teacher effectiveness affect student achievement. Using teacher effectiveness as measured by TVAAS, the study explained the effect of teacher effectiveness sequences on the TCAP Competency test, Tennessee's high stakes student test; it also described the pass/fail rates attributable to teacher effectiveness sequences at various assumed cut scores. By varying the assumed score requirement for Competency passage, it was possible to determine the effect of teacher sequence on pass/fail rates, adding to the body of knowledge available concerning the setting of cut scores.

Since policy makers in most states are currently participating in standards setting, the timeliness of this study increased its significance. Typically associated with standards setting is the assumption that some type of high stakes test will be used to measure a student's attainment of the standard, thus implying that some students may experience a positive or negative consequence as a result of the test. Since the empirical evidence of this study indicated a positive relationship between teacher effectiveness and academic achievement of at-risk populations, the measurement of teacher effectiveness became a critical determinant in pass/fail probability beyond the demonstration of student competency. Thus, by measuring student competencies against learning standards with no thought to equity of educational delivery, policy makers may have produced dire consequences for some low achieving students.

Testing policies which ignore possible teacher inequity in the measure of student performance raise an issue of fairness for students not previously explored. Linking educational delivery to student performance established an additional dimension relevant to the potential biases associated with testing. Thus, this study provided policy makers concerned with equity in student opportunity empirical evidence of the impact of teacher effectiveness sequences on student achievement and added to the body of knowledge concerning standard setting.

Assumptions

TVAAS estimates of teacher effects are valid, reliable unbiased estimates of the effects of teachers on academic progress of students (Harville). They indicate the amount of change in student scale scores during a year that is *beyond* the district average change *and* attributable to a particular teacher; TVAAS estimates may be either positive or negative, depending on the teacher's relative effectiveness or ineffectiveness. For instance, a teacher with an estimated effect of –6 would have taught children who scored, on average, six points below expectation on the TCAP achievement scale based on their previous history. Similarly, a teacher with an estimated effect of +6 would have taught children who scored six points above expectation based on their previous history (Sanders, Saxton, & Horn, 1997). Generally ineffective teachers are ineffective with all sub-populations of students and generally effective teachers are effective with all sub-populations of students (Sanders & Rivers, 1996).

The TCAP Competency mathematics subtest provides a valid measure of the knowledge and skills students should have mastered during their elementary school years provided that it was administered in a standardized setting with appropriate modifications. A committee of K-12 educators from across the state selected the objectives included in the test. The State Testing and Evaluation Center developed the test from items written by Tennessee teachers (State Testing and Evaluation Center, 1997).

The District A and B teachers included in the study were representative of the variability of their districts at their respective grade levels. The slight understating of children of color in the subset of data used in the study did not affect the study outcome. Prior achievement levels of the students with complete data represented the variability of all students taking the competency test in the ninth grade for Districts A and B.

Summary

The contemporary thinking in public education establishes standards of excellence for graduates of public education. However, the educational infrastructure that would support students of all ability levels in reaching these high standards may be lacking in the area of teacher effectiveness. Little is known about the differentiation in student consequences for those who spend entire grades, sometimes multiple grades, with very ineffective or very effective teachers. Additionally, as the score for passing is raised, the magnitude of the impact on a student's passing probability due to the effectiveness of his/her prior teachers is unknown.

The primary focus of this study was to discern the relationship between student Competency math scores and teacher effectiveness and the way it affected lower achieving students in particular. Chapter II provides a review of pertinent literature concerning students deemed to be at-risk for academic failure, teacher assignment policies and their impact on student achievement, and a history of methodology for measuring improvement in student achievement. Chapter III describes the methodology and the rationale for using it, the delimitations, and the possible limitations of the study. Chapter IV reports and interprets the results of the analyses. Chapter V provides a discussion of the findings and their implications.

Definitions

California Test of Basic Skills (CTBS/4) © -- a nationally normed test developed by CTB/MacMillian McGraw Hill.

Cumulative Teacher Effect - "the accumulation of measurable effects of teachers on students' learning in the years after the students have left their tutelage." (Sanders and Rivers, 1996)

Cut Score – a score that differentiates between student failure and success, e. g., the current Competency test cut score is 70.

Estimable Function – "a linear function of the parameters is defined as estimable if it is identically equal to some linear function of the expected value of the vector of observation." (Searle, 1971). The use of estimable functions in the context of this study enables the evaluation of the relationship between various combinations of the predictor variables (i.e., fourth grade prior achievement scores, teacher effectiveness, etc.) and the dependent variables (competency scores, probabilities of passing). This approach allows the results to be presented in a way that is more easily interpreted.

Mixed Model Methodology – "a statistical process which enables a repeated-measures, multivariate response analysis allowing the inclusion of all of the information available for each student regardless of the degree of missing information." (Sanders, Saxton, and Horn, 1997)

District Effect – "a measure of the influence of a school district or district on indicators of student learning." (Sanders and Horn, 1995)

School Effect – “a measure of the influence of a school on indicators of student learning.” (Sanders and Horn)

Teacher Effect - “a measure of the influence of a teacher on indicators of student learning.” (Sanders and Horn)

Teacher Effectiveness Sequence - the level of effectiveness of a student’s four individual teachers ordered by grade.

Tennessee Comprehensive Assessment Program (TCAP) – the assessment program mandated for K-12 students in Tennessee (achievement, writing, competency, and high school end of course); more specifically, the anacronym TCAP is commonly associated with achievement testing.

Tennessee Value Added Assessment System (TVAAS) – “a statistical process which provides measures of influence that school districts, schools, and teachers have on indicators of student learning.” (Sanders and Horn)

CHAPTER II

REVIEW OF THE LITERATURE

Research in three key areas influenced the development of this study: (1) the commonly held understanding of students at risk for academic failure, (2) teacher assignment policies and practices and their impact on student achievement, and (3) the development of methodology for measuring improvement in student achievement. The first section of this chapter traces the educational researcher's attempt to grapple with variation in student achievement across ethnicity, gender, and socioeconomic variables, and the ensuing conflict between the educator and the economist. The second section focuses on the ways teachers are assigned to schools and students are assigned to teachers within schools and the impact of these policies and practices. The final section examines the improvement of statistical measurement through the development of more sophisticated methodology and technology.

Students at Risk for Academic Failure

The link between earnings and prior education has instigated some of the research in this country concerning student at-riskedness. For some time, economists have viewed public education's inability to keep more students connected to school until they reach some level of Competency as a failure of public education (Hanushek, 1996). Educators have countered with claims that they are powerless to impact academic failure without increased spending. Increased spending across the past thirty years, however, has not resulted in "measurable increases in student performance" (Smith, Scoll, & Link, 1996).

The most blatant lack of student success would be found, of course, in students who experience academic failure and leave school before completion (Kronick and Hargis, 1990). Coleman concluded in the early 1960's "schools have great uniformity insofar as their effect on the learning of pupils is concerned." He attributed variation in student achievement primarily to home and community influences and to the attitudinal variation of individual student populations resulting from the two. This review suggests that this finding from so renowned a source remained virtually unchallenged for the next thirty years. Only in this decade have researchers begun to chisel at this finding as a potentially invisible barrier to student achievement (Sanders & Rivers, 1996 & Sanders, Horn, & Wright, 1997).

"Dropout" was first coined to describe the social and economic problem in the 1960s: dropouts were primarily male, living in poor inner-city communities, prone to juvenile delinquency, and typically psychologically troubled (Dorn, 1996). Derived initially from the demographics common to dropouts and ultimately expanded to include characteristics of families, schools and communities, the probabilities associated with these indicators reflected the students' likelihood for academic failure (Waggoner, 1991).

Peng and Lee (1992), using eighth grade data from the National Educational Longitudinal Study of 1988 and a follow-up in 1990, investigated the effect of interaction of various characteristics on at-risk probabilities: students with single parents, parents who often indicated a lack of education themselves, families with low incomes, families who spoke little English, students in schools with over 50% participation in federal meal programs, students with limited supervision after school, students in urban and rural public schools. Through this analysis they established a basis for determining at-risk students--children from impoverished communities and children with multiple risk characteristics, particularly when the children were minority students. The work of Peng and Lee suggested that some parent/family characteristics might, in reality, be indicators of a lack of available resources. Although critics might question that mere availability would impact at-risk children, Orfield, et. al., (1997) report that demand among urban minority families for quality educational opportunities increases with availability.

Teacher Assignment Policies and Practices and their Impact on Student Achievement

Bewildered by factory delivery organizational models founded on calendar years and grade levels and overly sensitive to the negative consequences of tracking, the frustrated education community has sometimes held itself harmless to criticisms of the negative effects of schools and teachers on at-risk students. More frequently they have focused on ways to acclimate the student to the school culture rather than ways to accelerate the at-risk student's academic achievement (Cuban, 1989, Kronick & Hargis). Administrators may have sought solace in the statistics which define the improbability of educating the "at-risk," allowing their classrooms to become the dumping ground for incompetent or less effective teachers and administrators (Bridges, 1996). According to Bridges, relatively ineffective teachers tended to be assigned to students characterized by a minimal public voice, students typically thought of as at-risk by the education community. Schools characterized as having large at-risk populations frequently reported one or more of the following characteristics: high student transfer rates, high percentages of students eligible for free/reduced meals, or high percentages of minority students.

Bridges' findings of teacher placement were collaborated by Ferguson (1990), Ingersoll and Gruber (1996) and the National Commission on Teaching and America's Future (1996). Ferguson, based on research linking Texas teacher Competency test data and student achievement, reported that children of color in this state tended to have teachers with weaker language skills. The consequence of the teachers' weaker skills was more than 25% difference in student performance of comparable groups in math and reading between black and white students and 20% in Hispanic and white students. Although they did not address the consequences of out-of-field teaching, Ingersoll and Gruber showed from national data that more than one fourth of all high school students taking core courses were likely to have a math teacher without at least a college minor in mathematics or mathematics education. Low-income or urban schools were less likely to have teachers with basic qualifications than affluent schools. They further showed that seventh and eighth grade students were more likely to have underqualified teachers than high school students. Lower achieving students were assigned more frequently to underqualified teachers than higher achieving students, but Ingersall and Gruber did not conclude that minorities were more often served by underqualified teachers than classes of predominately white students. The National Commission on Teaching and America's Future (1996), however, concluded that the probability of a minority student's math or science teacher having earned a teaching license and a degree in math or science was less than 50%.

Aside from the Ferguson research on the effect of inferior teachers on student learning, Bodenhausen (1988) measured the effects of teachers with strong academic backgrounds and multiple years of experience. Bodenhausen was able to show that students taking Advanced Placement courses under the tutelage of teachers with strong academic backgrounds in majors in the subjects they were teaching had a better probability of scoring at proficient levels on the Advanced Placement tests. Further students with teachers with at least 10 years of experience tended to score higher than students of less experienced teachers. Challenging one characteristic of at-riskness, Bodenhausen found no effect on student Advanced Placement test results when minority students were over-represented in the classes of teachers with strong academic backgrounds.

Summers and Wolfe (1977) looked at teacher preparation and years of experience and their effect on student learning. They maintained that students at risk because of low family income were more likely to benefit most from teachers graduating from colleges considered to be more effective in preparing quality teachers. Generally, however, there was a negative relationship between teacher experience and the achievement of lower-achieving students. One possibility for the negative relationship of teacher experience and student achievement could be that less effective, more experienced teachers tended to be assigned to schools with lower-achieving students. Krei (1993) provided insight into staffing at low-income schools. Students in low-income schools were more likely to receive inexperienced teachers or teachers with experience who were considered to be less qualified. She (1998) further concluded that staffing policies and practices continued to be a detriment to the assignment of quality teachers to schools for students without a public voice, because experienced teachers from urban schools tended to seek appointment to suburban schools as their seniority allowed their transfer requests.

From exploratory research into pupil assignment within schools, Monk (1987) reported there was a reluctance of administrators to acknowledge the damage to students for having experienced ineffective teachers. The more popularly held belief was that although teachers might vary in strengths, they were equivalent over all. When pressed about their assignment of students to seemingly incompetent teachers, principals indicated such concessions as "easy classes, small classes, etc." to accommodate the teacher. Within the denial of potential harm to students was the concession, however, that only one child per family was typically assigned to the same incompetent teacher.

Development of Methodology for Measuring Improvement in Student Achievement

Although the complexity of teacher effectiveness has been linked to a variety of possible indicators—teacher language skills, academic background, experience--historically, educators have taken issue with methodology that measured teacher effectiveness based on student performance on standardized tests (Popham, 1997). They have cited the statistician's inability to remove the various confoundings of socioeconomic status, ethnicity, gender, parenting variables, etc., from the measures. Further, educators have contended these confoundings made public reporting unfair to low-scoring schools with high percentages of students eligible for free and reduced meals, those with diverse student populations, etc., because raw score means did not adequately reflect what the students had achieved (Kortez, 1996). Raw scores have typically been reported in aggregated form in cross-sectional analyses.

Another concern associated with test data usage is the validity of the measurement of standardized tests. According to Stiggins (1994), educators have asserted that the limitations of paper and pencil testing lack the robustness to measure "authentic" learning and raised questions concerning the alignment between actual test content and curricular objectives in mandated standardized testing. In actuality, Stiggins maintains that paper and pencil testing can be used to measure lower to intermediate levels of learning and to provide specific identification of weaknesses in knowledge and skill mastery that prohibit student ability to demonstrate master of higher levels of learning. Stiggins contends that one explanation of the problem lies in the under-preparation of teachers in mastering effective assessment practices.

In initiating statewide achievement testing in 1991, Tennessee policy makers saw fit to minimize as much as possible the argument over alignment between testing and curricular alignment. A team of educators from across the state was selected to review various tests available, comparing them to the curricular objectives in the five subject areas to be tested. The California Test of Basic Skills (CTBS/4) © norm-referenced portion of the TCAP achievement test was recommended by the committee of educators for adoption by the state because it most closely represented the state's curricular objectives (Background of the Tennessee Comprehensive Assessment Program, 1988).

The validity of the chief criticisms for using student achievement in the evaluation of districts, schools, and teachers is challenged by more sophisticated statistical techniques than are required for reporting raw scores and by a more realistic understanding of the application of paper and pencil testing (Stiggins; Sanders & Horn, 1995). Raudenbush and Bryk (1988) reported that longitudinal analysis measuring schooling effects made gain versus position the desirable outcome, strengthening causal inferences of research based on nonexperimental design. Current attention to longitudinal analyses of student achievement data continue through improved technology and a reassessment of the appropriateness of quantitative methodology for education research (e. g., Ross, Sanders, Wright, & Stringfield, 1998).

Tennessee's Value-Added Assessment System (TVAAS) is an example of quantitative methodology for estimating the effects of school districts, schools, and teachers on individual student achievement from standardized achievement test scores (Bock and Wolfe, 1995, Darlington, 1997, & Harville, D. A., 1995). TVAAS estimates the schooling effects without the biases commonly associated with raw test score means. Application of TVAAS methodology removes the commonly recognized confoundings by allowing each student to serve as his/her own blocking factor (Analyses reported in the Graphical Summary of Educational Findings from the Tennessee Value-Added Assessment System (TVAAS), 1995 show no correlation between the means and socioeconomic status or ethnicity). The simultaneous estimation of the five subject effects produces unbiased estimates for each individual subject. Further, the more sophisticated methodology of TVAAS accommodates students with missing data in the estimations of schooling effects (Typically, the methodologies incorporated into educational research have not accommodated the utilization of partial data on individual students). Since "at-risk" students are more likely to miss tests because of sporadic school attendance, their influence on the results of some investigations of schooling effects has been minimized or excluded. Because all available student data is used in the TVAAS model to estimate schooling effects (district, school, and teacher), the estimates it provides are impacted less by poor attendance of sub-populations. The preciseness of the estimates is further enhanced by the annual recalculation of the effects of prior years as the most current data is incorporated into the TVAAS database (Sanders, Saxton, and Horn, 1997).

In a preliminary dissertation research project conducted on the math test results in grades three-five for two large school districts, teacher effect on student math achievement levels was both measurable and cumulative up to two years beyond the initial teacher treatment (Sanders & Rivers, 1996). This study used a simplistic version of gain per teacher for ranking and categorizing teacher effectiveness rather than the more sophisticated, and possibly more conservative, TVAAS teacher effectiveness estimates. Sanders and Rivers also showed that students with similar prior-achievement levels benefit similarly from the same level of teacher effectiveness regardless of student ethnicity. In the two districts that were a part of the study, minority students were under-represented by about 10% in the classrooms of very effective teachers and similarly over-represented in the classrooms of very ineffective teachers. A follow-up study by the Dallas Public School District using their own test data confirmed the findings (Jordan, Mendro, & Weerasinghe, 1997). Sanders, Horn, and Wright (1997) have shown that, holding student ability constant, teacher effectiveness is the single largest factor examined thus far affecting student achievement. The measurable cumulative teacher effects suggest that high stakes standards hold students accountable for the effectiveness of their teachers. Is it possible that some degree of student success on these tests—or the lack of it—can be traced to the sequence of teaching effectiveness a student encounters? The limited research into the impact of teacher effectiveness and its latent consequence for students suggests the problem to be addressed by this study.

CHAPTER III

METHODS AND MATERIALS

That teacher effectiveness has a residual and latent effect on student Competency math scores is the research hypothesis for this study. This chapter outlines the methodology and describes the data used in the analyses. The first section delimits the data, both the students and teachers to be included and the measurement methodologies previously designated as standard procedure by Tennessee's accountability legislation. The second describes the study's assumptions. The third outlines the methodology and the rationale for its selection. The final section describes the limitations of the study.

Delimitations

Students and Teachers.

This study was limited to the first-time ninth grade students from two large school districts with known Competency math scores from the fall 1997 test administration; one district will be denoted as District A and another as District B. These students would have been and fifth graders in 1994. Data for the study included ninth grade Competency math scores for these students and fourth grade achievement test scale scores.

The teachers included in the study were the mathematics teachers of the students described above. Students of teachers for whom TVAAS estimates could not be obtained were excluded from the study. The teachers included from Districts A and B would have taught the students mathematics during their fifth, sixth, seventh and/or eighth grade years and have had at least one year of measurable teacher effect in mathematics prior to teaching the children.

Measurement Methodologies.

The three measurements for this study included two independent estimates of student achievement/mastery in mathematics (TCAP math achievement scale scores for grade four and TCAP Competency math scores for grade nine) and the TVAAS quantitative estimates of teacher effectiveness. The TCAP Competency test is administered to students for the first time in the fall of their ninth grade year. Although the test consists of a mathematics and language arts subtest, only the results of the mathematics subtest administration in the fall of 1997 were considered for this study.

Independent of the TCAP Competency test, Tennessee students in grades three through eight were tested each year to measure their achievement in math, reading, language arts, science, and social studies. The fourth grade mathematics scores were used as a covariate for various parts of the study.

Although the tests were developed and administered independently, there was a strong correlation between the Competency and achievement tests. According to the State Testing and Evaluation Center (1997), the correlation between student performance on the spring (8th grade) TCAP achievement math test and the fall (9th grade) TCAP Competency math subtest was .80 (p=.001).

The TCAP achievement test norm referenced scale scores for individual students have been merged at the student level into the Tennessee Value-Added Assessment System (TVAAS) longitudinal database. Currently, this database includes individual student records from administrations of the TCAP achievement test from 1991 (mandated for grades two through eight) through 1998 (mandated for grades three through eight beginning with this year). The CTBS/4 and various hybrids of CTBS/4 provided the first seven years of data, and the CTB-McGraw Hill *TerraNova* provided the eighth year. Although prior years of CTBS/4 were equated to the *TerraNova* for future reporting years, the fourth grade scores used in this study were actual CTBS/4 scale scores.

The TVAAS database provides the basis for the state's estimation of district, school, and teacher effects in these same subjects. TVAAS reporting at each level was required by Tennessee's Education Improvement Act of 1992. District effects were first publicly reported in 1993, and school effects were first reported in the fall of 1994. Beginning with 1994, data connecting teachers to individual students were collected, allowing individual teacher effects to be estimated annually from that time. This allows for the cross-referencing of teacher effectiveness estimates to a teacher's individual students in each of five possible subjects: math, reading, language arts, social studies, and science. The first teacher effect reports were released in 1996. Unlike district and school effects, teacher effects reports were exempted from public scrutiny by statute to protect the rights of individual teachers (Ceperley and Reel, 1997).

TVAAS teacher effects are reported for individual years and in three-year averages and were available for four years. Each year's estimates are recalculated annually, using all available data from both current and previous years, to provide the most accurate measurement of teacher contribution to student progress. The inclusion of data from previous years ensures that teachers benefit from the most accurate estimate of their individual contribution. To minimize bias potential, the most recently calculated TVAAS estimates of math teacher effects that *did not include* the years *during or after* the students were under the tutelage of these specific teachers were used for this study. Thus, the teacher estimates selected for this study were based on measurements of student achievement before the students in the sample were assigned to teachers.

Methodology

The student level information necessary for these analyses was acquired from multiple sources and prepared for these analyses by The University of Tennessee Value-Added Research and Assessment Center. This preparation required the merging of fall 1997 student Competency data from Districts A and B with achievement and teacher effectiveness estimates for the same students. The Competency data contained the math scores, district and school numbers, the student's names, social security numbers, grades, birth dates, a ninth grade repeater indication (ninth grade repeaters were excluded from the study), gender, ethnicity, and a special education indicator. The Competency student records were matched at student level (name, social security number, and birth date) with records in the TVAAS database. By searching the entire state database, it was possible to locate TVAAS records for students who might have moved from other districts prior to taking the ninth grade test, minimizing the number of incomplete records. Mathematics scale scores for grade four from the TVAAS database were added to the Competency data. Students were ranked within each district and assigned a quartile rank based on their fourth grade performance on the TCAP mathematics test (e. g., the top 25% of the students were given the highest rank, the second 25%, the second highest, etc.). This quartile rank added to the student records as an indicator of prior student achievement level. Additionally, each student's record was encoded with a success variable for each of the following assumed cut scores: 60, 65, 70, 75, 80. For example, if a student's earned Competency mathematics score was 73, the success variables for 60, 65, and 70 were coded with a "1" to reflect success, while the success variables 75 and 80 were coded with a "0" to reflect failure.

The TVAAS numerical teacher identifiers for mathematics teachers for grade levels five through eight (beginning with 1994) were added to the individual student records. The grade and numerical teacher identifiers were used to match TVAAS teacher effect estimates to the individual Competency records by student-year. Students with no teacher identifier were ultimately dropped from the analyses. The teacher effect estimates for each grade for each school district were ranked and divided into quartiles with the lowest quartile having a teacher effectiveness rank of "1," the middle quartiles a rank of "2," and the highest quartile a rank of "3." The teacher effectiveness ranks for the appropriate grade/year were added to the individual student Competency records to facilitate both the analyses and the reporting of results.

Ascertaining the Relationship between Student Competency Math Scores and Teacher Effectiveness.

A scatter plot of the Competency mathematics scores versus the fourth grade mathematics TCAP achievement scores for each district was produced. Additionally, a univariate procedure was used to determine the normality of each data set. The scatter plot suggested the need for both a linear and quadratic relationship between the fourth grade mathematics TCAP achievement score and the Competency mathematics score. Thus, the fourth grade achievement quartile rank was included as a class variable to subsequent models in the analyses.

The base model for ascertaining the relationship between Competency math scores and teacher effectiveness included (1) fourth grade mathematics achievement scores (covariate), (2) the fourth grade student quartile rank (within district) as a discrete class variable, (3) the district, (4) the interaction of fourth grade score and district, (5-8) the fifth, sixth, seventh, and eighth grade TVAAS estimates of teacher effectiveness, and (9,10) the interactions of fourth grade scores with fifth and sixth grade TVAAS teacher estimates. Prior analyses had shown no significant district and TVAAS teacher estimate interaction effects, and this absence of interaction allowed for the simultaneous inclusion into one base analysis both districts' Competency scores. Additionally, there were no significant two-way interactions between fourth grade scores and seventh and eighth grade TVAAS estimates, nor were there significant two-way interactions of TVAAS estimates across grade (fifth and sixth, sixth and seventh, or seventh and eighth). These variables were eliminated from the final model for predicting Competency math scores (statistical model included in Appendix A).

The analysis of variance resulting from fitting the base model to the data produced a solution vector that was non-full rank. Since a non-full rank model estimates linear combinations of parameters rather than the parameters themselves, making the solution vector non-unique, it was necessary to specify a series of estimable functions to be estimated as the model was fit to the data. For the evaluation, various independent variables were set at predetermined levels to provide the Competency means necessary for answering the questions of the research hypotheses (A sample estimable function is included in Appendix B). This process was chosen to enable graphic and tabular views, thus simplifying the interpretation.

The distributions of the continuous variables were evaluated to ascertain the various combinations of student scores and levels of teacher effectiveness to be included in the estimable functions. These means were incorporated into estimable functions and applied to the solution vector from the model above. For evaluation and reporting purposes, fourth grade prior-achievement levels corresponding to prior student quartiles were chosen (667.8, 701.1, 724.6, and 766.4). Three levels of teacher effectiveness by grade (grade five: -6, 0, 8; grade six: -6, .3, 7.5; grade seven: -6, -.3, 7.5; and grade eight: -6, 0, 6) were selected for the evaluation to classify varying levels of teacher effectiveness within a grade. Estimates of Competency means for each level of fourth grade prior achievement with each category of teacher effectiveness for grades five through eight were produced. Additionally, other functions were used to provide estimates of the cumulative effect of four consecutive very ineffective (lower quartile) elementary teachers, four consecutive average (second and third quartiles) teachers and four consecutive very effective teachers (top quartile) on ninth grade mathematics Competency scores for each of the four groups.

<u>Exploring Ethnicity and Gender Interaction with Teacher Effectiveness</u>.

One of the research questions required the evaluation of the existence of ethnic by gender by teacher effect interaction. Ethnicity, gender, district and ethnicity interaction, ethnicity and teacher estimate (grades five, six, seven, and eight), and fourth grade score and ethnicity and teacher estimates for grades five and six were added to the model above. Again a series of functions were included to produce estimates of Competency means for the sequences of teacher effects for each combination of teacher effectiveness-ethnic-fourth grade prior-achievement rank quartile.

<u>Unforeseen Implications of Raising Cut Scores</u>.

The purpose of this series of analyses was the estimation of a student's probability for passing the test as the assumed cut score varied from 60 to 80 in 5-step increments (the current cut score for the test is 70). Again the data from the two districts were analyzed simultaneously with a series of SAS © GLM © (SAS Institute Inc., 1997) procedures. A success variable (embedded in the student records in prior steps as "1" for passing and "0" for failing at each proposed cut score) replaced the competency math score as the dependent variable in the base model described above. In these analyses, the same series of functions used with the base model provided estimates of success probability means for each teacher effectiveness and prior-achievement interaction for each grade. They also provided a means of evaluating the cumulative effects of teachers on the passing probabilities of students experiencing four consecutive very ineffective, four consecutive moderately effective, and four consecutive very effective teachers at each prior-achievement level.

Limitations

Many students who took the Competency test had not been included in all previous administrations of the achievement tests and thereby had incomplete information. Therefore, these students were not included in the analyses because these analyses required complete information of the students included. It is recognized that this is a truncated data set with disproportionately less lower achieving students and that this may have produced a dampening of the effects.

CHAPTER IV

RESULTS

This chapter describes the results of the analyses used for the study. The first section shows the results of the base model analysis (ANOVA table, solution vector, and estimable functions). The second section reports the results of tests for ethnic and gender interaction with teacher sequences. Finally, the third section reports the changes in probabilities at passing related to teacher effect when cut scores are varied. In all instances, p-values of 0.05 or less were considered significant.

Base Model Results

Since the primary emphasis of this study was to ascertain the residual and cumulative effects of teachers on Competency scores, effects of interest were the TVAAS Teacher Effectiveness Estimates for grades five through eight and the interactions of fifth and sixth grade teachers with fourth grade scores. The resulting analysis of variance table appears in Table 1. The model explained approximately 49.6% of the variation in the Competency scores (F = 213.20, p-value = 0.0001).

The partial sums of squares showed significant F-statistics (Table 1) for teacher effects at all grades, confirming the presence of a residual teacher

Table 1. Source of Variation in Competency Math Scores.

Dependent Variable: Competency Math Score		(Type I)			
Source of Variation	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model Type	12	286213.03	23851.09	213.20	0.0001
Residual	2599	290760.61	111.87		
Corrected Total	2611	576973.64			
	R-Square				Competency Math Mean
	0.4961				76.68

		Partial (SAS Type III)			
Source of Variation	DF	Sum of Squares	Mean Square	F Value	Pr > F
4th Grade Score	1	14536.02	14536.17	129.93	0.0001
4th Grade Quartile	3	7095.27	2365.09	21.14	0.0001
District Number	1	1775.69	1775.69	15.87	0.0001
4th Grade Score * District Number	1	1458.85	1458.85	13.04	0.0003
5th Grade TVAAS Teacher Effectiveness Estimate	1	979.26	979.26	8.75	0.0031
6th Grade TVAAS Teacher Effectiveness Estimate	1	1657.99	1657.99	14.82	0.0001
7th Grade TVAAS Teacher Effectiveness Estimate	1	1972.42	1972.42	17.63	0.0001
8th Grade TVAAS Teacher Effectiveness Estimate	1	8214.85	8214.85	73.43	0.0001
4th Grade Score * 5th Grade TVAAS Estimate	1	725.73	725.73	6.49	0.0109
4th Grade Score * 6th Grade TVAAS Estimate	1	1250.91	1250.91	11.18	0.0008

effect long after students leave their classrooms: fifth grade--*F* = 8.75, p-value = 0.0031; sixth grade--*F* = 14.82, *p*-value = 0.0001; seventh grade--*F* = 17.63, *p*-value = 0.0001; eighth grade--*F* = 73.43, *p*-value = 0.0001. The significant *F*-statistics for the partial sums of squares for teacher estimates indicated the presence of the residual effects free of partial confounding by other variables. With the fifth grade significance, this study confirmed the presence of a residual teacher effect on student achievement four years after students leave fifth grade (preliminary research (Sanders & Rivers) had confirmed a two-year residual effect).

The significant interaction (Table 1) between fourth grade scores and fifth and sixth grade teachers indicated a differential effect of teachers on varying student achievement levels (*F* = 6.49, *p*-value = 0.0109; *F* = 11.18, *p*-value = 0.0008, respectively). This interaction will be further explored in the discussion of estimated means.

The solution vector for this analysis (see Table 2) provided an early indication of the magnitude of the individual teacher effects. As reported earlier, the effectiveness estimates for very ineffective fifth grade teachers average –6, while very effective teachers average +8. Students with very ineffective fifth grade teachers are predicted to score as much as 28 points lower on Competency than students with similar prior-achievement levels who benefit from very effective fifth grade teachers (-12 to +16). This difference is calculated by ignoring the added interaction effect and using the extremes of –6 and +8 in conjunction with the fifth grade teacher effect estimate of 2.0390.

Table 2. Solution Vector for Competency Math Score.

Parameter		Estimate	T for HO: Parameter=0	Pr>|T|	Std Error of Estimate
Intercept		-14.1798 B	-1.29	0.1957	10.9565
4th Grade Score		0.1311 B	9.14	0.0001	0.0144
4th Grade Quartile	0	-9.0635 B	-6.43	0.0001	1.4096
	1	-5.4074 B	-5.42	0.0001	0.9969
	2	-1.4174 B	-1.88	0.0606	0.7552
	3	0.0000 B	.	.	.
District Number	A	-31.3134 B	-3.98	0.0001	7.8598
	B	0.0000 B	.	.	.
4th Grade Score * District Number	A	0.0393 B	3.61	0.0003	0.0109
	B	0.0000 B	.	.	.
5th Grade TVAAS Teacher Effectiveness Estimate		2.0390	2.96	0.0031	0.6892
6th Grade TVAAS Teacher Effectiveness Estimate		2.4091	3.85	0.0001	0.6258
7th Grade TVAAS Teacher Effectiveness Estimate		0.1605	4.20	0.0001	0.0382
8th Grade TVAAS Teacher Effectiveness Estimate		0.2907	8.57	0.0001	0.0339
4th Grade Score*5th Grade TVAAS Estimate		-0.0024	-2.55	0.0109	0.0009
4th Grade Score*6th Grade TVAAS Estimate		-0.0029	-3.34	0.0008	0.0009

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

These teacher estimates are very conservative since the individual effects of the teachers were calculated *prior to* the year the students were under the tutelage of the teachers; thus, the students in the sample in no way impacted the teacher estimates used in the analyses.

The sixth grade teacher's level of effectiveness also produced a sizable effect on the Competency scores. From Table 2, the sixth grade teacher effect estimate is 2.4091. This parameter multiplied by the means for very ineffective and very effective sixth grade teachers, -6 and +7.5 respectively, predicts the difference in scores for very ineffective and very effective teachers will average 32 points.

Although the parameters estimated for seventh and eighth grade teachers were much smaller, 0.1605 (se = 0.0382) and 0.2907 (se = 0.0339) respectively, the impact was important because of the student stakes associated with the test. Although the estimated effects were very small, the standard errors of the estimates confirmed the stability of the estimates. These minimal effects would be more critical for students whose scores fall very near the cut score required for passing.

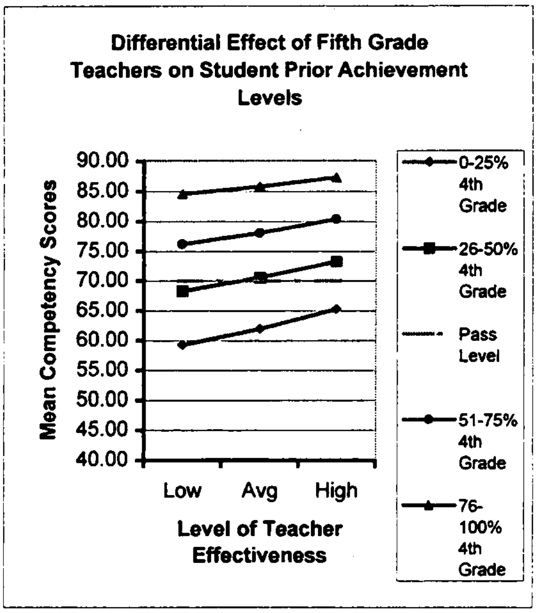
The observed significant interaction between fourth grade scores and fifth and sixth grade teacher effectiveness levels (Table 1) suggested the inclusion of a series of functions (Table 3) to evaluate the estimated Competency means as influenced by differential teacher effects and varying levels of student previous achievement. For example, f_x 1 provides the Competency mean estimate, the related t-test for the H(O) and its probability,

Table 3. Estimable Functions for Competency Mathematics Score.

f_x #	Parameter	Estimate	Std Error of Estimate
1	Q1 Low 5th Teach	59.212	0.627
2	Q1 Avg. 5th Teach	61.943	0.524
3	Q1 High 5th Teach	65.281	0.805
4	Q1 Low 6th Teach	58.949	0.633
5	Q1 Avg. 6th Teach	61.956	0.521
6	Q1 High 6th Teach	65.392	0.671
7	Q1 Low 7th Teach	60.849	0.553
8	Q1 Avg. 7th Teach	61.764	0.520
9	Q1 High 7th Teach	63.017	0.613
10	Q1 Low 8th Teach	60.069	0.558
11	Q1 Avg. 8th Teach	61.813	0.521
12	Q1 High 8th Teach	63.557	0.560
13	Q1 4 Low Teach	53.641	0.733
14	Q1 4 Avg. Teach	62.038	0.523
15	Q1 4 High Teach	71.808	0.939
16	Q2 Low 5th Teach	68.319	0.504
17	Q2 Avg. 5t Teach	70.550	0.421
18	Q2 High 5th Teach	73.277	0.566
19	Q2 Low 6th Teach	68.153	0.495
20	Q2 Avg. 6th Teach	70.559	0.420
21	Q2 High 6th Teach	73.308	0.493
22	Q2 Low 7th Teach	69.481	0.472
23	Q2 Avg. 7th Teach	70.396	0.421
24	Q2 High 7th Teach	71.648	0.518
25	Q2 Low 8th Teach	68.700	0.479
26	Q2 Avg. 8t Teach	70.444	0.421
27	Q2 High 8th Teach	72.188	0.456
28	Q2 4 Low Teach	63.321	0.618
29	Q2 4 Avg. Teach	70.617	0.420
30	Q2 4 High Teach	79.089	0.673
31	Q3 Low 5th Teach	76.199	0.482
32	Q3 Avg. 5t Teach	78.081	0.401
33	Q3 High 5th Teach	80.372	0.485
34	Q3 Low 6th Teach	76.111	0.472
35	Q3 Avg. 6t Teach	78.086	0.400
36	Q3 High 6th Teach	80.344	0.448
37	Q3 Low 7th Teach	77.029	0.462
38	Q3 Avg. 7t Teach	77.944	0.402
39	Q3 High 7th Teach	79.196	0.494
40	Q3 Low 8th Teach	76.248	0.465
41	Q3 Avg. 8t Teach	77.992	0.402
42	Q3 High 8th Teach	79.736	0.435
43	Q3 4 Low Teach	71.619	0.612
44	Q3 4 Avg. Teach	78.127	0.400
45	Q3 4 High Teach	85.672	0.587
46	Q4 Low 5th Teach	84.530	0.632
47	Q4 Avg. 5th Teach	85.771	0.433
48	Q4 High 5th Teach	87.288	0.470
49	Q4 Low 6th Teach	84.556	0.601
50	Q4 Avg. 6th Teach	85.769	0.434
51	Q4 High 6th Teach	87.157	0.493
52	Q4 Low 7th Teach	84.748	0.516
53	Q4 Avg. 7th Teach	85.663	0.440
54	Q4 High 7th Teach	86.916	0.498
55	Q4 Low 8th Teach	83.968	0.503
56	Q4 Avg. 8th Teach	85.712	0.439
57	Q4 High 8th Teach	87.456	0.464
58	Q4 4 Low Teach	80.666	0.826
59	Q4 4 Avg. Teach	85.780	0.430
60	Q4 4 High Teach	91.681	0.565

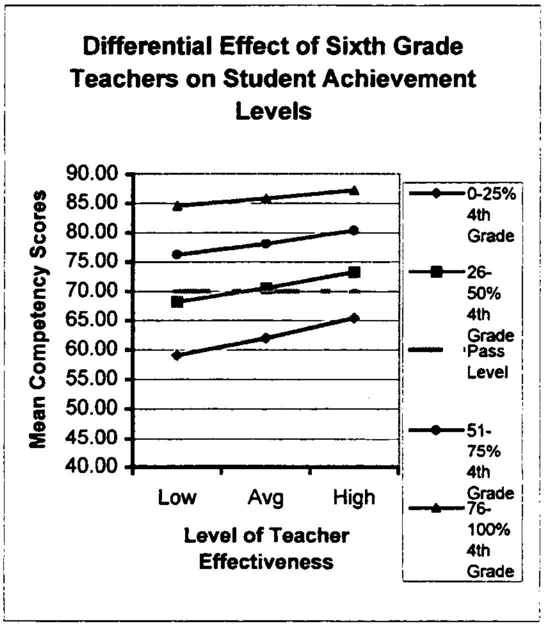
and the standard error of the estimate for the lowest quartile of prior achievement (Q1) students experiencing the least effective fifth grade teachers. These means were estimated holding all teacher effects constant except the effect(s) being considered and assuming that teachers for grades other than that being considered were average in effectiveness. From these means, it was possible to discern the magnitude of the interaction effect on varying achievement levels of students. For instance, mean Competency scores for students of a very low prior-achievement level ranged from 59 to 65 (se = 0.63 to 0.81), depending on the level of teacher effectiveness the students encountered in fifth grade. Similarly, mean Competency scores for students in the 26-50% prior-achievement level varied from 68 to 73 (se = 0.50 to 0.57) as the level of teacher effectiveness increased from very ineffective to very effective. The passing cut score of 70 is required for a student to receive a high school diploma.

This interaction is illustrated graphically in Figure 1. For example, the slope of the line representing the first quartile (Q1) is steeper than that of the line representing the fourth quartile (Q4). The nonparallel lines representing each quartile of students indicate that the fifth and sixth grade teachers are not equally effective with all prior-achievement groups of students. The magnitude of the differences in estimated means for both fifth and sixth grade teachers diminished as prior-achievement level increased as a result of this differential effect of fifth and sixth grade teachers on prior-achievement



Standard Errors for Each 5th Grade Teacher Level

	Low	Avg	High
0-25% 4th Grade	0.63	0.52	0.81
26-50% 4th Grade	0.50	0.42	0.57
51-75% 4th Grade	0.48	0.40	0.49
76-100% 4th Grade	0.63	0.43	0.47



Standard Errors for Each 6th Grade Teacher Level

	Low	Avg	High
0-25% 4th Grade	0.63	0.52	0.67
26-50% 4th Grade	0.49	0.42	0.47
51-75% 4th Grade	0.47	0.40	0.45
76-100% 4th Grade	0.60	0.43	0.49

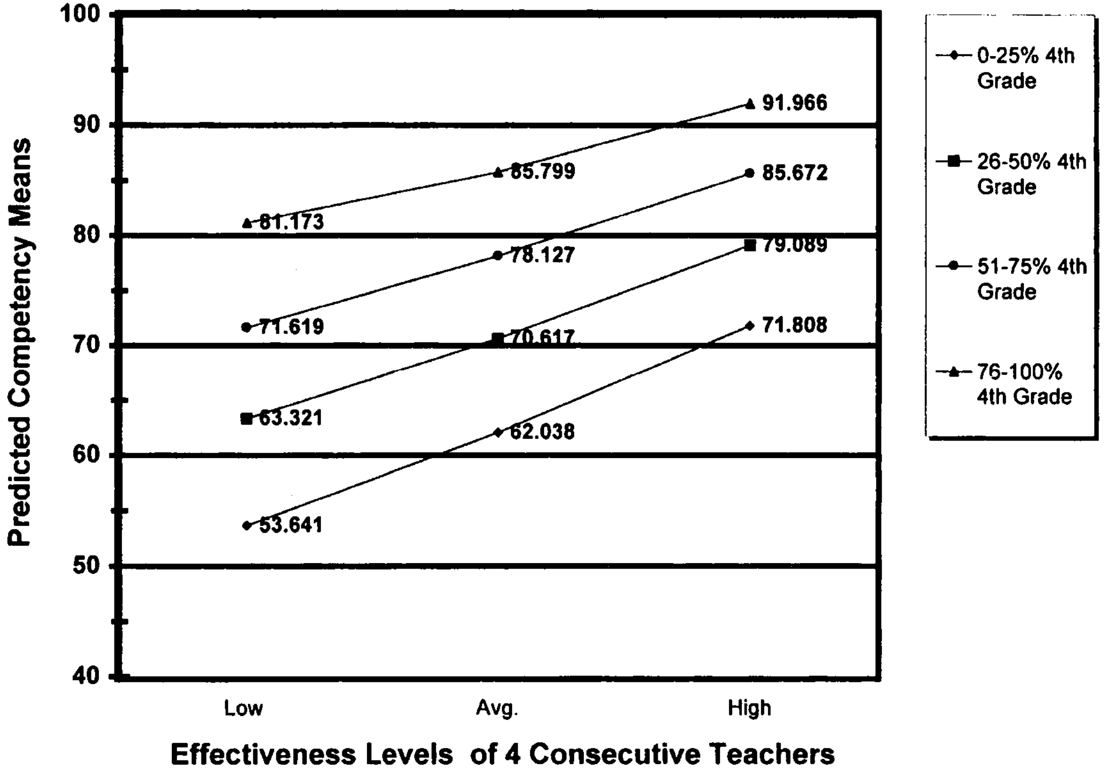
Figure 1. Teacher Effectiveness and Prior Achievement Interaction

levels. Thus, the lower achieving students were more susceptible to Competency failure as a result of their teachers' effectiveness.

The measurable effect of teacher effectiveness sequences on student Competency scores was determined by a comparison of the estimated means where a student's four consecutive teachers were of comparable effectiveness levels (Figure 2. Effects of 4 Consecutive Teachers of Similar effectiveness Levels on Fourth Grade Prior Achievement Levels). The Competency means estimated for these sequences for students of similar prior-achievement levels illustrated the potential for variability in students passing or failing scores resulting from the residual teacher effects accumulating across grades. In Figure 2, for example, the estimated Competency means of the students in the lowest quartile of fourth grade prior achievement vary from 53.6 (se = .733) for four very ineffective teachers to 71.8 (se = .939) for four very effective teachers. Additionally students in the 26-50% range of fourth grade prior achievement were predicted to have mean Competency scores of 63.3 (se = 0.62) if they experienced four very ineffective teachers in grades five through eight and 79.1 (se. = 0.67) if they experienced four very effective teachers in these same grades.

Tests for Ethnic and Gender Effects

This step in the analyses required the testing for possible ethnic and gender differential effects on student prior achievement at various sequences of



Standard Errors for Each Teacher Level

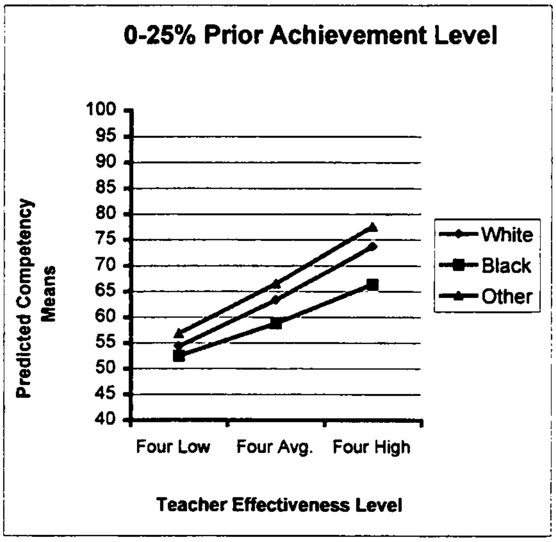
	Low	Avg	High
0-25% 4th Grade	0.813	0.525	0.939
26-50% 4th Grade	0.618	0.420	0.673
51-75% 4th Grade	0.612	0.400	0.587
76-100% 4th Grad	0.826	0.430	0.565

Figure 2. Interaction of 4 Consecutive Teachers of Similar Effectiveness Levels with 4th Grade Prior Achievement Levels

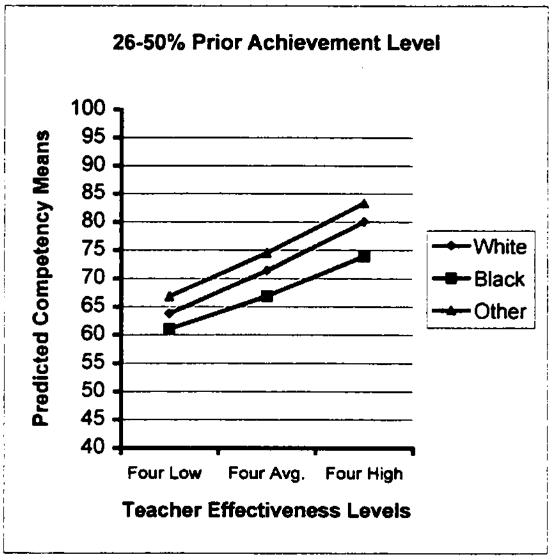
teacher effectiveness. This required the addition of ethnicity and gender variables and the logical interactions of district and ethnicity, ethnicity and gender, teacher estimates (grades 5-8) and ethnicity, fourth grade score, 5^{th} and 6^{th} grade teacher estimates and ethnicity, and teacher estimates (grades 5-8) and gender were added to the base model. The partial sums of squares showed a significant effect for ethnicity and the interactions of district and ethnicity and grade seven teacher effectiveness and ethnicity (F = 27.80, 4.94, and 3.67, respectively; p-value=0.001, 0.0072, and 0.0256). This differential effect appeared less significant, however, than the interaction of teacher effectiveness and fourth grade prior-achievement level. Clearly, lower achieving children in all ethnic groups benefit dramatically from sequences of highly effective teachers. See Figure 3 for a comparison of extreme levels of effectiveness by prior-achievement groups.

Predictions for Passing Probabilities

To determine the effect of cut score variation on passing probabilities for students of similar ability within specific teacher sequences, the base model was fitted to the data to predict a success variable at each assumed cut score: 60,65, 70, 75, 80. A combined ANOVA table for these analyses is presented in Table 4, with corresponding solution vectors in Table 5. The F-statistics increased rapidly and the R-Square for individual models increased as the cut score was raised, indicating that the passing probability grew increasingly more

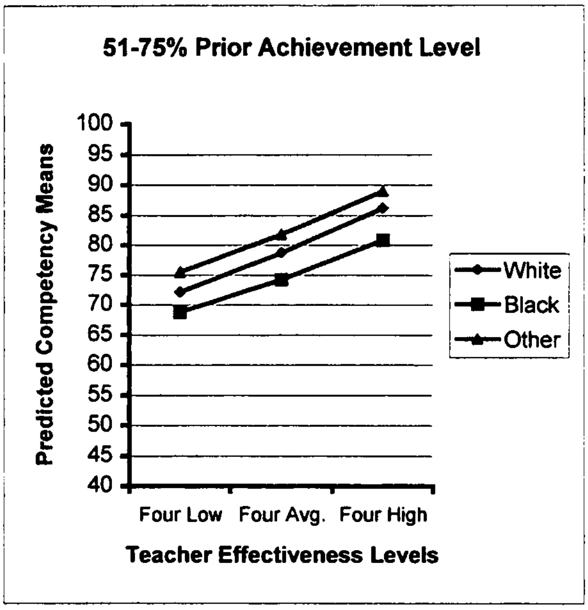


Standard Errors 0-25% Prior Achievement Level			
	Low	Avg	High
White	0.94	0.58	1.10
Black	1.04	0.68	1.54
Other	4.00	1.38	4.17

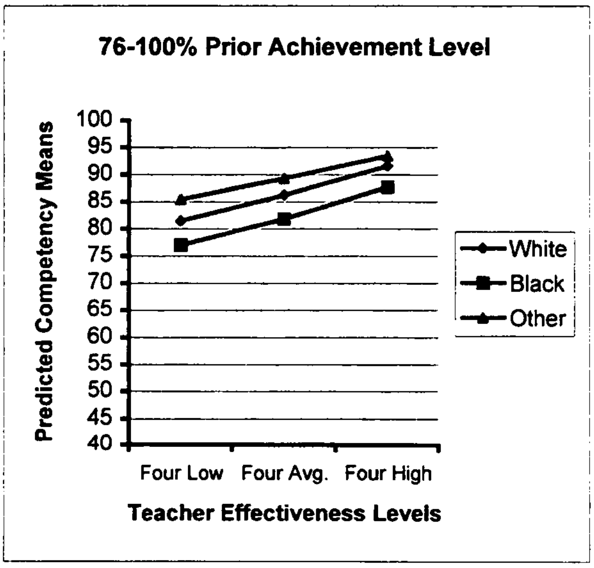


Standard Errors 26-50% Prior Achievement Level			
	Low	Avg	High
White	0.74	0.46	0.79
Black	0.99	0.66	1.28
Other	3.22	1.31	2.95

Figure 3. Four Consecutive Teachers' Effect on Prior Achievement Groups by Ethnicity. (a) Prior Achievement Quartiles 1 and 2.

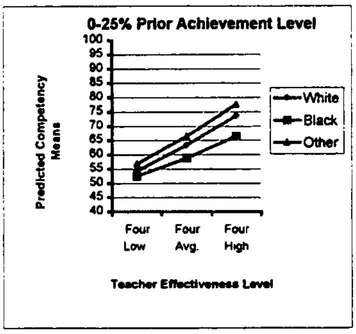


Standard Errors 51-75% Prior Achievement Level			
	Low	Avg	High
White	0.69	0.42	0.65
Black	1.11	0.67	1.35
Other	2.91	1.32	2.33

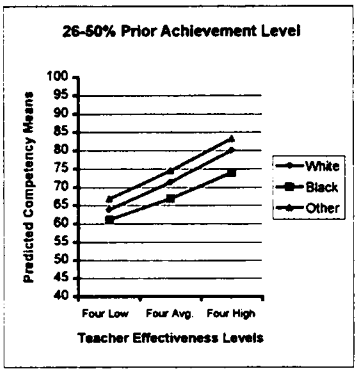


Standard Errors 76-100% Prior Achievement Level			
	Low	Avg	High
White	0.88	0.44	0.61
Black	1.55	0.71	1.81
Other	2.99	1.33	2.36

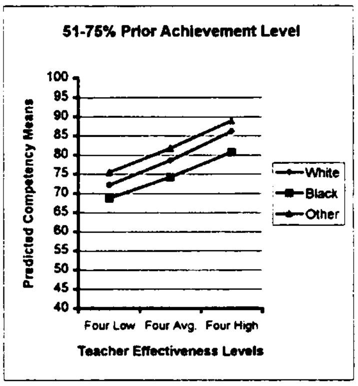
Figure 3. (Continued) (b) Prior Achievement Quartiles 3 and 4.



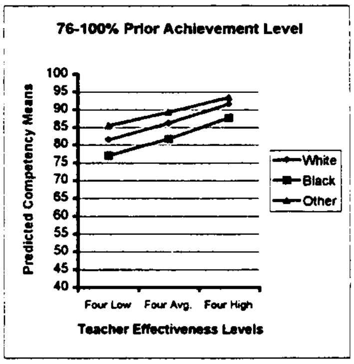
Standard Errors 0-25% Prior Achievement Level			
	Low	Avg	High
White	0.94	0.58	1.10
Black	1.04	0.68	1.54
Other	4.00	1.38	4.17



Standard Errors 26-50% Prior Achievement Level			
	Low	Avg	High
White	0.74	0.46	0.79
Black	0.99	0.66	1.28
Other	3.22	1.31	2.95



Standard Errors 51-75% Prior Achievement Level			
	Low	Avg	High
White	0.69	0.42	0.65
Black	1.11	0.67	1.35
Other	2.91	1.32	2.33



Standard Errors 76-100% Prior Achievement Level			
	Low	Avg	High
White	0.88	0.44	0.61
Black	1.55	0.71	1.81
Other	2.99	1.33	2.36

Figure 3. (Continued) (c) All Prior Achievement Quartiles.

Table 4. ANOVA Table for Predicting Success Probabilities at Various Cut Scores

Dependent Variable: Success at Cut Scores 60, 65, 70, 75, 80

		Cut Score = 60			Cut Score = 65			Cut Score = 70			Cut Score = 75			Cut Score = 80		
Source of Variation	DF	Mean Square	F Value	Pr > F	Mean Square	F Value	Pr > F	Mean Square	F Value	Pr > F	Mean Square	F Value	Pr > F	Mean Square	F Value	Pr > F
Type I Sums of Squares																
Model	12	7.93	82.12	0.0001	11.39	96.02	0.0001	14.58	108.1	0.0001	17.77	116.07	0.0001	20.09	126.83	0.0001
Residual	2599	0.10			0.12			0.13			0.15			0.16		
Corrected Total	2611															
R-Square For Model		0.275			0.307			0.333			0.349			0.369		
Mean Probabilities for Cut Score		0.843			0.782			0.721			0.626			0.511		
Partial Sums of Squares (SAS Type III)																
4th Grade Score	1	4.61	47.67	0.0001	4.33	36.50	0.0001	3.86	28.65	0.0001	3.98	25.98	0.0001	5.96	37.64	0.0001
4th Grade Quartile	3	2.47	25.58	0.0001	3.35	28.24	0.0001	4.31	31.93	0.0001	4.28	27.96	0.0001	4.14	26.14	0.0001
District	1	3.64	37.65	0.0001	3.31	27.92	0.0001	2.48	18.37	0.0001	1.45	9.46	0.0021	0.14	0.87	0.3522
4th Grade Score * District	1	3.30	34.19	0.0001	2.83	23.87	0.0001	2.02	14.98	0.0001	1.07	6.98	0.0083	0.06	0.36	0.5510
5th Grade TVAAS Teacher Effectiveness Estimate	1	2.20	22.76	0.0001	1.39	11.69	0.0006	0.45	3.34	0.0677	0.25	1.62	0.2034	0.11	0.72	0.3960
6th Grade TVAAS Teacher Effectiveness Estimate	1	3.03	31.38	0.0001	3.13	26.37	0.0001	2.47	18.35	0.0001	0.94	6.11	0.0135	0.14	0.91	0.3401
7th Grade TVAAS Teacher Effectiveness Estimate	1	0.14	1.43	0.2320	0.07	0.61	0.4331	0.46	3.40	0.0651	1.95	12.74	0.0004	1.99	12.54	0.0004
8th Grade TVAAS Teacher Effectiveness Estimate	1	1.92	19.88	0.0001	4.35	36.68	0.0001	6.39	47.36	0.0001	7.09	46.29	0.0001	7.66	48.34	0.0001
4th Grade Score * 5th Grade TVAAS Estimate	1	2.03	21.03	0.0001	1.24	10.42	0.0013	0.34	2.53	0.1115	0.14	0.93	0.3354	0.04	0.25	0.6168
4th Grade Score *6th Grade TVAAS Estimate	1	2.72	28.18	0.0001	2.78	23.45	0.0001	2.10	15.54	0.0001	0.67	4.38	0.0364	0.06	0.35	0.5553

Table 5. Solution Vector for Success Variables.

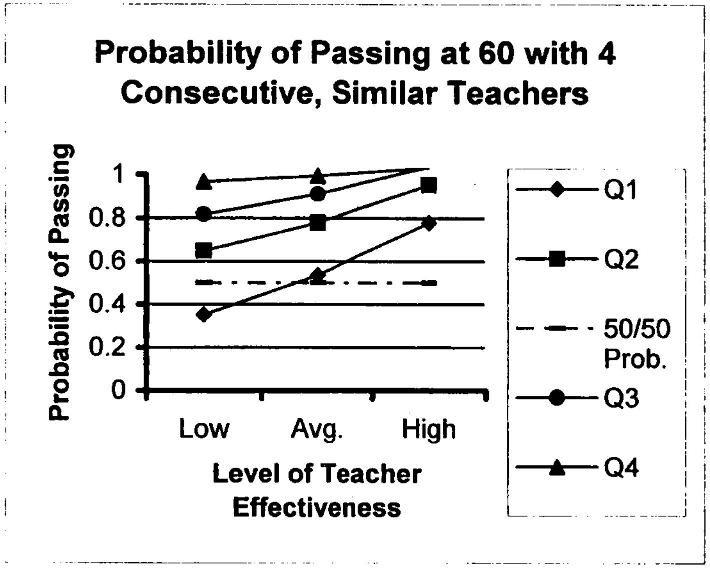
Parameter		Cut Score = 60 Estimate	Cut Score = 60 Std Error of Estimate	Cut Score = 65 Estimate	Cut Score = 65 Std Error of Estimate	Cut Score = 70 Estimate	Cut Score = 70 Std Error of Estimate	Cut Score = 75 Estimate	Cut Score = 75 Std Error of Estimate	Cut Score = 80 Estimate	Cut Score = 80 Std Error of Estimate
Intercept		-0.3537 B	0.3220	-0.3442 B	0.3568	-0.3663 B	0.3805	-0.6000 B	0.4054	-1.4203 B	0.4122
4th Grade Score		0.0017 B	0.0004	0.0017 B	0.0005	0.0017 B	0.0005	0.0020 B	0.0005	0.0029 B	0.0005
4th Grade Quartile	0	-0.1947 B	0.0414	-0.2944 B	0.0459	-0.3694 B	0.0489	-0.3782 B	0.0522	-0.3501 B	0.0530
	1	-0.0397 B	0.0293	-0.1327 B	0.0325	-0.1941 B	0.0346	-0.2798 B	0.0369	-0.3128 B	0.0375
	2	0.0285 B	0.0222	-0.0078 B	0.0246	-0.0388 B	0.0262	-0.0860 B	0.0279	-0.1250 B	0.0284
	3	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.
District Number	A	-1.4173 B	0.2310	-1.3524 B	0.2560	-1.1698 B	0.2729	-0.8943 B	0.2908	-0.2752 B	0.2957
	B	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.
4th Grade Score * District Number	A	0.0019 B	0.0003	0.0017 B	0.0004	0.0015 B	0.0004	0.0011 B	0.0004	0.0002 B	0.0004
	B	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.	0.0000 B	.
5th Grade TVAAS Teacher Effectiveness Estimate		0.0966	0.0203	0.0767	0.0224	0.0437	0.0239	0.0324	0.0255	0.0220	0.0259
6th Grade TVAAS Teacher Effectiveness Estimate		0.1030	0.0184	0.1047	0.0204	0.0931	0.0217	0.0572	0.0232	0.0225	0.0235
7th Grade TVAAS Teacher Effectiveness Estimate		0.0013	0.0011	0.0010	0.0012	0.0024	0.0013	0.0051	0.0014	0.0051	0.0014
8th Grade TVAAS Teacher Effectiveness Estimate		0.0044	0.0010	0.0067	0.0011	0.0081	0.0012	0.0085	0.0013	0.0089	0.0013
4th Grade Score*5th Grade TVAAS Estimate		-0.0001	0.0000	-0.0001	0.0000	-0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
4th Grade Score*6th Grade TVAAS Estimate		-0.0001	0.0000	-0.0001	0.0000	-0.0001	0.0000	-0.0001	0.0000	0.0000	0.0000

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations.
Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

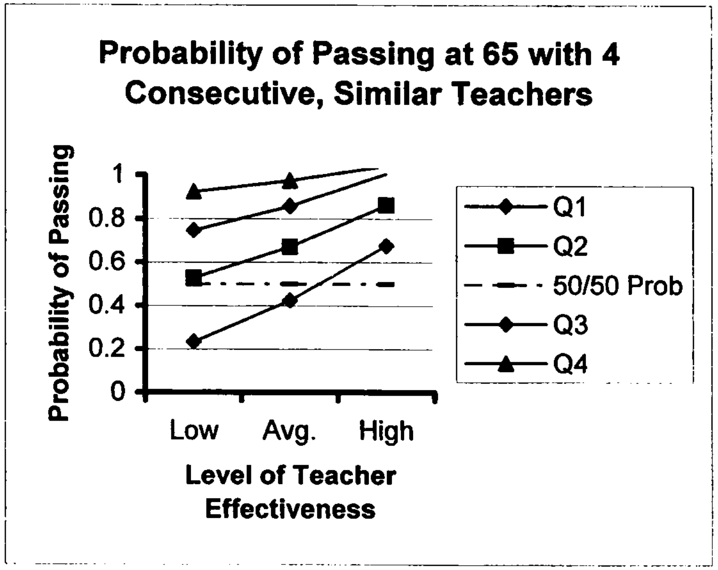
dependent on the variables of the model. The mean probabilities for passing that were generated decreased from .843 (average probability of passing at 60) to .511 (average probability of passing at 80). The average percentage of students passing at 60 was 84%, while the number passing at 80 was 51%, a drop of about one third of the students.

As evidenced by the partial sums of squares of Table 4, fifth grade teachers, although initially were highly significant, gradually lost their impact on passing probabilities when the cut score reached 70. Sixth grade teachers remained highly significant through a cut score of 70, trailing off to non-significance at 80. Seventh grade teachers did not contribute to success probabilities until the cut score reached 75, but continued when the cut score was raised to 80. Eighth grade teachers, as would be expected, had a measurable influence on passing probability across all scores.

The graphs in Figure 4 illustrate the dramatic effect of teacher effectiveness sequences on student passing probabilities. Clearly, all children benefit from highly effective teachers, but the lower 50% of the children appear to benefit most. Looking at the 50% probability of passing line in the center of each graph, it is apparent that a change in level of effectiveness of teacher sequence can increase the probability of passing to at least 50/50 chance for at least one quartile of students for each cut score. Additionally, the differential effect of teacher effectiveness on student prior-achievement levels is obvious from Figure 4. Beginning with an assumed score of 60, this differential effect continues through all scores until the cut score is set at 75.



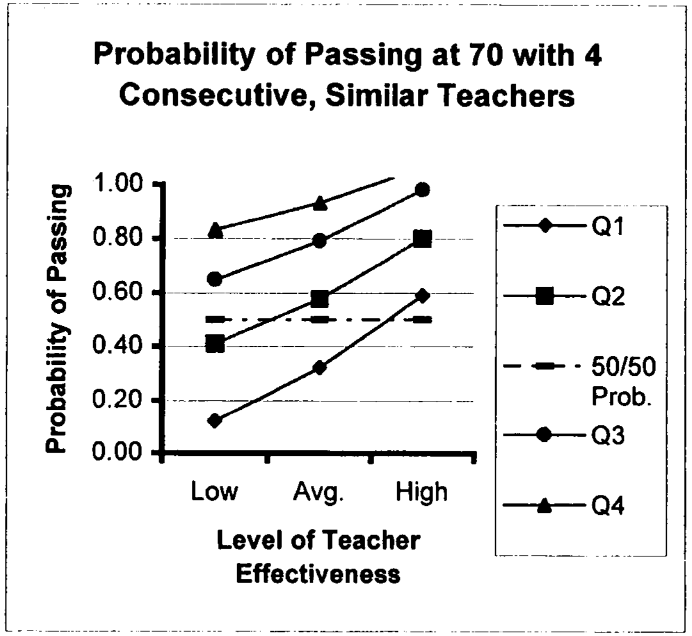
Standard Errors At Cut Score of 60			
	Low	Avg	High
Q1	0.022	0.015	0.028
Q2	0.018	0.012	0.021
Q2	0.018	0.012	0.018
Q4	0.024	0.013	0.018



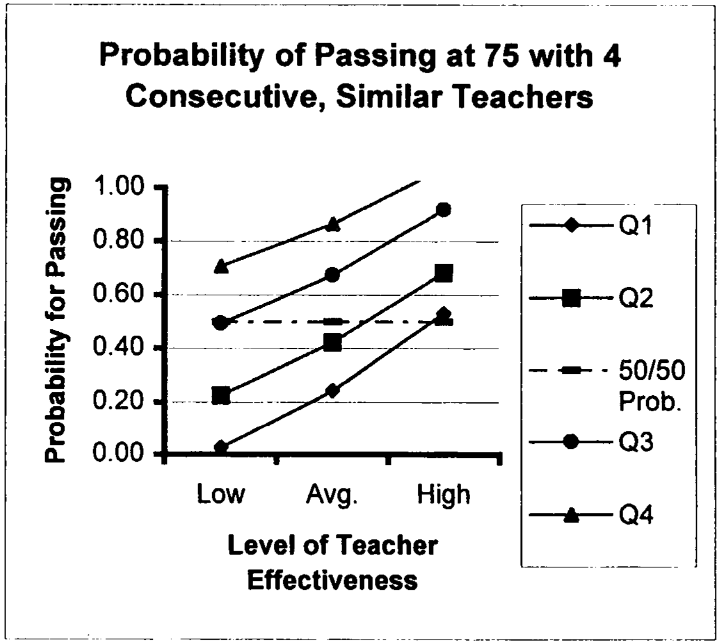
Standard Errors At Cut Score of 65			
	Low	Avg	High
Q1	0.024	0.017	0.032
Q2	0.020	0.014	0.023
Q2	0.020	0.013	0.020
Q4	0.027	0.014	0.020

Figure 4. Effect of Varying Cut Scores on Probabilities for Passing.

(a) Assumed Scores of 60 and 65.

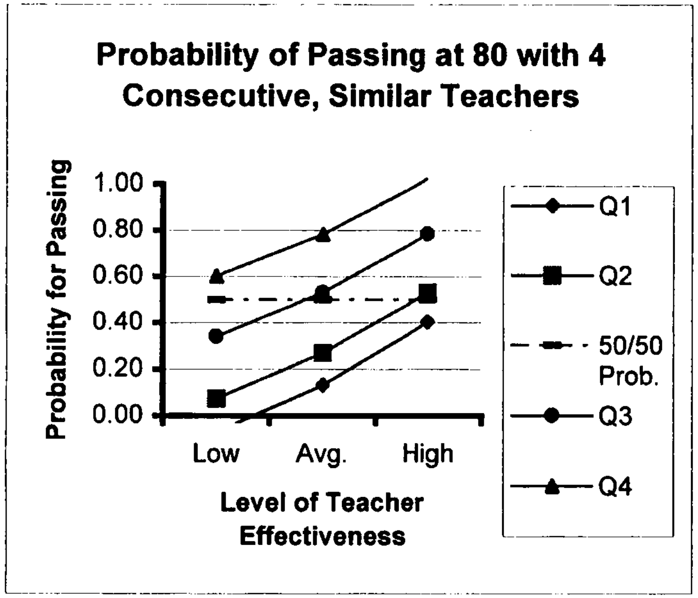


Standard Errors At Cut Score of 70			
	Low	Avg	High
Q1	0.025	0.018	0.034
Q2	0.021	0.015	0.024
Q2	0.015	0.021	0.014
Q4	0.029	0.015	0.021



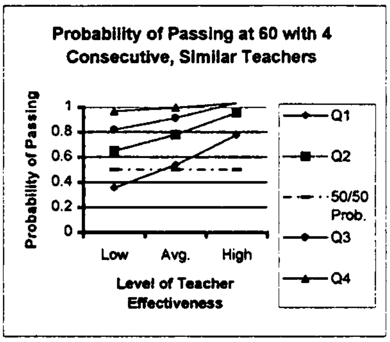
Standard Errors At Cut Score of 75			
	Low	Avg	High
Q1	0.027	0.019	0.036
Q2	0.023	0.016	0.026
Q2	0.022	0.015	0.023
Q4	0.031	0.016	0.022

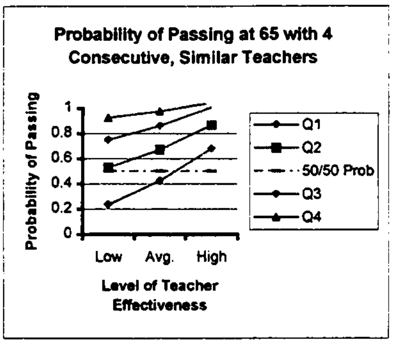
Figure 4. (Continued) (b) Assumed Scores of 70 and 75.



Standard Errors At Cut Score of 80			
	Low	Avg	High
Q1	0.028	0.02	0.036
Q2	0.023	0.016	0.017
Q2	0.017	0.015	0.023
Q4	0.031	0.017	0.022

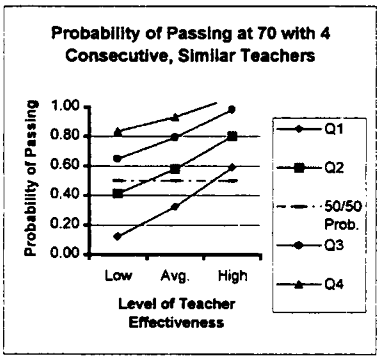
Figure 4. (Continued) (c) Assumed Score of 80.

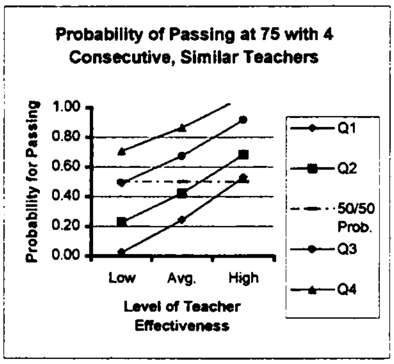




Standard Errors At Cut Score of 60			
	Low	Avg	High
Q1	0.022	0.015	0.028
Q2	0.018	0.012	0.021
Q2	0.018	0.012	0.018
Q4	0.024	0.013	0.018

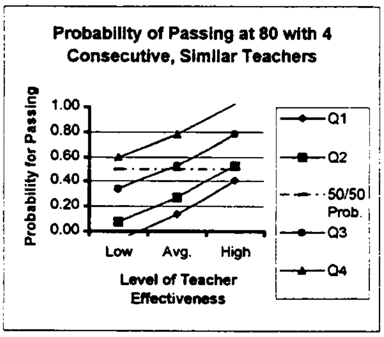
Standard Errors At Cut Score of 65			
	Low	Avg	High
Q1	0.024	0.017	0.032
Q2	0.020	0.014	0.023
Q2	0.020	0.013	0.020
Q4	0.027	0.014	0.020





Standard Errors At Cut Score of 70			
	Low	Avg	High
Q1	0.025	0.018	0.034
Q2	0.021	0.015	0.024
Q2	0.015	0.021	0.014
Q4	0.029	0.015	0.021

Standard Errors At Cut Score of 75			
	Low	Avg	High
Q1	0.027	0.019	0.036
Q2	0.023	0.016	0.026
Q2	0.022	0.015	0.023
Q4	0.031	0.016	0.022



Standard Errors At Cut Score of 80			
	Low	Avg	High
Q1	0.028	0.02	0.036
Q2	0.023	0.016	0.017
Q2	0.017	0.015	0.023
Q4	0.031	0.017	0.022

Figure 4. (Continued) (d) Assumed scores of 60, 65, 70, 75, and 80 are provided for trend analysis.

CHAPTER V

DISCUSSION AND IMPLICATIONS

Summary of the Study

When a child fails a high stakes test, no one takes into account the quality of that child's educational experience prior to invoking whatever negative consequences are to be associated with his/her failure. The assumption that effective educational delivery is a given constant may not be appropriate. The problem addressed by this study is the way in which teacher assignment might affect students' ability to pass the TCAP Competency test. If the effectiveness of the child's previous teachers is measurable when the test is administered and the degree of effectiveness is the deferential factor in determining the pass/fail rate for similar populations of students, then perhaps the school district bears some responsibility for the child's failing score. On the eve of higher standards in Tennessee, the impact of raising the cut score on the current test to assumed higher levels merits additional attention. The latent and cumulative role of teacher effectiveness in determining student achievement is unknown.

The purpose of this study was to estimate the effect of a student's series of effective/ineffective teachers on his/her TCAP Competency math score. Incidental to the series determination was the estimation of the effective of individual teachers on the scores. The estimates of teacher effectiveness were from years prior to the year the student was assigned to the teacher (thus eliminating potential bias of measuring the effectiveness of the teacher with the scores of students included in the analyses). Varying the assumed cut scores for this test allowed the measuring of the impact of teachers and teacher sequences on changing the scores. The study adds to the body of knowledge concerning teacher effectiveness and sequences of teacher effectiveness and provides valuable information for policy makers about educators' liability in student success/failure on high stakes student tests.

Findings

The research hypothesis for the study was: teacher effectiveness has a residual and latent effect on student Competency math scores. Three research questions guided the study.

I. What is the measurable effect of teacher effectiveness sequences on student Competency math scores?

(1) A model which included fourth grade scores as a covariate, fourth grade quartile rank as a class variable, the continuous variables of TVAAS teacher effects estimates for grades five through eight, and interaction terms for fourth grade math scores and teacher estimates for grades five and six explained 49.6 of the variation the Competency scores.

(2) Residual elementary teacher effects, estimated as an intermediate step, were measurable on students' ninth grade Competency math scores. Partial sums of squares for teacher effects for grades five through eight were significant, indicating an effect for each grade free of partial confounding by other variables.

(3) There was a significant interaction between prior-achievement level and both fifth and sixth grade teachers, indicating a differential effect of teachers on prior achievement levels. Additionally, the Competency means estimated by functions at mean prior-achievement levels (for each quartile) for each level of teacher effectiveness gave an indication of the variation in effect across prior achievement levels. The different levels of teacher effects produced more variation for lower achieving students.

(4) The effects were cumulative, with no significant interaction among teachers to indicate a compensatory effect; the magnitude of the cumulative effects was evidenced by the Competency means estimated for each quartile of student prior achievement for each teacher effectiveness level.

II. What are the interactions among gender, ethnicity and prior levels of student achievement and sequences of teacher effectiveness?

(1) When added to the model, partial sums of squares showed a significant effect for ethnicity and the interactions of district and ethnicity and grade seven teacher effectiveness and ethnicity. Estimated student Competency means for extremes of teacher effectiveness show that children of

color appeared to respond to increased teacher effectiveness at a slightly lesser rate than their white peers of comparable prior-achievement levels.

(2) None of the interactions between gender and teacher effects by grade were significant. There was no difference in the way males and females respond to ineffective and effective teachers.

III. How does varying the cut score required for passing the Competency test affect the mean probability of passing for students within specific teacher sequences?

(1) As the passing cut score was raised from 60 to 80, the *F*-statistics for the model estimating Competency scores from prior achievement and teacher effectiveness estimates for grades five through eight increased from 82.12 (*p*-value=0.0001) to 126.83 (*p*-value=0.0001), indicating the increasing need for teacher effectiveness for all students. Partial sums of squares were reported for the following, indicating no partial confounding with other variables. Fifth grade teacher effects for all students were significant at assumed scores of 60 and 65; sixth grade effects at 60 through 75; seventh grade effects at 75 and 80; eighth grade effects at 60 through 80. Additionally, with the significant interactions of 4th grade score and fifth and sixth grade teacher effects at assumed scores of 60 and 65, the differential effect of teachers with prior achievement groups was apparent. The sixth grade interaction remained significant for cut scores of 70 and 75.

(2) Functions estimating passing probabilities for the 0-25% prior achievement subgroup showed extreme variation in passing probabilities for

four consecutive teachers of similar effectiveness. Passing probabilities for the 26-50% prior achievement subgroup showed similar results. As the standard for passing a high stakes tests was raised, it became highly improbable that lower achieving students had any chance at all of passing unless they had experienced the most effective teachers available.

Conclusions

The results of this study suggest three general areas of potential impact:

(1) If teacher effects are residual and cumulative and measurable on Tennessee's high stakes student graduation test, then Tennessee's current high stakes student test is holding students accountable for something beyond their control, and the punitive policies associated with it should be reviewed for appropriateness.

(2) If lower achieving students are more adversely affected by these teacher effects than their higher achieving peers, a high stakes Competency test for lower achieving children offers more punitive potential for them than a similar test for higher achieving children. The significant interaction between prior-achievement levels and fifth and sixth grade teachers, whereby lower achieving children were more vulnerable to failure of the Competency test because of the relative ineffectiveness of their former teachers, has implications for those responsible for resource allocation. Before children can be held accountable for their learning, they must have adequate opportunities. The

variability in teacher effectiveness observed in this study suggests some lower achieving students may have failed because they were lacking in the opportunities provided them. Ironically, it was for the same lower achieving students that high stakes testing was put in place, the idea being that improved teaching and learning would trickle down from the testing. Now it appears that it is these children who suffer the most punitive damage for something that is beyond their control—the effectiveness of their teachers. For lower achieving children, especially, the necessity of effective teachers to prepare them academically for such a test is not only desirable; it is a critical determinant in their ability to pass.

(3) Additionally, the measurable effects of the broad variability in teacher effectiveness suggests a need for improved teacher preparation and additional staff development opportunities for practicing teachers and perhaps the addition of accountability measures for both.

Recommendations

The recommendations below are based on the conclusions of this study and have potential impact for three broad areas with some degree of overlap: student testing and staffing patterns, teacher quality, and educational research.

Student Testing and Staffing Patterns.

(1) Since the residual and cumulative teacher effects of elementary teachers are measurable on Competency student math scores at grade nine, policy makers should rethink the testing requirement as a prerequisite for high school graduation. This measurable effect of teachers on student performance adds a significant issue to the fair testing conundrum, the issue of equity in effective educational delivery. Under current policy, there is no adjustment in the testing process for the inability of the educational system to provide students with adequate learning opportunities. Students should be held harmless from ineffective teaching.

(2) Although there was no difference in the way males and females responded to ineffective and effective teachers, children of color appeared to respond to increased teacher effectiveness at a slightly lesser rate than their white peers of comparable prior-achievement levels. The magnitude of this differentiation paled, however, when students of ineffective teacher sequences were compared to those of very effective teachers. Thus, educational policy makers and educational practitioners should look very closely at the policies regulating teacher assignment to schools, making sure that schools with high percentages of ethnically diverse students receive at least their fair share of effective teachers. Additionally, the policies regulating student assignment to individual teachers merit a second look, so that children of color are not under represented in very effective teachers' classrooms.

(3) As the standard for passing a high stakes test is raised, teacher effectiveness becomes more critical to the success of all students, especially lower achieving ones. Educational policy makers who rush into setting high standards with only a superficial understanding of the consequences, although well meaning, are ill advised. Decisions concerning high stakes standards for students should be continually re-evaluated based on a method of measurement of the schools and teachers that insures that all students have effective teachers available to them. This process would protect both the student and the taxpayer.

(4) Additionally, a careful look at the sub-populations of students to be impacted by higher cut scores and the practices that allocate resources to these sub-populations is critical. As the standard for passing a high stakes test is raised, it becomes highly improbable that lower achieving students have any chance at all of passing unless they have the most effective teachers available. If there is no probability of success for some students, perhaps the expectation is unrealistic. At the very least, the policies governing resource allocation and the support systems currently in place for this sub-population of at-risk students require rethinking.

Teacher Quality.

The findings concerning the residual and cumulative effect of elementary teachers bear a message to policy makers responsible for teacher preparation and teacher post service development. Both processes have the potential for improving teacher quality, but policy makers need further evaluation of these areas. A rethinking of the resources available for both and the policies governing the accountability measures for both is certainly in order. One immediate consideration might be a focus on measurement interpretation. If prospective and current teachers learn to interpret measurements of their effectiveness as it relates to the achievement of their students, they will better understand where they are effective and where they are ineffective.

Educational Research.

In the area of educational research, two areas of need are obvious: (1) additional focus on what makes teachers and principals effective, and (2) the continued measurement of student achievement each year in each grade.

(1) Although with some degree of reliability, it is possible to identify relatively ineffective and effective teachers and schools, there has been little research beyond the identification. An immediate area of need would be additional research into what causes the variability. For instance, is there a differential effect of colleges of education on teacher preparation? Or, what areas of commonality are discernable in effective teaching? Additionally, an exploration of other subject areas, e. g., language arts, to measure the residual and cumulative teacher effects on student achievement in these areas would be appropriate.

(2) This study would not have been possible without the longitudinal data available from Tennessee's current accountability legislation and the tax dollars invested in accountability. The expense of providing longitudinal measurement of student achievement to gather the data necessary to adequately assess effective teaching and effective programs may seem prohibitive. In reality, it is less than ½% of the annual expenditure of educational dollars per student. Appropriate use of the information it provides would benefit policy makers and researchers and offset the cost. Adequate measurement for the purpose of ascertaining what is effective and why it is effective may be less expensive than

failed programs and failed students. It provides a wealth of information necessary for researchers to determine the lasting benefits found in sound educational practice.

REFERENCES

REFERENCES

Background of the Tennessee Comprehensive Assessment Program. (1988). An Internal Working Paper. Knoxville: University of Tennessee State Testing and Evaluation Center.

Bock, R. D. & Wolfe, R. (1996). Audit and review of the Tennessee Value-Added Assessment System (TVAAS): Final Report, Part 1. In Bock, R. D., Wolfe, F., & Fisher, T. H., A Review and Analysis of the Tennessee Value-Added Assessment System (Nashville: Tennessee Comptroller of the Treasury, Office of Education Accountability), pp. 69-72.

Bodenhausen, J. (1988, April). Does the academic background of teachers affect the performance of their students? Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 293 836)

Bridges, E. M. (1996). Evaluation for tenure and dismissal. In Millman, J. & Darling-Hammond, L. (Eds.), The New Handbook of Teacher Evaluation (pp. 147-157). Newbury Park: Sage Publications.

Ceperley, P. E., & Reel, K. (1997). The impetus for the Tennessee Value-Added Accountability System. In Millman, J. (Ed.), Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure? (pp. 133-136). Thousand Oaks, California: Corwin Press, Inc.

Coleman, J. S., Campbell, E. Q., Hobson, C J., McPartland, J., Mood., Alexander M., Weinfeld, F. D., and York, R. L. (1966). Equality of Educational Opportunity. Washington: U. S. Government Printing Office.

Cuban, L. (1989). At-risk students: What teachers and principals can do. Educational Leadership, 46 (5), 29-32.

Darlington, 1997. The Tennessee Value-Added Assessment System: A challenge to familiar assessment methods. In Millman, J. (Ed.), Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure? (pp. 163-168). Thousand Oaks, California: Corwin Press, Inc.

Dorn, S. (1996). "Social dynamite." In Creating the Dropout: An Institutional and Social History of School Failure (pp. 65-80). Westport, CT: Praeger Publishers.

Ferguson, R. F. (1990, December). Racial patterns in how school and teacher quality affect achievement and earnings. Cambridge, MA: John F. Kennedy School of Government, Sponsoring agency, Meadows Foundation. (ERIC Document Reproduction Service No. ED 347 217)

Graphical Summary of Educational Findings from the Tennessee Value-Added Assessment System (TVAAS), 1995. (1995). Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Hanushek, E. A. (1996). Outcomes, costs, and incentives in schools. In Hanushek, E. A. & Jorgenson, D. W. (Eds.), Improving America's Schools: The Role of Incentives, (pp. 29-52). Washington, DC: Natinal Academy Press.

Harville, D. A. (June 6, 1995). A Review of the Tennessee Value Added Assessment System (TVAAS). Iowa: Iowa State University.

Ingersoll, R. M., & Gruber, K. (1996, October). Out-of-Field Teaching and Educational Equality. Washington, D. C.: American Institutes for Research in the Behavioral Sciences, Sponsoring agency, National Center for Education Statistics. (ERIC Document Reproduction Service No. ED 402 302)

Krei, M. S. (1993). The political economy of teacher allocation: Influences and constraints on teacher recruitment, retention, and dismissal in low-income urban schools. Unpublished doctoral dissertation, Columbia University.

Krei, M. S. (1998, March). Intensifying the barriers: The problem of teacher allocation in low-income urban schools. Urban Education 33(1), 71-94.

Kortez, D. (1996). Using student assessments for educational accountability. In Hanushek, E. A. & Jorgenson, D. W. (Eds.), Improving America's Schools: The Role of Incentives, (pp. 171-195). Washington, DC: Natinal Academy Press.

Kronick, R. F., & Hargis, C. H. (1990). The visible and the invisible. In Dropouts: who drops out and why—and the recommended action (pp.23-31). Springfield, Illinois: Thomas Books.

Monk, D. H. (1987). Assigning elementary pupils to their teachers. The Elementary School Journal, 88(2), 467-187.

National Commission on Teaching & America's Future. What Matters Most: Teaching for America's Future. New York: Author, 1996.

Orfield G., Arenson, J., Jackson, T., Bohrer, C., Gavin, D., & Kalejs, E. (1997). City-suburban desegregation. Parent and student perspectives in metropolitan Boston. Cambridge, MA: Harvard Civil Rights Project, Sponsoring agency, John D. and Catherine T. MacArthur Foundation. (ERIC Document Reproduction Service No. ED 414 391).

Peng, S. S., & Lee, R. M. (1992, April). Measuring student at-riskness by demographic characteristics. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 347 679)

Popham, W. J. (1997). The moth and the flame: Student learning as a criterion of instructional competence. In Millman, J. (Ed.), Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure? (pp. 133-136). Thousand Oaks, California: Corwin Press, Inc.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. Review of Research in Education, 15, (pp. 423-479).

Ross, S. M., Sanders, W. L., Wright, S. P., & Stringfield, S. (June 1998). The Memphis restructuring initiative: Achievement results for years 1 and 2 on the Tennessee Value-Added Assessment System (TVAAS). A Special Report Prepared for Memphis City Schools. Memphis: The Center for Research in Educational Policy.

Sanders, W. L., & Horn, S. (1995). Educational assessment reassessed: The usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes. Educational Policy Analysis Archives, 3(6).

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

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Assessment System: A quantitative, outcomes-based approach to educational assessment. In Millman, J. (Ed.), Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure? (pp. 137-162). Thousand Oaks, California: Corwin Press, Inc.

SAS Institute Inc. (1997). SAS/STAT* Software: Changes and Enhancements through Release 6.12 (pp. 349-359), Cary, NC: SAS Institute Inc.

Searle, S. L. (1971). Linear Models (p. 180). New York, New York: John Wiley & Sons, Inc.

Smith, M. S., Scoll, B. W., & Link, J. (1996). Research-based school reform: The Clinton administration's agenda. In Hanushek, E. A., & Jorgenson, D. W. (Eds.), Improving America's Schools: The Role of Incentives (pp. 9-28). Washington, DC: National Academy Press.

State Testing and Evaluation Center. (1997). TCAP/CT Technical Manual. Knoxville: University of Tennessee.

State Testing and Evaluation Center. (1998). Statewide summary of student test results, 1997-98. An Internal Report. Knoxville: University of Tennessee.

Stiggins, R. J. (1994). Student-Centered Classroom Assessment. New York: Macmillan College Publishing Company.

Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? American Economic Review, 67, 639-652.

Minutes of the Tennessee State Board of Education, September 1998. (1998). Internal Document. Nashville: State Board of Education.

Waggoner, D. (1991). Who are the undereducated youth. In Undereducation in America: The Demography of High School Dropouts (pp. 19-43). Westport, CT: Auburn House.

APPENDICES

Appendix A

$$
\begin{aligned}
y_{ijk} = \; & a + b_1\,(4^{th}\text{ grade score}) + (4^{th}\text{ grade quartile})_i + \text{district}_j \\
& + b_2\;(4^{th}\text{ grade score} * \text{district}_j) + b_3\,(5^{th}\text{ grade TVAAS estimate}) \\
& + b_4\,(6^{th}\text{ grade TVAAS estimate}) + b_5(7^{th}\text{ grade TVAAS estimate}) \\
& + b_6\,(8^{th}\text{ grade TVAAS estimate}) \\
& + b_7\,(4^{th}\text{ grade score}*5^{th}\text{ grade TVAAS estimate}) \\
& + b_8\,(4^{th}\text{ grade score}*6^{th}\text{ grade TVAAS estimate}) \\
& + e_{ijk}
\end{aligned}
$$

where,

y_{ijk} = Competency score of the kth student in the jth district in the ith quartile of prior achievement

a = Intercept

b_1 = partial regression coefficient for fourth grade scores

$(4^{th}$ grade quartile$)_i$ = effect of 4^{th} grade quartile of prior achievement

district$_j$ = district effect

b_2 = partial regression coefficient of the interaction of 4^{th} grade quartile of prior achievement * district$_j$

b_3 = partial regression coefficient for 5^{th} grade TVAAS estimate

b_4 = partial regression coefficient for 6^{th} grade TVAAS estimate

b_5 = partial regression coefficient for 7^{th} grade TVAAS estimate

b_6 = partial regression coefficient for 8^{th} grade TVAAS estimate

b_7 = partial regression coefficient of the interaction of 4^{th} grade score * 5^{th} grade TVAAS estimate

b_8 = partial regression coefficient of the interaction of 4^{th} grade score * 6^{th} grade TVAAS estimate

e_{ijk} = random error

Appendix B

Estimable Function to Estimate Competency Means for 26-50% Group of Prior Achievement Assuming Sixth Grade Teacher Was Very Ineffective

Model	Value
Intercept	1
4th Grade Score	701 (mean 4th grade score of this quartile)
4th Grade Rank	0 1 0 0 (2nd position coded with 1 indicates 2nd quartile)
District	.5 .5 (District effects are averaged across)
4th Grade Score * District	350.5 350.5 (District effects are averaged across)
TVAAS estimate for 5th grade teacher	0 (Held at average)
TVAAS estimate for 6th grade teacher	-6 (Lowest mean effectiveness score
TVAAS estimate for 7th grade teacher	0 (Held at average)
TVAAS estimate for 8th grade teacher	0 (Held at average)
4th Grade Score * TVAAS estimate for 5th grade teacher	0 0 0 (Held at average)
4th Grade Score * TVAAS estimate for 6th grade teacher	0 -4206 0 (Worst 6th grade teacher * 2nd Quartile of Students)

VITA

June C. Rivers Sanders was born in Hickman County, Tennessee on June 12, 1947. She attended Centerville Elementary School and graduated from Hickman County High School in May 1965. She began her studies at Middle Tennessee State University in August of that year, graduating with a Bachelor of Science in Mathematics and minors in secondary education and English in August 1968. She earned a Masters in Education from Trevecca Nazarene College in August 1990 and was admitted to graduate school in the College of Education at the University of Tennessee in 1995.

She has worked in a variety of public and private sector positions: Tennessee Eastman Company, computer programmer analyst; R. J. Reynolds, Inc., computer programmer analyst; Lowndes County, Georgia, high school math and science teacher; Centerville Church of Christ, director of Day Care Center; Industrial Metals Corporation, administrative assistant; Hickman County Schools, teacher and central office administrator; Tennessee Department of Education, director of Tennessee's school nutrition program and executive administrative assistant to the commissioner of education; University of Tennessee State Testing and Evaluation Center, educational consultant. She is currently the director of the University of Tennessee Test Processing Services.